DEVELOPMENT OF LARGE SCALE STRUCTURED LIGHT BASED MEASUREMENT SYSTEMS

By

Chi Zhang

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Electrical Engineering

2012
ABSTRACT

DEVELOPMENT OF LARGE SCALE STRUCTURED LIGHT BASED MEASUREMENT SYSTEMS

By

Chi Zhang

The development of large scale structured light based dimensional measurement systems is introduced in this study. Based on different applications, Two different structured light systems are developed: 3D inspection system which emphasized on accuracy and fast environmental reconstruction which emphasized on measurement speed.

Quality inspection is a process to evaluate the quality of a manufactured work piece’s 3D shape. Compared with traditional Coordinate Measurement Machine (CMM), structured light based optic measurement has the merits of fast speed, high inspection sample rate, and overall simplicity of operation. The measurement result is further compared with the Computer Aided Design (CAD) model to evaluate its quality. There are basically three main issues, when the system scale is enlarged.

Calibration of such a large system with long standoff distance, large Field Of View (FOV), deep Depth Of View (DOV), and multi-sensor is a challenge task. In order to maintain high accuracy, an innovative 3D sensor model with less calibration parameters was developed. Instead of employing the incident light in projector frame, 3D point recovery was achieved within the camera frame via plane induced parallax information, in which projector’s intrinsic and orientation are avoided in the 3D model. Second, structured light systems is usually weak against surface optical properties. Material exhibiting different color textures, reflection ratios, and especially a mixture between specula reflection and diffuse reflection, fails the optic sensor to correctly acquire incident light information. Traditional structure light method is only valid for parts with diffuse reflection property. A robust surface
decoding method for industry application was developed. Monochromatic light was utilized against different object color textures. The illumination of the projector was automatically adjusted based on the test material. Furthermore, an extrapolation model to solve the internal reflection problem and a sub-pixel interpolation model to increase measurement accuracy were proposed too. At last, registration among sensor frame to the common was achieved based on a modified Iterative Closest Point (ICP) method. The final point cloud joined by each individual point cloud could represent the 3D shape of a large work piece. Object-oriented on-line calibration was developed based on ICP and Geometry Dimension and Tolerance (GDT) information to register the final point cloud and the CAD model.

Environmental reconstruction for navigation is a process to quickly acquire surrounding information around the vision sensor and presents a 3D fusion display for the operator. Traditional navigation system only employed a camera to get 2D information. Structured light system based on infrared light could quickly rebuild objects depths and fuse it into the displayed images without influence the operator’s normal vision. The sensed points on the objects are highlighted by the color-codes: from red to blue, which are used to indicate the object distances. In order to measure the moving objects, an one-shot surface coding algorithm was developed. Only one projection image is needed to acquire the 3D information. The codeword is determined in a single pattern because the code of each primitive depends on the values of the primitive and its neighbors.

In order to reconstruct the environment without blind areas, an omni-direction panoramic structured light sensing system was developed to increase FOV. Hyperbolic mirrors are put in front of a projector and a camera. 3D reconstruction model was build up associated with the hyperbolic mirror. At last, a 360 image fused with depth information is achieved by the designed system.

In summary, the study developed large scale structured light systems based on two applications: accurate inspection for quality control and fast environmental reconstruction for robot navigation.
# TABLE OF CONTENTS

## LIST OF TABLES

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>viii</td>
</tr>
</tbody>
</table>

## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ix</td>
</tr>
</tbody>
</table>

## 1 Introduction

1.1 Applications and Motivations ........................................... 1
1.2 Background and Literature Review ...................................... 3
1.2.1 Implementation of a Measurement System .......................... 3
1.2.2 Review of 3D Inspection Sensors Based on Optical Methods ...... 6
1.2.3 Sensor Calibration ..................................................... 8
1.2.4 Surface Coding via Structured Light ............................... 13
1.2.5 3D Data Registration .................................................. 17
1.2.6 Omni-directional 3D Reconstruction Vision Systems .......... 18
1.3 Difficulties and Challenges ............................................. 23
1.4 Objectives of the Work .................................................. 25
1.5 Organization of the Study .............................................. 26

## 2 Development of a 3D Sensor Based on Structured Light

2.1 Introduction of Imaging Devices ........................................ 29
2.1.1 A Pinhole Model of an Image Device .............................. 29
2.1.2 Homography Matrix from Single View Geometry .................. 31
2.1.3 Epipolar Geometry and Fundamental Matrix ....................... 32
2.1.4 Plane Induced Parallax in an Imaging System .................... 35
2.2 Optimize the 3D Sensor Model via Parallax Information ........... 36
2.2.1 Problem Statement ..................................................... 36
2.2.2 Recording the Reference Information with Homography Matrix .. 38
2.2.3 Correspondence between Measurements ............................ 40
2.2.4 Working Principle of the New Sensor Model ....................... 41
2.3 Chapter summary ......................................................... 45

## 3 Sensor calibration

3.1 Precision Analysis of the Given Sensing System .................... 46
3.1.1 Uncertainty Analysis of an Imaging Device ....................... 48
3.1.2 Simulation of the System Error Model ............................. 50
3.1.3 System Precision Test Results ...................................... 52
3.2 Sensor Calibration ....................................................... 53
3.2.1 Calibrate Homography Matrix between the Camera Image Plane and the Reference Plane ............................................. 53
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>The development of an Omni-direction 3D camera for mobile robot</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td>6.1 The Working Principle of the Omni-directional 3D Sensor</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td>6.2 Calibration of the Omni-directional 3D Camera</td>
<td>130</td>
</tr>
<tr>
<td></td>
<td>6.3 Error Analysis for the Designed 3D camera</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td>6.4 One-shot Surface Coding for the Sensor</td>
<td>133</td>
</tr>
<tr>
<td></td>
<td>6.4.1 Image warping in an omnidirectional camera</td>
<td>134</td>
</tr>
<tr>
<td></td>
<td>6.4.2 One-shot projection pattern design for the omnidirectional 3D camera</td>
<td>135</td>
</tr>
<tr>
<td></td>
<td>6.4.3 Epipolar constraints and the decoding method</td>
<td>136</td>
</tr>
<tr>
<td></td>
<td>6.5 Chapter Summary</td>
<td>137</td>
</tr>
<tr>
<td>7</td>
<td>Current Experimental Results and Data Analysis</td>
<td>146</td>
</tr>
<tr>
<td></td>
<td>7.1 The Rapid Inspection System (RIS)</td>
<td>146</td>
</tr>
<tr>
<td></td>
<td>7.1.1 The Configuration of the System</td>
<td>146</td>
</tr>
<tr>
<td></td>
<td>7.1.2 The Inspection Result via Single-shot Coding Algorithm</td>
<td>148</td>
</tr>
<tr>
<td></td>
<td>7.1.3 The Inspection Result via Multi-shot Coding Algorithm</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>7.1.4 The Registration Result from Four Pairs of Sensor Frames</td>
<td>152</td>
</tr>
<tr>
<td></td>
<td>7.1.5 Inspection for Different Work Pieces</td>
<td>154</td>
</tr>
<tr>
<td></td>
<td>7.2 The Navigation System for Single Direction</td>
<td>155</td>
</tr>
<tr>
<td></td>
<td>7.2.1 The Prototype of the Navigation System for Single Direction</td>
<td>156</td>
</tr>
<tr>
<td></td>
<td>7.2.2 The Fusion Result for the System</td>
<td>156</td>
</tr>
<tr>
<td></td>
<td>7.2.3 Integration the Single Direction Sensor with a Mobile Robot</td>
<td>158</td>
</tr>
<tr>
<td></td>
<td>7.3 The Omni-direction Navigation System</td>
<td>159</td>
</tr>
<tr>
<td></td>
<td>7.4 Chapter Summary</td>
<td>164</td>
</tr>
<tr>
<td>8</td>
<td>Conclusions</td>
<td>170</td>
</tr>
<tr>
<td></td>
<td>8.1 Conclusions</td>
<td>170</td>
</tr>
<tr>
<td></td>
<td><strong>BIBLIOGRAPHY</strong></td>
<td>173</td>
</tr>
</tbody>
</table>
LIST OF TABLES

3.1 The standard deviation results for all sensors . . . . . . . . . . . . . . . 52
4.1 Pixels binarization without projection mask . . . . . . . . . . . . . . . . 94
4.2 Pixels binarization without projection mask . . . . . . . . . . . . . . . . 97
7.1 The number of the point clouds in different experiments . . . . . . . 151
LIST OF FIGURES

2.1 The pinhole model of a camera. (For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation) ......................................................... 30

2.2 (a) The homography matrix between the image plane and the world plane. . 32

2.3 The epipolar geometry between two image planes. .............................. 33

2.4 The pencil of epipolar lines in each image plane on the epipole. ............. 35

2.5 A plane W induced parallax in camera image plane. .......................... 36

2.6 Traditional 3D reconstruction model. .............................................. 37

2.7 Recoding reference planar information. .......................................... 39

2.8 Correspondence between reference recording and work piece inspection. . 40

2.9 The optimized 3D model. .............................................................. 41

2.10 3D sensor model in the parallel configuration. .................................. 45

3.1 The difference between precision and accuracy. ............................... 47

3.2 (a) High accuracy with low precision; (b) low accuracy with high precision. . 48

3.3 Camera error model. ................................................................. 49

3.4 Projector error model. ............................................................... 49

3.5 Error analysis results: height errors when input error equals 0.1, 0.2, 0.3, 0.5 pixel. ................................................................. 51

3.6 (a) The illustration of homography matrix calculation; (b) The used chessboard. 54

3.7 The light array passing two pinhole models. .................................... 56

3.8 Multi-plane calibration strategy. ................................................... 59

3.9 The distribution of $p_c^r(h_1, u, v)$ for each layer. .......................... 60

3.10 The second order polynomial to approximate the residual function for a pixel 62
4.1 The pattern generation procedure. ........................................ 67
4.2 An example of occlusion. .................................................. 68
4.3 The comparison of different primitives. ............................... 68
4.4 The primitive value. ......................................................... 69
4.5 Image processing: (a) contour extraction; (b) location detection; and (c) principal axis identification. .................................. 71
4.6 Determination of the codeword for each primitive. .................. 71
4.7 Traditional encoding and decoding. ...................................... 74
4.8 The flow chart of the encoding and decoding strategy. .............. 76
4.9 The captured picture when projecting an image with intensity 100. . 80
4.10 The photo response curve for 5 points in the last figure. ........... 81
4.11 The proper projection values in the camera frame. .................. 82
4.12 (a) The homography mapping among the projector image plane, the camera image plane and the inspection table, (b) the correspondence pixel $P_P$ changes according to the $P_W$. ........................................ 83
4.13 a) The initial projection mask; (b) the dilated projection mask; (c) one example of a projection pattern. ................................. 84
4.14 (a) A received image of projected pattern; (b) a received image for the inverse pattern; (c) interpolation of the edge location; (d) estimation of the edge location in a strong illumination condition. ......................... 85
4.15 The final projection mask. .................................................. 87
4.16 The received image without the projection mask. .................... 89
4.17 The received image with the initial mask. ............................. 90
4.18 The received image with the final mask. ............................... 91
4.19 (a) A selected detail view of the black part when projecting pattern without the mask; (b) the same view when projecting the inverse pattern; (c) the intensity curve around point $B / B'$. ........................................ 92
4.20 (a) A selected detail view of the shiny part when projecting pattern without the mask; (b) the same view when projecting the inverse pattern; (c) the intensity curve around point $C / C'$. ........................................... 93

4.21 (a) A selected detail view of the black part when projecting pattern with the mask; (b) the same view when projecting the inverse pattern; (c) the intensity curve around point $B / B'$. ........................................... 95

4.22 (a) A selected detail view of the shiny part when projecting pattern with the mask; (b) the same view when projecting the inverse pattern; (c) the intensity curve around point $C / C'$. ........................................... 96

5.1 ICP algorithm .......................................................... 105

5.2 System registration gauge ........................................... 108

5.3 Sensor 1 measurement result ....................................... 108

5.4 The center of each marker ........................................... 109

5.5 The rough alignment between sensor 1 and CMM data ........ 110

5.6 The fine alignment between sensor 1 and CMM data .......... 111

5.7 System calibration results .......................................... 111

5.8 Space partition by the octree space method ..................... 113

5.9 The node tree of the octree method ............................... 113

5.10 Octree for CAD model .............................................. 114

5.11 Octree for the final point cloud ................................... 115

5.12 The difference between two algorithms ......................... 116

5.13 Check fixture and the stopper bar touching datum points of the work piece .................................................. 118

5.14 (a) CAD model with datum points. (b) the registration result based on datum points .................................................. 125

5.15 The color coded error distribution map for a piece of windshield ............................................................... 125

5.16 The fusion result for navigation application ..................... 126
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1</td>
<td>The working principle of the designed omni-direction system.</td>
<td>138</td>
</tr>
<tr>
<td>6.2</td>
<td>Calibration of the omni-directional camera.</td>
<td>139</td>
</tr>
<tr>
<td>6.3</td>
<td>Calibration of the omni-directional projector.</td>
<td>140</td>
</tr>
<tr>
<td>6.4</td>
<td>Received image from the world frame: (a) the received image; (b) Illustration of the optical path.</td>
<td>141</td>
</tr>
<tr>
<td>6.5</td>
<td>The object depth error from the resolution limitation.</td>
<td>142</td>
</tr>
<tr>
<td>6.6</td>
<td>The omnidirectional camera resolution: (a) the resolution curve in image frame; (b) the resolution curve in camera frame.</td>
<td>143</td>
</tr>
<tr>
<td>6.7</td>
<td>Image warp for a panoramic view.</td>
<td>144</td>
</tr>
<tr>
<td>6.8</td>
<td>The designed one-shot projection pattern.</td>
<td>144</td>
</tr>
<tr>
<td>6.9</td>
<td>The epipolar geometry of the omnidirectional system.</td>
<td>145</td>
</tr>
<tr>
<td>7.1</td>
<td>RIS system configuration.</td>
<td>147</td>
</tr>
<tr>
<td>7.2</td>
<td>The inspected pillar with surface pre-treatment.</td>
<td>149</td>
</tr>
<tr>
<td>7.3</td>
<td>The inspected point cloud of the pillar.</td>
<td>149</td>
</tr>
<tr>
<td>7.4</td>
<td>The 3D point clouds for the work pieces.</td>
<td>150</td>
</tr>
<tr>
<td>7.5</td>
<td>The comparison between CMM data and the RIS data.</td>
<td>152</td>
</tr>
<tr>
<td>7.6</td>
<td>The CMM error map.</td>
<td>154</td>
</tr>
<tr>
<td>7.7</td>
<td>The first tested work piece with bright color and more complex shape.</td>
<td>155</td>
</tr>
<tr>
<td>7.8</td>
<td>The corresponding CAD model for the first work piece.</td>
<td>156</td>
</tr>
<tr>
<td>7.9</td>
<td>The point cloud for the first work piece.</td>
<td>157</td>
</tr>
<tr>
<td>7.10</td>
<td>The error map for the first work piece.</td>
<td>158</td>
</tr>
<tr>
<td>7.11</td>
<td>The configuration of the navigation system for single direction.</td>
<td>159</td>
</tr>
<tr>
<td>7.12</td>
<td>The tested objects.</td>
<td>160</td>
</tr>
<tr>
<td>7.13</td>
<td>The captured image when projecting near infrared light pattern.</td>
<td>161</td>
</tr>
</tbody>
</table>
7.14 The fusion result for the sensor. ................................. 162
7.15 The integrated mobile robot with the designed sensor. ............ 163
7.16 The description of the integrated system. ................................ 164
7.17 The result of designed system. ........................................ 166
7.18 The developed omnidirectional 3D sensor. ............................. 167
7.19 Examples of the projected pattern on the left corn of the room. .... 167
7.20 The received image by the omnidirectional camera. .................. 168
7.21 The fused result with the object depth. ................................. 169
7.22 The unwarped fusion result. ............................................. 169
CHAPTER 1

Introduction

1.1 Applications and Motivations

The development of rapid and accurate measurement systems is an active research area. Among current 3D measurement techniques, structured light based 3D vision has made significant strides and has been utilized in many applications, such as manufacturing quality control, 3D profile reconstruction, reverse engineering and environmental reconstruction. According to the system performances, the applications can be roughly separated into two branches: accurate inspection and fast reconstruction.

Inspections for industry applications usually emphasize the result accuracy. A structured light based inspection system acquires the reflection lights from the test work pieces and rebuilds the 3D results by a set of points, named point cloud. The point cloud is further compared with Computer Aided Design (CAD) model to find manufacturing distribution error in the application of quality control. The color-code error distribution map is evaluated by the acceptable tolerance, work pieces exceeding the tolerance are considered as unqualified.

Fast reconstructions, such as robot navigation, usually consider the acquisition speed as the most important parameters. Its accuracy requirement is relative lower than inspection’s. However, the navigation application is very sensitive with the measurement speed. The
The system should be able to quickly rebuild a moving object. The developed system employs the infrared light to sense the objects and uses one-shot coding method to quickly rebuild the scene 3D information. The reconstruction result is further fused with the camera captured image for the operator to get 3D information.

Currently there are still some limitations preventing the further applications of the technology.

First, a large scale of the sensing system for manufacturing industry is needed to inspect big work pieces. Traditional inspection machines, such as CMM, are very sensitive with the inspection scale. The price could increase dramatically once the scale increases. Such an imaging system with large Field Of View(FOV), long standoff distance, random image noise from environment, and registration between different sensors coordinates greatly arises the difficult of the system implementation and calibration.

Second, vision based measurement system always suffers from the surface optical properties. Work pieces with various properties, such as bright, white, dark, black parts, and especially the surfaces exhibiting a mixture between specular and defused reflection, greatly influence the vision system. Manufacturing industries need to pre-treat the surface into a diffused reflective one, and then utilize the optic system to measure the work pieces, which increases the cost and the processing time. Therefore, a system capable against optic challenging objects is needed for industry application.

Third, there lacks of a quickly on-line navigation system for mobile robot. Since the mobile base is moving all the time, the system should reconstruct the environment based on only one projection. Traditional structured light systems project multi-image patterns could fail to this application. Furthermore, the results should also simultaneously satisfy the efficacy, reliability, high density, and accuracy.

At last, an omni-direction 3D navigation sensor is needed for the operator to get 360 degree environmental information. Without the blind area of the vision system, the operator could control the mobile robot and view the surroundings more efficiently. The structured
light system could simultaneously compensate the mirror distortion, sense 360 degree, re-
construct objects in the environment and fuse the 3D information into the 2D image.

Aiming to these demands, this study is trying to develop proper methodology and reliable
machine vision based systems for those applications.

1.2 Background and Literature Review

1.2.1 Implementation of a Measurement System

The development of a structured light based measurement system can be described as the
following parts: 3D recovery model for a traditional active vision system, calibration of the
system sensor, encoding and decoding for the structured light, registration between different
work frames and omni-direction system.

The imaging sensing device, a camera for example, is usually simulated as a pin-hole
model. Objects in the 3D world frames maps into the image frame by a $3 \times 4$ perspective
matrix. Only 2D information is preserved through the mapping, and the limitation of the
last degree of freedom is missing. The recovery from an imaging pixel forms a linear array
instead of a point. In order to accomplish the 3D reconstruction task, additional constraint is
needed to fix last degree of freedom. The 3D recovery model, as the starting point to design
a structured light, build up the work principle of the system sensors. Aided from another
devices, such as a laser, a projector, or another camera, the third constraint equation is able
to be build up. The 3D recovery model will control the following implementation steps of
the system, such as configuration, calibration, and so on. An simple, precise, and easy to be
implemented 3D recovery model plays an important role in the system design.

After the 3D recovery model is fixed, the next step is to find proper way to calibrate the
system. A structured light system usually consists of a projector and a camera. Calibration
tasks usually involve both devices’ the intrinsic and extrinsic parameters. The intrinsic
parameters controls the mapping from its image frame to its own sensor frame, and the
extrinsic parameters transfer the sensor frame into the world frame. Together there are
totally 22 parameters for both devices. If the 3D model is properly designed, less parameters
could be calibrated.

Structured light system comes from stereo vision system by replacing one camera with
a projector. Stereo vision system suffers from correspondence problem: only corresponded
pixels in both imaging devices could build up the 3D points. Without feature aided, stereo
vision system could not establish the correspondence. Therefore, only a few points having
clear features could be recovered. Structured light system replaces one camera with a projec-
tor. The projected pattern is properly design to encode certain a codeword into the incident
light. The camera records the reflection light and extracts the codeword. Since the incident
light and the reflection light preserves the codeword, pixels sharing the same codeword in
both imaging devices are corresponded. A proper designed projection pattern could gener-
ate the correspondence without the influence from the textures and geometries of the test
objects. Therefore, a robust encoding and decoding strategy is critical for a structured light
system. According to the number of projections, the encoding and decoding methods can be
separated into two categories: one-projection based method and multi-projection method.
One-shot projection method’s acquisition time is much shorter than multi-projection method,
and therefore can be used to measure moving objects. However, its result point density and
its accuracy is less than multi-projection method. Thus, one-shot projection method is usu-
ally used for 3D reconstruction. Multi-projection method is mainly utilized to inspect stable
work pieces. With multiple projections, the method could inspect optically-challenging ob-
jects without pre-treatment, such as objets with different color, different reflection ratios, or
material exhibiting a mix between diffuse reflection and optic reflection. Two encoding and
decoding methods are introduced separated in the study.

After 3D point cloud is recovered in each individual frame, usually there are some post
processing tasks to obtain the final results. If the system contains more than one pairs of
sensors, registration between each sensor frame and a common frame is needed to merge all
small point clouds into the final one. For quality control application, the final point cloud needs to compare itself with CAD model, so addition registration between the common frame and the CAD frame is also needed. For navigation application, the final point cloud should be fused into the 2D image display. The third dimension, object depth: distance from objects to the camera, is fused into the image via color codes. Distance from far to near is assigned by different colors, from blue to red. Therefore, registration from 3D world frame to the image frame is needed to finish the fusion.

Omni-direction systems usually has different work principles with the traditional structured light systems, because of the additional convex shape mirror in front of the cameras and the projectors. Its literature view is introduced in a separated section in this chapter.

From the view of applications, structured light systems can be roughly separated into two branches: accurate inspection and the fast environment reconstruction. From the of development, the difference between two applications lies in the encoding and decoding strategy. Accurate inspection usually employees multi-shot projection to acquire test objects information to obtain a high dense and accurate point cloud. Fast environment reconstruction on a mobile robot usually adopts single-shot projection to record surround information to increase sample speed.

Once the applications’ scale increases, especially for the FOV of the system, both branches also utilize different methods to enlarge detection range. In order to increase of FOV of the system, accuracy inspection puts more multiple sensors to cover a big work piece. Traditional methods, such as enlarging the standoff distance, could also increase the FOV of the system. However, the resolution area of an imaging device, the area one camera / projector pixel covers in the world frame, also increases. The deviation within the resolution area could not be detected by traditional methods, and thus the results accuracy is reduced. Therefore, multi-sensor strategy could enlarge the FOV and maintain high accuracy at mean time.

For the application of navigation, there is limited space on the mobile robot. So it is very difficult to put multiple sensors on the robot. In stead of multiple sensors, convex shape
mirrors are put in front of the projector and the camera to enlarge the FOV. The method could also detect 360 degree of view without blind area in the limited space. Since the accuracy of navigation is much small than the quality inspection, the error within resolution area can be tolerated.

In the following parts of this chapter, the literature is mainly follow the development steps of a structured light system: 3D recovery model, system calibration, encoding and decoding for the structured light, registration between different work frames and at last the omni-direction system. Special attentions are given to two major applications: quality inspection and environment reconstruction.

1.2.2 Review of 3D Inspection Sensors Based on Optical Methods

Quality inspection is a process of finding the deviation between the work pieces and CAD model by a set of specifications [1]. The whole process can be divided into two parts: the measurement of the work pieces and the generation of the error distribution map.

Basically, existing measurement techniques can be separated into to categories: contact based and non-contact based. A CMM is a traditional system commonly used for dimension inspection based on contact method. A touching probe is employed to contact the work piece surface and obtain the coordinates of the points. Through the accuracy of the a CMM is very high, the nature of point-wise contact based method limits both CMM’s measurement speed and the capability of inspecting complicated work pieces. Non-contact measurement via machine vision [2], with the merits of fast speed, high inspection sample rate, and overall simplicity of operation, becomes a popular alternative inspection method for CMM. To many applications in the manufacturing industry, inspection of a large part requires multiple 3D vision sensors to cove the whole piece, in which additional registration between each sensor frame to the common frame is needed to complete the final point cloud. At last the final measurement result, the point cloud representing the test work piece, is compared with the Computer Aided Design (CAD) model to find the error distribution map. The color-coded
error distribution map will be further evaluated with the acceptable tolerance by requirement. Only the work pieces within the manufacture tolerance can be delivered to the customers. Furthermore, a machine vision system is also widely utilized in different industry fields [3] [4].

The problem of retrieving 3D information form an object by vision method starts from the basic imaging device: a camera. A camera, which is usually simulated by a pinhole model, presents a perspective mapping from 3D world frame into the image frame. Each pixel in the image frame recovers a light array passing the optic center instead of a point in the world frame. In order to achieve the 3D recovery, additional constraints are needed to limit the last degree of freedom along the light array. Stereo vision [5] was developed in 3D recovery based on two cameras viewing the same object from two different view points. The correspondence pixels from two cameras build two light arrays by pinhole approximation. The intersection point between two corresponding light array recovers the 3D coordinates in the world frame. Sometimes, the middle point of the common perpendicular of the two corresponding light arrays is utilized as the 3D reconstruction point due to the calibration error.

Stereo vision is also called as passive vision, since none of the sensors emit any kind of radiation but instead rely on detecting the radiation reflection by the objects. Despite it was widely used in computer vision, it presents several drawbacks in 3D reconstruction. The first one is the correspondence problem. In other words, determining relationships from pixels of different views is not a trivial step. The second one is resolution of 3D points. The technique can only reconstruct points with unique features. Therefore, dense reconstructions is difficult to obtain. In order to overcome both drawbacks, two solutions are developed when dense reconstructions are required. The first one is adding an active imaging device, such as a laser or a projector, projecting a group of pre-defined image patterns, which is also called structured light. The codeword carried by each pixel from the projector is designed to be unique. Therefore the correspondence in both cameras is able to established. The limitations of the this method lies in: a) the Field Of View (FOV) is the overlap of three
imaging devices, b) the Depth Of View (DOV) is minimum of three image devices, and c) additional device increases the cost of the system.

Facing the limitation of the first solution, the second one is proposed by simply replacing one camera by a projector in the stereo vision. The projector provides both correspondence and the incident light array equation. Treating the projector as a pseudo camera, 3D reconstruction can also achieved by the intersection between corresponding light arrays. Since projector emits lights, the system is also called structured light based active vision system. In this study, 4 pairs of structured light based active vision sensors are employed to build a large scale measurement system as a prototype for an industry quality control system.

1.2.3 Sensor Calibration

Calibration of the 3D sensor is crucial for a measurement system to get accurate results. Since a 3D sensor consists of a camera and a projector, its calibration can roughly separated into two categories: camera calibration and projector calibration. The pinhole model approximation of each image device adopts intrinsic and extrinsic parameters to construct the light array equation from a given pixel. Intrinsic parameters transfer the image pixel into a linear light array in the camera frame. Extrinsic parameters transfer the camera frame into the world frame on the inspection table. Conventional calibration methods calculate the intrinsic and extrinsic parameters of a camera or a projector.

Camera calibration has been studied for decades, With the aid from the correspondence between a calibration gauge and pixel locations in the image, parameters of a camera can be regressed. In the calibration process, the calibration gauge plays an important characteristic function. A large scale 3-dimensional gauge with high accuracy usually costs a lot. Based on the dimension of the gauge, calibration techniques can be roughly classified into the follow categories:

1) 3D reference object based calibration. Camera calibration is performed by observing the gauge whose geometry in 3-D space is known with accurate precision [6]. This kind
of approaches usually require an expensive calibration apparatus, because it costs a lot to manufacture an gauge with accurate 3D coordinates, especially for a large scale one.

2) 2D plane based calibration which is also the most common used method. In order to reduce the cost of the gauge, calibration algorithm is optimized to obtain the camera intrinsic and extrinsic parameters on planar objects. Tsai [7] adopted a planar gauge to obtain camera parameters via two steps. In the initial step, constraints between parameters, such as a rotation matrix is orthonormal, are not enforced, but a quantity that simplifies the analysis and leads the tasks to linear equations via Radial Alignment Constraint (RAC). Solution of the linear equations is calculated by least square method and is utilized as the starting values for the final optimization in step two. In the second step, the rest of parameters are obtained via non linear optimization Levenberg-Marquardt Algorithm (LMA) [8] that finds the best fit between the observed image points and the predicted points by the target model.

Tsai’s method is known as two limitations: a) the solution procedure, however, fails if the camera plane is nearly parallel with the calibration board. Tsai et al. [9] also devised an algorithm that handled the exactly parallel case. Years later, Zhuang and Wu [10] extended Tsai’s method to calibrate a camera in a near parallel configuration when camera intrinsic parameters were known. Luo et al. [11] proposed another algorithm to calibrate a camera when intrinsic parameters were unknown. b) The second limitation is that the translation between the calibration planes is needed to regress camera parameters, which results an elaborate setup. Different from Tsai’s method, Zhang [12] proposed a new planar based calibration method, in which the knowledge of the plane motion is not necessary. The orthogonal constraints from the rotation matrix were utilized as the calibration conditions. The homography matrix between image plane and calibration plane was factorized into two parts: intrinsic parameter and a matrix consisting of two columns from rotation and translation vector. An analytical solution can be calculated as initial condition and numerical solution can be found iteratively through LMA.

3) 1D calibration. In order to further reduce the requirement of the calibration gauge,
camera calibration based on 1D gauge is proposed. The calibration results can reach fair
level of accuracy. Zhang [13] first proposed to calibrate a camera with a 1D object. The
solution can be found if one point is fixed. A closed-formed solution is developed if six
or more observations of such 1D object are made. Qi et al. [14] extended Zhang’s 1D
calibration method, and calibrate a camera with aid from a gravity driven 1D object.

4) Another groups calibration are called semi-calibration or self-calibration. Such kind
of calibration usually assumed some parts of the parameters are known, and can calibration
the rest parameters without any requirement of any reference gauge. Therefore, sometimes
this kind of calibration methods are also called 0D calibration. Semi-calibration usually
adopts epipolar geometry constraints of the same object from two views. Essential matrix
was first mentioned by H. C. Longuet-Higgins [15]. If two camera’s intrinsic parameters are
known and normalized, the extrinsic parameters can be solved up to 4 analytical solutions
by SVD method. The essential matrix can be determined from eight-point algorithm [17].

If the transformation between two positions is known, a camera’s intrinsic parameter can
be estimated by fundamental matrix from epipolar constraints [16] [17].

The history of projector calibration is much shorter than camera calibration. After
structured light based 3D reconstruction model was developed, people began to investigate
different methods on projection calibration. Compared with a camera, the calibration of
a projector is always more difficult and the process is more complicated, since it is not
a information receiving device. The dimensions of the gauge can not be directly applied
to the calibration process and the projected pattern needs to employee another imaging
receiving device, such as a camera, to acquire. Therefore, the projector calibration error is
usually bigger than the camera’s, and becomes the major source of errors in a structured
light system. Drarei et al. [18] treated a projector as a pseudo camera with the aid from a
 calibrated camera and then utilized camera 2D calibration method to calibrate the projector.
The homography matrix between a projector pixel location and 3D points regressed from the
camera is used to regress the projector parameters. However, this kind of methods brings
the camera calibration error into the projector calibration, which double the final error. Zhang et al. [19] first adopted structured light to set up a link table between a projector and a camera, and then applied traditional 2D camera calibration methods to regress projector parameters. The 3D coordinates of the points on the gauge is not regressed from the camera, so the camera calibration error will not participate into the projector calibration. However, the calibration steps are very complicated. The planar gauge was placed around 20 different positions and 3D points on each position were recorded by horizontal and vertical Gray Code and Phase Shifting (GCPS) structured light. Since GCPS is not robust against image noise, the miss-linked location by GCPS also costs calibration error. Eipoplar constraints describes the internal geometric relationship between two views and be welled used in self-calibration. Homography matrix which is used to calibration a camera by 2D gauge also receives a big success in a camera calibration. Combining these two, epipolar constraints and homogrpahy matrix, Zhang et al. [20] calibrated the relative rotation and translation between a camera and a projector. But the algorithm requests the intrinsic parameters need to be known first. In order to reduce the error from the projector calibration, Shi [21] did not regress the pinhole model parameters of the projector but created a link table between camera light array and projector light array with respect to each level of height within the whole work space. The results are accurate, but the method is limited by the size of link table, which means it can not be used for large scale system.

Convectional calibration methods of a structured light system calculate both equipments intrinsic and extrinsic parameters, leaves a challenge task to calibrate a projector. Although projector calibration can be done via a complex way, the solution either contains too much error or time consuming. This thesis brings the camera planar calibration method and epipolar information into 3D reconstruction model to reduce the parameters in 3D model. The novel strategy is that the camera calibration information based on planar method can be saved as the reference information. The acquisition information of the test object is used to compare with the pre-saved reference information. The corresponding pixel deviation
between reference and a test object in camera image is applied to reconstruct the 3D point cloud, in which projector’s intrinsic and orientation parameters is avoided in calculation. Therefore, the projector calibration is not necessary for 3D reconstruction anymore. The correspondence function provided by the structured light is not only used to set up correspondence between a camera and a projector, but also establishes corresponding pair of pixels in a camera between two measurements: the camera pixel recording reference information and the corresponding camera pixel recording tested surface information.

Error analysis is another important topic in the calibration of a given 3D surface profile measurement system. As early as 1980s, when the active vision system was invented, the error analysis problems have been pointed out. People did various kinds of error analysis to evaluate the system from different aspects and error sources. The error model developed by Trobina et al. [22] for an active range sensor was mainly concentrated on the errors from the camera perspective geometry. Yang et al. [23] proposed a more detailed mathematic model for each calibrated parameter in the 3D reconstruction. A relative error model [24] was presented for calibration of an active triangulation sensing model. Uncertainties analysis and their effects in 2D camera calibration [25] have also been discussed. In summary, people mainly concentrated on the error analysis related with device parameters, sensor model, and calibration methods in the past. As the scale of a 3D profile sensing system increases, the consistence of the system cost by the random image noise also needs to be evaluated. In practical applications, especially for industry applications, consistence decides whether the measurement results are reliably or not. When the FOV and standoff distance of the system increase to a scale that consistence error is no longer small enough to ignore, consistence analysis and test need to be done before the calibration starts and a poor consistence test result requires the redesign of the system. In this study, the consistence performance of the design system is analyzed, simulated, and tested. Tradeoff values are chosen for the standoff distance and inspection area by considering consistence error.
1.2.4 Surface Coding via Structured Light

The encoding and decoding procedure of the structured light is one of the most critical factors for an optic measurement system. The biggest challenge is the capability of simultaneously satisfying the efficiency, reliability, and accuracy requirements. The projected patterns are deformed by the free-formed test surface. Therefore, the codeword is designed to be against this distortion and at mean time the encoded pattern needs to carry as many codewords as possible. The corresponding decoding on-line image calibration algorithm should dynamically calibrate the received image against the distortion and image noise to extract the codeword. Since these two procedures are always be coupling with each other, the literature review of both functions are explained simultaneously.

Several pattern strategies have been proposed to make a fast 3D inspection using a single or a few patterns, which can be categorized into two types in terms of test objects: one-shot pattern for moving objects and multi-pattern for static objects.

The moving objects can only be inspected by one-shot pattern algorithm, otherwise, the system may acquire incorrect pattern code due to the movements of the parts. The major challenge for a one-shot pattern is the capability of compressing enough codewords. People employed different strategies for it, such as color and spatial-neighborhood patterns. In the color method, the codeword is determined in a single pattern because the color of different primitives, such as dots, lines, grids, are different. The decoding calibration can separate each codeword via color filter. Spatial-neighborhood can be encoded because the code of each primitive (symbol) depends on the value of the primitive and its neighbors. Usually both strategies are combined together to create a one-shot pattern. Pattern with round disks, small color dots on the black background, was developed [26] [27]. Decoding function first separated each image primitive, and employed color information to assign unique codeword for primitive. The codeword density can be increased via strips design [28] [29] or grids design [30] [31]. The drawback of color strategy is that the color contrast ratio can be affected by the color reflectivity, inherent texture of the work piece, and ambient lights,
which results in less reliability than monochromatic light, such as black-white light.

In this study, a new structured light pattern based on the spatial-neighborhood strategy was proposed [32]. The primitive of the pattern is based on the corner of the black-white checkerboard pattern, which is also called the X-point. The X-point provides more robust and accurate position compared with a conventional pattern in which the primitive is based on symmetrical symbols such as disk, strip, etc. The orientation of the X-point is used to encode the primitive of the pattern. Moreover, the point cloud density in different regions of the inspected part can be adjusted. This strategy is very useful for applications, in which high point density is only required for some critical regions such as connection interfaces.

Multi-shot strategy is only available for stable objects but it is more robust against image noise than one-shot method. After detecting the illumination problems, such as shadow problem, under-/over-exposure problem and internal reflection problem, the projection pattern could be readjusted by a proper feedback. The recent research is mainly focused on how to improve the inspection accuracy and point cloud density. Among many strategies, Gray Code [33] combined with Phase Shifting, such as sinusoid wave [34], received a good result and was utilized in many areas. After successfully inspecting diffused based material via multi-shot method, people started to apply the multi-shot method to detect optical-challenging objects without pretreatment. As a pioneer, Nayar [35] [36] built upon a shape-from-intensity method and iteratively approximate the 3D shape of the test objects. But the method is only to model relatively simple geometries instead of inspections. Later, people began to investigate more robust methods against optically-challenging objects. Literature shows that there are approximately four categories that people applied to increase the measurement performances: directly coding method, color invariance based method, polarization based method, and adaptive projection method.

In directly coding methods, people tried to develop encoding patterns which were robust against image noises. Instead of phase shifting method, digitized Gray Code Line Shifting (GCLS) [37] [22] attempted binary square wave as the projection patterns. Binary wave
achieved better performance than sinusoid waves. Different from phase shifting method, decoding process was changed from analog sample into digitized sample. Codeword 1 is encoded as intensity 255 and 0 is encoded as intensity 0 in the projector frame. High image intensity is decoded back as '1', and low image intensity is considered as '0' in the camera frame.

However digitized directly coding method is still weak against shiny material. The un-illuminated area can also result in high image intensity in the camera frame due to the internal reflection. Low reflection coefficient could also cause low image intensity in camera frame even high illumination pattern is projected. Combined binary pattern with color information [38], some shiny objects were successfully measured. The authors hypothesized that the effects of internal reflection and ambient light in the scanning process could be eliminated via color invariants. The inter-reflection is not preserved under bas-relief (GBR) transformation. However, the method could not inspect the black material which could absorbs all incident colors. Furthermore, under-/over-exposure problem was not mentioned in the study. Therefore, the capability of the method is limited.

Based on the fact that specular reflection changed the polarization direction of the incident light, polarizers were added in front of the cameras/projectors to filter out the internal reflection light [39][40]. Therefore, shiny objects could be measured in their systems. However, without the pre-knowledge of test material optic properties, surface normal, and exact number of internal reflections, authors had to manually and carefully adjust the polarizers to a proper angle to filter out the internal reflection light. Once the position of the part changes, or a new part was put into the system, polarizers had to be re-adjusted [41]. Therefore, the system is not very flexible.

Adaptive coding methods could dynamically modify projection illumination based on the scene, which can avoid the over-under exposure problem and be further used to eliminate the internal reflection issues. The method is more robust than the polarization based method, since no modification is needed if a new part is inspected. However, the biggest challeng-
ing is the correspondence problem between the camera and the projector. The projector illumination needs to be adjusted but the receiver is the camera not the projector. The correspondence, which is usually called crosstalk function, is needed to create the feedback from the camera to the projector. After feedback information is obtained, the illumination of the structured light could be dynamically adjusted to measure optically-challenging objects.

However, the crosstalk function is a dilemma for a structured light system: structure light is used to find out correspondence, but the correspondence is needed to create a scene based structured light. Koninckx et al. [42] proposed a scene adapted structured light system to avoid under- / over- exposure problems. Not only the reflection coefficient by scene is considered, but also the nonlinear responses for both imaging devices are simulated. However, the approximate knowledge of the geometry has to be pre-known to solve the dilemma. Without the foreknowledge, the method is relatively lack of flexibility for different work pieces. P. Sen [43] estimated crosstalk function by turning on each projector pixel each time, and also applied an adaptive subdivision to speed up the algorithm. However, the authors only use the method to dynamically control the illumination condition but not to reconstruct the given scene. Furthermore the number of the projecting and recoding loops is too many. Therefore, the efficiency is very low. An adaptive corresponding algorithm was developed [44] [45] by Xu et al. to inspect objects with strong internal reflection. The authors classify the captured areas as certain areas and uncertain areas. They iteratively peel out the certain areas and enhance the illumination condition on the uncertain regions. Objects with strong internal reflection can be measured in their system without any pre-knowledge of the test geometry. While the limitation is that its iteration numbers is not fixed. It may cost hours to inspect a scene [44] [45] and thus the algorithm efficiency is relatively low.

In summary, a robust, flexible, and efficient inspection system for the optically-challenging objects is needed for quality inspection in industrial field. In this study, a robust encoding and decoding algorithm to measure optically-challenging objects is proposed. Monochromatic WB stripes are applied to avoid the influence from test objects’ colors. A scene based
projection mask is utilized to adjust the projection illumination so that the under- / over-exposure problem and shadow problem can be eliminated. Combined with the projection mask, an extrapolation model is developed to estimate the edges between high and low illuminated regions. Therefore, shiny objects could also be inspected in a structured light system.

### 1.2.5 3D Data Registration

Registration of point clouds is to find Euclidean motions among a set of point clouds or, between a point cloud and the model set of a given part, respect to a common reference frame. For a large scale inspection system, two categories registrations are needed:

1) Registration of multiple point clouds into one common frame.

2) Registration between the final point cloud and the CAD model.

Registration problem can be roughly considered to be two parts: coarse alignment and fine registration. Coarse alignment determines an initial condition and fine registration estimates a best solution to be close to the truth as much as possible. Without coarse alignment, fine registration is time consuming and often fails to find a solution. Methods for coarse alignment include principle component analysis (PCA) [46], point signature [47], and image transformation [48], etc.

One of the most famous fine registration methods is the Iterative Closest Point (ICP) algorithm [49] [50]. ICP is based on the search results of pairs of nearest points in the two data sets to estimate a coordinate transformation. This iterative process executes until result reaches the pre-defined threshold. Many improvements and modifications of the basic ICP algorithm have been developed: Johnson et al. [51] adopted color information combined with Euclidean distance to set up correspondence in order to improve the registration results. Instead of searching the distance between corresponding points, Pottmann used the distance between the surfaces as a registration evaluation standard [52]. Gregory investigated the Euclidean invariant features in generalization ICP in order to decrease the probability of
being trapped in a local minimum [53]. Moreover, ICP method needs a lot of processing time to find the corresponding closest points. In order to deal with this problem, people tried to increase the speed of the algorithm. Timothee used multi-resolution way to speed up the ICP searching process [54].

ICP family is a good algorithm in fine registration, and the merit of this algorithm is that the total linear least square error will be minimized after the final alignment. In this study, ICP based method is utilized to solve the registration of multiple point clouds into one common frame. However, this merit will change into a drawback in error map registration between final point cloud and CAD. The error between work piece and CAD model should not be minimized by registration process. The reduced error map can not reflect the correct manufacture error, therefore, an object-oriented registration method based on ICP is proposed to solve the problem.

1.2.6 Omni-directional 3D Reconstruction Vision Systems

The capability of vision based control is highly related with information provided to the operator / computer. An omnidirectional 3D camera is a system that provides rich 3D information with a wide rang viewing and thus becomes a valuable choice for such applications. An omnidirectional camera offers a 360 Field Of View (FOV) and a complete information about the environment surrounding the vision sensor. The corresponding realtime 3D perception algorithm could further extend the capability of the omnidirectional camera to measure the depth information. Applications, such as robot navigation, remote manipulation, auto-driving system, map reconstruction, and pedestrian detection, demand this kind of vision sensor to quickly and accurately collect necessary information. With the aid of an omnidirectional 3D camera, the navigated robot could be easily avoid collisions; the remote controlled robot arm could freely pick and place parts all around; map reconstruction for an unknown scene could be achieved within one pass, and the pedestrians approaching from the rear could also be detected. Aiming on this demand, we develop a low cost structured light
based omnidirectional 3D camera, consisting of a projector, a camera, and two hyperbolic mirrors. A panoramic view is provided around the sensor without any blind areas, and the depths from the sensor to the objects are measured within one-shot projection.

Basically there were two ways to obtain omnidirectional images [55]. The first approach was to generate the omnidirectional images through the image mosaic method. The method either employed multiple cameras viewing different directions or utilized one rotating camera [56] to obtain multiple images of the environment. The final omni-directional image was the merged result using the mosaicing method. 3D reconstruction by such methods was usually achieved by rotating a stereo vision system consisting of two cameras [57] [58]. In each rotation angle, the system was considered as a normal stereo vision system and acquired the 3D coordinates. An omni-directional 3D rebuild was achieved by scanning the environment a full 360 degrees. For such a system, it could not build all the environment at same time and the mechanic rotation could also influence the accuracy of the result.

The second approach is to use an omni-direction camera [59] [60] [61] to obtain panoramic vision. An omni-directional camera usually consists of a convex shape mirror, such as a hyperbolic, parabolic, conic, or sphere mirror, and a normal camera. By viewing the back image of the mirror, a panoramic vision can be obtained. The math models [62] [63] and the configurations [64] have been discussed a lot in the past.

An omnistereo vision system could be formed to obtain 3D coordinates by employing two omni-direction cameras. Via vertically-aligning two omnidirectional cameras [65], an omnistereo system was successfully designed and was implemented. A 3D sample point was rebuild by the intersection of two corresponding light arrays from two calibrated omnidirectional cameras. Monocular vision could also rebuilt 3D information but less commonly compared with stereo vision method. In order to reconstruct an environment map, an omnidirectional camera was put onto a mobile robot [66] [67]. Via the estimation of the robot motion, the feature points in the scene could be reconstructed. However, both systems usually suffered from the correspondence problem and both measurement accuracy and result
density were limited by this correspondence issue.

Facing on the correspondence problem, a structured light based omnidirectional vision sensor was developed [68] [69]. The authors put a laser facing to a conic mirror and then formed their structured light emitter. An omni-direction camera and such a light source were put vertically to form their omni-direction structured light based 3D sensor. The laser was split into a light ring by the conic mirror and the light ring was emitted to the environment. The omnidirectional camera captured the light ring and the radius of the light ring was utilized to calculate the depth information. Since there was only one light ring in the image, correspondence problem did not exist in their system. The scene points illuminated by the light ring cloud be detected in realtime. However, the system could only scan a thin ring from the scene and limited 3D information was provided. In order to sample more points from the scene, Ukida et al. [70] developed a structured light aided omnistereo vision sensor. Three projectors shooting in 3 different directions were added to an omnistereo vision system. With the aid from the structured light, a more accurate and more dense 3D reconstruction result was achieved. However, 3 projectors did not cover the whole FOV of the omnidirectional camera, several blind areas were failed to obtain 3D information.

In short, there is a demand to design an omnidirectional 3D sensing system capable to fully measure a 360 degree surrounding environment with a high dense sample result. In this paper, a structured light based omnidirectional 3D camera, consisting of a projector, a camera, and two hyperbolic mirrors, was developed. The hyperbolic mirror and the projector formed an omnidirectional projector. The omnidirectional projector and the omnidirectional camera were aligned vertically and cloud sample dense 3D points around all 360 degree surroundings. The system can measure the environment without any blind areas.

Besides the high density 3D sample results, the measurement speed is also very important. A lot of applications, such as navigation, remote manipulation, pedestrian detections and et al., did not only require measuring the environment but also needed a realtime sampling method. Based on the number of projection images, structured light systems could
be roughly separated into two categories: one-shot projection system and multi-shot projection system. Once the imaging sensor was moving or a moving object was approaching, multi-shot method failed to handle such cases.

The work [68] [69] that Gluckman et al. developed could measure the scene in realtime, but its point density was very limited. The system [70] that Ukida et al. developed applied several multi-projection methods to set up correspondence. The measurement could not be finished in real time. In order to sample dense 3D points in realtime, a robust one-shot surface coding method was needed. In the past, people only developed different kinds of one-shot coding methods for a normal structured light system. Kinect system developed by Microsoft could detect 3D in realtime and obtained dense samples, but such a system’s view angle was so limited compared with omnidirectional system. Much less information was presented to the operator compared with an omnidirectional 3D camera. Color based and spatial-neighborhood based coding methods were two most common one-shot methods for a conventional structured light system. In the color method, the codeword was determined in a single pattern because the color of different primitives were different. The drawback of the color strategy was that the color contrast ratio could be affected by the background colors, which resulted in less reliability than with monochromatic light, such as black-white light.

Spatial-neighborhood [71] using monochromatic light was more robust than the color based method. The codeword of each primitive (symbol) depended on the value of the primitive and its neighbors. Each codeword was unique from the others so that correspondence could be established. However, this method was not suitable for an omnidirectional sensor. The geometric relationship could be destroyed by the convex shaped mirrors. The projection pattern was first warped by the hyperbolic mirror in front of the projector and then was further warped by the hyperbolic mirror in front of the camera. Without knowing two warping functions, the spatial-neighborhood method could not directly applied to the omnidirectional 3D sensor. Therefore, existing solutions for one-shot surface coding were not suitable for an omnidirectional 3D system.
Besides the structured light systems, ladar system [72] could also get realtime 3D results from the scene. However, such a system could only sample limited number of 3D samples. Furthermore, the cost of the system is also much more expensive than our system.

In order to meet this demand, a one-shot surface coding for a structured light based omnidirectional 3D system was developed to measure the environment in realtime. Light rings embedded with different angular frequencies and intensities were emitted from the omnidirectional projector to the environment through the hyperbolic mirror. Taking advantage of the epipolar constraints, the invariance of the phase angle and angular frequency were utilized to establish the correspondence and then were trangulized to obtain the depth information. Compared with traditional one-shot method, the contribution was the one-shot design for an omnidirectional projector and the challenge was the pattern design against both mirror distortion and scene distortion. The codeword should be designed based on the invariance from scene distortion and the two warps that were affected by the hyperbolic mirrors were also needed to develop the pattern.

In order to obtain an accurate 3D profile reconstruction, it was also of great importance to calibrate the geometrical relation between the optical components of omnidirectional stereo vision. Generally, a uniform theoretical model was used to present the geometric relationship between the 3D point in the scene and the corresponding 2D point in the image plane of the camera or projector. Then, the model parameters were estimated through several features with known locations in the scene [73] [74]. The calibration accuracy, however, was limited by the imprecise alignment of optical components and residual errors due to the manufacturing error of the off-the-shelf optical components. To improve the accuracy, the vector-based calibration strategy was proposed to uniformly calibrate both camera and projector, where an incident vector for a pixel of the projector and a reflected vector for that of the camera were assigned respectively.

In summary, The development of the omnidirectional 3D camera contributed in the following aspects:
1) Developed and configured a structured light based omnidirectional 3D camera. It was capable to measure 360 degree environment with dense 3D samples.

2) Developed and implemented a one-shot algorithm for such a 3D camera. 3D samples could be measured in real time. It was essential when the sensor was located on a moving base.

3) Developed a vector based calibration for such a system. The proposed method was able to calibrate an omnidirectional projector, which people have not discussed yet based on my literature review.

1.3 Difficulties and Challenges

The objective of this study is to develop large scale structured light based measurement systems, which include topics not only in the development of theoretical methodology but also application implementation. Although several 3D sensors have been developed in past, this study aims at the development of large scale optic systems for accurate inspection and fast environment reconstruction.

The main challenges involved in this study include:

1) Calibration tasks for a large scale structured light system. It is very difficult to maintain high inspection accuracy when the system scale increases. One pixel error, 3D reconstruction error cost by one image pixel, in a large scale system is much bigger than the error in a small one. A complete calibration methodology for a large scale structured light sensor, including 3D model optimization and precision analysis, needs to be developed. Both difficulty during the calibration implementation and the calibration requirement arise when the scale is enlarged. Traditional calibration tasks involves too much parameters in their 3D recovery model, 22 parameters including both imaging devices' intrinsic and extrinsic parameters. Especially the projector parameters is the major source of calibration error.

2) Consistence analysis because of the large scale of the system. Conventional structured
light inspection system usually much less FOV. The influence costed by the random image noises can tolerated in small systems. However, for a large scale system, the random image noises begin to affect the consistence of the system, which is a novel problem for this field.

3) Robust encoding and decoding algorithm is needed inspect optically-challenging objects. Without pretreatment, the reflection ratios for a large work piece can vary a lot. The incident illumination could be dynamically change correspondingy. Otherwise, it is a dilemma to adjust camera aperture. The dark part requires a big aperture and the shiny part needs a small aperture. Optical challenging objects, ranging from shiny, dark, and even blank, should be inspected simultaneously. Shiny objects could cost the mis-coding problem due to the internal reflection problem – the reflection light from a point in the scene is not only related with its own incident light but also with the incident lights in its neighborhood. The un-illuminated area can also result in high intensity in the camera image because of internal reflection.

4) Registration between each sensor frame to the common frame. A large inspection system employees more than one pair of sensors to cover a work piece. The point cloud saved in each individual sensor frame should be merged into a common frame. Sensor calibration error could enlarge the registration error, and make the final result unacceptable. Since the test work piece may not necessarily have any features to establish the correspondence. And furthermore, very Each sensor frame shares very small overlaps. The registration is a challenge task for multi-sensor inspection system.

5) On-line registration calibration between the final point cloud and CAD model within limited time. The color-coded error distribution map is the final purpose of a quality inspection system, which is based on the registration between the final point cloud and CAD. A valid registration tells the distributions of manufacturing errors. Failure to gain a correct error map costs expensive to the manufacturing.

6) Robust one-shot encoding and decoding algorithm is needed for environment reconstruction. Since the system needs to measure moving objects, and thus is sensitive to the
acquisition time. All necessary information for 3D reconstruction must be obtain within one projection and acquisition. The encoding and decoding algorithm should simultaneously satisfy the requirement of accuracy, robustness, and efficiency.

7) 360 degree navigation system with realtime 3D perception ability should be build up in a limited space on a mobile robot.

1.4 Objectives of the Work

1) The development of the new 3D sensor model with less calibration parameters. Compared with camera calibration, projector calibration is more complicated and has more calibration errors. Therefore, a novel sensor model which can avoid the projector intrinsic parameters and orientation should be developed. 3D point is not calculated through traditional triangularization method but through the plane induced parallax information.

2) Analysis, simulation, and verification of the precision of the system was done for the large scale structured light system. The effect of random image noise was analyzed and numerical evaluated. The precision of the designed and implemented system reached the system requirement.

3) A robust surface coding method was developed to inspect the optically-challenging objects. Monochromatic White and Black (WB) stripes instead of color patterns are utilized to eliminate the effect from material's texture and color. A projection mask, dynamically adjusting the illumination of a projector, is applied to avoid under- / over-exposure issues: incident lights shooting onto the shiny or high reflection area are reduced and lights onto the shadow or low reflection area are enhanced. An extrapolation model combined the projection mask successfully solves the internal reflection problem. A pattern of WB stripes and its inverse are sequentially projected onto the test material. The edges between white and black stripes are estimated through an extrapolation model even internal reflection occurs.

4) A registration method was developed to merge multi-sensor frame into a common
frame, and thus a final point that cloud could represent the entire work piece was derived. Based on a pre-measured painted windshield, four individual sensor frame were transferred into a common frame.

5) A color-code error distribution map was created to evaluate a given work piece to finish the quality inspection task. People could easily find out whether a work piece is qualified or not.

6) A new structured light pattern based on the spatial-neighborhood strategy was proposed. The primitive of the pattern was based on the corner of the black-white checkerboard pattern, called the X-point. The X-point provides more robust and accurate position compared with a conventional pattern in which the primitive is based on symmetrical symbols. The orientation of the X-point is used to encode the primitive of the pattern.

7) Adding hyperbolic mirrors in front of the camera and the projector, an omni-directional 3D optical sensor for mobile robot navigation was developed. Its corresponding sensor model, coding method, and calibration were developed.

1.5 Organization of the Study

The development of a structured light based 3D measurement system roughly consists of 3 steps: the working principle of the 3D sensor, calibration of the sensor, and surface coding method for the system. For an accurate inspection system, multiple sensors are utilized to enlarge the system. Therefore, additional registration is needed after each sensor rebuild each individual point cloud. For a fast navigation system, a convex shaped mirror based omni-direction imager is used to enlarge the system field of view. Therefore, omni-direction 3D reconstruction navigation system is explained after the introduction of a traditional single direction navigation system.

Chapter 2 explains the working principle of the 3D sensor. The corresponding sensor calibration is given in chapter 3. Surface coding emphasizing single-shot or multi-shot al-
gorithms are introduced in chapter 4. Chapter 5 provides the registration algorithm to
transfer data in different coordinates. Chapter 6 gives the development of omni-direction 3D
rebuild system. Current experimental results are given in chapter 7. Chapter 8 presents the
conclusions, discussions and further works.
CHAPTER 2

Development of a 3D Sensor Based on Structured Light

The 3D shape of a work piece is conventionally measured by a Coordinate Measurement Machine (CMM) via adopting a touching probe onto the inspected surface and acquiring coordinates. The nature of point-wise contact based method limits both CMM’s measurement speed and the capability of inspecting complicated work pieces. Non-contact optic measurement method via machine vision, with the merits of fast speed, high inspection sample rate, and overall simplicity of operation, becomes a popular alternative inspection method for CMM.

The implementation of a structured light based inspection system is highly related with its scale. A successfully designed and implemented system is capable to measure big work pieces and remain high accuracy. A large scaled structured light system with long standoff distance, large Field Of View (FOV), deep Depth Of View (DOV), and multi-sensor raises the inspection error, and thus lift the difficulty of its calibration. The final inspection result for such an inspection system carries the sensor calibration error, random image noise, and multi-sensor registration error. Traditional 3D reconstruction model requires intrinsic and extrinsic parameters of both the camera and the projector. In order to reduce sensor calibration
error, an optimized 3D reconstruction sensor model with fewer parameters is developed. By defining a reference plane in the sensor model, the intrinsic parameters and the orientation of the projector can be avoided in the 3D recovery calculation.

This chapter describes an innovative 3D reconstruction sensor model with less parameters. Such a sensor model successfully solve a practical problem in industry application again large scale.

2.1 Introduction of Imaging Devices

A Structured light based 3D scanner usually consists of a projector and a camera. The camera obtains a perspective mapping from the world frame to the image from. The projector provides additional constraints to limit the last degree of freedom in the perspective geometry, and then achieve the 3D reconstruction. In order to optimize the 3D recovery, background information of the imaging devices are given in this section: pinhole model approximation, homography matrix between two planes, epipolar geometry between two pinhole model, parallax imaging, lens distortion, Field of View (FOV) and Depth Of View (DOV).

2.1.1 A Pinhole Model of an Image Device

An imaging device, such as a projector or a camera, can be approximated as a pinhole model. Since there is no difference between a camera pinhole model and a projector pinhole model, this section introduces a camera pinhole model as an example.

The geometric model of pinhole camera consists of an optic point $O_c$ and an image frame $O_{Iuv}$ on the focal plane $I$. The fundamental property of the perspective geometry is that an image point $p = [u_p, v_p, 1]'$ is collinear with the $O_c$ and its corresponding point $P = [x_P, y_P, z_P, 1]'$. The point $O_c$ is called optical center, and the camera frame is $O_cX_cY_cZ_c$ with the axis $O_cZ_c$ perpendicular to the image frame $I$. $O_I$ is the intersection point between axis $O_cZ_c$ and plane $I$, which is called the principal point. Rotation matrix $R_c$ and translation
vector $t_c$ provide the geometric relationship between the world frame $O_WX_WY_WZ_W$ and the camera frame $O_cX_cY_cZ_c$. The light array passing any points in world frame into the image plane can be denoted as a $3 \times 4$ perspective transformation matrix $C_c$.

$$\lambda_c p = C_c p$$  \hspace{1cm} (2.1)

, in which $\lambda_c$ is a scale factor. The Matrix $C_c$ can also be factored into three parts: intrinsic parameter $K_c$, rotation matrix $R_c$, and the translation vector $t_c$ in world frame, and Eq. 2.1 can be transferred into:

$$\begin{bmatrix} u_p \\ v_p \\ 1 \end{bmatrix} \cong K_c \begin{bmatrix} R_c & t_c \end{bmatrix} \begin{bmatrix} x_p \\ y_p \\ z_p \\ 1 \end{bmatrix}$$  \hspace{1cm} (2.2)

$$K_c = \begin{bmatrix} f_u & \alpha & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$  \hspace{1cm} (2.3)

Symbol $\cong$ means the equation equals up to a scale factor. $f_u$ and $f_v$ in $K_c$ are the focus lengths along the $u$ direction and $v$ direction in image frame respectively. $\alpha$ is the skew
factor of a camera. For a square pixel camera, \( f_u = f_v = f \) and \( \alpha \) is 0.

The original point of the camera frame \( O_c \) can also be described by rotation matrix \( R_c \) and \( t_c \).

\[
\vec{t}_c = -R_c\vec{O}_c \tag{2.4}
\]

Substituting Eq. 2.4 into Eq. 2.2, the perspective transformation of the pinhole model can be expressed as:

\[
\begin{bmatrix}
u_p \\
v_p \\
1
\end{bmatrix} \cong K_c R_c \begin{bmatrix}
I, & -\vec{O}_c
\end{bmatrix} \begin{bmatrix}
x_P \\
y_P \\
z_P \\
1
\end{bmatrix} \tag{2.5}
\]

Back projection from Eq. 2.5, each light array from a pinhole model can be denoted as Eq. 2.6. \( L \) denotes as the light array passing point \( P \) and \( \lambda_c \) is the scale factor.

\[
L = \vec{O}_c + \lambda_c (K_c R_c)^{-1} p_p \tag{2.6}
\]

In order to reconstruct a light array from a given pixel \( p_p \) of a pinhole model. The parameters in the linear equation: intrinsic parameters, rotation matrix and translation vector are needed.

### 2.1.2 Homography Matrix from Single View Geometry

Planar objects, as one of the most common entities, is widely utilized in machine vision. The link between the planar object and their corresponding image points under perspective projection is described in this section.

Homography matrix describes the projective geometry of a camera image plane and a world plane. Image of points on the image plane are related to corresponding points on world plane by homography matrix as shown in figure 2.2. The planar surface normal is defined as the \( z \) axis of the world frame, and the \( z \) values of all points on the plane equal to 0. Light array passing the pixel \( p \) intersects the world plane at point \( P \). The homography
matrix between two planes can be derived from the camera perspective mapping by Eq. 2.7. The \(i^{th}\) column of the rotation matrix \(R_c\) is denoted by \(r_i\).

![Diagram of image plane and world plane](image)

From Eq. 2.7, the definition of the image plane and the world plane can be expressed as a \(3 \times 3\) nonsingular matrix:

\[
\lambda \begin{bmatrix} u_p \\ v_p \\ 1 \end{bmatrix} = K_c \begin{bmatrix} r_1 & r_2 & r_3 & t_c \end{bmatrix} \begin{bmatrix} X_p \\ Y_p \\ 1 \end{bmatrix} = K_c \begin{bmatrix} r_1 & r_2 & t_c \end{bmatrix} \begin{bmatrix} X_p \\ Y_p \\ 1 \end{bmatrix}
\] (2.7)

From Eq. 2.7, the definition of the image plane and the world plane can be expressed as a \(3 \times 3\) nonsingular matrix:

\[
H = K_c \begin{bmatrix} r_1 & r_2 & t_c \end{bmatrix}
\] (2.8)

From the definition in Eq. 2.8, homography matrix \(H\) can be factorized into three parts: the intrinsic parameters \(K_c\), the first two columns \(r_1\) and \(r_2\) of the rotation matrix \(R_c\) and the translation vector \(t_c\). Utilizing the factorization information, homography matrix is usually used to calibrate a camera or a projector by regressing into the \(K_c\), \(R\) and \(t\) ??.

### 2.1.3 Epipolar Geometry and Fundamental Matrix

A typical structured light sensor consists of a camera and a projector. The epipolar geometry is the intrinsic projective property between two the views. The most important property
of the epipolar constraint is that it is independent of scene structure, and only depends on
the imaging devices’ internal parameters and relative poses. The fundamental matrix $F$
encapsulates such an intrinsic geometry. Fundamental matrix $F$ is $3 \times 3$ matrix of rank 2.

![Figure 2.3: The epipolar geometry between two image planes.](image)

Consider the case shown in figure 2.3, $P$ is a point on the tested surface; $p_c$ and $p_p$ is the
projections of the point $P$ on the image planes $I_c$ and $I_p$ respectively; $O_c$ and $O_p$ are the
optical centers of the camera and the projector. Point $P$, $O_c$ and $O_p$ form a plane, which is
defined as the epipolar plane. The intersection of the epipolar plane with $I_c$ is the epipolar
line $l_c$, and the intersection between the epipolar plane with $I_p$ is $l_p$. The position of point
$P$ varies for different tested surface, but the correspondence point $p_p$ in $I_p$ is always on the
epipolar line $l_p$, which is the most important properties of the epipolar geometry. Each focal
point projecting on the other image plane forms point $e_c$ and $e_p$, named epipolar or epipolar
point.

The point correspondence between $p_c$ and $p_p$ is:

$$p_c F p_p = 0 \tag{2.9}$$

, in which $F$ presents the $3 \times 3$ fundamental matrix between two pinhole model.
The correspondence between the pixel and its epipolar line is:

\[ l_p = F p_c \]  
(2.10)

\[ l_c = F^T p_p \]  
(2.11)

The epipolar lines in both image are related by:

\[ l_p = F [e_c]_\times l_c \]  
(2.12)

\[ l_c = F^T [e_p]_\times l_p \]  
(2.13)

, in which \([e]_\times\) is skew symmetric matrix of an epipolar \(e\).

The epipolar plane is formed by an arbitrary point from scene and two optical points of two pinhole models. Illustrated by figure 2.4, points \(P_1, P_2,\) and \(P_3\) forms three eqipolar planes with the optical points \(O_c\) and \(O_p\). The epipolar planes intersect \(I_c\) at line \(l^1_c, l^2_c\) and \(l^3_c\) and intersect \(I_p\) at line \(l^1_p, l^2_p\) and \(l^3_p\). Since all the eqipolar planes share with the baseline \(O_cO_p\), and \(e_c\) and \(e_p\) are the intersection points of baseline with \(I_c\) and \(I_p\) respectively. Therefore all the epipolar lines in \(I_c\) intersections at point \(e_c\) and all the epipolar lines in \(I_p\) intersections at point \(e_p\). \(e_c\) and \(e_p\) is the null space vector of fundamental matrix \(F\) and \(F^T\) respectively.

The epipolars in both images are:

\[ F e_c = 0 \]  
(2.14)

\[ F^T e_p = 0 \]  
(2.15)

The fundamental matrix \(F\), presenting the epipolar geometry in two pinhole models, is widely used solving correspondence problem in stereo vision. It is also one of the easiest
parameters to be calibrated, since the calculation of fundamental matrix does not require any foreknowledge of the scene. No calibration gauge is needed.

2.1.4 Plane Induced Parallax in an Imaging System

In the figure 2.5, Point $P$ in the scene projects itself as $p_c$ and $p_p$ in image plane $I_c$ and $I_p$. $O_c$ and $O_p$ are the two optical points of two pinhole models. The epipolar in each image frame are $e_c$ and $e_p$ respectively. $l_c$ and $l_p$ are the corresponding epipolar lines in both image planes. The ray $PO_p$ intersects the plane $W$ at point $P_p^W$ and ray $PO_c$ intersects the plane at point $P_c^W$. The images of point $P$ and $P_p^W$ are coincident points at $p_p$ in the image plane $I_p$. $p_c^W = (H_c^W)^{-1} P_p^W = H_p^c p_p$ is the projection of point $P_p^W$ on the image plane $I_c$. The vector between points $p_c$ and $p_c^W$ in the camera image plane is the parallax related with the homography induced by the plane $W$. If the point $P$ is in the other side of the plane $W$, and then the point $p_c^W$ will be on the other side of point $p_c$ along the epipolar line $l_c$.

Point $p_c$ can be expressed by:

$$p_c = p_c^W + \rho e_c = H_p^c p_p + \rho e_c$$  \hspace{1cm} (2.16)

Since $p_c$, $p_c^W$ and $e_c$ are collinear on $l_c$, the scalar $\rho$ and $e_c$ form the parallax vector relative
Figure 2.5: A plane $W$ induced parallax in camera image plane.

to the homography $H_p^c$ induced by the plane $W$. $\rho$ is viewed as depth information related with the plane $W$. If $\rho = 0$ then point $P$ is on the plane $W$. The sign of $\rho$ indicates the side between point $P$ and plane $W$. if $\rho$ is smaller than 0, and then point $P$ is behind plane $W$. Otherwise, point $P$ is before plane $W$.

2.2 Optimize the 3D Sensor Model via Parallax Information

2.2.1 Problem Statement

Traditional 3D reconstruction model based on structured light utilized the intersection between incident light and reflection light in the epipolar plane. As shown in figure 2.6, this constraint is interpreted geometrically in terms of the rays in space corresponding to the two image points. $p_p$ and $p_c$ are two projection points from $P_w$ in the projector plane $I_p$ and camera plane $I_c$ respectively. $O_p$ and $O_c$ are optical points of the projector and the camera respectively. The incident light from $p_p$ is named as $L_p$ and the $L_c$ denotes the reflection light. From the epipolar constraints, $p_p$, $O_p$, $P_w$, $O_c$ and $p_c$ are in the same plane. $P_w$ is the
intersection point of $L_p$ and $L_c$. Since $p_p$, $p_c$ and $P_w$ constructs a triangle in this plane, this method is called triangulation method.

![Diagram of traditional 3D reconstruction model](image)

Figure 2.6: Traditional 3D reconstruction model.

The triangulation 3D reconstruction model requires to generate the back projection equations for the correspondence pixels:

$$L_c = \vec{O}_c + \lambda_c (K_c R_c)^{-1} p_c \quad (2.17)$$

$$L_p = \vec{O}_p + \lambda_p (K_p R_p)^{-1} p_p \quad (2.18)$$

The solution of Eq. 2.17 and Eq. 2.18 is the coordinate of point $P_w$. Due to the calibration error or image noise, $L_c$ and $L_p$ may not be in the same plane resulting that there is no intersection point between this two rays. The middle point of common perpendicular is viewed as the solution in such a case.

This traditional structured light 3D model is known at least two limitations:
1) The first one is the large amount of parameters to be calibrated. It includes the intrinsic and extrinsic parameters of both imaging devices, together 22 parameters. The more parameters to be calibrated, the more errors induced and the less accuracy is received in the result point cloud.

2) The second one is the projector calibration. Due to the property of the projector, not serving as an information receiving device, its calibration is more complicated than a camera and contains more errors than a camera.

This thesis combines the camera planar calibration method into 3D reconstruction to optimize the number of unknowns in the equation. The novel strategy is that the camera calibration information based on planar method can be saved as the reference information. The acquisition information of the test object is used to compare with the pre-saved reference information. The corresponding deviation between reference and test object is applied to reconstruct the 3D point cloud, in which only few parameters need to be calibrated. Therefore, the projector calibration is not necessary for 3D reconstruction anymore. The correspondence function provided by the pattern design is not only used to set up correspondence between a camera and a projector, but also establishes corresponding pair of pixels in camera between two measurements: the camera pixel recording reference information and the corresponding camera pixel recording tested surface information.

2.2.2 Recording the Reference Information with Homography Matrix

If the camera intrinsic $K_c$ and extrinsic parameters, $R_c$ and $t_c$, are calibrated corresponding to plane $W$ through Eq. 2.8, the light array vector from projector can be calculated without projector intrinsic parameters and its rotation matrix. In figure 2.7 $L_p$ coming from the projector intersects reference plane $W$ at point $P_r$ and is reflected into the camera at $p_c^r = [u_c^r, v_c^r, 1]^T$ by the light array $L_c^r$. $t_{pc}$ is the baseline distance between the projector and the camera. By vector relationship, each light ray from the projector can be represented by
reference point $P_r$ and light array $L^r_c$:

$$\overrightarrow{OP_r} = \overrightarrow{OC_r} - \overrightarrow{OC}$$  \hspace{1cm} (2.19)

Since the camera was already calibrated with the plane $W$ via the homography matrix $H^W_c$. Point $P_r$ can calculated through the definition of $H^W_c$: $P_r = H^W_c p^r_c$. Substitute $P_r$, $p^r_c$ and camera calibration information into Eq. 2.6, the depth information $\lambda_{P_r}$ of point $P_r$ can be calculated by:

$$\lambda_{P_r} = \frac{|P_r - O_c|}{|\overrightarrow{O_cP_r}|}$$  \hspace{1cm} (2.20)

Once the camera’s parameters are well calibrated with the reference board, $\overrightarrow{O_cP_r}$ can be considered as known. Vector $t_{pc}$ is the only value to calculate the linear equation of vector
From the projector, which also results that the intrinsic and orientation parameters of the projector are avoided in reconstructing light vector coming out from a projector. The plane information was saved as reference information and replaced projector parameters in 3D reconstruction.

2.2.3 Correspondence between Measurements

![Diagram](attachment:figure_2.8.png)

Figure 2.8: Correspondence between reference recording and work piece inspection.

Image patterns designed in projector plane carry unique information from pixel to pixel. In the reference recording process, codeword carried in \( L_P \) establishes the correspondence with light \( L_r^c \). In the test surface measurement, codeword carried in \( L_P \) establishes the correspondence with light \( L_s^c \). Combing with two measurements, the correspondence between \( L_r^c \) and \( L_s^c \) is set up in two measurements as shown in figure 2.8.
In figure 2.8, a tested surface $S$ in blue color is located on the reference board. $L_p$ from the projector is shot onto the surface at point $P_s$ instead of $P_r$ on the reference board and is reflected into the camera image plane $I_c$ at point $p^s_c = [u^s_c, v^s_c, 1]^T$ by the light array $L^s_c$. From the codeword provided in the light $L_p$, the correspondence between $L^r_c$ and $L^s_c$ is established. The Correspondence pixels $p^r_c$ and $p^s_c$, which are both in camera frame, are used to reconstruct point $P_s$. In contrast, traditional method constructs point $P_s$ needs correspondence pixels in different image frames: projector frame $I_p$ and $I_c$, which results calibration in both image devices is needed instead of one.

2.2.4 Working Principle of the New Sensor Model

![Figure 2.9: The optimized 3D model.](image)
Shown in figure 2.9, the reference reflection light \( L_r^c \) can be written as:

\[
L_r^c = \vec{O}_c + (K_cR_c)^{-1}\lambda_rP_r^c
\]

(2.21)

The reflection reference point \( P_r \) on the reference \( W \) is:

\[
P_r = \vec{O}_c + (K_cR_c)^{-1}\lambda_P^cP_r^c
\]

(2.22)

\( K_c \) is the camera intrinsic parameter. \( R_c \) is the rotation matrix, and \( O_c \) is the coordinate of the camera optic point. \( \lambda_r \) is the depth along the array \( L_r^c \) and \( \lambda_P \) is the depth of point \( P_r \) along \( L_r^c \).

The surface reflection light \( L_s^c \) denotes:

\[
L_s^c = \vec{O}_c + (K_cR_c)^{-1}\lambda_sP_s^c
\]

(2.23)

The incident light \( L_p \) is:

\[
L_p = \vec{O}_p + (K_pR_p)^{-1}\lambda_pP_p^p
\]

(2.24)

The projection of projector optical point \( O_p \) in camera image plane \( I_c \) at epipolar \( e_p^c = [u_e, v_e, 1]^T \). The coordinate of \( O_p \) can be expressed in camera frame by:

\[
O_p = \vec{O}_c + (K_cR_c)^{-1}\lambda_O^cO_p^c
\]

(2.25)

, in which \( \lambda_{O_p} \) the depth along array \( O_pO_c \).

Since point \( O_p \) and \( P_r \) can be expressed in the camera frame, substitute Eq. 2.25 and Eq. 2.22 into the following:

\[
L_p = O_p + \lambda \left( \vec{O}_rP_r^c \right)
\]

(2.26)

The incident light, which previously be expressed in the projector frame by Eq. 2.24, can be written in camera frame by:
\[ L_p = O_c + (K_c R_c)^{-1} \left[ \lambda O_p e^p_c + \lambda \left( \lambda P_r p_c - \lambda O_p e^p_c \right) \right] \] (2.27)

Through Eq. 2.23 and Eq. 2.27, an equation of depth \( \lambda_s \) can be written by:

\[ \lambda_s p_c^s = \lambda O_p e^p_c + \lambda \left( \lambda P_r p_c - \lambda O_p e^p_c \right) \] (2.28)

Extend Eq. 2.28, we can get:

\[ \lambda s u_{ps} = \lambda O_p u_e + \lambda \left( \lambda P_r u_{pr} - \lambda O_p u_e \right) \] (2.29)

\[ \lambda s v_{ps} = \lambda O_p v_e + \lambda \left( \lambda P_r v_{pr} - \lambda O_p v_e \right) \] (2.30)

When the depth \( \lambda_s \) equals the following, the surface point \( P_s \) can be derived.

\[ \lambda_s = \lambda P_s = \frac{\lambda O_p \lambda P_r (u_e v_{pr} - v_e u_{pr})}{\lambda P_r (u_{ps} v_{pr} - v_{ps} u_{pr}) - \lambda O_p (u_{ps} v_e - v_{ps} u_e)} \] (2.31)

Transfer the Eq. 2.31 into the following:

\[ \lambda_s = \frac{\lambda O_p \lambda P_r \begin{vmatrix} u_e & u_{pr} \\ v_e & v_{pr} \end{vmatrix}}{\lambda P_r \begin{vmatrix} u_{ps} & u_{pr} \\ v_{ps} & v_{pr} \end{vmatrix} - \lambda O_p \begin{vmatrix} u_{ps} & u_e \\ v_{ps} & v_e \end{vmatrix}} \] (2.32)

The Eq. 2.32 can be simplified into a determinant form:

\[ \begin{vmatrix} u_e & u_{pr} & u_{ps} \\ v_e & v_{pr} & v_{ps} \\ \frac{1}{\lambda O_p} & \frac{1}{\lambda P_r} & \frac{1}{\lambda P_s} \end{vmatrix} = 0 \] (2.33)

Substitute \( \lambda_s \) into the Eq. 2.23, The point of \( P_s \) can be calculated:

\[ P_s = O_c + (K_c R_c)^{-1} \frac{\lambda O_p \lambda P_r (u_e v_{pr} - v_e u_{pr})}{\lambda P_r (u_{ps} v_{pr} - v_{ps} u_{pr}) - \lambda O_p (u_{ps} v_e - v_{ps} u_e)} p_c^s \] (2.34)
From Eq. 2.34, the 3D reconstruction is all done in the camera frame via the pixel location deviation between two measurements: reference recording and surface measurement. The parameters that need to be calibrate are: $O_c, K_c, R_c, e_c, \lambda_{Op}$. The parameters $\lambda_{Pr}, u_{pr}$ and $v_{pr}$ are calculated though the reference recording process.

**Parallel Configuration**

In last section, 3D reconstruction has been achieved only the camera frame by adopting the pixel deviation between two measurement as the system input. Projector calibration can be avoided, resulting that less parameters is utilized in the 3D recovery model. If the baseline $O_pO_c$ is configured parallel with the reference plane $W$, the 3D recovery can be even more easier via similar triangle theory. Shown in figure 2.10, The baseline $O_pO_c$ is parallel with the reference recording plane $W$, so translation vector $t_{pc}$ can be consider as one scale value as $d$.

Both imaging devices, the projector and the camera, share with the same standoff distance $S$. $P^s_r$ is the intersection point between plane $W$ and extension of line $O_cP_s$, and $P^s_r$ can be consider the projection of $P_s$ on plane $W$ when viewing from the camera. $h_{Ps}$ denotes the height of point $P_s$. Since line $O_pO_c$ is parallel with $P^s_rP_r$, triangle $P_sO_pO_c$ is similar with the triangle $P_sP^s_rP_r$. The ratio of similar triangles is constant:

$$\frac{h_{Ps}}{S - h_{Ps}} = \frac{P^s_rP_r}{d}$$

(2.35)

$P^s_rP_r$ denotes the length between point $P^s_r$ and $P_r$ in the reference plane and it can be transferred into the camera image frame by:

$$P^s_rP_r = |P^s_r - P_r| = |H^w_c (p^s_c - p^r_c)| = |H^w_c p^s_{pr}|$$

(2.36)

The point $P_s$ can calculated through the following equations:

$$h_{Ps} = \frac{(P^s_rP_r) S}{d + P^s_rP_r} = \frac{|H^w_c p^s_{pr}| S}{d + |H^w_c p^s_{pr}|}$$

(2.37)
2.3 Chapter summary

For a large scale structured light sensor with multi-sensor, large field of view, and long standoff distance, the work principle of a 3D measurement sensor is critical for the system. It could greatly affect the system performances. A novel 3D reconstruction model is developed. Compared with the traditional structured light sensors, the intrinsic and rotation of the projector are avoided in the 3D reconstruction model. Since the projector calibration error is the major error source, the developed 3D sensor cloud further reduce the calibration error and improve the system accuracy.
CHAPTER 3

Sensor calibration

Sensor calibration is to regress of the parameters in the 3D reconstruction model. It usually consists of two parts: the consistence evaluation of the system and the calibration method to regress of parameters. The two problems left to be solved from Eq. 2.37 and Eq. 2.38 in sensor calibration are summarized as:

1) Error analysis and consistence evaluation of the input $p_{sr}^c$ based on the system configuration;

2) Camera calibration and regression of translation vector $S$ and $d$.

3.1 Precision Analysis of the Given Sensing System

Precision is also called the consistence of the system. People always confuse it with the concept of accuracy. In fact these two are totally different concepts. Precision is a value to evaluate the system’s repeatability. The accuracy is a value describing the difference between the ground truth and the mean of the multiple measurements. The difference between these two concepts are shown in figure 3.1. $y$ axis denotes the probability density and $x$ axis denotes the measurement value. Precision can be viewed as the standard deviation of multiple measurements, which represents the repeatability. Accuracy is the deviation between reference value (ground truth) and the mean of multiple measurement. There is a
significant difference between these two values: the accuracy of the system can be improved by calibration, but the precision can not. Because precision is mainly affect by the random image noise from the environment.

![Figure 3.1: The difference between precision and accuracy.](image)

Two examples are given in figure 3.2. In case (a), system’s accuracy is high but with low precision. The difference between ground truth (the center of the rings) and the mean of all the dots is small. The system can be considered as a well calibrated system with high accuracy. However, Industry can not use such a system, since neither measurement is able to provide correct information. The difference between the ground truth is almost random, which means the system can not improved by calibratoin. In case (b), System’s accuracy is low but with high precision. The difference between ground truth and the mean of all the dots is big. The system can be considered as a poor calibrated system with high precision. Industry is able to utilize such a system by simply adding an offset into the measurement result.

Calibration can not improve the precision. In contrast, poor precision may significantly lower the calibration accuracy. The calibration process utilizes the measurement of the gauge to regress of the model parameters. If the measurements have too big standard deviation, the calibrated parameters will contain big errors. In the calibration process, mean of multi-measurement of a given calibration gauge is used to regress of parameters. In the industry
3.1.1 Uncertainty Analysis of an Imaging Device

Consistence performance of a 3D inspection system is mainly from the random image noise from the surrounding environment. The random image noise increases corresponding with the standoff distance $S$. Larger inspection range results a longer standoff distance. Traditional inspection field ranging from 10,000 to 50,000 $mm^2$ has relatively lower image noise, which can be ignored in measurement process. When a sensor’s Field Of View (FOV) reaches about 2,400,000 $mm^2$, uncertainty analysis of random noise needs to be considered. Proportional with the FOV, the standoff distance is almost 10 times larger than the traditional application, resulting big random image intensity noise from environment. Larger FOV also means that each pixel covers bigger area. Same level of sub-pixel variation in the camera image gives bigger 3D reconstruction error.

In figure 3.3, The pattern is shot from a projector to a flat object at point $X_i$. Pixel $I_i$ is the projection from the point $X_i$ in the world frame. Due to the random noise, $I_i'$ is applied instead of $I_i$ to construct the light array which intersect the plane object at point $X_i'$. The angle between $I_i'X_i'$ and $I_iX_i$ is $\theta$. The focus length is $f$, and $S$ denotes the standoff distance. The error between $X_i$ and $X_i'$ is defined as $e$:

Figure 3.2: (a) High accuracy with low precision; (b) low accuracy with high precision.

application, a proper precision result is critical for the system. Calibration accuracy is bonded by the system precision.
Figure 3.3: Camera error model.

Figure 3.4: Projector error model.
Same principles with a camera model, the error model of a projector is also approximately proportional with the standoff distance $S$, which is shown in figure 3.4. Longer standoff distance has worse system precision. An inspection system with a given design parameter standoff distance $S$ needs to measure the precision to decide whether it needs further adjust or not.

3.1.2 Simulation of the System Error Model

From Eq. 2.37, once the input $p_{sr}^c$ has a certain number of uncertainties, the height error will be:

$$
\Delta h_{Ps} = \left| \frac{S}{d + \frac{d}{1 + \frac{1}{H_{Wc}^w P_{sr}^c}} - \frac{S}{d + \frac{d}{1 + \frac{|H_{Wc}^w (p_{sr}^c + \Delta p_{sr}^c)|}}}} \right|
$$

(3.2)

Homography matrix $H_{Wc}^w$ is the mapping from the image plane to the reference in Eq. 2.36. $P_r^S P_r$ is the distance in the reference plane, and $p_{sr}^c$ is the pixel distance in the camera image plane. In the current system, the camera is configured almost parallel with the reference plane to reach big FOV. So we can assume that the resolution of the camera is a constant value $k$ which transfers the pixel length to the physical length on the reference plane. So the simulation formula can be approximately expressed as:

$$
\Delta h_{Ps} = \left| \frac{S}{d + \frac{d}{k |p_{sr}^c|} - \frac{S}{d + \frac{d}{k |p_{sr}^c| + k |\Delta p_{sr}^c|}}} \right|
$$

(3.3)

, in which $k$ represents the resolution scalar factor to convert the unit from pixel into millimeter. Assuming the system has 0.1, 0.2, 0.3, 0.5 pixels’ error, the error in height are shown in the figure 3.5. $S$ is designed as 1600 millimeters. $d$ is designed as 500 millimeters, and $k$
equals 0.7 mm/pixel. The $y$ axis in figure 3.5 denotes the height error amount and the unit is millimeters. The $x$ axis is input value of the sensor model: $p_{sr}^c$, the deviation in the camera image frame between point $p_s^c$ and point $p_r^c$. As the error increases from 0.1 to 0.5 pixels, the error of height increases. In each sub-figure in figure 3.5, the error curve drops as the input $p_{sr}^c$ increases. $p_{sr}^c$ in the sensor model is costed by the height deviation between point $P_s$ and reference plane. Large $p_{sr}^c$ means that point $P_s$ has a big height from the reference in figure 2.10, which results the standoff distance between the sensor to the point $P_s$ becomes shorter. From the Eq. 3.1, a shorter standoff distance results smaller error.

![Graphs showing error analysis results](image)

Figure 3.5: Error analysis results: height errors when input error equals 0.1, 0.2, 0.3, 0.5 pixel.
Table 3.1: The standard deviation results for all sensors

<table>
<thead>
<tr>
<th>Sensor one</th>
<th>Sensor two</th>
<th>Sensor three</th>
<th>Sensor four</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.17 mm</td>
<td>0.08 mm</td>
<td>0.08 mm</td>
<td>0.13 mm</td>
</tr>
</tbody>
</table>

3.1.3 System Precision Test Results

An inspection gauge is measured 10 times. The standard deviation is used to describe the precision of the system. The mean of the height map in each pixel is calculated by:

\[
\bar{h}(u, v) = \frac{1}{n} \sum_{k=1}^{n} h_k(u, v)
\]  

(3.4)

in which \(u\) and \(v\) stand for the column and row location of a pixel; \(n\) is the total measurement times, and \(k\) is the measurement order. The standard deviation of each pixel is:

\[
Std(u, v) = \sqrt{\frac{1}{n} \sum_{k=1}^{n} (h_k(u, v) - \bar{h}(u, v))^2}
\]  

(3.5)

The standard deviation for each couple of projector and camera can be calculated by:

\[
Std_s = \frac{1}{n_u \cdot n_v} \left( \sum_u \sum_v Std(u, v) \right), \quad u = 1, ..., n_u; \quad v = 1, ..., n_v
\]  

(3.6)

where \(n_u\) and \(n_v\) are maximum row and column number of the camera resolution. In our application, they are 2044 and 2002, respectively. Current four precision results for four sensor are given in table ??

52
3.2 Sensor Calibration

3.2.1 Calibrate Homography Matrix between the Camera Image Plane and the Reference Plane

Apply the definition of homography matrix in Eq. 2.7:

\[
\begin{bmatrix}
u_i \\ v_i \\ 1
\end{bmatrix} = H_W^{c} \begin{bmatrix}
X_i \\ Y_i \\ 1
\end{bmatrix}, P_i = [X_i, Y_i, 1]^T \in W, p_i = [u_i, v_i, 1]^T \in I_c, i = 1, \ldots, n
\] (3.7)

in which \( p_i \) and \( P_i \) are a pair of corresponding pixel in image plane and point in reference plane. \( n \) is the number of corresponding pairs used to calculate the homography matrix.

Convert the \( H \) matrix in vector form as \( \vec{h} = (h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{31}, h_{32}, h_{33})^T \), and substitute it into the Eq. 3.7. The homogeneous equations for \( n \) points become \( Ah = 0 \), with \( A \) the \( 2n \times 9 \) matrix:

\[
A = \begin{pmatrix}
u_1 & v_1 & 1 & 0 & 0 & 0 & -u_1 X_1 & -v_1 X_1 & -X_1 \\
0 & 0 & 0 & u_1 & v_1 & 1 & -u_1 Y_1 & -v_1 Y & -Y_1 \\
u_2 & v_2 & 1 & 0 & 0 & 0 & -u_2 X_2 & -v_2 X_2 & -X_2 \\
0 & 0 & 0 & u_2 & v_2 & 1 & -u_2 Y_2 & -v_2 Y_2 & -Y_2 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
u_n & v_n & 1 & 0 & 0 & 0 & -u_n X_n & -v_n X_n & -X_n \\
0 & 0 & 0 & u_n & v_n & 1 & -u_n Y_n & -v_n Y_n & -Y_n
\end{pmatrix}
\] (3.8)

The solution of \( Ah = 0 \), can follow the standard procedure which can minimize the algebraic residuals \( |Ah| \). Via singular value decomposition (SVD), the null space vector \( h \) is the solution which can minimize the residuals.

Fig. 3.7 (a) shows the calibration method to calculate the homography matrix between the chessboard and the camera image plane. Fig. 3.7 (b) gives a top view of the chessboard.

The camera frame is \( O_c x_c y_c z_c \). \( O_c \) is the optic center of the camera and \( z_c \) axis is perpendicular to the image plane \( I_C \). The frame \( O_w x_w y_w z_w \) is the world frame. The surface
normal of the chessboard is defined as the $z$ direction of the world frame. Both frames follow the right-hand rule. The homography matrix can be viewed as a part of the perspective projection matrix $C_c$ without the third column.

3.2.2 New Calibration Tasks of the Camera

The limitation of the traditional homography based calibration method lies in the regression process from homography matrix to the intrinsic and extrinsic parameters of the image devices. The regression solution goes into ill-conditioned when imaging devices are parallel with the inspection plane. The intrinsic and extrinsic parameters are used to construct a
light array from a given pixel such as the array in Eq. 2.17. If the light arrays can be expressed in different ways, the regression process can be avoided.

Define a $3 \times 3$ non-singular matrix $H$, and multiply the pixel $p_c$ with $(H^{-1}H)$, Eq.2.17 is changed into:

$$L_c = \tilde{O}_c + \lambda_c (K_c R_c)^{-1} \left(H^{-1}H\right) p_c$$ \hspace{1cm} (3.9)

The equation of the light $L_c$ can be transferred into

$$L_c = \tilde{O}_c + \lambda_c (HK_c R_c)^{-1} (Hp_c)$$ \hspace{1cm} (3.10)

The new calibration tasks of the Eq.3.10, which is mathematically not difference with Eq.2.17, are changed from finding $K_c, R_c$ and $O_c$ into:

1) Choose a proper $3X3$ non-singular matrix $H$ and find the matrix multiplication result of $(HK_c R_c)$.

2) Find the translation vector $\tilde{O}_c$;

The forward transformation of Eq.3.10 can be expressed as:

$$H \begin{bmatrix} u_c \\ v_c \\ 1 \end{bmatrix} \cong HK_c R_c \begin{bmatrix} I, & -\tilde{O}_c \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$ \hspace{1cm} (3.11)

Defining $p'_c = (u'_c, v'_c, 1)^T = Hp_c$ and $K_c^N R_c^N = HK_c R_c$. Eq.3.11 is changed into the following:

$$\begin{bmatrix} u'_c \\ v'_c \\ 1 \end{bmatrix} \cong K_c^N R_c^N \begin{bmatrix} I, & -\tilde{O}_c \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$ \hspace{1cm} (3.12)

, in which $K_c^N$ is intrinsic parameter of a new virtual pinhole model, $R_c^N$ is the new rotation matrix of the new pinhole model, and $p'_c$ is the new pixel location on the new image plane.
In Fig. 3.7, Optic point of the camera denotes point $O$, image plane $I_c$ has the distance of $f$ to the optic point. The virtual image plane denotes $I'_c$, with distance $f'$ to the optic point. Homography matrix between image plane $I_c$ to the $I'_c$ is the $3 \times 3$ matrix $H$. The intrinsic parameter of both models are $K_c$ and $K_N^c$ respectively. The light array $L$ can be both constructed from pixel $(u, v)$ or by the pixel $(u', v')$. If the homography matrix between two pinhole models is known, the light array equation can be reconstructed in both ways depending which one is easier to calculate.

### 3.2.3 The Choice of Non-singular Matrix $H$ in New Calibration Tasks

By factorization the homography matrix into the traditional format of Eq. 2.8, camera calibration \cite{75} can be processed by iterations to find $K$, $r_1$, $r_2$, $r_3 = r_1 \times r_2$, $\vec{t}_c$, and $t_c = -R_c \vec{O}_c$ to fit parameters existed in the light array equation. Traditional calibration method stops factorizing homography matrix here. However, the homography matrix $H_w^c$ can be continually factorized:
In the world frame, \( \vec{i} = [1, 0, 0] \) and \( \vec{j} = [0, 1, 0] \) are two unit vectors along \( x \) and \( y \) directions respectively. \( \vec{O}_c = [x_c, y_c, z_c] \) is the coordinate of the optic point \( O_c \) in the world frame. Eq. 3.13 transfers into the following form:

$$ H^c_{W} = K_c \begin{bmatrix} R_c \vec{i} & R_c \vec{j} & R_c (-\vec{O}_c) \end{bmatrix} $$  \hspace{1cm} (3.13)

, in which \( \vec{i} = [1, 0, 0] \) and \( \vec{j} = [0, 1, 0] \) are two unit vectors along \( x \) and \( y \) directions in the world frame respectively. Eq. 3.13 transfers into the following form:

$$ H^c_{W} = K_c R_c \begin{bmatrix} 1 & 0 & -x_c \\ 0 & 1 & -y_c \\ 0 & 0 & -z_c \end{bmatrix} $$  \hspace{1cm} (3.14)

matrix \( \begin{bmatrix} 1 & 0 & -x_c \\ 0 & 1 & -y_c \\ 0 & 0 & -z_c \end{bmatrix} \) seems a normal \( 3 \times 3 \) matrix. In fact it is an inverse matrix of new pinhole model’s intrinsic parameter matrix:

\[
\begin{bmatrix} 1 & 0 & -x_c \\ 0 & 1 & -y_c \\ 0 & 0 & -z_c \end{bmatrix} = (z) \begin{bmatrix} -z_c & 0 & x_c \\ 0 & -z_c & y_c \\ 0 & 0 & 1 \end{bmatrix}^{-1} \hspace{1cm} (3.15)
\]

Define a new pinhole model’s intrinsic parameter matrix as:

$$ K^N = \begin{bmatrix} -z_c & 0 & x_c \\ 0 & -z_c & y_c \\ 0 & 0 & 1 \end{bmatrix} $$  \hspace{1cm} (3.16)

The final factorization of the homography matrix can go to:

$$ H^c_{W} = K_c R_c (-z) \begin{bmatrix} -z_c & 0 & x_c \\ 0 & -z_c & y_c \\ 0 & 0 & 1 \end{bmatrix}^{-1} $$  \hspace{1cm} (3.17)

The calibration board and the optic point form a new virtual pinhole model, in which new intrinsic parameters are expressed by the coordinate of the optic point \( O_c \) in world frame. The absolute value of \( z_c \) can be considered as new focus length. The image center is \( [x_c, y_c]^T \). The new pinhole model can be viewed as a ‘square pixel camera’, since focus lengths along \( x \) direction and \( y \) direction are the same and skew factor equals to 0. Since the new pinhole model is defined in the world frame, the rotation matrix between the world
frame and the new camera frame is an $3 \times 3$ identity matrix. Substituting Eq. ?? into Eq. ??, the relationship between the image pixel and the point on the calibration board is expressed by the following equation:

$$K^N I \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \cong K_c R_c \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

(3.18)

Choose $H_c^w = (H_{wc}^c)^{-1}$ as the 3X3 non-singular matrix $H$ in calibration tasks in section III and substitute it into Eq. 3.11:

$$\begin{bmatrix} u'_c \\ v'_c \\ 1 \end{bmatrix} \cong \begin{bmatrix} -z_c & 0 & x_c \\ 0 & -z_c & y_c \\ 0 & 0 & 1 \end{bmatrix} R^{-1} K^{-1} K R \begin{bmatrix} I, & -t_c \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

(3.19)

The reconstruction of the light array changes into the final form:

$$L_c = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} + \lambda_c \begin{bmatrix} -z_c & 0 & x_c \\ 0 & -z_c & y_c \\ 0 & 0 & 1 \end{bmatrix}^{-1} H_c^w \begin{bmatrix} u_c \\ v_c \\ 1 \end{bmatrix}$$

(3.20)

The forward transformation matrix becomes:

$$H_c^w \begin{bmatrix} u_c \\ v_c \\ 1 \end{bmatrix} = \begin{bmatrix} u'_c \\ v'_c \\ 1 \end{bmatrix} = \begin{bmatrix} -z_c & 0 & x_c & 0 \\ 0 & -z_c & y_c & 0 \\ 0 & 0 & 1 & -z_c \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = C'_c \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

(3.21)

in which $-z_c$ equals to the standoff distance $S$. The new perspective matrix $C'_c$ only contains the translation vector $O_c$. $H_c^w$ is the transformation from a pixel $p_c$ in the old image plane to a pixel $p'_c$ in the new image plane.

$$H_c^w \begin{bmatrix} u_c \\ v_c \\ 1 \end{bmatrix} = \begin{bmatrix} u'_c \\ v'_c \\ 1 \end{bmatrix}$$

(3.22)

After the homography matrix $H_c^w$ was chosen as the $3 \times 3$ nonsingular matrix in the new calibration task, the multiplication of $(H_c^w K_c R_c)$ can be easily expressed by three
elements of $\vec{O}_c$. And now the only unknown is $\vec{O}_c$ in the camera pinhole model equation. The traditional linear or nonlinear regression to find intrinsic and extrinsic parameters to rebuild the light array equation is avoided. The new image plane is mapped into a plane perfect parallel with the reference plane.

### 3.2.4 Calibration of Camera Translation Vector $O_c$ and Baseline Distance $d$

![Multi-plane calibration strategy](image)

**Figure 3.8: Multi-plane calibration strategy.**

A multi-plane pixel to pixel calibration strategy has been developed for current system. The chessboard shown in figure 3.6 is lifted up multiple times and its correspondingly position is recorded for calibration. In order to eliminate the random image noise in recording, all the recordings in the following calibration process are measured 5 times and the mean of measurement results is used in data regression.

$$\bar{I} (u, v) = \frac{1}{m} \sum_{k=1}^{m} I_k (u, v), m = 5$$  \hspace{1cm} (3.23)

$I_k(u, v)$ in Eq. 3.23 represents the intensity value of a pixel $[u, v]^T$ of the $k^{th}$ recording.

The calibration procedure can be summarized as follow:

1) Place the chessboard as reference board underneath of a sensor, and the homography matrix $H_c^{w}$ and lens distortion are calculated via the corners coordinates $[x_w^r, y_w^r, 0, 1]^T$ in
the world frame and their corresponding pixel locations \([\overline{u}^r_c, \overline{v}^r_c, 1]^T\).

2) The Grey Code Line Shifting (GCLS) is projected onto the chessboard and is captured by the camera as the reference cording.

3) Lift the chessboard at height \(h_1\) called layer 1. Substitute the point \(P_w = [x^i_w, y^i_w, h_i, 1]^T\), the corresponding camera pixel \([\overline{u}^l_c, \overline{v}^l_c, 1]^T\), and \(H^w_c\) into Eq. 3.21.

4) Project GCLS onto the broad and find the correspondence by the codeword explained in figure 2.8. The pixel distance \(p_{sr}^c(h_1, u, v)\) is calculated based on the correspondence in each pixel \(u = 1, ..., n_u, v = 1, ..., n_v\). Since every point on the board has the same height, \(p_{sr}^c(h_1)\) is:

\[
p_{sr}^c(h_1) = \frac{1}{n_u n_v} \sum_u \sum_v p_{sr}^c(h_1, u, v)
\]

where \(n_u\) and \(n_v\) are the column number of the camera and row number, respectively. Each layer's \(p_{sr}^c(h_1, u, v)\) distribution is approximated as a Gauss curve shown in figure 3.9, which is due to the random image noises and some implementation errors.

Substitute \(H^w_c\) and \(p_{sr}^c(h_1)\)

5) Repeat the step 3) and 4) 9 times and each time at different height.

6) Solve Eq.3.21 by the collected data: Transfer the Eq.3.21 into the following:
\[
\begin{bmatrix}
0 & -\left(P_i^w\right)^T & \overrightarrow{u}_{c_i}^T \left(P_i^w\right)^T \\
(P_i^w)^T & 0 & -\overrightarrow{v}_{c_i}^T \left(P_i^w\right)^T \\
-\overrightarrow{u}_{c_i}^T \left(P_i^w\right)^T & \overrightarrow{v}_{c_i}^T \left(P_i^w\right)^T & 0
\end{bmatrix}
\begin{bmatrix}
\overrightarrow{c}_c^{1T} \\
\overrightarrow{c}_c^{2T} \\
\overrightarrow{c}_c^{3T}
\end{bmatrix} = 0
\] (3.25)

where \(\overrightarrow{c}_c^{i}\) is the \(i\)-th row of camera perspective matrix \(C'_c\). The form of Eq. 3.25 can be transformed into \(A\overrightarrow{c} = 0\). For a set of \(n\) pairs of correspondence points, a \(3n \times 12\) matrix \(A\) by stacking up Eq. 3.25 for each correspondence. \(\overrightarrow{c}\) is the right null-space vector, and can be solved by SVD method.

7) Substitute \(H_c^w\) and \(\overrightarrow{p}_c^{sr}(h_1)\) into Eq. 2.37, and change the format into:

\[
\begin{bmatrix}
h_1 & \overrightarrow{p}_c^{sr} (h_1) \\
\vdots & \vdots \\
h_n & \overrightarrow{p}_c^{sr} (h_n)
\end{bmatrix}
\begin{bmatrix}
d \\
S
\end{bmatrix} =
\begin{bmatrix}
-h_1 \overrightarrow{p}_c^{sr} (h_1) \\
\vdots \\
-h_n \overrightarrow{p}_c^{sr} (h_n)
\end{bmatrix}
\] (3.26)

where \(\overrightarrow{p}_c^{sr} = H_c^w \overrightarrow{p}_c^{sr}\). \(S\) and \(d\) are solved by least square method.

8) Residual function compensation for each pixel: The sensor model requires that the baseline is parallel with reference plane. The implementation error makes the real case is a little off from the ideal case. In order to further decrease the errors, the residual compensation function for each pixel of the camera is approximated by a second order polynomial. The Residual function for each pixel is the difference between:

\[
e (u, v) = h_n (u, v) - \frac{S}{d} \left(1 + \frac{1}{|H_c^w \overrightarrow{p}_c^{sr} (h_n, u, v)|}\right)
\] (3.27)

Since the difference between the ground truth and the system measurement value, the residual function can be approximately by Taylor expansion. A second order polynomial is used to approximate each pixel’s residual function in Eq. 3.28.

\[
e (u, v) = A (u, v) |H_c^w \overrightarrow{p}_c^{sr} (h_n, u, v)|^2 + B (u, v) |H_c^w \overrightarrow{p}_c^{sr} (h_n, u, v)| + C (u, v)
\] (3.28)
Figure 3.10 illustrates a result of residual function approximation of a given pixel \( u = 80, v = 2 \). The \( x \) axis is the value of pixel distance \( |p_c^o r(u, v)| \) when measuring different heights, and \( y \) axis is the difference between the ground truth and measurement results (unit in millimeter). The blue curve is the residual between ground truth and measurement result, and the red one is the second order polynomial approximation. From the figure, we can see that the error reaches 1 millimeter before the error compensation and becomes less than 0.2 millimeter after compensation.

![Figure 3.10](image)

**Figure 3.10**: The second order polynomial to approximate the residual function for a pixel

### 3.3 Chapter Summary

A complete calibration method is developed for the structured light sensor. All the parameters in the sensor model can be successfully regressed. In order to further increase the
accuracy, the residual function for each camera pixel is approximated by a second order polynomial. The consistence of the developed system is analyzed, simulated and tested. The ill-posed condition problem in camera calibration problem is avoided by applying homography mapping.
CHAPTER 4

Surface Coding for a Structured Light System

The encoded patterns, which are called structured light, are projected from the projector onto the test work pieces and are recorded by the camera. The codewords carried in reflection lights are extracted in camera frame within the decoding process. With the aid from structured light, the camera and projector pixels sharing same codeword are corresponded to obtain 3D coordinates of the test work piece. The measurement result, the point cloud representing for the work piece, can be further used for different applications, such as quality inspection for industry, mobile robot navigation, and so on.

4.1 Introduction of the Surface Coding Methods

The tasks for the surface coding process is to set up the correspondence between image frames: the camera frame and the projector frame. The projected codewords should be preserved from the effect of the test objects distortion, such as distortion from their geometries, textures, and optical properties. The corresponding decoding process should correctly extract codeword. The efficiency, robustness, density, and its accuracy are the most important characters to evaluate the part.
The efficiency of the coding methods is usually reevaluated by the system acquisition time. The codewords could be encoded and decoded with less projection patterns. Therefore, the system could finish the measurement process with less time. The robustness of the coding method is the ability to avoid the albedo affect from the tested objects. The reflection light from the tested objects to the camera is highly related with the surface optical properties. The textures, colors, and reflection coefficients could disturb the decoding process. The encoding and decoding process should be designed against the optical properties of the objects. The density of the encode codewords controls how dense the measurement point cloud is. More codewords could sample more 3D points from the objects and thus obtain more geometry information of the part. A dense sample result is critical for industry inspection application, as errors could be reflected easily by a high density point cloud. The codewords are encoded / decoded in the projector frame / camera frame respectively. It is very important to accurately extract the codewords from both image frames. The image coordinates of the codewords are used to reconstruct the 3D measurement result. Therefore, the accuracy in extracting codewords determine the accuracy of the measurement result.

According to the acquisition time, the coding methods could be roughly separated into two categories: multi-projection and single-projection. Usually a multi-projection method takes more time than a single-projection method, as more projecting-recording loops are contained in a multi-projection method. A single-projection method encodes the codewords within one projection pattern, while a multi-projection method creates the codewords via several patterns. The efficiency of a coding method is highly related with the capability that the system could measure. A moving object can only be measured by single-projection method. However, the disadvantage is that its robustness, density, and accuracy are usually lower than the multi-projection method. Therefore, any applications that emphasize on the measurement time not the accuracy prefers to utilize the single-projection method.

Multi-shot strategy is only available for stable objects but it is more robust against image noise than one-shot method. After detecting the illumination problems, such as shadow
problem, under- / over- exposure problem and internal reflection problem, the projection pattern could be readjusted by a proper feedback. Since more projections are applied to the objects, more codewords can be encoded into the patterns. Therefore, the advantages of the multi-projection method are the robustness, density, and accuracy. The multi-projection is usually used in inspection application for quality control in industrial application.

Emphasizing on different performance factors, the coding methods are explained in two categories: single-shot projection method and multi-shot projection method.

4.2 Single-shot Coding Method

4.2.1 Pattern Generator

The patterns based on spatial neighbors can be generated shown in figure 4.1. The generation procedure starts with the top left $3 \times 3$ random matrix. Then, the window moves to the next position. The Hamming distances between the new primitive codeword and all of those previously generated are calculated, where the Hamming distance is defined as:

$$h_{(uv,u'v')} = \sum_{k=1}^{9} \delta_k, \delta_k = \begin{cases} 0 & \text{if } p_{uv} = p_{u'v'} \\ 1 & \text{otherwise} \end{cases}$$ (4.1)

It is shown in figure figure 4.1 that we can obtain better results when the Hamming distance is greater than 3. If all the distances are no smaller than a design threshold, the primitive is accepted and the next iteration was performed until all the gaps are filled. If not, a new primitive is assigned. The window is successively shifted from the left to the right with a step size of $\Delta x$ pixels, from the top to the bottom with a step size of $\Delta y$ pixels. The shifting steps can be adjusted to satisfy the requirement of construction resolution as long as the primitive can be identified clearly in the camera image plane. For example, some critical region of the inspected part need higher resolution than other regions on the part.
4.2.2 Primitive Design

Designing a good pattern is critically important for achieving accurate correspondence with the optical triangulation technique, especially for one-shot method. The primitive design should satisfy the following constraints:

(a) monochromatic light and
(b) robust and accurate detection.

Taking the monochromatic light into account, the symbols should not contain color coding information. Hence, the symbols with geometrical feature will be adopted in this study, instead of the traditional color based coding patterns. Therefore, the image features should be accurately extracted. The center symmetric primitive, such as circle, disk etc., is widely used for fringe patterns and the intensity centroid of the symbol is regarded as the symbol.
location. However, it is inaccurate when shadows and occlusions occur.

As shown in figure 4.2, an inspected part with a step in the scene hides part of the projected pattern from the camera. Hence, a partial occlusion affects the centroid position. In this design, the strategy determines the symbol location with the corner of high contrast checkerboard. Because it only uses the local information of this hidden pattern, it does not influence the accuracy performance.

Figure 4.2: An example of occlusion.

Figure 4.3: The comparison of different primitives.
Thus, portions of the X-point region hidden by smudges or occlusions are disregarded with no significant impact on accuracy as shown in figure 4.2.

Additionally, the disk in an image does not contain any geometrical information other than their center location. The X-points has both a location and an orientation. This additional characteristic is used to discriminate the different primitives in pattern. As shown in figure 4.4, the red arrow denotes the orientation of the primitive and the corresponding value. The line connecting the neighbor primitives in the horizontal direction is considered as the base axis. The angles between the principle axis of the primitives and the base axis are $0^\circ$, $45^\circ$, $90^\circ$, and $135^\circ$ and the corresponding values are 0, 1, 2, and 3, respectively.

### 4.2.3 Primitive Detection and Codeword Decode

Once the pattern is projected on the scene, a single frame is captured by the camera. The image processing is simple because the primitives are designed on the black background and separated enough. There is a strong gradient in the image intensity around the primitive boundary. Hence, the contour of the primitive is easily detected and can be implemented in real-time. Additionally, it is less sensitive to reflectivity variation and ambient illumination than threshold based segmentation.

Consequently, the moment and principal axis method is adopted to recognize the primitive. The moment of the geometrical primitive is represented as:

$$M_{ij} = \int_{-\infty}^{\infty} x^i y^j f(x, y) dx dy$$  \hspace{1cm} (4.2)
The coordinate of the mass center is denoted as:

\[ X_m = M_{10}/M_{00}, Y_m = M_{01}/M_{00} \]  

(4.3)

A central moment is basically the same as the moment just described except that the values of \( x \) and \( y \) used in the formula are displaced by the mean values:

\[ CM_{ij} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - X_m)^i(y - Y_m)^j f(x, y) dxdy \]  

(4.4)

The angle between the principal axis and \( x \) axis is:

\[ \alpha = \frac{1}{2} \tan \left( \frac{2CM_{11}}{CM_{20} - CM_{02}} \right) \]  

(4.5)

Hence, the mass center of the contour is detected by Eq. 4.3 and is regarded as the initial rough location of the X-point. Then, the final location of the X-point is determined by the Harris algorithm within a local region for corner detection. Consequently, the principal axis is extracted by Eq. 4.5. In fact, two perpendicular axes can be extracted: the long and short axes. The long axis is regarded as the principal axis. Then, orientations of the principal axis and the base axis are compared to determine the value of the primitive. If difference between the principal and base axis is lower than an experimentally predefined threshold, then this primitive is determined. The image processing for the pattern with \( 5 \times 5 \) primitives is shown in figure 4.5.

Once each primitive value is determined, the next step is to find out the eight corresponding neighbor primitives to determine its codeword. The closest primitives in the eight directions are easily found with the help of the direction of the long axis. Next, the codeword of this primitive depends on its value and its neighbor values. For example, a primitive codeword is determined as figure. 4.6.

When each value of the primitive on the camera image plane is obtained, the corresponding pixel matching will be performed. The left upper primitive on the camera image plane
Figure 4.5: Image processing: (a) contour extraction; (b) location detection; and (c) principal axis identification.

Figure 4.6: Determination of the codeword for each primitive.

is selected as the matching window to find the corresponding primitive on the projector image plane. Then, the matching windows both on the camera and projector image plane are shifted to the next primitive from left to right, and from top to bottom. The procedure will be performed by a cursive search algorithm until all corresponding primitives are found.
4.3 Multi-shot Projection Coding Method

4.3.1 Problem Statement

Although some commercial 3D area sensors have been developed and been utilized in quality inspection, the technology still need to conquer several challenges to extend its further applications. One of the biggest challenges is its weakness against the optical properties of the test objects. Optically-challenging objects, such as painted parts with different color, materials with different reflection coefficients, and shiny punched metal work pieces, have to be pre-treated into diffuse reflection surface before inspection. The textures and colors of objects could influence a color based structured light. The difference of reflection coefficients can result in under- / over- exposure problem, which brings out a dilemma to adjust the camera aperture: a bright part requires a small aperture, while a dark area needs a big aperture. The shiny work pieces induce the internal reflection problem: a reflection light from a scene point is not only related with its incident light but also the incident lights in the neighborhood. A low illuminated area could be lighted up by a bright area in its neighborhood and thus results a wrong codeword. In summary, there is a demand of a structured light system that is fully capable to measure optically-challenging objects and to cope with the following problems:

1) Color and texture problem: An Object has different colors and textures. A color based codeword could be influenced from the test object and therefore generates a wrong codeword.

2) Shadow problem: A directly illuminated area can be viewed darker than an indirectly illuminated part. As a consequence, a high illuminated area could be considered as small intensity region and low illuminated area could be interpreted as high intensity region.

3) Under-/over- exposure problem: The intensities of the reflection lights from same illumination can vary a lot because of the different reflection coefficients. If reflection coefficients of the work pieces have significant differences, it is a dilemma to adjust the camera aperture.
4) Internal reflection: The reflection of light from a point in the scene is not only related with its own incident light but also with the incident lights in its neighborhood. The un-illuminated area can also result in high intensity in the camera image because of internal reflection.

Aiming to solve problems mentioned above, the paper developed a novel encoding and decoding method of a structured light based inspection system capable to sample dense 3D point clouds from optically-challenging objects. In order to successfully inspect the test objects, the developed structured light should simultaneously have the following properties:
1) A color invariance encoding and decoding algorithm.
2) Dynamical control the projector illumination to avoid under- / over- exposure problem and shadow problem.
3) Robust decoding against the internal reflection problem.

The symbols used in this part are explained here: $I$ means a pixel intensity. The subscript $P$ and $C$ respectively represent a projector and a camera. $(u, v)$ is usually used as a pixel location in the projector frame, and $(u', v')$ is used as a pixel location in the camera frame. For instance, $I_P(u, v)$ means the intensity value at pixel location $(u, v)$ in the projector frame.

4.4 Encoding and Decoding Strategy

3D reconstruction through structured light system requires the establishment of the correspondence table between two image devices: a projector and a camera, in which encoding and decoding play important roles. Digitized binary monochromatic patterns which are robust against image noise were utilized as encoded projection images. Traditional approaches assume that the inspected objects are perfect diffuse reflection. The strategy of the encoding and decoding is shown in figure 4.7. The first step is the structured light encoding in projec-
tor frame: codeword 0 is encoded as intensity 0 and codeword 1 is transferred into intensity 255. In inspection process, the patterns are projected into the scene, and are reflected into the camera. In the decoding process, a low intensity camera pixel is decoded as 0 and a high intensity pixel is translated into 1. Pixels in both frames shared with the same codeword are corresponded.

![Diagram](image)

**Figure 4.7: Traditional encoding and decoding.**

Once the inspection object is changed into an optically-challenging one, the encoding and decoding become more complicated. The intensity of a projection pattern needs to rapidly compensate the albedo effects from the optically-challenging objects. The so-called density adjusted projection patterns are the combination between GCLS and an intensity mask matrix $M_P$. The intensity of a projected pixel $I_P(u, v)$ can be expressed by the following formula:

$$I_P(u, v) = I_P^{GCLS}(u, v) \cdot M_P(u, v)$$  \hspace{1cm} (4.6)

$I_P^{GCLS}(u, v)$ denotes the intensity of a traditional GCLS pattern at location $(u, v)$ in a projector frame. $M_P(u, v)$ can be considered as a weighted coefficient for the code $I_P^{GCLS}(u, v)$: if $M_P(u, v) = 0$, the pixel is turned off; if $0 < M_P(u, v) < 1$, the intensity of the projection is reduced; if $M_P(u, v) = 1$, the intensity of the pixel remains the same; if $M_P(u, v) > 1$, the intensity of the projection is enhanced. Internal reflection phenomena can be eliminated if its incident light is reduced. The camera could view a dark / shadow region clearly if the incident lights are enhanced. In summary, the projection mask
works as a feedback function from the camera receiver to the projector.

In order to get the scene adapted mask matrix $M_P$, the camera frame photo response function $f_{ph}$ and the crosstalk $f_{cr}$ from the camera frame to the projector frame are needed. The camera frame photo response function $f_{ph}$ in Eq. 4.7 represents the relationship between the projection $I_P(u, v)$ intensity and captured intensity $I_C(u', v')$, which is mainly costed by the albedo effect of the scene.

A proper value $I_P$ of the projection intensity is first derived from the inverse camera photo response function to ensure the camera pixel $(u', v')$ can receive a suitable intensity. The crosstalk function $f_{cr}$ in Eq. 4.8 allocates this proper value $I_P$ into the correspondence projector pixel $(u, v)$ to ensure the radiance from the projector pixel $(u, v)$ does go into the camera pixel $(u', v')$. After this two tasks are accomplished, the projection mask can be derived.

$$I_C(u', v') = f_{ph}(I_P(u, v))$$  \hspace{1cm} (4.7)

$$(u', v') = f_{cr}(u, v)$$  \hspace{1cm} (4.8)

Considering the camera as the feedback receiver, the projected intensity $I_P(u, v)$ can be expressed by:

$$I_P(f_{cr}^{-1}(u', v')) = f_{ph}^{-1}(I_C(u', v'))$$  \hspace{1cm} (4.9)

In this study, we normalized the projection and capture chain by the following assumption: a projection of intensity 255 will result in intensity 255 in the camera frame. Therefore, the intensity mask $M(u, v)$ equals:

$$M(u, v) = M(f_{cr}^{-1}(u', v')) = \frac{f_{ph}^{-1}(255)}{255}$$  \hspace{1cm} (4.10)

Since the intensity mask $M(u, v)$ should be dynamically adjusted by the captured value $I_C(u', v')$, $f_{cr}$ is first utilized to transfer the camera pixel location $(u', v')$ into projector pixel
location \((u, v)\). \(f_{ph}^{-1}(255)\) specifies a desired projection value, which can make the received value as 255. The ratio between the desired projection value and 255 is the projection intensity mask value.

The flow chart of the inspection process is displayed in figure 4.8. It first derives the intensity mask and then applies it to the structured light and at last finishes the 3D reconstruction.

There are generally three steps for the inspection process:

(1) Initial projection to derive the camera photo response function, the initial guess of the crosstalk function and the density mask.
(2) Advanced projection of the intensity adjusted pattern to modify the crosstalk function and update the density mask.

(3) Final projection to obtain the correspondence and finish 3D reconstruction.

The details of steps are explained in the following sections.

4.4.1 Derivation of the Camera Photo Response Function for a Given Scene

The high illumination intensity reflection light can cost the camera saturation problem – some highly saturated pixels spill over and affect neighbor pixels’ intensity value, such as a shiny spot in a picture. On the other hand, low illumination intensity reflection light may cause the camera fail to capture the incident light. In order to simultaneously capture the reflection lights from both high and low reflective surfaces, a proper feedback algorithm is needed. A photo response function reflects the radiometric chain from the projector image frame to the camera image frame. The inverse function of the photo response function is used as the feedback function to eliminate scene albedo effects. A photo response function basically contains three parts: radiance of a given projector pixel $f_P$, reflection radiance from incident radiance $f_R$, and intensity of a camera pixel from in the reflection radiance $f_C$. In this study, all of three functions are approximated as linear functions.

The radiance $R_P(u, v)$ from a projector pixel $I_P(u, v)$:

$$
\begin{cases}
R_P(u, v) = f_P (I_P(u, v)) = k_1(u, v)I_P(u, v) + k_2(u, v) \\
0 \leq I_P(u, v) \leq 255
\end{cases}
$$

(4.11)

The reflection radiance $R_C(x, y, z)$ at a scene point $(x, y, z)$ from the incident radiance $R_P(u, v)$ can be written as:

$$
R_C(x, y, z) = f_R (R_P(u, v)) = k_3(x, y, z)R_P(u, v) + k_4(x, y, z)
$$

(4.12)

The camera Intensity $I_C(u', v')$ from the reflection radiance $R_C(x, y, z)$
\[
\begin{align*}
I_C(u',v') &= f_C(R_C(x,y,z)) = k_5(u',v') IP(u,v) + k_6(u',v') \\
0 &\leq I_C(u',v') \leq 255
\end{align*}
\] (4.13)

where \(k_1(u,v), k_2(u,v), k_3(x,y,z), k_4(x,y,z), k_5(u',v')\), and \(k_6(u',v')\) are unknown coefficients in the linear functions. The radiometric chain through different projector pixel, different scene point, and different camera pixel varies via the subscripts of the coefficients.

Eq. 4.11 and Eq. 4.12 are substituted into Eq. 4.13 and then the equation is treated as a linear function since all three equations are linear functions:

\[
\begin{align*}
I_C(u',v') &= k_7(u',v') IP(u,v) + k_8(u',v') \\
0 &\leq I_C(u',v') \leq 255 \\
0 &\leq IP(u,v) \leq 255
\end{align*}
\] (4.14)

\(k_7(u',v')\) and \(k_8(u',v')\) are the combination of the previously coefficients. The radiometric chain problem is treated as a black box problem. Only \(k_7(u',v')\) and \(k_8(u',v')\) are needed to adjust the projection intensity through received intensity. Coefficients from \(k_1\) to \(k_6\) are not necessary for the photo response function.

In order to solve the photo response function in Eq. 4.14, two tasks need to be accomplished:

(a) The coefficients \(k_7(u',v')\) and \(k_8(u',v')\). These two are used to find out what the suitable projection intensity is for a camera pixel \((u',v')\) to avoid under- / over- exposure.

(b) The crosstalk function between projector pixel \((u,v)\) and camera pixel \((u',v')\). This function is utilized to find out where the suitable projection intensity should be located in the projector frame.

Basically, the solution of the first task answers what the projection value is, and the second one answers where this value should be located. Since it is very difficult to solve both tasks simultaneously, task (a) is first peeled by projecting single intensity images: all the projector pixels equal to same intensity value \(IP\). Since all the pixel values in the projector frame are same, the crosstalk can be ignored. No matter which pixel projector is crosstalked with the camera pixel \((u',v')\), the intensity is always the same. Equation 4.14 can be changed
Single intensity images with the value equaling 0, 50, 100, 150, 200, 250 are projected to the test objects. Figure 4.9 gives an example when an 100 intensity image was projected. 5 arbitrary points from A to E in the received image are selected and marked as different colors to check their intensity variations. Their received intensities are given in figure 4.10. x axis is the projection intensity in projector frame, and y axis represents the received intensity in the camera frame. The color lines represent the photo response curves for different points in the scene. A is a point on a black object, so its received intensity value is always very low. Even projecting intensity 255 onto the point, its received value is still a little less than 50. D and E are two traditional shiny points, their correspondence camera pixels can be saturated very easily. C can be viewed as a 'normal' diffuse point, neither too shiny nor too dark. B is a special case for points on a black object. When the incident angle and reflection angle reach certain condition, black object can also result in 'mirror-like' effect, very sensitive to the incident light. The parameters $k_7$ and $k_8$ can be regressed after the projected intensities and received intensities are substituted into the Eq.4.15.

The result of the proper projection values in the camera frame is given in figure 4.11. Similar as a film of a camera, the brighter the camera receives, the darker the proper projection value is.

### 4.4.2 Initial guess of the crosstalk function

The result given in figure 4.11 only solves how big the projection value should be, it does not answer where projection value should be located. $I_P$ is calculated by Eq. 4.14 but the index number $u$ and $v$ in projector frame are missing.

The initial guess of the crosstalk function assumes that the test scene is a flat surface.
on the inspection table. The crosstalk between two image frames becomes a homography mapping in Eq. 4.16.

$$P_P = H_W^P P_W = H_W^P H_C^W P_C = H_C^P P_C$$  \hspace{1cm} (4.16)

As shown in figure 4.12 (a), $I_P$ and $O_P$ form a pinhole model of a projector; $I_C$ and $O_C$ form a pinhole model of a camera. $P_P = (u_P, v_P, 1)$, $P_W = (x_W, y_W, 1)$, and $P_C = (u_C, v_C, 1)$ are the points on projector image plane, the inspection table, and the camera image plane, respectively. $H_W^P$, $H_C^W$ and $H_C^P$ are three $3 \times 3$ homography matrixes between two of the planes.

### 4.4.3 Advanced Projection Pattern Generation

Combining the initial guess of the cross talk function and the camera photo response function, the initial projection mask can be derived via Eq. 4.10. The result is shown in figure 4.13 (a). However, in reality the real scene is different from the initial assumption, and the location
Figure 4.10: The photo response curve for 5 points in the last figure.

of the corresponding projector pixel will not be exactly at \( P_P = (u_P, v_P) \) but at somewhere near to this location. Shown in figure 4.12 (b), when the real scene is higher than the assumption, the real location will be \( P_{P1} \). If the real scene is lower than the assumption, the real location could be at \( (P_{P2}) \). Assuming the possibility of the real location is a Gaussian distribution around pixel \((u_P, v_P))\), the projection mask is dilated by a certain radius and is displayed in figure 4.13 (b). A combination between GCLS pattern with the projection mask is shown in figure 4.13 (c).
4.4.4 Interpolation Model for Diffused Reflection Region and Extrapolation Model for Internal Reflection Region

Both crosstalk function and the correspondence for 3D recovery need the binarization of the received images into codeword 1 and 0. The difference between these binarizations lies in the accuracy requirements. Binarization in pixel level accuracy is enough for a crosstalk function to control the illumination, but the correspondence for 3D recovery needs to reach sub-pixel level accuracy. However, in the other aspect the illumination condition in finding out correspondence is better than the condition in finding out the crosstalk function, because the illumination condition is further improved after the crosstalk function is calculated. In all, a robust binarization method is critical in the measurement of optical-challenging objects. An extrapolation model for internal reflection region and an interpolation model for diffuse
reflection region are proposed.

Binarization methods of the received image can be roughly classified into three branches. 1) Naive binarization: based on fixed intensity threshold. 2) Adaptive binarization: based on different thresholds for different pixels, in which a pixel’s threshold is determined by the overall weight of its neighborhood. 3) Dual projection binarization method: a pixel’s binarization is based on whether the projection or its inverse is brighter [34] [45]. Method three is more robust against image noise and has been widely used in 3D measurement.

In this study, a modified dual projections robust binarization method is used to establish the crosstalk function. In stead of unitizing the stripes for line shifting, the edge between every two stripes is employed. Sub-pixel interpolation for the edge location is added to increase measurement accuracy. An estimation algorithm for the edge location is also added when the camera saturation or the internal reflection occurs. As shown in figure 4.14 (a) and (b), images of a stripe pattern and its inverse are used to do binarization analysis. Black and
Figure 4.13: a) The initial projection mask; (b) the dilated projection mask; (c) one example of a projection pattern.
Figure 4.14: (a) A received image of projected pattern; (b) a received image for the inverse pattern; (c) interpolation of the edge location; (d) estimation of the edge location in a strong illumination condition.

White stripes are the projection pattern, and the highlighted blue lines indicate the edges. The edge falls into a camera image pixel. Therefore, an interpolation of sub-pixel model is needed to improve the measurement performance.

Figure 4.14 (c) illustrates the binarization method and edge interpolation for the diffuse reflection surface. $y$ axis indicates the received intensity in camera frame, and $x$ axis is the image coordinates in the vertical direction. The red line indicates the projected pattern and the blue line means the inverse pattern. $A$, $B$, $C$ and $D$ are the four intensity values to do binarization and edge interpolation. Since intensity $A$ is bigger than its inverse $B$ by a certain threshold, pixel $n$ is assigned as 1; for the same reason, pixel $n + 1$ is decoded as 0. Assuming that the intensity change within a pixel is a linear function, the location of the edge $X$ can interpolated by intensities $A$, $B$, $C$ and $D$. Although the linear assumption may not be perfect for all the cases, the error can still be bounded within one pixel. The model in
Figure 4.14 (c) is utilized in both calculating the crosstalk and the 3D correspondence, since it can go to sub-pixel accuracy.

\[ X = n + \frac{A - B}{(C - B) - (D - A)} \]  \quad (4.17)

In the advanced projection, the over exposure problem and the internal reflection problem can not be totally avoided. The calculation model of the edge location is changed into extrapolation from interpolation, which is shown in figure 4.14 (d). In the over exposure condition, the incident light intensities of the white stripes are bigger than 255. So the blue line raises from point \( B \) and reaches the maximum intensity 255 before reaching point \( C \). The red line also drops from dashed line into the solid line. As a consequence, point \( A \) changes into \( A' \) and \( C \) moves into \( C' \). \( BC' \) and \( DA' \) will not intersect each other until their extensions merge.

If the internal reflection happens, things could become more complicated. Point \( C' \) could illuminate its neighborhood \( C'' \) by internal reflection and causes line \( C'B \) move backwards to \( C''B' \). Point \( A' \) illuminates the point \( D \) and make the line \( A'D \) move forward to \( A''D' \). Therefore, the step distance \( S \) between point \( B' \) and \( D' \) changes into a number bigger than 1 pixel. If we view from a received image, it means that the white stripe erodes the dark stripes. The white stripes become wider and the dark stripes become narrower. Extrapolation model assumes that the intensity change within several pixels is a still linear function, so the stripe edge \( X \) can be derived via Eq. 4.18.

\[ X = n + \frac{S (A'' - B')}{(C'' - B') - (D' - A'')} \]  \quad (4.18)

The assumption that extrapolation used is stronger than the interpolation model assumed. The linear assumption is not only within one pixel but also within several pixels. In reality, linear function approximation within one pixel only contains little error, but the linear function approximation within several pixels can produce a certain amount of errors. When the width \( S \) increases, the estimation location of the \( X \) contains more errors. Since
the error for extrapolation model can not be bounded within one pixel, this model is only used in approximating the crosstalk function but not in correspondence calculation.

Figure 4.15 shows the final projection mask after crosstalk function is derived. The intensity mask could perfectly control the projection intensity to ensure the received intensity between under-exposure and over-exposure. It could also minimize the internal reflection by reducing the incident light intensity in the shiny area. From the figure, we could also see that the projection mask in the blind areas contains a lot of noise, for instance the zoomed parts (a) and (b). If we compare same parts with the original image ?? (a), we could see that these areas are the blind areas, either the projector can not illuminate these locations or the camera can not view the projected patterns there. Since there is always one image device failing to detect the blind area, correspondence pixels can not be generated to rebuild 3D points. Therefore, the projection mask error in these areas will not affect the 3D calculation results.

![Figure 4.15: The final projection mask.](image-url)
4.5 Verification of the Projection Mask Performance

As shown in figure ?? (a), three work pieces: one black part, one dark part, and one shiny part, are used to test the designed system. In order to verify the idea of this study, two aspects of received results are given and discussed in this part: a) the verification of the projection mask and b) the verification of the interpolation and extrapolation models.

4.5.1 the Verification of the Illumination Adjustment via Projection Mask

In this study, a dynamic scene based projection mask for adjusting the incident light intensity is proposed against optic challenging parts. In order to verify this idea and validate the mask performance, the received images before and after the mask are provided.

Figure 4.16, 4.17 and 4.18 illustrate the effect of projection mask in capturing the images of the scene. Figure 4.16 is the received image without the projection mask. There are several areas still having over-exposure and internal reflection problem. Since the aperture of the camera is adjusted to the maximum, almost all the work pieces are looked as shiny parts. Figure 4.17 is the captured image after the initial projection mask is added. Because the assumption that scene is flat is too strong, the projection mask could not adjust the incident light perfectly and some areas are still shiny. Figure 4.18 is the result after the final projection mask is applied. From the image we could see that all the incident lights are dynamically adjusted and the shiny phenomena are disappeared.

4.5.2 the Verification of the Interpolation and Extrapolation models

Figure 4.16, 4.17 and 4.18 only provide the general performance of the projection mask. More details intensity information are given to illustrate the followings:
1) Verification of the linear interpolation model which is utilized to get sub-pixel accuracy. The interpolation model employs the intersection point to find the intensity edges. The model needs to be proved by the real received data.

2) Demonstration of the extrapolation model which is proposed to explain the internal reflection and over-exposure issues. The real intensity distribution in the internal reflection region should prove the extrapolation model.

3) Verification of the result of the extrapolation model. The extrapolation model is not only applied to prove the internal reflection but also utilized to find the stripe edges in the shiny region. The corresponding experimental result needs to be provided.

4) The study proposed that the adaptive projection mask could reduce / eliminate internal reflection. The received intensity distributions before / after projection mask need to be compared to demonstrate the idea.
Figure 4.19(a) and 4.19(b) are two detailed views of a black part when stripe pattern and its inverse without projection mask are projected respectively. The internal reflection and the reflection coefficient of the black work piece are both low. Therefore, these detail views are selected to verify the sub-pixel linear interpolation model. The widths of W/B stripes $AB$ are the same with the widths of B/W stripes $A'B'$ equaling to 39 pixels. The high illuminated stripe does not erode the low illuminated stripe, which indirectly proves the internal reflection is small and thus the surface could be viewed as diffuse reflection. The pixels along the vertical line shown in figure 4.19(a) and 4.19(b) are scanned. The binarization results of the selected pixels are given in table 4.1. The pixel’s intensity curves around point $B / B'$ along the vertical direction are shown in figure 4.19(c) to verify the interpolation model. The red curve is the intensity distribution along red line in figure
4.19(a), and the blue curve is the intensity distribution along blue line in figure 4.19(b). It provides more clear information about the intensities variations. If the red curve is above the blue curve by a certain threshold, the pixels are decoded as "high" equaling as 1, otherwise the pixels are classified as "low" equaling as 0. The intersection point means edges between high and low illuminated regions. The distribution of the real received intensity is similar with the internal reflection developed in figure 4.14 (c). The pixel intensities at pixel location 899 are 61 and 34 respectively; and the pixel intensities at location 900 are 29 and 64 respectively. The location of the intersection point (x axis coordinate) is 899.4, after applying the model in Eq. 4.17. Comparing intensity distributions in figure 4.19(c) and the curves in figure 4.14 (c), the interpolation model can be verified.

In order to check the over-exposure and internal reflection phenomena, detail views of
Figure 4.19: (a) A selected detail view of the black part when projecting pattern without the mask; (b) the same view when projecting the inverse pattern; (c) the intensity curve around point $B / B'$. 
Figure 4.20: (a) A selected detail view of the shiny part when projecting pattern without the mask; (b) the same view when projecting the inverse pattern; (c) the intensity curve around point $C / C'$.
Table 4.1: Pixels binarization without projection mask

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>High</th>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pixel</td>
<td>821-860.1</td>
<td>860.1-899.4</td>
<td>899.4-938.7</td>
<td>938.7-975</td>
</tr>
</tbody>
</table>

the shiny part is displayed in figure 4.20(a) and 4.20(b). As explained in the extrapolation model, if the internal reflection and over-exposure issues happen, the $W$ stripe will erode the $B$ stripe, resulting that the $W$ stripe’s width is bigger than the $B$ stripe’s. The pixel distance $CD$ is 33, and the pixel distance $C'D'$ is 38. The width difference is the major harm from the internal reflection and the over-exposure issues. More clear information cloud be found in figure 4.20(c). Instead of an intersection point, the red curve and the blue curve share a common short line segment. The pixels within the common short line are always highly illuminated and can not been classified by traditional methods, which results the failure to sample 3D coordinates for shiny regions. The extrapolation sub-pixel model in figure. 4.14 (d) fits such a situation. The intensity values of the pixel 1292 and 1293 on the blue curve are 226 and 255 respectively; and the intensity values of the pixel 1294 and 1295 on the red curve are 255 and 155 respectively. Applying the Eq. 4.18, Strip edge could be estimated at location 1293.7 by extrapolation. The distributions in figure 4.20(c) could verify the extrapolation model in figure 4.14 (d). The extrapolation model result can successfully separate the high / low illuminated regions. The high / low illuminated region is decoded into 1 / 0.

As a comparison, the same views but with projection mask are shown in figure 4.21 and 4.22. Figure 4.21 is applied to verify that the illumination variation costed by the projection mask will not affect the pattern edge detection in the diffuse reflection region. Figure 4.22 is employed to show that illumination variation costed by the projection mask reduce / eliminate the internal reflection problem in the shiny region. Figure 4.21(a) and 4.21(b) are the two detail views of the black part when the adaptive mask is utilized. The stripe width
Figure 4.21: (a) A selected detail view of the black part when projecting pattern with the mask; (b) the same view when projecting the inverse pattern; (c) the intensity curve around point $B / B'$.
Figure 4.22: (a) A selected detail view of the shiny part when projecting pattern with the mask; (b) the same view when projecting the inverse pattern; (c) the intensity curve around point $C / C'$.
$AB$ still equals to $A'B'$ as 39 after projection mask is applied. The binarization results are shown in table 4.2. The result is almost the same with table 4.1. Figure 4.21(c) is the intensity curves around point $B / B'$. The pixel intensities at location 899 are 80 and 28 respectively; and the pixel intensities at location 900 are 25 and 62 respectively. Applying the model in Eq. 4.17 again, the intersection pixel location is 899.6. Compared to results from figure 4.19, we could conclude that the illumination variation costed by the projection mask will not affect the decoding in the diffuse reflection region.

Figure 4.22(a) and 4.22(b) are two detail views of the shiny part when projecting stripe pattern and its inverse with mask. The stripe widths of 'W' / 'B' are same: $CD = C'D' = 35$. Compared to result from figure 4.20, we could conclude that the internal reflection is eliminated. Same conclusion can be drawn from figure 4.22(c). After the illumination is adjusted by the mask, there is an intersection between two curves. Instead of applying the extrapolation model, the interpolation model can be utilized to find the edge. The intensities at pixel 1293 are 133 and 80 respectively; and the intensities at pixel 1294 are 83 and 141 respectively. The intersection point can be found, 1293.5, by Eq. 4.17.

Table 4.2: Pixels binarization without projection mask

<table>
<thead>
<tr>
<th>Pixel</th>
<th>Low</th>
<th>High</th>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>821-860.2</td>
<td>860.2-899.6</td>
<td>899.6-938.7</td>
<td>938.7-975</td>
</tr>
</tbody>
</table>

In summary, the illumination variation costed by the projection mask will eliminate internal reflection in shiny region without any bad effect to the diffuse region.

4.6 Chapter Summary

In this chapter, two surface coding methods are introduced: single-shot algorithm and multi-shot algorithm. single-shot algorithm could greatly reduce the acquisition time and thus obtain the ability to measure a moving object. The difficult of this method is to encode all
the codeword within one image. A spatial-neighborhood method was developed for single-shot method, the mass center and orientation of each primitive was utilized as its value each. Each primitive’s value associated with its eight neighbors’ value forms a codeword. Compared with traditional color based codeword, this method is more robust, as it can get ride of the color effect from the scene.

The disadvantages of the single-shot method were the poor density and the weakness against optical properties of the test objects. A multi-shot method was developed to inspect optically-challenging objects without surface pre-treated. Compared with single-shot method, it is more robust and dense but less efficient. A projection mask that dynamically compensate the albedo effects from the scene was developed. A proper feedback was used to control the illumination of the projector. Therefore, both shiny and black materials were able to be measured simultaneously. An extrapolation model combined projection mask was applied to avoid the internal reflection issues. Compared with the previous methods, the proposed method was much more robust, efficient, and flexible.
CHAPTER 5

Registrations between Different Coordinates after 3D Reconstruction

5.1 Registration between different 3D coordinates

The previous chapters describes the sensor working principle and its calibration. The measured data are usually given as "point cloud", which can be compared to the designed surface shape, such as the Computer Aided Design (CAD) model, for error analysis. The deviation between the point cloud and CAD can be displayed by a color-coded error map. The error distribution map is used to evaluate the part quality by comparing the results with acceptable Geometry Dimension and Tolerance (GDT).

In order to finish the tasks of quality inspection, there are still two questions remained:

1) Registration between sensor frame and common frame: Each point cloud is obtained in individual sensor frame. They need to be stitched together for representing the whole work piece It means all the point clouds need to be represented in one common frame. The transformation matrix between each sensor frame to the common frame is first calculated and then is applied to update each point cloud. The updated point clouds merges into one entire point cloud which can represent the whole work piece. The final point cloud can be
further compared with CAD model to find out the color coded manufacture distribution error map. Since each sensor frame is steady with each other, the transformation between sensor frame to the common wont not change after registration calibration. Therefore, this registration is also called as off-line registration calibrated.

2) Error map generation. The mathematical definition of the deviation between data sets: the triangulated CAD model and the point could. The deviation along the surface normal of each triangulated surface of CAD and the closest point from point cloud is applied to describe the error distribution. The color code from blue to red is used to visualize the distribution, in which blue denotes the CAD is lower than the point cloud and red means the CAD is higher than the point cloud. Error map generation requires:

a) the on-line registration between the final point cloud and the CAD model;

b) the definition of the deviation between two data sets.

Since the position of the work piece is not fixed, the registration matrix between the final point cloud and the CAD model is changing from time to time. The registration matrix needs to be calculated after inspection within a limited time. Therefore it is also called on-line registration calibration. The deviation between final point cloud and the CAD model is defined by the distance between closest point in point cloud and the triangulated surface in CAD. If the triangulated CAD model is above the closest point in final point cloud, the deviation is defined as negative, indicated as blue color; in the other side, the deviation is considered as positive, indicated as red color.

5.2 Off-line Registration Calibration from Sensor Frame to Common Frame

The tasks in this off-line registration calibration are to find the rigid transformation matrix to update all point clouds in their individual frames into a common frame.

The result point cloud $P^S_j$ in each sensor fame can be written as:
\[ P^S_j = \left\{ p^S_j(x_i, y_i, z_i) \in R^3 \right\}, j = 1, 2, 3, 4 \] (5.1)

The registration to transfer four point clouds into a common frame is defined as:

\[ P^C = \bigcup_{j=1}^4 \left( T^C_{S_j} P^S_j \right) \] (5.2)

\(T^C_{S_j}\) is the transformation matrix from sensor frame \(j\) to the common frame. The task is to find the transformation matrix \(T^C_{S_j}\) for each sensor frame. In order to solve \(T^C_{S_j}\) in Eq. 7.3, three tasks need to be accomplished.

1) Two point clouds \(P^S_j\) and \(P^C\) need to be derived first. Point set \(P^C\) can be determined by pre-measured gauge, and \(P^S_j\) is the sensor measurement result.

2) The correspondence pairs from two data sets are needed. The well-known ICP algorithm is used to get correspondence after the rough alignment. The closest points are assumed as a corresponding pairs between two data sets. The final alignment can be found iteratively between two sets.

3) Derive the rigid transformation matrix \(T^C_{S_j}\) after the correspondence is established. The rigid transformation matrix consists of two parts: the rotation transformation matrix \(R\) and translation vector \(t\). Matrix \(R\) is a unit orthogonal matrix, with three degree of freedoms. The derivation of \(R\) is a little complicated to follow the unit, orthogonal, and three degree of freedoms constraints.

### 5.2.1 Calculation of Rotation Matrix via Quaternion Algorithm

A quaternion is an extension of a complex number, which is defined in:

\[ q = q_0 + q_1 + jq_2 + kq_3 \] (5.3)

, where \(i, j, k\) are three orthogonal 3D vectors as the bases of 3D Euclidean space. \(i^2 = j^2 = k^2 = ijk = -1\). A unit quaternion \(q\) has norm of one and can be expressed by:
\[ q = \cos \left( \theta/2 \right) + \sin \left( \theta/2 \right) \vec{V} \]  

(5.4)

, in which \( \theta \) is the rotation angle and \( \vec{V} = [x_V, y_V, z_V]^T \) is a unit vector representing the rotation axis. Then, a unit quaternion \( q \) represents an arbitrary rotation with parameters \( V \) and \( \theta \) in 3D space. A quaternion based rotation matrix, therefore, becomes:

\[
R_q = \begin{bmatrix}
q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1q_2 - q_0q_3) & 2(q_1q_3 + q_0q_2) \\
2(q_1q_2 + q_0q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2q_3 - q_0q_1) \\
2(q_1q_3 - q_0q_2) & 2(q_2q_3 + q_0q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2
\end{bmatrix}
\]  

(5.5)

A point \( P = [x, y, z]^T \) can be considered as a pure quaternion \( (q_0 = 0) \). Its rotation can be expressed as:

\[ R_q P = q \circ P \circ q^* \]  

(5.6)

, where \( \circ \) denotes a quaternion multiplication and \( q^* \) is the conjugate of the quaternion \( q \).

The rigid transformation between different frames can be solved in two steps via quaternion algebra:

\[
\begin{bmatrix}
x_c \\
y_c \\
z_c
\end{bmatrix} = R_q \begin{bmatrix}
x_s \\
y_s \\
z_s
\end{bmatrix} + t = q \circ \begin{bmatrix}
x_s \\
y_s \\
z_s
\end{bmatrix} \circ q^* + t
\]  

(5.7)

, in which \( P_s = [x_s, y_s, z_s]^T \) is a point in sensor frame, and \( P_c = [x_c, y_c, z_c]^T \) is a point in common frame. We can obtain:

\[
(P_c + P_s) \times \eta - (t) \times \eta - t = P_s - P_c
\]  

(5.8)

, where \( (\cdot ) \times \) represents a skew symmetric matrix and \( \eta = \tan(\theta/2) \vec{V} \).

Let \( M_i \equiv P_s^i - P_s^{i+1} \) and \( N_i \equiv P_c^i - P_c^{i+1} \), Eq. 5.8 changes into:

\[
(M_i + N_i) \times \eta = M_i - N_i, \ i = 1, 2, ..., n - 1
\]  

(5.9)
, in which \( n \) is number of points in both data sets. After \( \eta \) is solved, the rotation axis \( \vec{V} \) and the rotation angle \( \theta \) can be obtained by:

\[
\vec{V} = \frac{\eta}{||\eta||} \quad (5.10)
\]

\[
\theta = 2a \tan \left( \frac{\eta_{\text{max}}}{V_{\text{max}}} \right) \quad (5.11)
\]

Therefore, the rotation matrix \( R_q \) and the translation vector \( t \) are found by:

\[
R_q = \begin{bmatrix}
    x_V \lambda \theta + c \theta & x_V y_V \lambda \theta - z_V s \theta & x_V z_V \lambda \theta + y_V s \theta \\
    x_V y_V \lambda \theta + z_V s \theta & y_V \lambda \theta + c \theta & y_V z_V \lambda \theta - x_V s \theta \\
    x_V z_V \lambda \theta - y_V s \theta & y_V z_V \lambda \theta + x_V s \theta & z_V \lambda \theta + c \theta
\end{bmatrix} \quad (5.12)
\]

\[
t = P_c - R_q P_s \quad (5.13)
\]

where \( \lambda \theta = 1 - \cos(\theta) \), \( c \theta = \cos(\theta) \), and \( s \theta = \sin(\theta) \). In order to calibrate a rigid transformation \( T^c_s \), the 3D coordinates of correspondence points in both sensor frame \( s \) and common frame \( c \) need to be determined. Coordinates of points in the common frame \( c \) can be predetermined by making a calibration gauge, and the points in the sensor frame are the measurement results of the gauge after the sensor is calibrated. The correspondence between two data sets: data in common frame and data in sensor frame is established by Iterative Closet Point (ICP) method.

### 5.2.2 Iterative Closest Point (ICP) based Registration

Based on the accuracy of the merge result, registration can be roughly considered to two parts: coarse alignment and fine adjustment. Coarse alignment derive the transformation matrix based on a few feature points in two data sets. Therefore, the result accuracy is not very high. It is usually used for the applications with low accuracy requirement, such as navigation or initial alignment for fine adjustment. The ICP fine alignment algorithm [?] [?] is based on searching of the pairs of nearest points in both data sets and estimates
a coordinate transformation to align them together. The whole process is iteratively executed until the registration error reaches the threshold, in which the overall registration error is minimized.

ICP algorithm is shown in figure 5.1 as a recursive process. The algorithm is to find the rigid transformation matrix \( T_p^q \) between set \( P \) with \( N_p \) points and set \( Q \) with \( N_q \) points. \( X_p \) and \( X_q \) are two points from set \( P \) and \( Q \), respectively. The algorithm includes:

Step 0) Generate the coarse alignment by a few correspondence features between to data sets. These features can be defined from artificial markers, colors, and curvatures. Update set \( T_0Q \rightarrow Q_0 \).

Each recursive loop of the ICP algorithm can be described as follows:

Step 1) Set up the correspondence pairs of points by minimizing the cost function, the Euclidean distance between every two points in two data sets. The Euclidean distance between an arbitrary point \( X_Q \in Q \) and all the points in the set \( P \) is calculated. The point from \( P \) which has smallest cost function, Euclidean distance, is considered as the correspondence point of point \( X_Q \). The algorithm goes from the first point in \( Q \) to the last one and create a list \( L_k \) with \( N_q \) pairs of correspondence points from two sets. \( k \) indicates the current loop number.

Step 2) Derive the rigid transformation matrix \( T_k \) which consists of rotation matrix \( R_k \) and \( t_k \), via quaternion method explained in previous section.

Step 3) Update the date set \( Q_k \) into \( Q_{k+1} \) via matrix multiplication: \( T_kQ_k = Q_{k+1} \).

Step 4) Calculate the average error \( e_k \) via:

\[
e_k = \frac{1}{N_q} \sum_{i=1}^{N_q} \sqrt{(T_kX_q^2(i) - X_p^2(i))} \tag{5.14}
\]

Step 5) If the decrement between \( e_{k-1} \) and \( e_k \) is smaller than \( Threshold1 \), or \( e_k \) is smaller than \( Threshold2 \), the algorithm output the results: \( P, Q_n, T_f = T_nT_{n-1}, ..., T_0, n \) means the total number of loops, and \( T_f \) means the final transformation matrix. Otherwise the algorithm goes to Step 1) and starts another loops. The algorithm was proved to converge.
Iterative Closes Point Algorithm
Registration

Set $P, N_p$ points
Set $Q, N_q$ points

Initial condition: coarse alignment

Set up correspondence pairs by closest points

Derive the rigid transformation matrix: $T_k$

Update data set $T_kQ_k -> Q_{k+1}$

Calculate The average error $e_k$

Threshold

No

Yes

Finish: $P, Q_n, T_f = T_n T_{n-1} ... T_0$

Figure 5.1: ICP algorithm

monotonously to the local minimum [49] [50].

ICP algorithm is widely used in point cloud registration. However, it can not be directly
applied for current application in this thesis due to several limitations and practical issues.

a) Correspondence selection due to point density. ICP algorithm assumes that two data set represent the same shape, and therefore the rigid transformation is able to be regressed via minimizing the cost function $e$. Point density of a data set represents the digitized sample rate of a given 3D shape. If the point densities of two data sets have significant difference, such as a measured point cloud and a triangulated CAD model, data set with less point density is preferred as set $Q$ in the ICP algorithm. $N_q$ pairs of correspondence points will be involved in the calculation. Otherwise, $N_p$ pairs of correspondence points are used, which greatly reduce the algorithm efficiency. Selecting $Q$ as the data set with high density also results that one point in $P$ is the corresponding point for multiple points in $Q$ and then may increase the error.

b) Interpolation points for the correspondence. Although two data sets sample the shape, the assumption that the closest point is the corresponding point is still too strong in some case. The corresponding point of a point in set $Q$ can be the interpolation point from set $P$ instead simply selecting the closest point from $P$. In order to further increase the accuracy of the algorithm, interpolation points may be needed.

c) The consistence of the registration. The coarse alignment is considered by the initial condition of ICP. Since ICP algorithm can only insure to reach the local minimum instead of global minimum, the final position varies when initial condition changes. ICP registration method based error map generation requires that the work piece’s initial position should be almost same. The error maps of the same work piece and same CAD should be same from time to time. If the standard deviation of the error maps exceed the threshold, the result error maps become meaningless.

d) The definition of the cost function. ICP algorithm utilizes the Euclidean distance between two points as the cost function, resulting that the closest points are the corresponding points to calculate transformation $T$. The cost function is not unique, the distance from point to plane for instance, can also be used as the converge criteria depending on the appli-
cation requirement. The minimization of the cost function results that the distance between points in set $P$ and planes in set $Q$ reaches the local minimum.

e) Calculation time of the algorithm. On-line calibration requires the registration between the final point cloud and CAD model can be finished within limited time. The searching for corresponding points in each loop requires $N_p \times N_q$ times calculation. If the size of the data is big, for instance around 100,000 points for each, the calculation time can reach more 30 minutes to complete the algorithm. Octree space labeling method and multi-resolution method are used to speed up the algorithm.

f) Objected-oriented registration. The registration between the final point cloud and the CAD model is different from the traditional registration application. Conventional registration assumes that two data sets are from the same shape, and then an algorithm that minimizes Euclidean distance can be applied. The assumption is broken in current application. The final point cloud does has manufacturing errors compared with CAD model, and furthermore finding these errors is the purpose of the whole system. The registration is defined by the factory based on how the work piece is assembled to the others. The connection parts, such as holes of screws, which is usually call datum points, should have more weight in the registration process. The new registration that provides additional constraints and weight is developed.

5.2.3 System Calibration Procedure

In figure 5.2, the painted windshield with white markers is used as the system registration calibration gauge. Each white dot marker’s center coordinate was measured by a CMM and was considered as the ground truth data for registration calibration. The rest part was painted into black to assure only white dots coordinates are inspected by RIS system. System registration calibration can be divided into four steps:

1) Obtain RIS measurement data: measure the system registration gauge, filter out the noise and save four measurement results in four sensor frames. Figure 5.3 is the measurement
result of sensor one, the coordinates of the left top part of the system registration gauge. RIS system measures a cluster of points for each marker on the gauge.

2) Interpolate the coordinate of the marker center: ICP algorithm can not be directly applied in current case due to the limitation b) explained in last section. Closest points in both data sets may not represent the corresponding points, therefore the corresponding point
can the interpolation among the 3D point cloud in the second data set. In order to further increase the calibration accuracy, the interpolation is first applied after the data acquisition of RIS sensor. The radius of all the circular markers is 5 millimeters. The coordinate of the center of each marker was interpolated after obtaining the raw RIS measurement data. Figure 5.4 is the top view of the sensor 1 measurement result. It also displays the interpolated centers of each marker. The green dots denote the measurement points, and the red stars are the interpolated marker center. As we can easily see from figure 5.4, the RIS sensor may not measure the center of the markers. Registration based on closest point can still generate a certain level of error. The interpolation point of each marker is able to represented the center and results better accuracy. The center of each cluster of green points was estimated by the conventional Hough transform.

Figure 5.4: The center of each marker.

3) Registration process: registration process is divided into two parts: rough alignment (initial alignment) and fine alignment. Rough alignment is to match the RIS data to the CMM data in a low accurate level. Fine alignment applies the result from the rough alignment as initial position. ICP [49] [50] method is applied to obtain a high accurate result.
Figure 5.5 shows the rough alignment result for sensor one. Figure 5.6 shows the fine alignment result for it. Red dots denote the RIS sensor one result, and blue ones denote the CMM result. Red dots and blue dots are overlayed with each other in fine alignment.

4) Repeat step 1) to 3) for the other 3 sensors.

After system registration calibration is done, four individual point clouds successfully join into one in a common frame shown in figure 5.7.
5.3 Registration between the Final Point Cloud and CAD Model

The last step of a quality inspection is to find the deviation between the final point cloud and the CAD model. The color code information is used to evaluate the manufacturing
error of the work piece. If the manufacturing error exceeds the tolerance, the work piece is considered as unqualified. In the registration step between point cloud and CAD, the fine alignment matrix is needed to match the measured data with CAD model. Current registration method can be directly applied for this type of application because:

1. Photogrammetry method is not feasible for each work piece on an assembly line.
2. Feature points do not exist on part surface.
3. The mechanic support can’t guarantee the same location and orientation of each work piece.
4. The online point cloud registration needs to be finished within limited time.

5.3.1 Octree Space labeling method and Multi-resolution Method based ICP

Registration between the final point cloud and the CAD model needs to be finished within a limited time. Therefore, several algorithms to speed up the registration are applied to the registration algorithm in this thesis. Traditional ICP method searches the corresponding point through whole data set $P$ for a point $P_q$ in set $Q$, which is not necessary. After the rough alignment, the corresponding point in $P$ stays within the neighborhood around point $P_q$. Searching a neighborhood around point $P_q$ instead of the entire point cloud can greatly save the algorithm execution time.

In order to limit the search region in set $P$, octree space labeling method is utilized. An octree is a tree data structure in which each internal node has exactly eight children. Octrees are most often used to partition a three dimensional space by recursively subdividing it into eight octants as shown in figure 5.8. The center of the space is considered as the fist node, and the space is divided into 8 parts according to the geometrical relationship with the node. Each subspace will be divided recursively into another 8 parts. The nodes of the octree is shown in figure 5.9. The first node in level one leads 8 children in level two, and each node in level two leads another 8 in level three.
In current application, both data sets, the point cloud and the CAD, are divided into 64 subsets by three levels of octree respectively. The subsets with same coding name in both data sets, such as the subset 000 in the CAD and the subset 000 in the final point, are considered as the corresponding subsets. The searching of corresponding points are conducted only
with in each corresponding subsets. Therefore, the execution time of algorithm can be cut approximately 63/64. The loop of the cctree algorithm for the CAD model is described figure 5.10.

Figure 5.10: Octree for CAD model.

\[ O_1 = [x_{o1}, y_{o1}, z_{o1}]^T, \] the geometry center of the data set, is first calculated by Eq. 5.15:

\[ O_1 = \frac{1}{N} \sum_{i=1}^{N} P_{CAD}^i, i = 1, \ldots N \tag{5.15} \]

The coordinates of a point \( P, x, y, \) and \( z \), are compared with \( O_1 \) step by step. 1 is returned if \( P' \) coordinate is not smaller than \( O_1' \) coordinate, otherwise 0 returns. By this way, every point goes into 8 different subsets: 000, 001, 010, 011, 100, 101, 110, and 111. The same loop goes into each subset again and result 8 \times 8 = 64 small subsets representing the CAD.

The final point cloud is also treated with similar algorithm and is first divided into 8 parts in loop 1 shown in figure 5.11. The only difference is that an offset value \( T \) is added to the separation criteria, the geometric center of the final point cloud. If the coordinate of
$P$ is smaller than $O_1 + T$, the algorithm returns 0; if the coordinate is bigger than $O_1 - T$, the algorithm returns 1. The points near the edge between two subsets can go to both sets. The difference is illustrated in figure 5.12. Each subset in the point cloud is bigger than the corresponding subset in the CAD. The neighbor subsets have a certain region of overlap. A point near the edge of a subset in CAD may result its corresponding points in the neighbor subset in point cloud. The corresponding subset in the final point should be enlarge to cover this case. Shown in figure 5.12, subsets in CAD does not have overlap, but the subsets in the point cloud have.

![Tree](image)

Figure 5.11: Octree for the final point cloud.

Traditional ICP method also searching the corresponding point one by one in final point cloud, which means the searching resolution is a constant value 1. At the begging of the recursive algorithm, it is not necessary to search every point. Searching every 4 or 8 points is enough to generally align two data sets. Resolution can be function of iteration number $n$. As $n$ increases, the searching resolution comes back to 1. Usually the iteration stop between
7 loops to 12 loops. Therefore, the resolution function used in this thesis is given in Eq. 5.16. The algorithm is also forced to run at least 4 loops to find the correct corresponding points.

\[
Res = \max \left( 1, 2^{(4-n)} \right)
\]  

(5.16)

5.3.2 Feature Weighted Registration

Usually, the measured work piece needs to be assembled with the other parts. Therefore, several user specified locations are used to connected as links to the other parts. Those points or parts, once defined, will significantly determine the transformation of a point cloud in its CAD frame. Points in the data sets weights different in the registration process. Traditional ICP methods weight all points same is improper in the error map registration.

Curvature

Curvature Information: Curvature information is widely used in feature-based matching. Usually people utilize mean curvature or Gaussian curvature information to find feature points in a 3-D data set. Those feature points are helpful to set up the correspondence since they can be easily identified. The points carrying relatively large curvature such as peak points or valley points are chosen as feature points. The method in this thesis estimates the invariant points from both data sets by curvature information. In quality inspection process, all work pieces share similar but not same geometry. Each work piece has some
deformation from the designed shape. From the manufacture process, it can be concluded that the simpler the part shape is, the lower the risk of manufacture error the part has. For instance, flat shape with zero curvature carries smallest manufacture error, because almost no manufacture is evolved. So flat parts can be viewed as the most likely invariant parts in the work piece and should be weighted more in the registration process. In contrast, complex geometric parts will be less trustable and weight less. If the industry does not provides the datum points / features, we consider that the parts carrying the least curvature weight most in the registration process.

**GDT and Measured Condition**

GDT is short for geometry dimension and tolerance. It is a standard method to specify the geometric form and tolerance information of work piece in industry which is used in quality inspection process. Only the error map is within the acceptable tolerance, the work piece can be considered as qualified. The error map generation is related with the registration process. Without correct registration, the error map can not reflect the part’s quality. The registration based on the data features is critical in the error map generation.

Check fixture is a properly designed support of the work piece, which guarantees the work piece’s position and orientation.

Figure 5.13 shows a picture of a check fixture for windshields. In order to avoid the position error, three or a few more widely spaced datum points stoppers are measured and fixed on the check fixture. The windshield will sit on the surface of check fixture with all stopper bars touching the windshield edge. However, the final position and orientation between a windshield and the check fixture are judged by eyes, which means both accuracy and precision can not be granted. The same work piece can provide different error distribution maps. The second disadvantage of the check fixtures is that each fixture only fits one kind of work piece. Different types of work pieces request a new check fixture, which costs a lot.

Straightness and flatness are perhaps the two most popular types of geometries to be
used as datum, because least manufacture involved in those parts. The installation points are chosen as datum points, since those points carry less curvature and less tolerance. In this Thesis, the edges touching the datums on the check fixtures will be weighted more in the ICP registration. The position and orientation of windshield corresponding the check fixture is used as an initial condition of the modified ICP algorithm.

5.3.3 The Object-oriented online Registration Model

Registration is the process to find out rigid transformation $T$ between two data sets $A$ and $B$ in two coordinates. The following parts are only concentrated on how to solve the equation $\{a_i\} = T\{b_i\}$ to get transformation $T$. The total square error $e^2$ is defined as

$$e^2 = \sum_{i=1}^{N_A} (a_i - Tb_i)^2$$

(5.17)

$a_i$ is a point from the CAD model $A$, and $b_i$ is the correspondent point in the measured data set $B$. 

Figure 5.13: Check fixture and the stopper bar touching datum points of the work piece.
In order to analyze and solve the online registration problem, the theoretical model is built up from ideal case to the real case.

**case 1**

Ideal case, which is the easiest. Given two data sets: \( A \{ a_i(x_i, y_i, z_i), a_i \in R^3 \} \), and \( B \{ b_i(x_i, y_i, z_i), b_i \in R^3 \} \) in two coordinate systems. Shape A is equal to shape B, Data set A is CAD, B is the measured data. The registration matrix \( T \) between two data sets needs to find out.

Solution: Since shape A equal shape B, the only error source for \( e \) is the registration error, which is named as \( e_{\text{Registration}} \).

The registration error for each pair of corresponding points in two data sets can be written in the Eq.5.18

\[
e_i = e_{\text{Registration},i} = a_i - T b_i
\]

(5.18)

Linear least square method is directly applied. By minimizing \( e^2 \), \( T \) can be solved by iteration. (ICP method)

**case 2**

Random noise \( e_{\text{Random}} \) is added in to data set B. The unknown registration matrix \( T \) needs to find out between two coordinate systems.

The error between each pair of corresponding points in two data sets has two parts: the registration error and the random error, so the equation can be written as Eq.5.19

\[
e_i = e_{\text{Registration},i} + e_{\text{Random},i} = a_i - T b_i
\]

(5.19)

Linear least square method can still be directly applied in the equation. Since linear least square method is good at solving the random noise, \( T \) can be found by iteration. (ICP method)
case 3

The deformation error $e_{\text{Deform}}$ is added into data set $B$, which can be viewed as the real case.

In this case, the error equation is:

$$
e_i = e_{\text{Registration},i} + e_{\text{Random},i} + e_{\text{Deform},i} = a_i - Tb_i$$

(5.20)

If linear least square method is still directly applied to equation by minimizing linear square of $e$ to find out $T$. The total square $e^2$ including the deformation error $e_{\text{Deform}}$ will be minimized in the registration procedure. Registration process changes the deformation information, which makes the error map un-trustable.

The solution: Since shape $B$ is a deformation form from $A$, there exist some parts $B_1$ that have same shape with the correspondent parts of $A$ denoted as $A_1$; there exist some parts $B_2$ that have small deformation with the correspondent parts of $A$ denoted as $A_2$; and there exist some parts $B_3$ that have medium or big deformation with the correspondent parts of $A$ denoted as $A_3$. Now data sets $A$ and $B$ will be separated into three subclasses: $A = \{A_1, A_2, A_3\}$, $B = \{B_1, B_2, B_3\}$.

3.1 First assume region $A_1$ in $A$ (CAD) is known from the GDT information In CAD, and $B_1$ is unknown in measured data $B$, and assume $B_1$ always exists. The error $e_1$ between $A_1$ and $B_1$ only have two parts: $e_{\text{Registration}} + e_{\text{Random}}$. Linear least square method is able to solve $T$ between $A_1$ and $B_1$. Based on the characteristics information carried between $A_1$ and $A$, a filter can be build up to estimate the possible data set $B_1^{\text{Possible}}$ of $B_1$ from $B$, which has the following relation $B \supseteq B_1^{\text{Possible}} \supseteq B_1$. The filter is built by the curvature and moment information $A_1$ carried in set $A$.

3.2 In this case, the existence of $B_1$ can not be guaranteed. The assumption that $A_1$ always has correspondent $B_1$ may not always be true. CAD model provides the information that $A_1$'s corresponding part $B_1$ should has low tolerance, however the work piece could be unqualified due to the manufacture process. By adding this constraints, case 3.1’s solution
matrix needs to verify before registration. The solution matrix $T$ was applied to calculate $e_1$. If $e_1$ is bigger than the acceptable tolerance carried by CAD, which means $B_1$ in fact does not satisfy the assumption, then this work piece will be considered as unqualified and thrown away. The post process’s verification result tells people any work piece challenges the assumption will be punished as an unqualified part to be thrown away, so that the solution strategy becomes un-challengeable.

3.3 $A_1$ in data set $A$ changed unknown too. None of any pre-knowledge about $A$ and $B$ is known.

Solution: curvature information in $A$ is used to estimate the $A_1$. And then $A_1$ is used to estimate $B_1$ to find $T$.

Assumptions: 1) During the manufacture process, the more complex the part is made, the more manufacture error. The less complex the part is made, the less manufacture error, the more reliable. 2) Geometric complexity is related with the curvature information: The less mean curvature point $a_i$ has, the less geometric complex the point has, the more likely $a_i$ belongs to $A_1$, $a_i \in A_1$. Based on this assumption, $A_1^{Possible} \supseteq A_1$ is estimated by curvature information.

Judged based on the curvature information, $A_1^{Possible}$, $A_2^{Possible}$, $A_3^{Possible}$ are estimated. Similar with case 3.1, replace $A_1$ by $A_1^{Possible}$, estimate $B_1^{Possible}$ by filter. After the equation $e_{Registration} + e_{Random} = \{a_i\} - T\{b_i\}$ are solved, $T_{k=1}$ is found. Similar with case 3.2, postprocess is carried out to calculate every point’s $e_i$ in $A_1^{Possible}$. If any point’s error is bigger than threshold, the point in $A_1^{Possible}$ will be thrown away. Iterative step $k$ goes to $k + 1$, New $B_1^{Possible}$ was re-calculated based on new $A_1^{Possible}$. This iterative steps will stop after $A_1^{Possible}$ converges to $A_1$ by comparing the post-processing verification results with the accepted tolerance.
5.3.4 The Methodology in Current Task

Notation: Before the error map generation, the transformation $T$ which aligns measured point cloud from sensor frame into CAD model in CAD frame is needed. The CAD model in the CAD frame can be denoted as:

$$CAD = \{p_{CAD}(x_i, y_i, z_i) \in \mathbb{R}^3\} \quad (5.21)$$

The datum points in the CAD model:

$$A_1 = \{p_{Datum}(x_i, y_i, z_i) \in \mathbb{R}^3\} \quad (5.22)$$

For different work piece $j$, its notation in part frame is:

$$W_j^p = \{p_{W,j,i}(x_i, y_i, z_i) \in \mathbb{R}^3\} \quad (5.23)$$

For measurement time $k$, the location of the work piece is denoted as $T_k$, this transformation matrix denote from part coordinate frame to world frame. So the description of the work piece $j$ in the location $k$ in world frame is:

$$W_{j,k}^w = T_k W_j^p \quad (5.24)$$

The measured point cloud of the work piece first stores in the common sensor coordinate frame can be denote as:

$$W_{j,k}^s = U_{i=1}^4(T_w^s)W_{j,k}^w = U_{i=1}^4(T_w^s)T_k W_j^p \quad (5.25)$$

The term $U_{i=1}^4(T_w^s)$ is to transfer four individual sensor frame point clouds into a final one. The inspection system used here has four 3-D sensors, each one measures a quarter of the work piece. The final point cloud needs to be transferred from four individual sensor frames.

The final registration process is to align the measured point cloud with the CAD model:
\[ W_{j,k}^{CAD} = TW_{j,k}^s = TU_{i=1}^4(T_w^s)T_kW_j^p \] (5.26)

In Eq.(5.26), there are two variables in the left side: \( W_j^p \) presents different work pieces, \( T_k \) presents different location of the work piece. The transformation matrix \( T \) needs to cover all these two variables. The algorithm is started by finding the possible datum points sets \( B_1^{Possible} \) in the measured data by applying a filter function to search through measured data set.

### 5.3.5 Registration Process

The filter is designed by the collection of datum points’ characteristics in CAD, including the curvature information and the moment information. After those information are captured, threshold tolerance is add on the filter. The threshold tolerance will control how many points can pass the filter into the set \( B_1^{Possible} \). The filter will go over the whole measured data set to get \( B_1^{Possible} \).

\[
Filter\ function\ f(p_i) = \begin{cases} 1, & K_i \in K_c \pm \tau_K, M_i \in M_c \pm \tau_M \\ 0, & otherwise \end{cases}
\] (5.27)

In the Eq.(5.27), \( p_i \) denotes the point \( i \) in measured data, \( K_i \) denotes the mean curvature of point \( i \), \( M_i \) denotes moment of point \( i \). \( \tau \) is the threshold tolerance. If the points satisfy the curvature and geometric condition, the filter function equals to 1, otherwise filter function equals to 0.

ICP is applied in the algorithm: Repeat the iteration process \( i = \text{from 1 to } i_{Final} \), until the total least square error reaches the threshold.

(a) Corresponding points: The CLOSEST points \( \min[(x_{a1} - x_{b1})^2 + (y_{a1} - y_{b1})^2 + (z_{a1} - z_{b1})^2] \) are considered as corresponding points between \( A_1 \) and \( B_1^{Possible} \).

(b) Minimizing the total square error, in order to find the rotation matrix \( R \) and translation vector \( t \).
(c) Updating the set $A_1^{(1)}$ to $A_1^{(2)}$ by $T$.

(d) Continue step (a) to (c) until the average total square error under threshold value.

5.3.6 Error Map Generation

After the point cloud is transferred into the CAD model frame, the distance $D$ between point cloud and CAD surface is calculated along the normal direction.

The deviation $\Delta$ can be described by Eq. 5.28, Where A represents CAD model and B represents a measured point cloud.

$$ \Delta : (A, B) \mapsto E $$

The error map $E$ includes a set of 3D points and their distances $D$ in a metric space to the closest triangle in CAD model $M$, as seen in Eq. 5.29

$$ D(b_i, A) = \min(d(p_i, T_j), i = 1, 2, \ldots m, j = 1, 2, \ldots n) $$

$$ E = \{ < b_i, D(b_i, A) > | i = 1, 2, \ldots m \} $$

The methodology explained earlier is implemented step by step. The first step is to calculate the datum points’ moment information and the characteristic information in CAD model. Then the estimated possible datum points in the measurement point cloud are found out. At last, ICP algorithm is applied between four datum points in CAD and the possible datum points sets in the measured point cloud. The registration matrix based on the invariance part of the windshield is finally derived. Figure 5.14 (a) is the CAD model surface with four datum points on it. Based on the four datum points and the methodology outlined before, the final alignment between CAD and point cloud is generated, shown in Figure 5.14 (b).

The final error map for a windshield is displayed in figure 5.15. Red color means that the CAD is higher than the point cloud. Green color means that both sets have same heights, and blue means that CAD is lower than the point cloud. From the color bar in the left side of the figure 5.15, the error is from from $-5$ millimeters to 6 millimeters.
5.4 Sensor Fusion

The 3D point cloud is not the only way to display the measurement result. In the navigation application, the 3D point cloud is fused into the 2D image. The distance between the object and the camera is first translated into a color code and then the color code is fused into the display. An example is given in figure 5.16. Since the 3D point cloud is calculated through
the corresponding image pixel, the transformation between the 3D point cloud and the 2D image can be directly obtained.

![Image](image.png)

**Figure 5.16:** The fusion result for navigation application.

### 5.5 Chapter Summary

After 3D point cloud is derived through the structured light sensor, there are usually several post processes. Registration or sensor fusion are two common tasks. Registration is to find out a rigid transformation matrix between two 3D frames and to put different point cloud into one frame. ICP based registration method is developed in this chapter. Through the system calibration, the final point cloud merges all the four individual point clouds and represents the entire work piece. Octree and multi-resolution is used to speed up the iteration speed. An object-oriented method is developed to register the final point cloud with the CAD model.
and thus an error distribution map could be calculated and displayed.
CHAPTER 6

The development of an Omni-direction 3D camera for mobile robot

6.1 The Working Principle of the Omni-directional 3D Sensor

The developed omnidirectional 3D camera consists of a projector, a camera, and two hyperbolic mirrors. Similar with traditional omnistereo systems, a projector replaced a camera and resulted in a structured light based system. The principle of the 3D camera is illustrated in figure 6.1. Under ideal condition, optical centers of the camera and the projector coincide with the focal points $F^2_{m1}$, $F^2_{m2}$ of the two hyperbolic mirrors, respectively. The projector is regarded as an inverse camera, that is, the projector maps a 2D pixel in the projector to a 3D array in the scene.

The optical path of the 3D camera can be described as follows: a light ray from the projector goes through the focal point $F^2_{m2}$ and then intersects with the hyperbolic mirrors at point $P_{m2}$. The hyperbolic mirror has a useful property that any light ray going towards one of the focal points will be reflected through the other focal point. Hence, the incident light ray will be reflected away from the other focal point $F^1_{m2}$. Then, it intersects with the
observed target with diffuse reflection property so that part of the light ray will be reflected to
the focal point $F_{m1}^1$ of hyperbolic mirror 1; thus, the light ray will be further reflected at point $P_{m1}$ of the hyperbolic mirror 1 to the camera through the other focal point $F_{m1}^2$

due to the property of the hyperbolic mirror. Therefore, under ideal condition, the relation
between any 3D point in the scene and the corresponding 2D point in the image sensor can be represented by a uniform model with a single viewpoint [62].

However, the above ideal mathematical model requires the lens’ optical center to be coincide with focal point of the hyperbolic mirror. This requirement, however, is difficult to be completely satisfied, especially for the projector since it cannot view the scene directly. So the uniform model for every pixel of the sensor will cause the residual error, resulting in incorrect 3D reconstruction.

To solve such a problem, a vector-based 3D reconstruction strategy which determines the ray vector for each pixel of the camera and projector in the scene was proposed and the corresponding look-up-tables (LUT) was established to calibrate the system. A light array $L_p(u_p, v_p)$ from a projector pixel $(u_p, v_p)$ intersects its corresponding light array from camera $L_c(u_c, v_c)$. The intersection point $P$ is the reconstruction point. Instead of reconstructing the light array from the image pixel, light arrays are directly rebuilt in the task space:

\[
L_p = P_w^1 + U_W p \lambda_p 
\]  

(6.1)

\[
L_c = Q_w^1 + V_W c \lambda_c 
\]  

(6.2)

$P_w^1$ is an arbitrary point on $L_p$ and $U_W$ is a unit vector representing its direction. $Q_w^1$ and $V_w$ in Eq. 6.2 have the similar meanings.
6.2 Calibration of the Omni-directional 3D Camera

First, the calibration of the omnidirectional camera was discussed using the vector-based calibration strategy. As shown in figure 6.2, $U_w$ is the incident vector in the world frame of a pixel $I_c$ in the catadioptric camera, which is our calibration parameter for each pixel. By moving a predefined reference board (white flat board) with several precise-located markers, the view vector $U_w$ can be specified by the two points $P_1^w$ and $P_2^w$. Although the calibration concept is straightforward, it is a challenge to calibrate all the pixels simultaneously.

To this end, an extra dioptric projector, as well as a dioptric camera, is introduced. The dioptric projector shoots two-dimensional encoded patterns onto the reference board; meanwhile, the dioptric camera records the resulting image of the board. In such a way, the phase value can be determined for each pixel in the dioptric camera. The origin of the reference board frame is set to coincide with one marker and the $xy$ axes are parallel to reference board. Obviously, the $z$ axis is perpendicular to the reference board. In this case, the coordinate $P_r$ in the reference board frame can be transformed to the corresponding $I'_c$ in the image plane frame of the dioptric camera by the homography matrix $H$:

$$P_r = H \times I'_c$$  \hspace{1cm} (6.3)

where $H$ can be specified by using the markers on the reference board since the coordinates of these markers in the image plane frame and reference board frame are both known. It should be pointed out that the location $P_r$ can be calculated within the sub-pixel by appropriate image processing.

The desired location $P_w$ of the point is measured in the world frame. Therefore, the location $P_r$ in the reference board frame should be further transformed to $P_w$ in the world frame:

$$P_w = RP_r + t$$  \hspace{1cm} (6.4)

where the $3 \times 3$ rotation matrix $R$ and $3 \times 1$ translation vector $t$ can be obtained through the known markers. Markers are measured using a high accuracy instrument (e.g. the
measuring arm or laser tracker), which is kept static during the calibration procedure so that the instrument frame is considered as the world frame. Whereas, the dioptic projector and the dioptic camera are moved to optimal position in sequence when the reference board is placed at different sides.

Furthermore, the illuminated patterns are simultaneously captured by the catadioptric camera so that the corresponding coordinate \( P^1_w \) in the world frame can be determined for every pixel in catadioptric camera. Similarly, the coordinate \( P^2_w \) can be obtained when the reference board is moved to another location. Then, the reflected vector \( U_w \) and a point \( P^1_w \) for each pixel in the world frame can establish a LUT, whose size is equal to the resolution of the catadioptric camera. Without loss of generality, any point viewed in the catadioptric camera can be presented by:

\[
P_w = P^1_w + \alpha U_w \tag{6.5}
\]

where \( \alpha \) is a scale factor; \( U_w = \frac{P^2_w - P^1_w}{|P^2_w - P^1_w|} \). It should be noted that the point viewed within the pixel in the catadioptric camera can be realized by using bilinear interpolation with four neighboring pixels.

On the other hand, the calibration of the catadioptric projector can be achieved in a similar way except with the absence of the extra dioptic projector. For this technique, encoded patterns are directly shot from the catadioptric projector and the dioptic camera records the phase distribution and calculates the corresponding locations \( Q^1_r \) and \( Q^2_r \) in the reference board frame. Similarly, markers on the reference board are measured by the high accuracy instrument so that we can obtain the locations \( Q^1_w \) and \( Q^2_w \) in the world frame.

Hence, any point on the incident light ray of the projector can be expressed by:

\[
Q_w = Q^1_w + \beta V_w \tag{6.6}
\]

where \( \beta \) is a scale factor; \( V_w = \frac{Q^2_w - Q^1_w}{|Q^2_w - Q^1_w|} \). Eventually, the incident vector \( V_w \) and a point \( Q^1_w \) for each pixel of the catadioptric projector can also establish a LUT, whose size is equal to the resolution of the catadioptric projector.
Once projector and camera are calibrated, the point on the observed target can be computed by the intersection of the two vectors of the projector and camera [76].

6.3 Error Analysis for the Designed 3D camera

The accuracy of 3D reconstruction is bounded by the resolution of the imagers. The resolution of a camera is the distance that one pixel covers at a certain amount of distance. Any variation within the distance can not be directly detected and thus bounds the accuracy of the designed system. Difference from a traditional camera, the resolution of an omnidirectional camera is not a constant value. As shown in figure 6.4(a), the resolution around the edge is much lower than the center of the image. The resolution decreases along the image radius. The reason is because of the curvature of the omnidirectional mirror. Since the curvature of the mirror is not a constraint value, the back image from image also shrinks at different ratios. Figure 6.4(b) illustrates the optical path of omnidirectional camera. The wall is approximated 2.5 meter from the sensor, and point $P$ moves along the wall. The corresponding pixel $p$ moves along the radius to the center. Since the resolution is different from pixel to pixel the detection accuracy of point $P$ is highly related with its position. If point $P$ is around the same height with the mirror, its resolution is low and its detection accuracy is high. Once the point $P$ moves lower, the resolution arises and the accuracy also drops.

\[
\begin{bmatrix}
  x \\
  y \\
  1
\end{bmatrix}_I \approx \frac{b^2 \left( -ez - a\sqrt{x^2 + y^2 + z^2} \right)}{b^2 z^2 - a^2 (x^2 + y^2)} \begin{bmatrix}
  fx & \alpha & u_0 & 2\epsilon v_0 \\
  0 & fy & v_0 & 2\epsilon v_0 \\
  0 & 0 & 1 & 2\epsilon
\end{bmatrix} \begin{bmatrix}
  x \\
  y \\
  z \\
  1
\end{bmatrix}
\]  

(6.7)

In error simulation, $P$ moves vertically from $300 mm$ to $-1500 mm$, and its corresponding resolutions in image frame and in world frame are shown in figure 6.6(a) and figure 6.6(b) respectively. Ideal omnidirectional camera model is applied in calculation through Eq 6.7. The parameters are used design values: $a = 18.75 mm$, $b = 18.75 mm$, $e = 40.0 mm$, $f_x = f_y = 250$, $u_0 = 320$, $v_0 = 240$, $X = 2500 mm$, $Y = 0$, $Z = -1500 300 mm$ and the camera pixels are
640 by 480. From the resolution curve in figure 6.6(a), the resolution is round $7 mm/pixel$ when the radius is 250 pixels, and the resolution arises to $25 mm/pixel$ when radius is 100 pixels. From the resolution curve in figure 6.6(b), the resolution reaches $7 mm/pixel$ when $z$ is 300$mm$ high. As the $z$ decreases, the resolution increase.

The omnidirectional projector resolution curve follows the same working principle with the omnidirectional camera. The only difference is that it was put upside down. The reconstruction error $\Delta d$ can be calculated by the resolution error. Shown in figure 6.5, $AP_1$ is the distance that one camera pixel covers, and $BP_1$ is the distance that one projector pixel covers. Since the triangle $ABP_2$ is similar with the triangle $OCOP_2$, $\Delta d$ can be easily calculated: $160 mm$.

When the object depth is around $4000 mm$, $\Delta d$ equals $256 mm$. If a 2K by 2K camera is used instead of the current one, $\Delta d$ equals $39 mm$ when object distance is $2500 mm$. $\Delta d = 62 mm$ when the object distance is $4000 mm$.

### 6.4 One-shot Surface Coding for the Sensor

A robust projection pattern is critically important for achieving accurate correspondence for a structured light based measurement system, especially for a one-shot method. The design should satisfy the following constraints:

(a) monochromatic light and

(b) pre-warping to cancel mirror distortion and

(c) invariance from scene variation.

Taking the monochromatic light into account, the designed primitives should not contain color coding information. The background colors will not influence a monochromatic light based system. Furthermore, the color based design pattern could not be applied to an infrared light source, but the monochromatic based design could. The final design of the omnidirectional 3D camera would employ near infrared light source instead of normal one.
Traditional one-shot pattern only works fine in a normal structured light system, it could not be used in omnidirectional structured light system without pre-warping. The convex shape mirrors in the omnidirectional sensor could distort both the geometries and the orientations of the primitives. Compared with traditional monochromatic light based patterns, light arrays are warped by the hyperbolic mirror twice in an omnidirectional system. Hence the geometrical features based pattern is even more difficult to be calculated compared with the traditional one. The wanted projection image should be first un-warped in order to cancel the mirror distortion. Since a projector can not receive information, its pre-warp function is much more difficult than a camera’s. Second, even the un-warping function is calculated, the projected geometrical light features are still further arbitrarily distorted by the unknown environment. The designs of the primitives should also be an invariance from this distortion and then a correct correspondence can be linked between two image frames.

6.4.1 Image warping in an omnidirectional camera

Both of the projector warp and the camera warp are the transformations between conic image to panoramic views [77]. Camera image warp has been discussed a lot in the past [77]. In essence, the warping function scans the conic image around the center and then horizontally allocates each part into a rectangle, shown in figure 6.7. The radius and the center location are needed to finish the warping process. The radius and the center can be easily obtained for an omnidirectional camera by image analysis, but the warp for the projector is much more difficult and has been discussed very little in literature. The center of the conic image can not be directly estimated, since a projector is not an information receiving device.

Through the LUT created in the calibration process, the projector can ’view’ the scene by a calibrated camera, which is illustrated in figure 6.3. By treating the omni-projector as a pseudo omni-camera, its warping function can also be obtained by LUT. Basically, the designed projection marks are first calculated on the reference board in the world frame and then are transferred back to the projector image frame via LUT. Hence, the conic image
center and its radius can be interpolated.

6.4.2 One-shot projection pattern design for the omnidirectional
3D camera

The task to design a one-shot projection is to assign each projected bright pixel / image
primitive / marker one unique codeword to distinguish each other. The unique codeword
will bring correspondence between two image frames.

A projection pattern with concentric circles is designed for the one-shot projection shown
in figure 6.8. After the center of conic projector image is derived, concentric circles with
different radii are utilized as the projection pattern. Each circle is assigned by a different
angular frequency and a projection intensity in order to distinguish each other. The intensity
functions of the concentric circles are described by Eq. 6.8 using the inverse Fast Fourier
Transform (FFT). The intensity function $I_p(r, \frac{N}{2\pi} \theta)$ is designed in the polar coordinate and
$r$ stands for the radius and $\theta$ is the angle. $N$ is the number points on the circle. $X(r, k)$ is
the assigned angular frequency for each circle and $I_r$ is the intensity coefficient for each
circle. Within each circle all of the bright pixels share the same angular frequency: $X(r, k)$. If
the image is scanned from the center to outside by a radius, the circles are labeled by
different angular frequencies and different projection intensities. If the image is scanned
along the tangent direction of each circle, each bright pixel can be labeled by a different
phase angle $\theta$. The angular frequency, projection intensities and the phase angle $\theta$ assign
each bright pixel one unique codeword.

$$I_p(r, \frac{N}{2\pi} \theta) = \begin{cases} 
\left(\frac{1}{N}\right) I_r \sum_{k=1}^{N} X(r, k) \omega_N^{\frac{N}{2\pi} \theta(k-1)}, & \text{if } r = R_1, \ldots, R_{10} \\
0, & \text{else} \end{cases} \tag{6.8}
$$

$$\omega_N = e^{(-2\pi i)/N} \tag{6.9}$$
\[ I_r = I_{R1}, ..., I_{R10}, \text{ when } r = R_1, ..., R_{10} \] (6.10)

### 6.4.3 Epipolar constraints and the decoding method

The task for a one-shot pattern decoding algorithm is to extract the codeword so as to establish the correspondence between a projector pixel and a camera pixel. Most importantly, the extracted codeword should be an invariance against the unknown environment distortion. The projected pattern is distorted by the unknown test scene and the distortion is used to reconstruct the 3D coordinates. Different object depths could bring about different distortions. Utilization of such distortions is the fundamental working principle of a triangular based 3D reconstruction. However, the drawback of the distortion is that it could also destroy geometric shapes and the locations of the pixels / primitives / markers. The designed codeword should avoid such a case. A robust encoding and decoding algorithm should utilize the invariance information against the distortion. In this paper, epipolar constraints [55] are utilized to extract the codeword and to find out the invariance.

As the objects depths vary, the corresponding camera pixel changes its position. However, there are still certain rules that it must follow. As shown in figure 6.9, a projector pixel \( I_p(r_p, \theta) \) emits a light array, intersects to the scene at point \( P \), and reflects into the camera image plane at point \( I_c(r_c, \theta) \). No mater where \( P \) is, \( I_p \) and \( I_c \) always share the same phase angle \( \theta \) and \( \theta \) is the invariance component from the scene variance. In this paper, each projection ring is labeled by different projection intensity and angular frequency and each pixel is assigned by an unique codeword: the phase angle. Therefore, the proposed one-shot pattern is capable of linking correspondence with high accuracy. The intensity of each arc / circle can been easily detected by the camera and its angular frequency can also be derived through FFT:
\[
X(r, k) = \sum_{j=1}^{N} I_c(r, j) \omega_N^{(j-1)(k-1)}
\]  

(6.11)

Combined its phase angle, a codeword of a bright pixel can be uniquely determined and be utilized to set up the correspondence.

### 6.5 Chapter Summary

This chapter presents a structured light-based omni-directional 3D environment reconstruction system using the projector, camera, as well as hyperbolic mirrors. The corresponding vector-based calibration strategy is proposed to improve the system accuracy.
Figure 6.1: The working principle of the designed omni-direction system.
Figure 6.2: Calibration of the omni-directional camera.
Figure 6.3: Calibration of the omni-directional projector.
Figure 6.4: Received image from the world frame: (a) the received image; (b) Illustration of the optical path.
Figure 6.5: The object depth error from the resolution limitation.
Figure 6.6: The omnidirectional camera resolution: (a) the resolution curve in image frame; (b) the resolution curve in camera frame.
Figure 6.7: Image warp for a panoramic view.

Figure 6.8: The designed one-shot projection pattern.
Figure 6.9: The epipolar geometry of the omnidirectional system.
CHAPTER 7

Current Experimental Results and Data Analysis

This chapter gives the experimental results and data analysis is given based on the results. According to the implementations of the systems, the measurement results are roughly separated into three categories:

1) The experimental results for the developed Rapid Inspection System (RIS) for free form surface.
2) The experimental results for navigation system in single viewing direction.
3) The experimental results for the omni-directional navigation system.

7.1 The Rapid Inspection System (RIS)

7.1.1 The Configuration of the System

Rapid Inspection System (RIS) for freeform surface, which is designed and implemented in Michigan State University Robotics and Automation Lab, is a 3D scanning system for manufacturing quality control. 3D point cloud and corresponding error distribution map with the Computer Aided Design (CAD) model will be generated after inspection process
has been conducted. Configuration of the system is illustrated in figure 7.1.

![Projector Camera](image)

**Figure 7.1: RIS system configuration.**

RIS is a multi-sensor system, having four pairs of cameras and projectors on the top. 3 inches × 3 inches firm Aluminum T slots are selected to build a frame structure to eliminate the vibration from the ground. Sensors are mounted on the upper beams, a workpiece, such as a windshield, is located on the lower holder. Each pair of sensors forms an active vision sensing system. 3D point cloud in each sensor frame is constructed by calculating the deformation of Gray Code Line Shifting (GCLS) pattern by the test surface. The LCD
projectors on the beam shoot image patterns on the tested object and the 2k × 2k high resolution cameras receive the reflection lights. The field of view for each sensor is 2,100 millimeters (L) × 1,600 millimeters (W) × 600 millimeters (H). In the left top red box of figure 7.1, it highlights and zooms the cross view of sensor 1. A LCD projector projects the light patterns and a CMOS camera receives the reflection lights.

The pixels on the projector frame and the camera frame denote as \( m_p : (u_p, v_p, 1)' \) and \( m_c : (u_c, v_c, 1)' \), respectively.

The result point cloud \( P^{S_j} \) in each sensor frame can be written as:

\[
P^{S_j} = \left\{ p^{S_j}_i (x_i, y_i, z_i) \in \mathbb{R}^3 \right\}, j = 1, 2, 3, 4
\]  

\[
P^{S_j} = f^j \left( m^j_p, m^j_c \right), j = 1, 2, 3, 4
\]

Function \( f^j() \) denotes the 3D reconstruction model of sensor \( S_j \) to recover the point cloud from corresponding pairs of pixels \( m^j_p \) and \( m^j_c \) in both image frames.

The registration to transfer four point clouds into a common frame is defined as:

\[
P^C = \bigcup_{j=1}^4 \left( T^C_{S_j} P^{S_j} \right)
\]

\( T^C_{S_j} \) is the transformation matrix from sensor frame \( j \) to the common frame.

### 7.1.2 The Inspection Result via Single-shot Coding Algorithm

As explained in the previous chapters, single-shot coding algorithm is capable to measure a moving object. Assuming the work pieces are moving along the assembling line, single-shot coding method is applied to inspection the parts. Since the single-shot is weak against the image noises, a pre-treated work piece is utilized: a spayed pillar part from automobile industry shown in figure 7.2. Its corresponding inspection result is given in figure 7.3. Although the method is weak against image noise and it could only obtain relatively low
point density, the method could still obtain the point cloud for a pre-spayed work piece. The point cloud is much lower than a multi-shot coding method, but is still higher than traditional CMM based inspection. Compared with multi-projection method and CMM, single-shot method could measure a moving work piece and its acquisition time is less than 1 second.

Figure 7.2: The inspected pillar with surface pre-treatment.

Figure 7.3: The inspected point cloud of the pillar.
7.1.3 The Inspection Result via Multi-shot Coding Algorithm

Multi-projection coding method is more robust against image noise, so it could inspect work pieces without surface pre-treatment. Figure 7.4 gives the 3D point cloud of three work pieces by the inspection system. From the result, we could conclude that the system could inspect the work pieces and provide dense 3D point clouds no matter what optic property is: black, dark, and even shiny.

![Figure 7.4: The 3D point clouds for the work pieces.](image)

The major advantages of the proposed method is to increase the completeness of the inspection results. The traditional inspection method can not fully inspect the optically-challenging objects without surface pre-treatment. In order to get a quantity evaluation of the completeness, two additional experiments were conducted. In the first experiment, the work pieces were powder sprayed. The camera and projector settings, such as camera aperture, projection brightness, and projection contrast, were properly readjusted to avoid under- / over- exposure issues. The number of the result 3D point cloud was considered as the ground truth. If any inspection result reached the same number, the inspection
completeness was considered as 100%. The comparison between our results and the first experimental results was to find out what is the percentage of the completeness we reached.

The second one is to inspect the same work pieces using naive structured light coding / decoding methods. The patterns without adaptive mask are projected. In this experiment, the camera aperture can not be adjusted perfectly. The black piece requires a big aperture, but the shiny piece needs a small aperture. The aperture is set around the middle: perfect to the dark piece, but too big to the shiny part and too small to the black part. The comparison between our results and the second experimental results is to find out how much the method improved.

The results are listed in table 7.1. The experimental results using our methods are listed as the third one.

<table>
<thead>
<tr>
<th></th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black one</td>
<td>30434</td>
<td>14793 (48.6%)</td>
<td>29526 (97.0%)</td>
</tr>
<tr>
<td>Dark one</td>
<td>77433</td>
<td>71024 (91.3%)</td>
<td>76533 (98.8%)</td>
</tr>
<tr>
<td>Shiny one</td>
<td>147755</td>
<td>126486 (90.7%)</td>
<td>143497 (97.1%)</td>
</tr>
</tbody>
</table>

The numbers in the second column in the table 7.1 are the results of the first experiment, and the number of points are considered as the ground truth. Its completeness is defined as 100%. The numbers in the third column are the results by the naive inspection method. From the results, we could see that more than half of the black work piece are failed to be reconstructed, and around one tenth of the other two pieces are failed to be rebuild either. The reason that black piece received poor performance is that the aperture size is adjusted relatively small. Therefore the inspection of the shiny part is relatively good but the black piece is too poor. The numbers from the third experiment shows the completeness of our method: all the parts can be inspected over 97%. Around 50% of the completeness is
improved for the black piece and around 7% of the completeness is improved for the other two. Because the projection mask could dynamically adjust the illumination, the system could simultaneously inspect both shiny part and black part.

7.1.4 The Registration Result from Four Pairs of Sensor Frames

A piece of painted windshield was inspected by the RIS system. Each pair of the 3D sensor covers around a quarter of the windshield. The final point cloud representing the windshield is the merged result from four individual point clouds in four sensor frames. The windshield is also measured by a CMM machine.

Since CMM has been utilized for years for its accuracy performance, the CMM data is utilized as ground truth. The difference between the CMM measurement data and the RIS measurement data is applied to evaluate the calibration result.

Figure 7.5: The comparison between CMM data and the RIS data.

Figure 7.5 shows both the measurements. The upper three point clouds with triangulated
surface are the RIS measurement result after aligned with CAD model with different view angles, and the lower three point clouds with low density are CMM measurement results with the same three view angles. The evaluation between these two sets employee the following formulas:

\[ M = \frac{1}{N_{CMM}} \sum_{i=1}^{N_{CMM}} \left| P^i_{RIS} - P^i_{CMM} \right| \quad (7.4) \]

\[ d^i = \left| P^i_{RIS} - P^i_{CMM} \right| = \sqrt{(X^i_{RIS} - X^i_{CMM})^2 + (Y^i_{RIS} - Y^i_{CMM})^2 + (Z^i_{RIS} - Z^i_{CMM})^2} \quad (7.5) \]

\[ P^i_{RIS} = \begin{bmatrix} X^i_{RIS} \\ Y^i_{RIS} \\ Z^i_{RIS} \end{bmatrix}^T \in M_{RIS} \quad (7.6) \]

\[ P^i_{CMM} = \begin{bmatrix} X^i_{CMM} \\ Y^i_{CMM} \\ Z^i_{CMM} \end{bmatrix}^T \in M_{CMM} \quad (7.7) \]

, in which \( M_{RIS} \) and \( M_{CMM} \) stands for the measurement results from RIS and CMM, respectively. The corresponding points are found via ICP method explained in the last chapter. The mean distance between CMM data and RIS data is about 0.15 millimeter. Considering CMM data as the ground truth, the calibration RIS system has around 0.15 millimeters error. Backprojection of the 0.15 millimeters error into the camera frames, is equivalent to 0.14 pixel calibration error. By comparison, conventional relatively calibration errors for a structured light system are around 0.02 pixel in camera and 0.29 pixel in projector respectively. Since in our sensor model projector’s intrinsic and extrinsic parameters do not participate in the calculation, 3D points in world frame are directly correspondence with projector pixels. Backprojected error in projector can be viewed as 0 pixel. The total errors in RIS can be viewed as 0.14 pixels.
The error map of the same piece of windshield from CMM is also applied to compare the calibration performance of the RIS system shown in figure 7.6.

Compared with the error map in last chapter 5.15, we can see that both error maps have similar distribution color coding. In the CMM error map, green color means that CMM measurement data has similar height with the CAD model. Blue color indicates that CMM data is lower than the CAD model. If CMM data is higher than the CAD, the color is red. The histogram in figure 7.6 shows that the error roughly ranges from $-5$ millimeters to 6 millimeters. Therefore, the magnitude of the error map is also similar with each other.

### 7.1.5 Inspection for Different Work Pieces

In order to test the capability of the inspection system, work pieces with different dimensions, geometric shapes, and optic surface properties are measured by the system. Figure 7.7 shows another tested work piece and its corresponding CAD model is given in figure 7.8. Figure 7.9 and 7.10 are the point cloud and error distribution map respectively. From the result in
this section and previous sections, it is demonstrated that the system is able to inspect work pieces without pre-treatment via multi-projection surface coding method.

Figure 7.7: The first tested work piece with bright color and more complex shape.

7.2 The Navigation System for Single Direction

The structured light system could also be utilized into the navigation application via the single-shot coding method. This section will provide some experimental result for such a system.
7.2.1 The Prototype of the Navigation System for Single Direction

Figure 7.11 shows the prototype of the navigation system for single direction. The system consists of a camera, a projector and an infrared light filter. The system field of view is about 400 mm by 400 mm at the distance around 1000 mm. The projector shots the single-shot pattern into the tested scene and the infrared light filter eliminates the visible light. The near infrared light pattern is projected onto the scene and is reflected into the camera.

7.2.2 The Fusion Result for the System

Figure 7.12 shows the tested objects in front of the sensor. The white object is a little further than the wooden one, and the wooden one forms a certain slop. The captured image from
the camera is shown in figure 7.13. Although people could not see the projected pattern, the camera could view both the near infrared light pattern and the scene at the same time. The experimental result is given in figure 7.14. The depth information is transferred into a color code and is fused into the image display. Conventional the operator could only get 2D information and the distance between the object and the camera has to be estimated. After the color-coded depth information is fused into the image, operator could control the mobile base more accurately. Furthermore, once the projection pattern is shotted onto a living object, such as a person or an animal, the infrared light pattern will not disturb them.
7.2.3 Integration the Single Direction Sensor with a Mobile Robot

After tested the developed sensor, the second step is to integrate it into a mobile robot. Figure 7.15 shows that the single direction sensor is put in front to the mobile robot and there is a robot arm above the sensor. Once the sensor returns the fusion results, different operations can be achieved based on 3D and 2D fusion information. The whole system can be described as figure 7.16. The system mainly contains Segway mobile platform, SCHUNK 7DOF LWA manipulator, near-infrared 3D sensor, phantom force reflect joystick, as well as tele-operation control computer. The mobile robot system work on two mode: the active and passive mode. The robot can mobile platform moves under the control of joystick by tele-operation, called passive mode. Similarly, the 7DOF manipulator can autonomously grasp
the target object in the workspace with the help of the 3D sensor perception at the active mode. The communication between the remote operator and robot system takes advantage of the internet.

Figure 7.17 illustrate the result of the integrated system. The robot arm is initially at a different position. After the fusion sensor returns the 3D coordinates of the desired part, the robot picks the part and lifts it up. The whole process demonstrates the feasibility of fusion sensor based control.

### 7.3 The Omni-direction Navigation System

In order to avoid the blind area, an omni-directional navigation system is also developed. Hyperbolic mirrors are added in front of the camera and the projector. Current experimental results are given in this section. The prototype of the omnidirectional sensor is implemented as illustrated in figure 7.18. The detail view of the developed omnidirectional 3D camera is in the right side of the figure. A projector associated with a hyperbolic mirror is used...
as an omnidirectional structured light emitter. A CCD camera (model M056-ZA180-002, resolution $640 \times 480$) with a hyperbolic mirror is used as an omnidirectional camera to view the panoramic view. Both imaging devices are put vertical and form the designed 3D omni-directional camera. The whole omnidirectional perception system was mounted on the back of the mobile base and there is a robot arm for operator to do the manipulation tasks.

The field of view (FOV) of the hyperbolic mirror is expected around 360-degree in horizontal plane with the upper elevation angle approximate 30-degree with respect to the horizontal plane. The mirror parameters $a$ and $b$ along the semimajor and semiminor axes
are $a^2 = 1248.2060$, $b^2 = 351.7940$, respectively.

In order to verify the performance of the designed 3D camera, the sensor is put into a room. The distance from the walls to the sensor is used to evaluate the developed 3D camera. The omnidirectional 3D camera was roughly located at the center of the room and obviously the room corners are further than the walls. Due to the limitation of the projector illumination the environment lights are turned off, when the structured light is projecting. The projected image is shown in figure 7.19. As explained in previous chapter, several concentric circles are projected from the projector and the radii of these light rings are used to perceive the object depths. From the figures 7.19, all the projected light rings are almost parallel in the task space. The positions of the rings change according to the object distance: once the distance increase, the rings lifts up; if the distance remains almost same, the rings also remain horizontal and are paralleled with each other. The radii distortions of
the projected rings are used to rebuild objects depths. From the figures, we can easily see that the rings lift up when they reaches the corners. The received image by the omnidirectional camera is also indicated by figure 7.20. Four walls are labeled by four directions: from east to north. The frequencies and the density variations for the light rings can be easily obtain to separated each light rings. The phase angle of corresponding pixels are preserved by epipolar constraints to separated each pixels along the rings. The radii of the rings are changing according to the distances. We can see how the ring radii change at the four corners of the room. Due to the low resolution of the camera, the radii changes around the south corners are not very big, but there are a clear radii changes around the north corners. It is because that the sensor is quite far way from the north wall but near to the south wall.

The second image is taken when the light is turned on and the fusion result is shown in figure 7.21. Different colors are assigned to the light rings to indicate the object depths. Blue color means the distance is far: around 4 meters and red means near: around 2 meters.
A panoramic view is also provided by figure 7.22 for a better display. From the image, we can see that the north wall is farther than the rest, since its color is assigned as blue. The ring patterns clearly raise at the corners indicating that the corners are further than the walls. The assigned colors also show that corner are further than walls. Because the figure 7.21 does not fully cover the conic image, there are several black regions in the un-warped panoramic view. One is above the person’s head in the figure 7.22 and two are in the top right and top left respectively.
7.4 Chapter Summary

The chapter provides all experimental results of the developed measurement systems. The developed RIS system could utilize the single-shot surface coding method to inspect a surface pre-treated work piece and obtains a fairly dense point cloud. With the aid of multi-shot coding method, the system could inspect optically-challenging objects without surface pre-treatment and obtains a high dense point cloud. Multi-sensor strategy is applied to enlarge the system field of view. A painted windshield is employed to test this mutli-sensor inspection system. The final point cloud is merged from 4 individual one. A comparison between CMM result is used to verify the result accuracy. Using single-shot coding method and infrared light, a structured light based navigation sensor is developed. A fusion image with color coded depth information is provided for the operator. Therefore, the operator can view the image and obtain the depth information at the same time. In order to avoid blind area, an omni-direction structured light based measurement sensor is developed. The experimental result for the prototype of the senor is provided. The unwrapped fusion result is given.
However, the accuracy of the measurement result still needs to be improved.
Figure 7.17: The result of designed system.
Figure 7.18: The developed omnidirectional 3D sensor.

Figure 7.19: Examples of the projected pattern on the left corner of the room.
Figure 7.20: The received image by the omnidirectional camera.
Figure 7.21: The fused result with the object depth.

Figure 7.22: The unwarped fusion result.
8.1 Conclusions

Based on the results and discussion presented in previous chapters, the following conclusions can be made:

1. A 3D sensor model with less calibration parameters is developed. Traditional triangular-based 3D reconstruction model employees the incident light from a projector and the reflection light to a camera to calculate the intersection point. The recovery of both lights demands parameters for these two devices, which in total is 22 parameters, to reconstruct a 3D point. Due to nature of the projector, not an information receiving device, projector calibration error becomes the major error sources of a structured light system. The novel 3D model optimizes the traditional model by reconstructing 3D point only in camera frame, resulting that the rotation and intrinsic parameters of a projector are avoided. The pixel deviation in camera image frame between two inspections: measurement of a plane and measurement a test, is employed to reconstruct the 3D coordinate. Two important concepts in multi-view perspective geometries, homography and epipolar, are utilized in the 3D model optimization.

2. A large scale structured light based inspection system for industry application is
developed and is successfully implemented. The system is capable to inspect a large moving
diffuse reflective object with one-shot algorithm and to inspect a large static object exhibiting
a mixture of both diffuse reflection and specular reflection. The inspection volume of the
system is much bigger than traditional 3D optic sensors’.

3. A multi-plane calibration strategy is developed to calibrate the new sensor model. In
order to further increase the accuracy of the system, the residual function for each pixel is
approximated by a second order polynomial. The calibrated sensor can reconstruct 3D point
cloud and represent the 3D shape of the test objects.

3. New surface coding methods are developed: one-shot algorithm for moving object and
multi-shot algorithm for free-formed object.

4. Registration between each sensor frame and common frame is developed based on a
modified ICP algorithm. Four individual point clouds finally merge into a entire one repre-
senting the whole piece of the object. An innovative object-oriented registration between the
final point cloud and CAD is developed to generate the error distribution map. Octree space
labeling and multi-resolution method are utilized to speed up the ICP based fine registration
process.

5. Inspection results are compared the CMM data, which represents the ground truth, to
valid the new sensor model and its corresponding calibration. The mean Euclidean difference
between 3D point clouds, the CMM data and the RIS data, is about 0.1 millimeter. CMM
data and RIS data exhibit the similar error distribution map for a same work piece.

6. The navigation system for the mobile robot control is developed via one-shot surface
coding method. The system can simultaneously provide the 3D image fused with the depth
information. Therefore, the operator could control the mobile robot more accurate.

7. An omni-direction system was developed. Its sensor model, and its corresponding
calibration, error analysis, and coding method are developed. Primitive results verified the
feasibility of the developed sensor.
BIBLIOGRAPHY
BIBLIOGRAPHY


177


