USE OF STRUCTURED LIGHT FOR 3D RECONSTRUCTION

SONG, Zhan

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy
in
Mechanical and Automation Engineering

© The Chinese University of Hong Kong

June 2008

The Chinese University of Hong Kong holds the copyright of this thesis. Any person(s) intending to use a part or whole of the materials in the thesis in a proposed publication must seek copyright release from the Dean of the Graduate School.

Acknowledgments

I would like to thank all those people who have helped me during the PhD study and thesis writing. Without them this thesis would not have been possible.

Firstly, I would like to express my gratitude to my PhD supervisor Professor Ronald Chung for having given me a chance to start my research work in computer vision. I am grateful for his encouragement and friendship over the years.

I would like to acknowledge the support of the members of my Ph.D. thesis committee: Professor Ru-xu Du, Professor Yun-hui Liu, and Professor Y.S. Hung of the University of Hong Kong. I appreciate their comments and suggestions on my research over the years.

Special thanks go to staff members Ms. Yuet-lin Kan, Maggie Chan, Joyce Wong, Winnie Wong, and Ms. Kit-wah Yau for their administrative assistance.

I would also like to thank all the friends I have had in the Computer Vision Lab of the Chinese University of Hong Kong, for creating a comfortable research atmosphere, and for their friendship. Special thanks go to Wei Wang, Yong He, Ding Yuan, Hong-Ming Chim, and Ming Zhao.

I want to thank the support from my whole family, my parents, my sister, my grandparents, and my parents-in-law.

At last, also the most important, I would like to express my gratitude and love to my wife Ming Ming for her patience over four years waiting in another city.

SONG, Zhan

13 June 2008

Abstract

The adoption of structured light illumination has been proven an effective and accurate visual means for 3D reconstruction. The system consists of a projector that illuminates controlled pattern or patterns to the target object, and a camera grabbing image or images of the illuminated object. Once correspondences between positions on the projector's pattern panel and positions on the camera's image plane are established, simple triangulation over light rays from the projector and the corresponding light rays to the camera would recover 3D information about the target object. Key issues involved in the approach include (1) *Calibration*: how the projector and camera can be calibrated so that metric measures about the object can be extracted from the image data; (2) *Image Feature Extraction*: what image features to use and how to extract them accurately from the image data; and (3) *Coding*: how the illuminated pattern can be designed so that each position of it embeds in the pattern a unique code which is preserved on the image side, so that position correspondences between the projector's pattern panel and the camera's image plane can be easily established. Each of the issues can affect the accuracy of the system. This thesis aims at providing improved solutions to each of these issues.

An accurate and convenient system for calibrating projector-camera system is presented. A consumer-grade LCD panel is used in place of the traditional printed pattern as the calibration plane. While patterns shown on the panel are used for camera calibration, when we turn the panel off (with its pose kept in space), patterns illuminated by the projector and reflected from it and captured by the camera can be used for the calibration of the projector. This way, patterns for calibrating the camera will not overlap with patterns for calibrating the projector, avoiding confusion in the image data. In addition, even a household-quality LCD panel has industrial-grade planarity. Experiments show that a setup as affordable as this can still have the system parameters calibrated in far less images with much higher accuracy.

On the image features for establishing correspondences between the projector's pattern panel and the camera's image plane, we propose to use rhombic pattern with binary (i.e., black and white) or colored elements as the projected pattern, and the grid-points between neighboring rhombic elements but not the centroids of the pattern elements themselves as the feature points. We show that, grid-point in the pattern owns the two-fold rotation symmetry, or so-called *cmm* symmetry, which is largely preserved on the image side after perspective projection and imaging. We propose a grid-point detector that exploits such a symmetry. By avoiding the direct use of raw image intensity, the detector is less sensitive to image noise and surface texture. Comparison with traditional operators shows its promising robustness and accuracy.

On the coding issue, we investigate a number of options. Coding can be established over time, and one widely used scheme in this direction is the adoption of Gray code in a series of binary patterns that are projected at different instants. We describe how the traditional Gray code patterns, if augmented by the use of strip shifting, can have the resolution of 3D reconstruction enhanced. The whole system setup of such a system is affordable, nonetheless experiments show that high accuracy can be achieved based on it. The disadvantage of such a system is mainly that multiple image captures are necessary for the operation.

Coding can also be established over the spatial domain (i.e., the plane of the 2D pattern). This corresponds to the use of color coded structured light in the literature. The advantage of such a system is that only one single image is necessary for 3D reconstruction. That makes it particularly suitable for real-time operation on objects that could move fast. We adjust a pseudorandom color pattern to one with rhombic pattern elements, and we use grid-points between the pattern elements, rather than the centroids of the pattern elements, as the image features, for largely enhancing not only the density but also the location precision of feature points. In addition, we show that not only are the grid-points useful, the grid-lines among the pattern elements can also be used to derive surface orientation information about the imaged object. The 3D position information as determined from the grid-points, and the 3D orientation as determined from the grid-lines, are information independently recovered, and they together would allow the 3D shape of the imaged object to be revealed in more complete form.

Finally, we explore how coding in structured light mechanism can be made even unnecessary. We adopt the above concept of recovering surface orientation from grid-lines, and show that by the use of a regular pattern, like a binary pattern with rhombic pattern elements, an orientation map about the imaged object can be recovered. Specifically, we show that here the correspondences over grid-lines between the projector's pattern panel and the camera's image plane can be approximated with a linear mapping and in turn boost the accuracy of surface normals calculation. We go on and show that, as long as no less than one reference point on the imaged object is available where absolute 3D is known, the above orientation map can even be converted to an absolute depth map by a simple integration process. Certainly with only one image capture as the input, with no coding injected into the projected pattern, the accuracy of 3D reconstruction is behind those of the time-coded (Gray code-based) and spatially coded (the color-code based) mechanisms. However, the need of a single image and the avoidance of the use of color (which can be confused by existing color on the imaged object) make it attractive for certain applications that demand only approximate shape of the object.

摘要

三維重構是計算機視覺領域的一個重要研究課題，其中的結構光技術因其構造簡單，精度和可靠性高而獲得了廣泛的應用和研究。結構光技術通過投射含有編碼信息的圖像，解決了立體視覺中的匹配問題。結構光系統主要包括三個核心問題：(1) 系統標定，即如何獲得投影儀和照相機的內部及外部參數；(2) 特徵檢測，即如何準確提取圖像中的特徵及其編碼信息；(3) 編碼，即如何組織投射出的光，從而使其包含盡可能多的有用信息。本文將針對以上三個問題進行深入的研究和探討，並提出改進的方案。

首先，本文針對目前投影儀－照相機結構光系統標定中存在的問題，提出了一種精確而方便的標定方法。與傳統的標定方法不同，我們使用了一種基於 LCD 模板的標定方案。首先通過 LCD 所顯示的高精度標定圖案對照相機進行標定，然後再通過標定好的照相機對投影儀進行標定，由於 LCD 具有很高的平面度並可以顯示很高精度的標定圖案，因而只需較少的標定圖像就可以獲得很高的系統標定精度，並可以有效地控制標定誤差從照相機到投影儀的擴散。

特徵提取是繼標定之後的又一重要步驟，圖像特徵的定位精度直接決定了系統最終的重構精度。本文提出了一種基於 *cmm* 對稱的網格點檢測算法，利用透視投影過程中，該點局部區域內 *cmm* 對稱特徵基本保持不變的特性，實現了網格點的精確定位。由於沒有直接使用傳統方法中的圖像灰度和顏色信息，從而使得該方法具有很高的抗圖像噪聲能力，即使在投影圖像發生嚴重變形和物體表面存在紋理的情況下，仍然能夠準確的進行圖像特徵的檢測。

在編碼方面，針對不同的編碼原理，我們進行了廣泛而深入的研究。首先，在傳統格雷碼的基礎上，我們在其後期編碼中加入移動條紋，通過對條紋邊緣的亞象素位

置進行編碼，從而提高了圖像特徵的定位精度，結合前文的高精度標定方法，使得我們在普通的投影儀－照相機系統的基礎上，實現了較高的重構精度。

時間域上的編碼方法能夠獲得較高密度的編碼點，但需要多張圖像資訊，這也降低了其處理效率，使其不適合有動態處理需求的場合。相比之下，空間域上的編碼方法將編碼過程限制在了單張圖像內，通過編碼點周圍顏色或特徵的不同組合，對其進行唯一的編碼。基於這種原理，本文提出了一種偽隨機彩色空間編碼方法，利用偽隨機矩陣的窗口特性，構造出了基於 $GF(4)$ 的偽隨機矩陣，並選擇菱形特徵為其基本元素，將各元素之間的網格點作為編碼點或者特徵點，利用其周圍元素的顏色序列對其進行編碼。對比傳統方法所使用的質心方案，我們所提出的設計方案不僅使可編碼特徵點數目提高一倍，而且精度也有很大的提高。我們進一步證明，網格點的法向量可以由形成該點的網格線的梯度和系統幾何參數來確定。這種法向量恢復方法不依賴三維深度信息，而其精度卻超過了僅僅使用三維深度信息所估計出的結果。利用我們所提出的空間編碼方案，通過單次投影和單張圖像，就可以恢復出編碼點處相互獨立的高精度的深度和法向信息。

最後，我們又探討了如何不使用編碼也能夠恢復出物體三維信息的問題。我們採用了一個普通的不包含任何編碼信息的菱形黑白圖案，採用上文所提出的法向恢復方法，我們就可以計算出網格點處的法向量，從而得到一個離散的法向圖。將其進行規則化處理，並引入可積和平滑約束之後，就可以通過傳統的積分方法獲得物體的三維形狀，然而這種基於積分的三維重構結果並不具有真實的空間尺度。我們進一步證明了，只需要給出一對匹配點（積分模型和真實模型），就可以將積分得到的三維資料轉化為具有真實尺寸的三維模型。對比基於三角計算的時間域和空間域上的編碼方案，其精度要略低；然而，該方法只使用一張圖像，而且避免了複雜的編碼和解碼過程，使得其可以應用於某些精度要求不高的場合。

Content

Abstract ..... **i**

摘要 ..... **iv**

Content ..... **vi**

List of Figures ..... **xi**

List of Tables ..... **xvi**

Chapter 1 ..... **1**

Introduction **1**

1.1 Background *1*

1.2 Motivation *4*

1.3 Objectives *5*

1.4 Organization of the Thesis *6*

Chapter 2 ..... **8**

3D Recovery **8**

2.1 Introduction *8*

2.2 3D Position Recovery *11*

2.2.1 Time of Flight 11

2.2.2 Laser Scanning 11

2.2.3 Moiré Method 12

2.2.4 Shape from Focus 13

2.2.5 Stereo Vision 14

2.2.6 Structured Light-based Systems 15

2.3 Structured Light Techniques for 3D Depth Recovery *16*
2.3.1 Epipolar Constrain 16
2.3.2 Triangulation 18
2.3.3 Pattern Codification Strategies 21
2.4 Surface Orientation Recovery *34*
2.4.1 Shape from Shading 35
2.4.2 Photometric Stereo 36
2.4.3 Shape from Polarization 38
2.4.4 Structured Light Projection Method 39
2.5 Summary *41*

Chapter 3 42

Calibration of Projector-Camera System 42

3.1 Introduction *42*
3.2 Previous Work *44*
3.3 Projector-Camera System Model *47*
3.4 Calibration using LCD Panel *48*
3.4.1 Camera Calibration 49
3.4.2 Projector Calibration 52
3.4.3 Calibration Flow 55
3.4.4 3D Triangulation 56
3.5 Experimental Results *57*
3.5.1 Lens Distortion 60
3.5.2 Re-projection Error 61
3.5.3 Planarity Check 63
3.5.4 Motion Measurement 65
3.6 Summary *68*

Chapter 4 69

Image Feature Extraction: the Grid-points 69

4.1 Introduction *69*

4.2 Previous Work *70*
4.2.1 Laplacian of the Gaussian Operator 71
4.2.2 Harris' Corner Detector 72
4.2.3 SUSAN Corner Detector 72
4.2.4 Spline Fitting Detector 73
4.2.5 Specialized Grid-point Detector 74
4.3 cmm-Symmetry based Grid-point Detector *74*
4.3.1 Hypothesis of Grid-point Positions 76
4.3.2 Confirmation of Grid-points and Position Refinement 77
4.4 Experimental Results *79*
4.4.1 Robustness Check under Extra Gaussian Noise 79
4.4.2 Color Pattern under Extra Gaussian Noise 81
4.4.3 Experiment with Spherical Surface 81
4.5 Summary *83*

Chapter 5 84

Coding over Time: Gray Code plus Strip Shifting Mechanism 84

5.1 Introduction *84*
5.2 Strip Shifting Combined with Gray Code Pattern *86*
5.2.1 Gray Code used with Strip Shifting Pattern 87
5.2.2 Strip Edge Detection 88
5.3 Experimental Results *89*
5.3.1 Calibration Accuracy 90
5.3.2 Metric Reconstruction Accuracy 90
5.3.3 Reconstruction of Free-form Shapes 91
5.3.4 Reconstruction of Texture and Reflective Object 94
5.4 Summary *95*

Chapter 6 96

Coding over Spatial Domain: Color Coding plus the use of Grid-points as Image Features 96

6.1 Introduction *97*

6.2 Determine Surface Normal from Image Tangents *98*
6.2.1 The Underlying Principle 98
6.2.2 The Path from Image Tangents to 3D Orientation 100
6.2.3 The Need of Correspondence 101
6.3 2D Rhombic Pseudorandom Pattern Design *101*
6.3.1 Background 102
6.3.2 Pseudorandom Sequence 103
6.3.3 Pseudorandom Array 104
6.3.4 Pattern Image Construction 106
6.3.5 Feature Definition 110
6.3.6 Grid-point vs. Centroid 112
6.3.7 Explicit 3D Reconstruction at Grid-points 114
6.4 Experimental Results *115*
6.4.1 3D Depth Recovery 115
6.4.2 Surface Orientation Recovery 118
6.4.3 Results on Bust and Human Face 119
6.5 Summary *123*

Chapter 7 124

Coding-Free Mechanism: Projection of Merely a Regular Pattern 124

7.1 Introduction *125*
7.2 Correction for Discrepancy from Perspective Projection *126*
7.3 Absolute 3D Determination from Normal Integration *127*
7.3.1 From Image Plane *x-y* to Object Plane *X-Y* 128
7.3.2 3D Recovery from Surface Orientation Map 130
7.3.3 A Weighted Procedure for Absolute Depth Recovery 133
7.4 Experimental Results *135*
7.4.1 Accuracy of Surface Orientation Recovery 136
7.4.2 Spherical Reconstruction 136
7.4.3 Reconstructions of Bust and Human Face 139
7.5 Summary *142*

Chapter 8 ..... **143**

Conclusion and Future Work **143**

8.1 Conclusion and Contributions *143*

8.2 Related Publications *146*

8.3 Future Work *147*

Bibliography **149**

List of Figures

Figure 2.1 Overview of the 3D recovery techniques that mentioned in this thesis. 10

Figure 2.2 Illustration of laser strip scanning. 12

Figure 2.3 Principle of shape from focus method. A sequence of images is obtained by shifting the object in z direction. The focused regions in each image were extracted by a focus judging function. Depth values at these focused regions can be calculated. 14

Figure 2.4. Illustration of stereo vision system. Given a 3D point M and its projection *m* in one camera, the correspondence problem is to find the other projection point *m*'. 15

Figure 2.5 Epipolar constrain between two camera image planes. 18

Figure 2.6 In real image, the image points y_1 and y_2 cannot be measured with arbitrary accuracy. Instead points y'_1 and y'_2 are detected and used for the triangulation. The corresponding projection lines do not, in general, intersect at point X. 20

Figure 2.7 Gray code pattern (top) and sinusoidal phase shifting (bottom) pattern. 22

Figure 2.8 Gray code (top) and line shift (bottom) pattern used in [Gühring 2000]. 23

Figure 2.9 Coded structured light patterns for spatial-temporal selection. 26

Figure 2.10 (Left) A grid pattern consisting of vertical and horizontal colored slits [Salvi 1998]. (Right) M-array color pattern with window property of 3×3 proposed in [Morano 1998]. 28

Figure 2.11 A pattern of vertical colored stripes with a horizontal sinusoid overlaid. 29

Figure 2.12 A 2D color pattern that designed for dynamic feature tracking task [Adan 2005]. 31

Figure 2.13 A 2D structured light pattern based M-array. The pattern element is defined to a small set of symbols associated to simple geometrical primitives [Albitar 2007]. 32

Figure 2.14 (Left) A color pattern designed (53×38) by [Desjardins 2007], it contains 3 colors and its window size is 3×3. (Right) A 38×212 tessellated pattern designed by [Chen 2008], 7 colors are used, and its codeword is formed by its 4-adjacents (i.e. top-left-bottom-right). 34

Figure 2.15 Photometric stereo methods. (Left) Surface normal is estimated from multiple illuminations. (Right) Image captured from different illumination directions. .. 38
Figure 2.16 Reflectance map in *p-q* space. (Left) Single image provide single *p-q* curve in the shape from shading method. (Right) Multiple images captured from different illumination directions are used to solve the unique solution in photometric method. 38
Figure 2.17 Projection of a stripe edge onto an object surface and reflected to the image plane. 40
Figure 3.1 The geometry between the camera, projector, and the world coordinate frames. With the geometry calibrated, depth value of any world point P can be calculated from the two corresponding points: (u^c, v^c) and (u^p, v^p) via triangulation. 48
Figure 3.2 Experimental system which consists of a DLP projector, a DSLR camera and a LCD display for system calibration. 58
Figure 3.3 (Top) The checkerboard pattern displayed on the LCD screen for camera calibration in our setup. (Bottom) Patterns projected onto the LCD screen for projector calibration. 59
Figure 3.4 (Left) The checkerboard pattern print of grey level 128 that was for camera calibration under the traditional calibration design. (Right) The reflected pattern on top of the above pattern, that was for projector calibration. 60
Figure 3.5 (Top) Radial distortion of camera. (Bottom) Radial distortion of projector. 61
Figure 3.6 RPE distribution: (a) Camera calibration result. (b) Projector calibration result, both from the use of LCD panel as the calibration object. (c) Camera calibration result. (d) Projector calibration result, both from the use of printed pattern as the calibration object. 62
Figure 3.7 (Top) Residual errors with respect to the fitted plane under the proposed calibration setup. (Bottom) Reconstruction of the plane with respect to the camera coordinate frame. The reconstructed plane had an orientation [-0.4611, 0.0259, -0.8870] with respect to the camera reference frame. 64
Figure 3.8 (Top) 11 reconstructed planes under the proposed calibration. (Bottom) Side view of the reconstructed planes. 66
Figure 4.1 Pattern with rhombic elements. (Left) The grid-point can be perfect located in the original pattern image. (Right) After projection, the grid-points are invisible. 75
Figure 4.2 Response value of the proposed operator at a real image. 78

Figure 4.3 Grid-point extraction results of three operators. (a) Result from Harris' corner detector. (b) Harris' corner detector on noisy image. (c) Result from SUSAN corner detector. (d) SUSAN corner detector on noisy image. (e) Result from the proposed corner detector. (f) The proposed corner detector on noisy image..... 80

Figure 4.4 Grid-point detection results on a color pattern image with extra Gaussian noise (σ = 0.03). (a) Result from the template method. (b) Result by the proposed detector. 81

Figure 4.5 Grid-point detection results on a spherical surface. (a) Result from the template method. (b) Result by the proposed grid-point detector. 82

Figure 5.1 The use of Gray code together with strip shifting for 3D reconstruction. (Top) A series of 8 Gray code patterns are used to construct 2^8 sub-regions each with a unique 8-bit long codeword. (Bottom) A strip pattern of the width 4 pixels is shifted 3 times to encode positions within each sub-region..... 88

Figure 5.2 (Left) Principle of zero-crossing edge detection. (Right) Sub-pixel edge detection from two complementary patterns. 89

Figure 5.3 *RPE* distribution of camera (left) and projector (right) under the calibration results. 90

Figure 5.4 Metric reconstruction quality evaluation experiment with a spherical object. (a) A ball imaged under strip pattern projection. (b) Encoded image with a periodic ambiguity of 8 pixels. (c) Final encoded image after ambiguity removal. (d) Reconstructed 3D model with the proposed pattern design..... 92

Figure 5.5 Quality evaluation of the reconstruction of a bust model. (a) Encoded image with a periodic ambiguity of 8 pixels. (b) Final encoded image after ambiguity removal. (c) Reconstructed 3D model without using strip pattern. (d) Reconstruction result with the improved pattern design..... 93

Figure 5.6 Quality evaluation of a ceramic teapot which have plentiful texture and reflective. (a) Object imaged under strip pattern projection. (b) Encoded image with a periodic ambiguity of 8 pixels. (c) Final encoded image after ambiguity removal. (d) Reconstructed 3D model with the improved pattern design. 94

Figure 6.1 Determining surface orientation from two image tangents, for object surface under pattern illumination. 99

Figure 6.2 A partial view of pseudorandom array with size of 65×63. (a) A part of the pseudorandom array, *2-3-2-2-1-3* indicates one unique codeword. (b) In the

pattern of image form, the codeword *2-3-2-2-1-3* is replaced with a color sequence: *G-B-G-G-R-B*. 108

Figure 6.3 A 2D pseudorandom color pattern with rhombic elements: the diagram of pattern composed of 65×63 elements which belong to red, green, blue, and black color. 109

Figure 6.4 Grid-point P_1 and P_2 have the same codeword: *c1–c2–c3–c4–c5–c6*. 111

Figure 6.5 The coding process of grid-point P_1. When P_1 is encoded, its nearest grid-point P_2 will share the same codeword (i.e. c1–c2–c3–c4–c5–c6) with P_1. 111

Figure 6.6 Detection of centroids and grid-points. (a) Human face under projection of rhombic color pattern. (b) The grey level image transformed from the color image. (c) Segmentation of pattern elements by a morphological method. (d) Result of centroid detection.. 113

Figure 6.7 Result of grid-point detection by the proposed detector, which has the grid-points divided into two types P1 and P2 (marked as red and green dots respectively).114

Figure 6.8 Reconstruction of a sphere and a ribbon-like surface. (a) A sphere with color pattern projected on it. (b) Reconstructed 3D surface. (c) A ribbon-like surface with color pattern projected on it. (d) 3D reconstruction result. 117

Figure 6.9 Recovered surface orientation map of a sphere (left) and ribbon-like surface (right). 118

Figure 6.10 Reconstruction of a bust. (a) A bust under pattern illumination. (b) Grid-point detection result. (c) Interpolated surface only using the recovered 3D points. (d) Interpolated surface orientation map only using the recovered surface normals at the grid-points. 120

Figure 6.11 Reconstruction of human face. (a) Human face without pattern illumination. (b) Human face with pattern projection. (c) Grid-point detection result. (d) Interpolated surface orientation map only using the recovered surface normals at the grid-points. 121

Figure 6.12 Reconstructed 3D model of a human face. (Top) Interpolated 3D model only using the reconstructed depth at the grid-points. (Bottom) 3D model with texture mapping. 122

Figure 7.1 Illustrations of three projection models: perspective, orthographic and weak perspective. 129

Figure 7.2 A regular rhombic pattern with three colored elements embedded. 135

Figure 7.3 Calculation of s_x and s_y, which transforms image plane x-y to object plane X-Y. 136
Figure 7.4 Reconstruction of a spherical surface by the proposed method. (a) Object surface under pattern illumination. (b) Grid-points detection result. (c) Reconstructed local surface orientations at the grid-points. (d) Integrated 3D surface without using reference points. 138
Figure 7.5 Reconstructed 3D model with absolute scale by the proposed method. 139
Figure 7.6 Reconstruction of a bust model by the proposed method. (a) Bust surface under pattern illumination. (b) Interpolated orientation map. (c) Reconstruction result with in absolute scale. 140
Figure 7.7 Reconstruction of a human face by the proposed method: (a) human face under pattern illumination; (b) grid-point detection results; (c) interpolated surface orientation map; (d) integrated 3D model; (e) reconstructed human face with metric measure. 141

List of Tables

Table 3.1 Calibration results from the use of LCD and printed pattern respectively (Unit: pixel). 67

Table 3.2 Measurement results of the various motion steps of a moving plane. (Unit: mm) 67

Table 6.1 Pseudorandom sequence and array size in GF(2), where 4≤m≤16: L represent the length of sequence, k1 and k2 indicate the unique window size, n1 and n2 indicate the array's dimension. 107

Table 6.2 Pseudorandom sequence and array size in GF(4), where 4≤m≤9: L represent the length of sequence, k1 and k2 indicate the unique window size, n1 and n2 indicate the array's dimension. 107

Table 6.3 A part of the pseudorandom sequence with length 4095 and the window size is 1×6. 109

Table 6.4 Accuracy comparison between some structured light systems. 116

Chapter 1

Introduction

Structured light-based system is an effective visual mechanism for 3D reconstruction. The core of its design is to replace one of the cameras in stereo vision by a projector, and by so doing convert the image in that camera as a projected pattern that is known and controllable. This way, the challenging correspondence problem in stereo vision can possibly be bypassed. The way to bypass the correspondence problem is to embed unique code to each of the positions of the projected pattern, and should such codes be preserved in the image of the camera, correspondences can be readily established between the projector (the first "camera") and the second camera. This chapter presents an overview of what the major issues are in the structured light-based mechanism, and what this thesis aims at addressing on the mechanism.

1.1 Background

Vision is the process of discovering what is present in the external world from images. As a frontier intellectual field, like any frontiers, it is exciting but disorganized; there is often no reliable authority to appeal to many useful ideas have no theoretical grounding, and some theories are useless in practice; developed areas are widely scattered, and often looks completely inaccessible from the other [Forsyth 2003].

Marr's book entitled *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* [Marr 1982] is a lucid presentation

which proposes a general theory of the visual processing stages up to (but not including) object recognition. The framework of this theory is based on three main symbolic representations of the visual world, which are computational level, algorithmic level, and mechanism level. Marr termed the two levels of computational and algorithmic, placing them above the third, mechanism level, which, in the study of the brain, refers to the neuroanatomy and neurophysiology of the mechanisms of perception, cognition, and action. The first and the second level are also named computational vision or model-based vision. And the third level can be named cognitive vision or task-based vision. A good example that Marr used to illustrate the idea of separate computational levels was binocular stereopsis [Marr 1976]. It is also well known as stereo vision, which retrieves depth information from the disparity between two corresponded image features.

Stereo vision has been widely researched in past decades. A complete framework has been found to depict the geometry between multiviews [Hartley 2004]. The most difficult problem in stereovision is to determine which pair of projective points represents the same object point in 3D space. This problem is known as the correspondence problem, which is the main limitation of stereo vision. Although a set of geometrical constraints, such as epipolar geometry is able to simplify the correspondence problem. However, it is not a definitive solution [Faugeras 1993]. For example, the correspondence problem cannot be solved when observing an object without textures, or features only appear in one of the images due to a surface occlusion, or points are multiply matched between the images.

Active stereo vision provides a solution to the correspondence problem. This technique is called active because it is based on changing the illumination of the scene by projecting structured light on it. The use of structured light solves the limitation of passive stereovision when observing non-textured objects. The core of this technique is to replace one of the cameras in stereo vision by a projector, and by so doing convert the image in that camera as a projected pattern that is known and controllable. This way, the challenging correspondence problem in stereo vision can possibly be bypassed. The way to bypass the correspondence problem is to embed unique code to each of the positions of the projected pattern, and should such codes be preserved in the image of the other camera, correspondences can be readily established between the projector and the camera. Once correspondences between positions on the projector's pattern panel and positions

on the camera's image plane are established, simple triangulation over light rays from the projector and the corresponding light rays to the camera would recover 3D information about the target.

Earlier structured light techniques were based on projecting some simple primitives such as a single dot or a single line of light, usually provided by lasers [Demeyere 2007]. The advantage of projecting such structured light primitives is that the correspondence problem of the illuminated points in the images is directly solved. Nevertheless, the number of correspondences in each image is very limited.

In order to increase the density of correspondences or pattern features, structured light patterns were introduced like arrays of dots, stripes, grids, or concentric circles. However, with this solution the identification of different pattern regions in the images becomes ambiguous so that the correspondence problem is not directly solved. This fact provoked the emergence of coded structured light. In this case, the projected patterns are coded so that each element of the pattern can be unambiguously identified in the images [Salvi 2004]. Therefore, the aim of coded structured light is to robustly obtain a large set of correspondences per image independently of the appearance or illumination condition of the object surface.

According to the coding schemes of the structured light pattern, it can be classified into temporal-based method and spatial-based method. The temporal coding schemes generate the codewords by projecting a sequence of patterns along time, so the pattern structure can be very simple, such as binary strips, lines etc. Since multiple patterns are used, dense surface points can be reconstructed by this means. Spatial coding scheme uses the codeword by its neighboring pattern elements. The coding process can be completed in a single pattern image. The pattern elements are usually designed as gray level, color, or a set of geometry primitives. To decode a spatial pattern correctly, not only the pattern feature itself but also its neighboring pattern elements should be extracted and recognized correctly. A high performance feature detector is demanded for this task. Due to the limits of projector and camera's resolutions, the number of pattern features is also limited. Only a sparse model with lower data density can be reconstructed by this method.

1.2 Motivation

Projector-camera structured light system has been widely used in industry and living because of its efficiency and convenience. There are three key issues involved in a projector-camera ranging system: (1) Calibration: how the projector and camera can be calibrated so that metric measures about the object can be extracted from the image data; (2) Image Feature Extraction: what image features to use and how to extract them accurately from the image data; and (3) Coding: how to design the illuminated pattern so that each position of it embeds in the pattern represent a unique code which is preserved on the image side. Each of the issues can affect the accuracy of the system.

Calibration is a crucial and usually the first step to any structured light system. With respect to the components that adopted by the system, the calibration means should be designed accordingly. In this thesis, we put our attentions to the most popular projector-camera based structured light system, which comprises one camera and one projector. After an investigation of existing calibration means, it is found that a more convenient and accurate calibration means for projector-camera based structured light system is quite needed.

When a pattern is projected on an object surface, the information that has been encoded inside it will be affected by a variety of factors unavoidably. For example, it can be affected by the surface shape, texture, illumination, occlusions etc. To decode an image with pattern projection, a robust and accurate feature detector is indispensably.

Pattern design is the kernel in a structured light system. According to the coding strategy, we can select to use a single or a series of patterns for the reconstruction. Obviously, the use of more patterns can encode more features and then generate denser 3D surface. However, the efficiency is sacrificed. The use of single pattern can deal with dynamic scenes or moving objects, but it can only produce few features or a sparse 3D model. How to make a trade off between efficiency and feature density is still an open problem in structured light domain.

When we talk about 3D recovery, we usually focused on how many and how accurate the 3D points can be achieved. The surface normal or orientation is often overlooked. It is usually considered as a by-product of the 3D position. Actually, surface orientation is

another important surface description as well as surface positions. It has been used to improve the precision of 3D geometry [Nehab 2005], to reconstruct the surface shape [Ikeuchi 1981], and even for objection recognition [Campbell 2001] etc. However, this important information is usually ignored in the structured light pattern design. There are still no related works about the structured light systems, which can recover both position and orientation independently and synchronously.

According to the 3D recovery principle, it can be categorized into two classes: triangulation-based and integration-based. Triangulation-based methods require precise correspondence between cameras or camera(s) and projector(s). The 3D information with absolute metric measure can be output. Structured light techniques can be concluded to this class. As for the integration-based method, it needs no correspondence. The 3D model is usually generated from the recovered surface orientation or image gradient maps. However, the integration is based on the image plane. That makes the integrated depth with no real metric meanings. How to reconstruct the absolute depth without using correspondences from single image is another open problem.

1.3 Objectives

The main objective of this thesis consists in developing a structured light system that can recover the 3D information, which includes not only surface position but also independent surface orientation via single pattern projection and image capture. All the three key issues: calibration, feature detection, and codification are involved.

In details, the objectives can be generalized as follows:

i. Design a more accurate and convenient calibration mechanism for the projector-camera based structured light system.

ii. Design a robust and accurate feature detector to decode the image with pattern projection.

iii. Develop a low-cost and high-accuracy dense 3D scanning system.

iv. Develop a mechanism for surface normal recovery without using 3D points.

v. Design a new single image based structured light pattern that permits more feature points to be encoded. Furthermore, at each encoded feature point, both its surface position and its surface normal should be recovered independently. Moreover, the accuracy of recovered surface normal should beyond that estimated only using the reconstructed 3D points.

vi. Propose a method without coding involved, but can recover the absolute 3D depth from single projection and imaging.

1.4 Organization of the Thesis

A comprehensive study of previous work on 3D recovery is addressed in Chapter 2. The survey is conducted from two aspects: 3D position recovery and 3D orientation recovery. A number of 3D position recovery methods are introduced, and the emphasis is put on structured light-based systems. The coding scheme is classified into temporal and spatial. The main coding strategies and recent progress are provided. In the review of surface orientation recovery, a number of methods such as shape from shading, shape from polarization, photometric stereo and structured light means are introduced and compared.

In Chapter 3, a novel calibration means for projector-camera structured light system is proposed. Firstly, previous work on this topic is reviewed and the existing problems are pointed out. Then a novel calibration method by the use of an LCD pattern is presented. Experimental results are offered to show its high accuracy and efficiency.

In Chapter 4, a novel grid-point detector is introduced. Grid-point in a rhombic pattern owns the two-fold rotation symmetry, or so-called *cmm* symmetry, which is largely preserved on the image side after perspective projection and imaging. We propose a grid-point detector that exploits such a symmetry. By avoiding the direct use of raw image intensity, the detector is less sensitive to image noise and surface texture. Experimental results from different feature detectors are also provided for comparison.

Chapter 5 describes a Gray code based temporal coding strategy. We describe how the traditional Gray code patterns, if augmented by the use of strip shifting, can have the resolution of 3D reconstruction enhanced. The whole system setup of such a system is affordable, nonetheless experiments show that high accuracy can be achieved based on it.

We show that, with the proposed design of calibration and coding scheme, even a structured light system made of household-quality projector and camera can achieve high quality 3D reconstruction.

In Chapter 6, we propose a new spatial encoded structured light pattern that can determine both surface position and orientation independently. The mechanism involves projection of a specially designed pseudo-random rhombic color pattern. Grid-points between pattern elements are defined as the pattern features. Each grid-point is encoded by the color profile of its neighboring pattern elements. We show that, with this design, the number of feature points can be doubled with respect to previous pattern, which defined the centroids of the pattern elements as the pattern features. The grid-point also with much higher location precision since it is insensitive to image noise and perspective projection. In addition, we show that the grid-lines among the pattern elements can also be used to derive surface orientation information about the imaged object. The 3D position information as determined from the grid-points, and the 3D orientation as determined from the grid-lines, are information independently recovered, and they together would allow the 3D shape of the imaged object to be revealed in more complete form. Extensive experiments including accuracy evaluation of both surface points and normals, reconstruction of free-from surfaces and real human face are provided.

Chapter 7 presents a coding-free scheme for absolute depth recovery. We adopt the above concept of recovering surface orientation from grid-lines, and show that by the use of a regular pattern, like a binary rhombic pattern, an orientation map about the imaged object can be recovered. Specifically, we show that here the correspondences problem can be avoided by making one specific design and one approximation. We go on and show that, as long as no less than one reference point on the imaged object is available where absolute 3D is known, the above orientation map can even be converted to an absolute depth map by a simple integration process. To boost the reconstruction accuracy, several colored elements are embedded into the binary pattern, and their centroids are reconstructed and used as the reference points. Experiments on different surface and real human face are provided to show its precision and feasibility.

Chapter 8 summarizes the contributions in this thesis and related publications. It further discusses how future work could extend on the results of this thesis.

Chapter 2

3D Recovery

3D information at a given surface point comprises not only its 3D coordinates but also its surface normal or orientation. Surface orientation can be estimated from the 3D position directly, but this will not bring new information that beyond 3D position itself. This chapter gives a comprehensive review of previous work on surface position and surface orientation respectively. We put our attention on the structured light system and the surface orientation recovery methods.

2.1 Introduction

With the development of science and technology, the need for 3D information is increasing dramatically. 3D reconstruction techniques have been introduced into more and more domains such as manufacturing, automation, medicine, biology, geography, entertainments, security, military etc. In manufacturing, the accurate measurement of 3D shapes can speed up and ensure product development and manufacturing quality [Moore 2001]. In automation and robotics, the 3D information can be used for control, obstacle detection, and guidance of intelligent robots [Loaiza 1999]. By scanning the human body, a 3D virtual model of the participator can be constructed in the computer [Treleaven 2007]. Then it can be used for various amusements such as computer games, special movie effects, and virtual dressing etc [Vassilev 2000]. In biology and medicine, 3D reconstruction of the organ or tissue can be used for pathology analysis, robotics surgery [Liere 2002] etc. In security, more and more 3D information have been introduced into

biometric like 3D face recognition [Alexander 2005]. It has been proven more accurate and robust than previous 2D image-based systems [FRVT 2006].

Varieties of methods for 3D recovery have been proposed in former works. According to measuring principles, it can be classified into contact and non-contact measuring methods. Coordinate-Measuring Machine (CMM) is the most popular and important means in contact measuring method. High-sensitive probes and sensors inside the machine can promise high accuracy. Since the measurement is operated point by point, that makes it low-efficiency [Haitjema 2001]. The contact of probe and object surface also limits its application, for example, the human face or mobile objects. The non-contact measuring means mainly refers to the optical measurement techniques. It usually projects a beam or a sheet of special light (laser or visible light) onto the inspected object surface. And then the reflected light will be captured by specialized sensors (e.g. CCD, CMOS). By analyzing the recorded signals, 3D distance between the object and sensor can be determined by respective mechanisms. The entire 3D model of object can be produced after a full scanning. According to the structure of the projected light, the optical measurement methods can be classified into time of flight, laser scanning, Moiré, stereo vision, shape from focus and structured light [Chen 2000] etc.

As the other principle surface descriptor, surface normal also plays important role in the 3D model. The orientation map can be used for precise 3D modeling [Nehab 2005], 3D registration [Fancourt 2007], 3D shape reconstruction [Chen 2004], and object recognition [Alessandro 2006] etc. Surface normal is a local description of surface property and is difficult to be measured directly. Traditionally, it is directly estimated from its surrounding 3D positions in a local area [Max 1999]. But that will not bring new information beyond the 3D positions. It can be considered as a by-product of 3D points. There are also some methods to recover surface normal without using the knowledge from 3D positions. For example, shape from shading, photometric stereo, shape from polarization, and structured light means etc. Figure 2.1 gives an overview of these techniques.

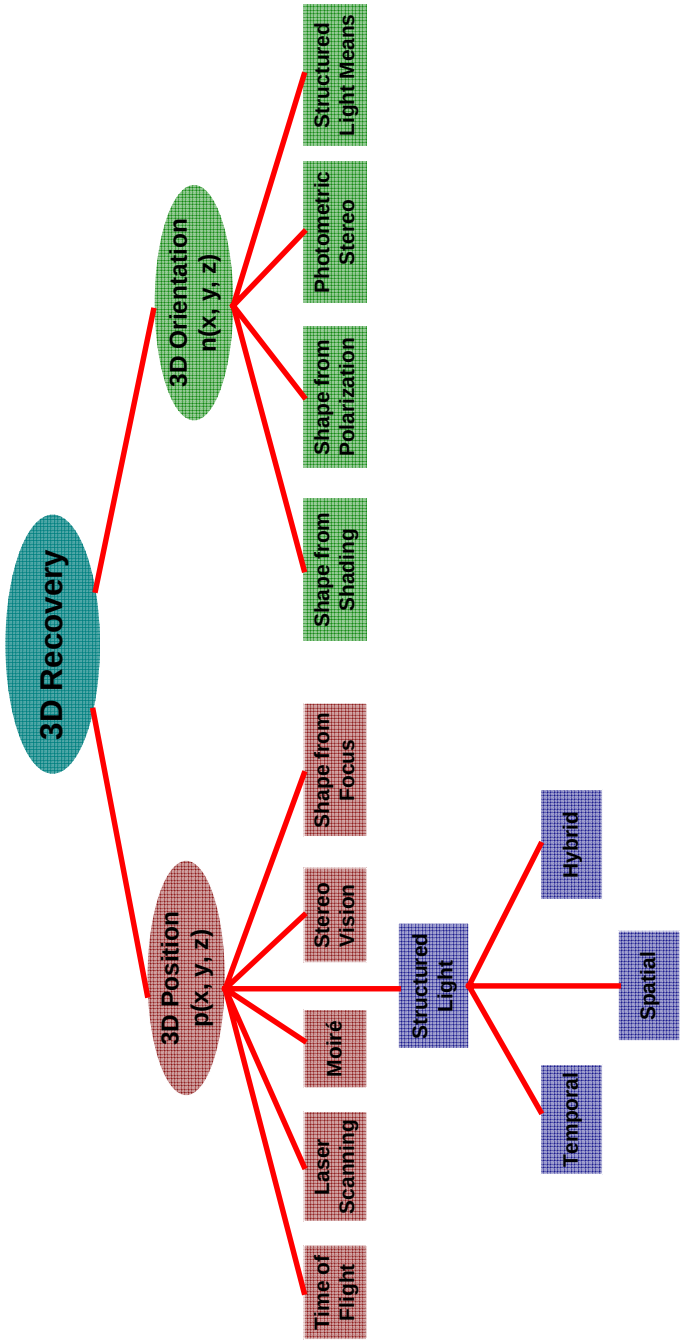


Figure 2.1 Overview of the 3D recovery techniques that mentioned in this thesis.

This chapter is organized as follows. In Section 2.2, a brief introduction of various 3D position recovery techniques is presented. In Section 2.3, structured light techniques including its principle, epipolar constrain, triangulation and a variety of different coding strategies are offered. In Section 2.4, previous work on image-based and structured light-based surface orientation recovery is introduced. A summary is offered in Section 2.5.

2.2 3D Position Recovery

This section gives a brief review of optical measurement techniques. According to the measuring principles, we classify them into time of flight, laser scanning, Moiré, shape from focus, stereo vision, and structured light systems.

2.2.1 Time of Flight

Time of Flight (TOF, or Pulse Echo) method is based on the direct measurement of the time of flight of a laser or other light source pulse. In the measurement, an object pulse is reflected back to the receiving sensor and a reference pulse is passed through an optical fiber and received by the sensor. The time difference between the two pulses is converted to distance. A typical resolution for the time of flight method is around 1 mm. With sub-picoseconds pulses from a diode laser and high-resolution electronics, sub-millimeter resolution is achievable. It's usually used in long-range measurement like geomorphology [Lange 2000].

2.2.2 Laser Scanning

A single laser beam projected on a 3D surface will produce a single light spot. Imagine that we have a 3D surface defined as function $z(x, y)$. When the illuminated spot is projected onto the surface at known x, y-coordinates and tracked by a camera, the surface height z at known coordinates in the x-y plane can be determined by measuring the geometric displacement of the image of the light spot on the CCD sensor of the camera and by solving the triangulation problem. Accuracy can be relatively high because the location of the light spot can be tracked with sub-pixel accuracy up to approximately 0.1 pixel, depending on the spot size, illumination, and surface smoothness. To speed up the scanning, a laser light sheet can be used to scan in one direction over the surface as shown in Figure 2.2. The measurement is done by detecting the laser line in the image

and simple geometric transformation. Moving or vibrating surfaces cannot be measured, as the scanning time is relatively large. Another factor that affects the measurement accuracy is the difference in the surface characteristic of a measured object from the calibration surface. Usually calibration should be performed on similar surfaces to ensure the measurement accuracy. In the work of [Sohmura 2004], a 3D laser scanner (VIVID-900) is used to scan a human face. And the smallest error of 0.2 mm was reported. In [Demeyere 2007], a laser strip scanning system was proposed to measure the spheres with radii of 6.5 cm and 13 cm. With a working distance of 2.5 m, the results show a general accuracy of 1% on the radius of the measured objects.

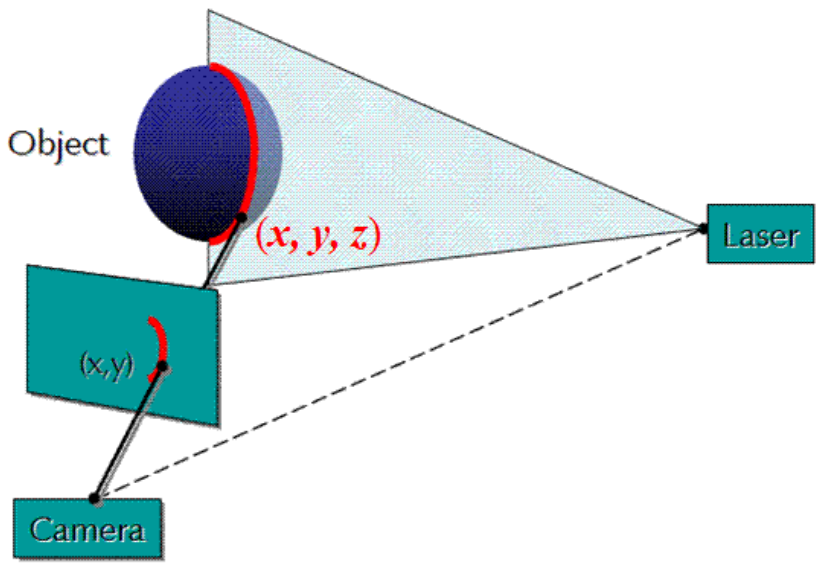


Figure 2.2 Illustration of laser strip scanning.

2.2.3 Moiré Method

The Moiré method can be divided into shadow and the more practical projection techniques [Asundi 1993]. The key to the Moiré technique is two gratings, one is a master grating, and the other is a reference grating, from which contour fringes can be generated and resolved by a camera. Increased resolution is realized since the gratings themselves do not need to be resolved by the camera. However, if the reference grating is computer generated, as in the logic Moiré method, the master grating must be resolved by

the camera. The penalties for the high resolution are the implementation complexity and the need for a high power light source as compared with a structured light technique. A comprehensive comparison of some high-speed Moiré contouring methods with particular stresses on sources of noise and system error functions can be found in [Harding 1998]. The typical measurement range of the phase shifting Moiré method is from 1 mm to 0.5 m with the resolution at 1/10 to 1/100 of a fringe.

2.2.4 Shape from Focus

The basic idea of this method is that objects at different distances from a lens are focused at different distances. Shape from focus is the problem of reconstructing the depth of the scene changing actively the optics of the camera until the point of interest is in focus. The change in the optics is obtained by changing either the lens position or the object position relative to the camera. The point in focus gives information about its depth through the thin lens Gaussian Law [Nayar 1994]. Shape from focus is a method that searches the camera parameters (lens position or focal length) that correspond to focusing the object. In the method, an unknown object is moved with respect to the imaging system and a sequence of images that correspond to different levels of object focus is obtained as shown in Figure 2.3. A focus measure is computed in the small image regions of each of the image frame in the image sequence. The value of the focus measure increases as the image sharpness or contrast increases and it attains the maximum for the sharpest focused image. Thus, the sharpest focused image regions can be detected and extracted. This facilitates auto-focusing of small image regions by adjusting the camera parameters so that the focus measure attains its maximum value for that image region. In addition, such focused image regions can be synthesized to obtain a large image where all image regions are in focus. Further, the distance or depth of object surface patches that correspond to the small image regions can be obtained from the knowledge of the lens position and the focal length that result in the sharpest focused images of the surface patches. It measures the amount of blur in an image and needs as low as only two images with different optical settings to obtain the depth map of an entire scene. In the work of [Subbarao 1995], a measuring system based focus image surface was proposed. A mean error of 12.4 mm or 2.07% at a working distance of 0.6 m was reported. There was a linear relation between the measurement error and the working distance.

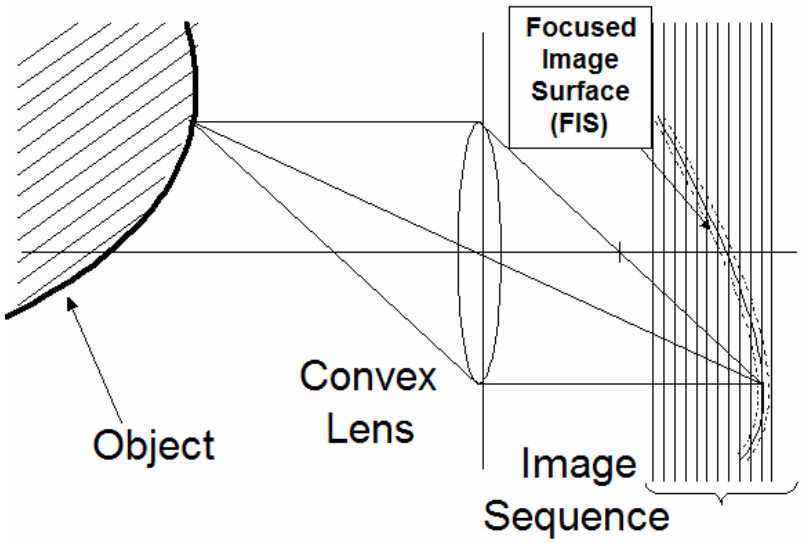


Figure 2.3 Principle of shape from focus method. A sequence of images is obtained by shifting the object in z direction. The focused regions in each image were extracted by a focus judging function. Depth values at these focused regions can be calculated.

2.2.5 Stereo Vision

Stereo vision has long been researched in computer vision. Principally, stereo vision used two cameras to take two pictures of the same scene like our eyes. A computer compares the images while shifting the two images together over top of each other to find the parts that match. The shifted amount is called the disparity and can be used to calculate the depth [Marr 1976]. Given two or more images of the same 3D scene, taken from different points of view, the correspondence problem is to find a set of points in one image, which can be identified as the same points in another image [Steinman 2000]. A human can normally solve this problem quickly and easily, even when the images contain significant amount of noise. In computer vision, the correspondence problem is studied for the case when a computer should solve it automatically with only the images as input. Once the correspondence problem has been solved, resulting in a set of image points, which are in correspondence, other methods, can be applied to this set to reconstruct the position of the corresponding 3D points in the scene as shown in Figure 2.4.

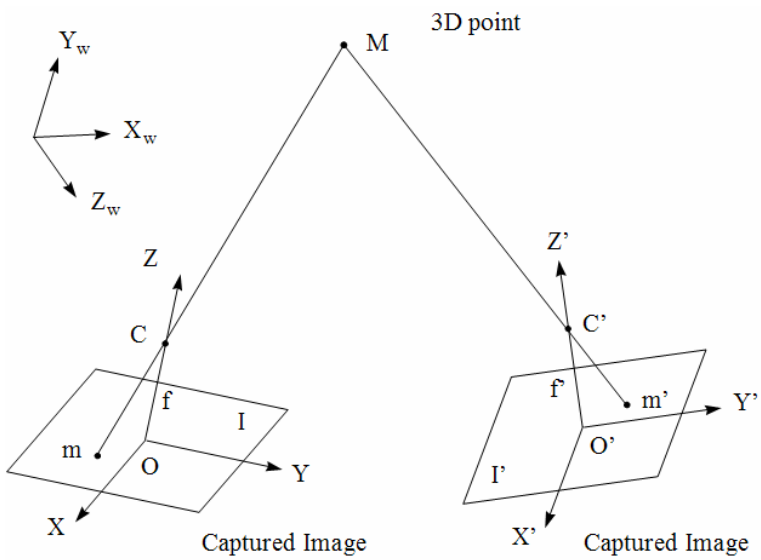


Figure 2.4. Illustration of stereo vision system. Given a 3D point M and its projection *m* in one camera, the correspondence problem is to find the other projection point *m*'.

The correspondence problem typically occurs when two images of the same scene are used, the stereo correspondence problem. This concept can be generalized to the three view correspondence problem or, in general, the *N*-view correspondence problem. In the general case, the images can come either from *N* different cameras that depict the same scene or from the same camera that is moving relative to the scene. An even more difficult version of the correspondence problem occurs when the objects in the scene can be in general motion relative to the camera(s) [Faugeras 1993]. A typical application of the correspondence problem occurs in image mosaicing or stitching [Levin 2004]. Given two or more images, which only have a small overlap, the image mosaicing aims to stitch these separate images into a larger composite image. In this case, it is necessary to be able to identify a set of corresponding points in a pair of images in order to calculate the transformation between them.

2.2.6 Structured Light-based Systems

The correspondence problem can be considerably alleviated by an active method. One of the most widely used active methods is based on structured light projection. Normally, only one camera is used to image the projection of a given pattern on the measuring

surface. 3D information manifests itself in the apparent deformations of the imaged pattern from the projected one. Analyzing these deformations, we can get information about the position, the orientation, and the texture of the surface on which the pattern has been projected. The 3D analysis can also be quite alleviated using coded structured light [Blais 2004]. This technique allows us to know for each imaged point, its original point on the emitted projector plane. Then, the correspondence problem can be directly solved, so it is not necessary to use geometrical constraints. Various codification strategies have been proposed in last decades. A good literature survey can be found in [Pagès 2003a] and [Salvi 2004].

2.3 Structured Light Techniques for 3D Depth Recovery

Structured light technique is based on the projection of a single pattern or a set of patterns onto the measuring surface, which is then viewed, by a single camera or multiple cameras. Coded structured light consists of replacing one of the two cameras by a device that projects a light pattern onto the measuring surface. The most commonly used devices are LCD or DMD video projectors, although previously the most typical were slide projectors. The patterns are specially designed so that codewords are assigned to a set of pixels. Each pixel has its own codeword, so there is a direct mapping from the codeword to the corresponding coordinates of the pixel in the pattern. The codewords are simply numbers, which are mapped in the pattern by using grey levels, color, or geometrical primitives. The larger the number of points that must be coded, the larger the codewords are. And the mapping of such codewords to a pattern becomes more difficult [Pagès 2003a].

This section gives a brief introduction to the structured light techniques, for example, the epipolar constrain, triangulation, and main coding strategies. A good classification and survey can be found in [Salvi 2004]. It summarized most structured light works before 2004. In this work, we will use more paragraphs on the recent works from 2005.

2.3.1 Epipolar Constrain

Epipolar geometry refers to the geometry of stereo vision. When two cameras view a 3D scene from two distinct positions, there are a number of geometric relations between the

3D points and their projections onto the 2D images that, for example, lead to constraints between the image points. These relations are derived based on the assumption that the cameras can be sufficiently well approximated by the pinhole camera model. Figure 2.5 depicts two pinhole cameras looking at point **X**. In real cameras, the image plane is actually behind the focal point, and produces a rotated image. Here, however, the projection problem is simplified by placing a virtual image plane in front of the optical center of each camera. $\mathbf{O}_L$ and $\mathbf{O}_R$ represent the optical centers of the two cameras. **X** represents the point of interest in both cameras. Points $\mathbf{x}_L$ and $\mathbf{x}_R$ are the projections of point **X** projects onto the image planes.

Each camera captures a 2D image of the 3D world. This conversion from 3D to 2D is referred to as a perspective projection and is described by the pinhole camera model. It is common to model this projection operation by rays that emanate from the camera, passing through its focal point. Note that each emanating ray corresponds to a single point in the image.

Since the two optical centers of the cameras are distinct, each optical center is projected onto a distinct point into the other camera's image plane. These two image points are denoted by $\mathbf{E}_L$ and $\mathbf{E}_R$ and are called epipoles. Notice that both epipoles $\mathbf{E}_L$ and $\mathbf{E}_R$ and both focal points $\mathbf{O}_L$ and $\mathbf{O}_R$ lie on a single line. The line $\mathbf{O}_L$-**X** is seen by the left camera as a point because it is directly in line with that camera's focal point. However, the right camera sees this line as a line in its image plane. That line ($\mathbf{E}_R$-$\mathbf{x}_R$) in the right camera is called an epipolar line. Similarly, the line $\mathbf{O}_R$-**X** seen by the right camera as a point is seen as epipolar line $\mathbf{E}_L$-$\mathbf{x}_L$ by the left camera [Faugeras 1993].

As an alternative visualization, consider the points **X**, $\mathbf{O}_L$ and $\mathbf{O}_R$ that form a plane called the epipolar plane. The epipolar plane intersects each camera's image plane where it forms the epipolar lines. The epipolar lines in the right image must intersect the right epipole independent of how **X** is chosen, and correspondingly for the left image. If the projection point $\mathbf{x}_L$ is known, then its projection line $\mathbf{x}_L$-$\mathbf{O}_L$ is known and, consequently, also the epipolar line $\mathbf{E}_R$-$\mathbf{x}_R$ is known, even though $\mathbf{x}_R$ is unknown. However, it must be the case that the projection of **X** in the right image, $\mathbf{x}_R$, lays on this particular epipolar line [Astrom 1999]. This means that for each point in one image the corresponding point in the other image must lie on a known epipolar line. This provides an epipolar constraint,

which corresponding image points must satisfy, and it means that it is possible to test if two points really correspond to the same 3D point. Epipolar constrains can in a convenient way be described by the essential matrix or the fundamental matrix related to the two cameras.

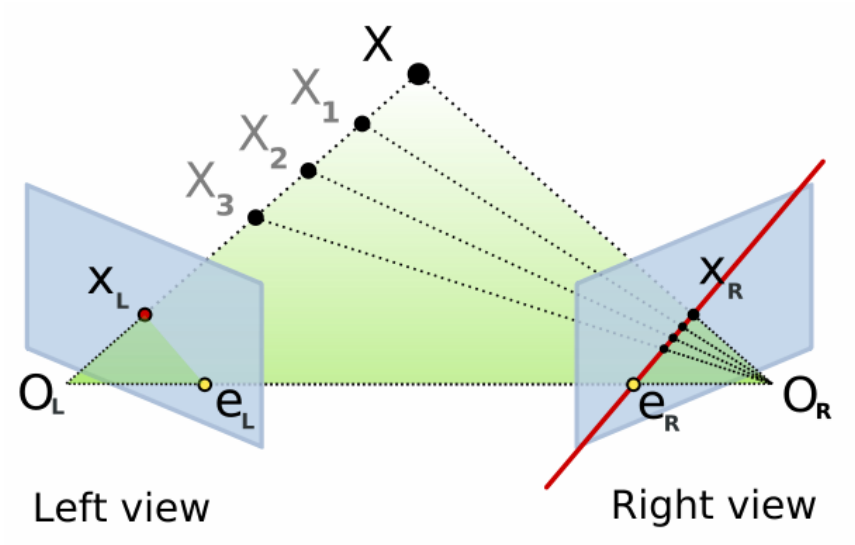


Figure 2.5 Epipolar constrain between two camera image planes.

If the points $\mathbf{x}_L$ and $\mathbf{x}_R$ are known, their projection lines are also known. If the two image points correspond to the same 3D point **X,** the projection lines must intersect precisely at **X**. This means that **X** can be calculated from the coordinates of the two image points $\mathbf{x}_L$ and $\mathbf{x}_R$. This method is also named triangulation.

2.3.2 Triangulation

In computer vision, triangulation refers to the process of determining a point in 3D space given its projections onto two, or more, images. In order to solve this problem it is necessary to know the parameters of the camera projection function from 3D to 2D for the cameras involved, in the simplest case represented by the camera matrices.

The triangulation problem is in theory trivial. Each point in an image corresponds to a line in 3D space such that all points on the line are projected to that first point in the image [Hartley 1997]. If a pair of corresponding points in two, or more images, can be found it must be the case that they are the projection of a common 3D point **P**. The set of

lines generated by the image points must intersect at **P** and the algebraic formulation of the coordinates of **P** can be computed in a variety of ways.

In practice, however, the coordinates of image points cannot be measured with arbitrary accuracy. Instead, various types of noise, such as geometric noise from lens distortion or feature detection error, lead to inaccuracies in the measured image coordinates. Therefore, the lines generated by the corresponding image points do not always intersect in 3D space [Hartley 2004]. The problem, then, is to find a 3D point, which optimally fits the measured image points. In the literature, there are multiple proposals for how to measure optimality and how to find the optimal 3D point. Since they are based on different optimality criteria, the various methods produce different estimates of the 3D point **x** when noise is involved.

As shown in Figure 2.6, the measured image points are y_1'and y_2' instead of y_1 and y_2. However, their projection lines do not have to intersect in 3D space or come close to **X**. In fact, these lines intersect if and only if y_1'and y_2' satisfy the epipolar constraint defined by the fundamental matrix. Given the measurement noise in y_1'and y_2' it is rather likely that the epipolar constraint is not satisfied and the projection lines do not intersect. This observation leads to the problem, which is solved in triangulation. Which 3D point **X'** is the best estimate of **X** given y_1'and y_2' and the geometry of the cameras? The answer is often found by defining an error measure which depends on **X'** and then minimize this error.

All triangulation methods produce **X'=X** in the case that $y_1 = y_1$'and $y_2 = y_2$', that is, when the epipolar constraint is satisfied (except for singular points, see below). It is what happens when the constraint is not satisfied which differs between the methods. A triangulation method can be described in terms of a function *f* such that:

$$\mathbf{X} \cong f(y_1^{'}, y_2^{'}, \mathbf{A}_C, \mathbf{A}_P) \tag{2.1}$$

where y_1'and y_2' are the homogeneous coordinates of the detected image points and $\mathbf{A}_C$ and $\mathbf{A}_P$ are the camera and projector matrices. **X** is the homogeneous representation of the resulting 3D point. The $\cong$ sign implies that *f* is only required to produce a vector, which is up to a non-zero scale since homogeneous vectors are involved.

Some of the methods fail to correctly compute an estimate of **X** if it lies in a certain subset of the 3D space, corresponding to some combination of y_1', y_2', $\mathbf{A}_C$ and $\mathbf{A}_P$. A point in this subset is then a *singularity* of the triangulation method. The reason for the failure can be that some equation system to be solved is under-determined or that the projective representation of **X'** becomes the zero vector for the singular points.

In the mid-point method [Beardsley 1994], each of the two image points y_1' and y_2' has a corresponding projection line, here denoted as $\mathbf{L}_1$' and $\mathbf{L}_2$', which can be determined given the camera matrices $\mathbf{A}_C$ and $\mathbf{A}_P$. Let d be a distance function between a 3D line and a 3D point such that $d(\mathbf{L},\mathbf{X})$ represents the Euclidean distance between **L** and **X**. The mid-point method finds the point **X'** which minimizes following function:

$$d(\mathbf{L}_1^{'},\mathbf{X})^2 + d(\mathbf{L}_2^{'},\mathbf{X})^2 \tag{2.2}$$

It turns out that **X'** lies exactly at the middle of the shortest line segment which joins the two projection lines. More details and discussions about 3D triangulation methods can be found in [Hartley 2004].

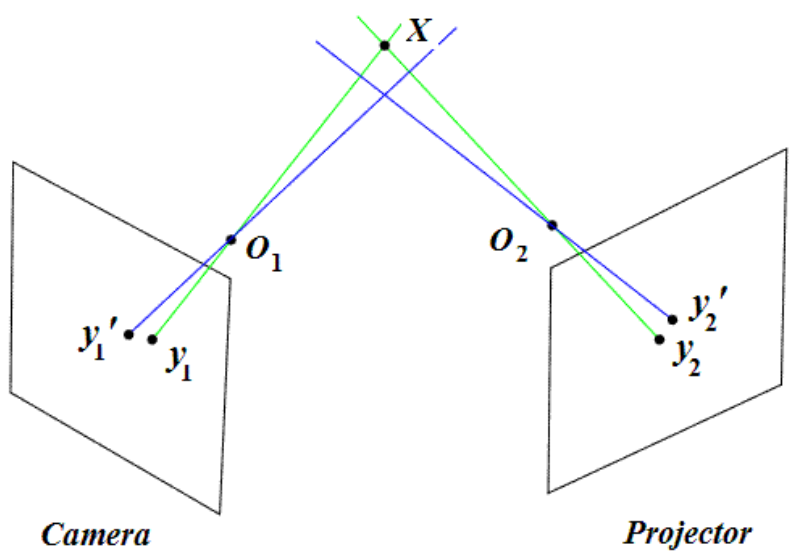


Figure 2.6 In real image, the image points y_1 and y_2 cannot be measured with arbitrary accuracy. Instead points y'_1 and y'_2 are detected and used for the triangulation. The corresponding projection lines do not, in general, intersect at point X.

2.3.3 Pattern Codification Strategies

Pattern projection techniques differ in the way in which every point in the pattern is identified, i.e. what kind of codeword is used, and whether it encodes a single axis or two spatial axes [Salvi 2004]. In fact, it is only necessary to encode a single axis, since a 3D point can be obtained by intersecting two lines (i.e. when both pattern axis are coded) or intersecting one line (the one which contains a pixel of the camera image) with a plane (i.e. when a single pattern axis is coded).

In this work, we follow the classification proposed by Salvi in [Salvi 2004]. The coding methods in structured light were classified into two classes: temporal and spatial codification. The temporal coding schemes generate the codewords by projecting a sequence of patterns along time, so the pattern structure can be very simple. Since multiple images are used, dense surface points can be reconstructed by this means. Spatial codification represents the codeword in a unique pattern. And its structure is often complex. Single or fewer pattern images are usually used in spatial codification. That makes its output 3D model with lower data density. In addition, there is also another hybrid class that combined the features of temporal and spatial coding strategies like [Ishii 2007] etc.

- *Temporal Codification*

Patterns based on Gray code, as well as binary and *n*-ary codes, have the advantage that the pixel codification is made punctually and no spatial neighborhood has to be considered in the codification. However, the discrete nature of such patterns limits the range resolution. Besides, phase shifting methods exploit higher spatial resolution since they project a periodic intensity pattern several times by shifting it in every projection. The drawback of these methods is the periodic nature of the patterns, which introduces ambiguity in the determination of the signal periods in the camera images. The integration of Gray code with phase shift method brings together the advantages of both strategies, i.e. the unambiguity and robust codification of pattern stripes of Gray code, plus the high resolution of phase shifting method. The combination of both techniques leads to high-accuracy 3D determination. However, the number of projecting patterns increases accordingly.

In [Bergmann, 1995], a Gray code pattern is projected in order to label the measuring surface regions where every period of a sinusoidal intensity pattern will be projected. Thus, the ambiguity problem between signal periods is solved. The sinusoidal patterns are represented with grey levels. A total number of four Gray code patterns are projected in order to label 16 different regions on the measuring surface as shown in Figure 2.7.

Then, the periodic intensity pattern is projected four times by shifting ¼ of the period, each time. For every given pixel in the camera image, the phase of the first periodic pattern projected to the corresponding surface point must be found. For this purpose, a classic four-step phase shift is applied as following:

$$\phi(x, y) = \arctan\left(\frac{I_2 - I_4}{I_3 - I_1}\right) \tag{2.3}$$

I_1, I_2, I_3 and I_4 are the grey levels of pixel (x, y) of the camera images corresponding to every one of the 4 projected shifted patterns. Once the phase of the given pixel is known, the period of the sinus where the pixel lies is obtained with the Gray code labeling. Therefore, the pattern stripe projected to a certain surface point can be precisely calculated.

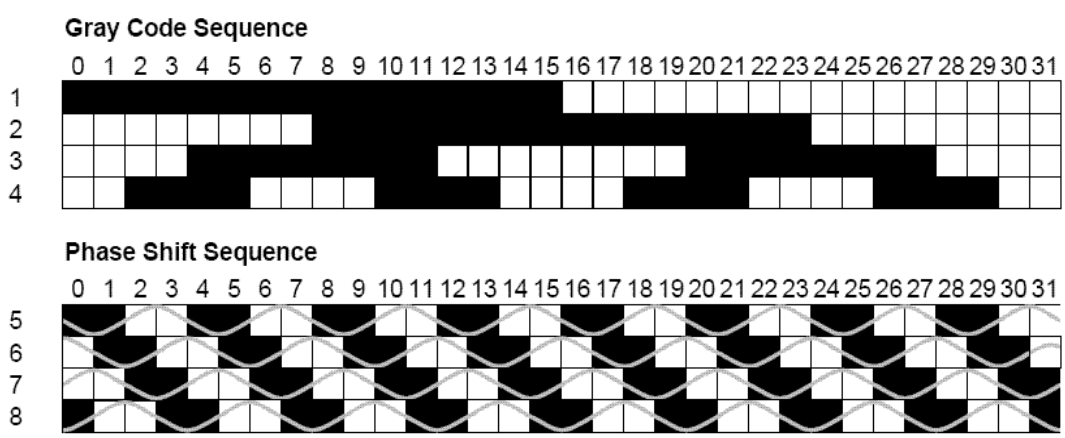


Figure 2.7 Gray code pattern (top) and sinusoidal phase shifting (bottom) pattern.

Gühring proposed substituting the phase shifting method for a new method called line shifting [Gühring 2000]. It was also combined with gray code. The author pointed out that Phase Shifting has a series of drawbacks. For example, when reconstructing surfaces

with non-uniform albedo the phase cannot be recovered precisely. Moreover, camera devices tend to integrate over a certain area so that pixel values are affected by its neighbors. To avoid this problem, camera resolution must be sufficiently higher than projector resolution. In order to avoid the problems of phase shifting, the author proposed substituting the sinusoidal periodic profile of such methods by a multi stripe pattern as shown in Figure 2.8. Gühring designed a 640×640 pattern for LCD projectors, where every sixth column is white and the remaining ones are black. By consecutively shifting the pattern 6 times, the whole resolution for every row of the pattern is covered. While repeating the process with row-encoded patterns, the entire resolution of the pattern is used. Since it is a periodic pattern, the projection of Gray code patterns is also required in order to solve the ambiguities that arise.

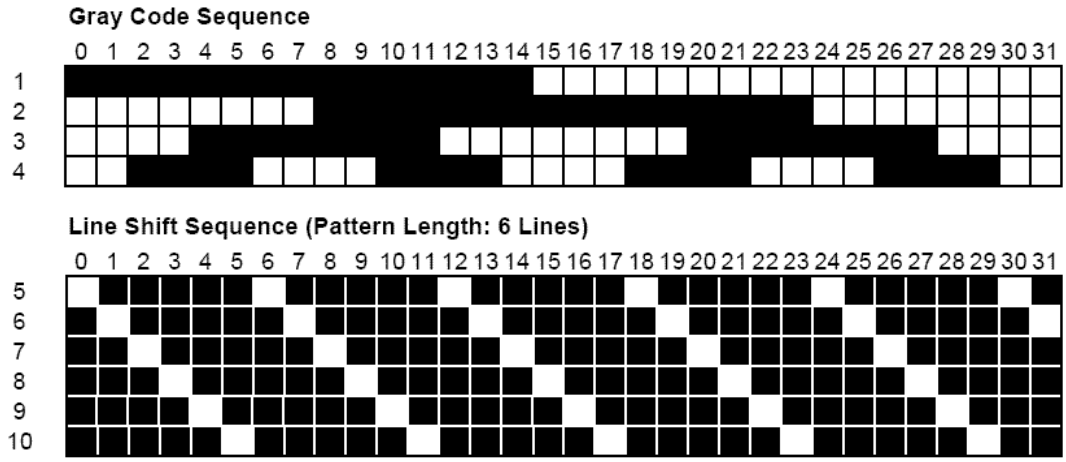


Figure 2.8 Gray code (top) and line shift (bottom) pattern used in [Gühring 2000].

In the method, five vertical Gray codes patterns are used to encode each sub-region; 6 vertical multi stripe patterns are used to encode the image points inside every sub-region. With regard to the Gray coded patterns, a total number of regions should be labeled, being the number of lines projected in every line shifting pattern. Thus, every pattern area where a line is shifted has its own label. For example, if patterns of 32 columns were projected, six lines would be shifted and, therefore, six bands of 6 pixels width should be labeled by the Gray code. Thus, three patterns of Gray code should be projected. However, in the transitions of each region, a decoding error implies a large measuring error of around one period. To avoid this problem, Gühring proposed to introduce an

over-sampling technique consisting of projecting an additional Gray coded pattern. In this way, thinner bands of pixels are labeled, and more robust decodification of the regions is obtained. The maximum error when calculating the global position of an illuminated line due to transition of a Gray codeword was 2 pixels.

As for the patterns containing line shifting, the peak position of every illuminated line was intended to be located in the camera images with sub-pixel accuracy. Since the intensity profile of such imaged lines presented a Gaussian distribution, the peak detector proposed by [Blais 1986] was used. The line shifting method had a similar or even better resolution than techniques based on phase shifting and more accurate measurements. The author developed a system set up based on both LCD and DMD projectors, obtaining similar results with an average error of 0.03 mm and a maximum deviation of 0.28 mm for both devices.

In [Koninckx 2006], a scene-adapted structured light system was proposed in order to overcome existing limitations of structured light 3D acquisition methods such as the colors, intensities, and shapes. Based on a crude estimate of the scene geometry and reflectance characteristics, the local intensity ranges in the projected patterns are adapted, in order to avoid over- and underexposure in the image. The estimated geometry also helps to limit the effect of aliasing caused by the sampling of foreshortened patterns. Furthermore, the approach also accounts for the adverse effects that small motions during scanning would normally have. Moreover, the approach yields a confidence measure at every pixel of the range image.

Ranges are acquired in a two-step procedure. First, the scene is analyzed. This comprises a crude estimation of scene geometry and measures to take surface reflectance into account. Then, the projection patterns are adapted to avoid under- or over-exposure, as well as aliasing problems. This leads to an iterative process, where the deviation between expected and observed patterns allows the system to refine the geometry and provides it with confidence measures. The process stops when neither the geometry nor the confidences are updated. The estimated geometry also facilitates the detection and compensation of motion between consecutive frames. On the photometric side, nonlinearities and crosstalk between the color channels are modeled for both the camera and projector. The outcome still is a very low cost system, based on one normal camera

and one LCD projector that now can deal with a wider category of objects (e.g. metal industrial work pieces) and that require fewer interactive settings. The experimental results with specular object proved its feasibility.

In [Ishii 2007], the author proposed a spatial-temporal coded structured light projection method that attempts to combine the advantages of both spatial and temporal patterns, high accuracy in the case of a static object and robustness in the case of a moving object. The projection of multiple coded structured light patterns that enable both encoding along the time axis and encoding in the spatial neighborhood. The adaptive selection of encoding types in every local image region by calculating the image features that dependent to the object motion. As a result, codewords along the time axis are determined by the regions where there are no motions, while codewords in the spatial neighborhood are determined by the region where the brightness changes dynamically. As an example of the coded patterns, Figure 2.9 shows the patterns when the number of bits of the space codes is set to $n = 8$. In the figure, four coded patterns at time $t = 0, 1, 2,$ and 3 are shown, and the numbers $k = 0\ldots 7$ at the bottom in each image indicate that the 1D pattern in the y direction is set to $G(k, y)$. It can be seen that the space code can be generated only by using spatial neighborhood information, because eight gray code patterns – $G(0, y), \cdots, G(7, y)$ – are periodically arranged in a single image. The projected pattern also shifts to the left after a certain period of time, and the gray code pattern on the same pixel changes $G(0, y)$, $G(1, y)$, $G(2, y)$, $\cdots$ every unit time. In this manner, a space code at each pixel along the time axis can be obtained.

The distance between the camera and projector is approximately 180 mm, and a background screen is placed at a distance of approximately 1100 mm from them. In this case, light patterns from the projector are projected over an area of 37×37 cm on the screen. The frame rate of camera and projector can reach 1000 fps. And the system can reconstruct a moving hand at a frame rate of 30 fps.

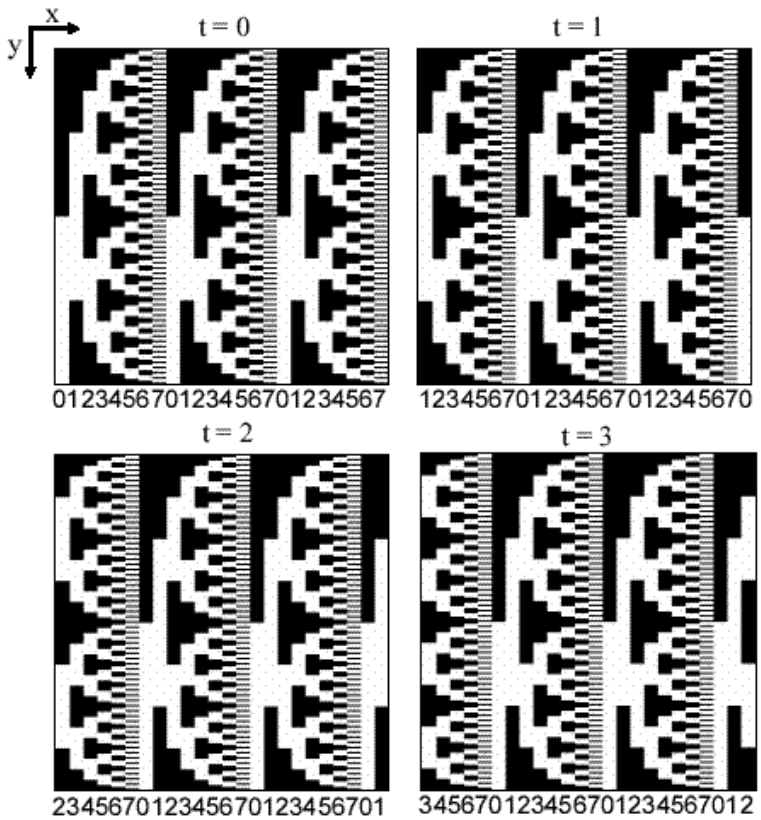


Figure 2.9 Coded structured light patterns for spatial-temporal selection.

- *Spatial Codification*

Spatial coding schemes aim to encode each pattern feature with a unique codeword in single pattern image. Perfect maps have been adopted in spatial codification for its unique window property. Mathematically, assume M be a matrix of dimensions $r \times v$ where each element is taken from an alphabet of k symbols. If M has the window property, i.e. each different sub-matrix of dimensions $n \times m$ appears only once, then M is a perfect map. If M contains all sub-matrices of $n \times m$ except the one filled by 0's, then M is called an *M-array* or *pseudorandom array* [MacWilliams 1976], [Etzion 1988]. This kind of array has been widely used in pattern codification because the window property allows every different sub-matrix to be associated with an absolute position in the array.

A binary pseudorandom array of 24×24 was proposed in [Morita 1988]. This array has the window property of 3×4. The M-array representation is made by painting black dots

on a white background, for the array elements corresponding to symbol. Two patterns are projected on the measuring surface: the first one contains all the possible black dots in order to locate their centers in the camera image. The second pattern is the M-array representation. Therefore, the method is restricted to static scenes.

A systematic process for constructing a maximum size array of $n \times m$ based on an alphabet of b symbols was proposed in [Griffin 1992]. The construction was based on certain restrictions: every element of the array has a unique codeword formed by its own value and the values corresponding to its four neighbors. Such an array is a special case of perfect maps, since it has window property of 3×3, but not all the possible windows appear.

In [Kiyasu 1995], a pseudorandom pattern is proposed to reconstruct the shape of specular polyhedrons. Normally, most coded structured light systems are not intended to reconstruct specular surfaces but Lambertian ones. A binary M-array represented with a grid of 18×18 circular spots with a window property of 4×2 was used. The planarity check shows a mean residual about 1 mm at a working distance of 500 mm.

A pattern consisting of a grid of thin vertical and horizontal slits was proposed in [Salvi 1998]. The author argued that grid coding is a better solution because the grid-points can be easily segmented. Furthermore, the neighborhood around a grid-point can be found by just tracking the edges of the grid. The pattern was designed as a 29× 29 grid using three different colors for horizontal slits and three more colors for vertical slits as shown in Figure 2.10. The color assignment was made by using the same *De Bruijn* sequence of order 3 (size of window property) for both rows and columns of the grid. The grid intersection points are reconstructed after decoding their position in the pattern using the window property.

In [Morano 1998], the author proposed an algorithm for constructing an M-array, fixing the length of the alphabet, the window property size, the dimensions of the array and the Hamming distance between every window. The author declared that, previous methods worked with a Hamming distance of one, which did not allow error correction. In fact, the arrays used by Morano are simply perfect sub-maps since not all the possible windows are included. The algorithm used to generate an array with fixed properties is based on a brute force approach. For example, when constructing an M-array based on

three colors with window property of 3×3 the following steps are taken: first, a sub-array of 3×3 is chosen randomly and is placed in the north-west vertex of the M-array that is being built. Then consecutive random columns of 1×3 are added to the right of this initial sub-array, maintaining the integrity of the window property of the array and the Hamming distance between windows. Afterwards, rows of 3×1 are added beneath the initial sub-array in a similar way. Then, both horizontal and vertical processes are repeated by incrementing the starting coordinates by one, until the whole array is filled. Whenever the process reaches a state where no possible elements can be placed, while accomplishing the global window property, the array is cleared and the algorithm starts again with another initial sub-array. The study of the performance of this algorithm showed that using M-array of 45×45, the window property is 3×3 if 3 colors are used as shown in Figure 2.10. The fixed Hamming distances between windows from the typical 1 up to 4 can be generated.

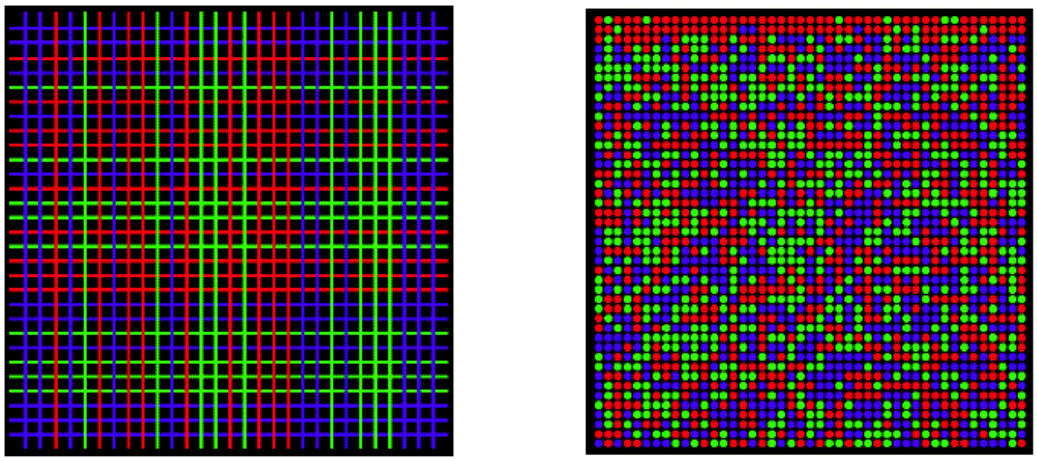


Figure 2.10 (Left) A grid pattern consisting of vertical and horizontal colored slits [Salvi 1998]. (Right) M-array color pattern with window property of 3×3 proposed in [Morano 1998].

In [Spoelder 2000], a binary pseudorandom array with size of 65×63 is used to reconstruct the shape of the cornea. Cyan and yellow were used to encode the binary values of the pattern. The pattern was designed as a checkerboard, where the white squares are used to represent the pattern element. The black squares are used to ease the pattern segmentation in the camera image. The cross-points of the checkerboard are detected by simple template operator. Due to the complex reflectance characteristics of

the cornea, many feature points cannot be extracted correctly. No accuracy evaluation is provided.

In [Fong 2005], a structured light pattern that combines features of color stripe and sinusoidal is proposed for deforming object reconstruction. In the pattern, the color stripes are used to get sparse but absolute depth values and the sinusoid is used to propagate these values and get a high-resolution range map. To guide the propagation, the image is firstly segmented into regions that do not contain any depth discontinuities. Compared to patterns used in previous work, which employ either color or luminance alone, this pattern is more robust because the color and sinusoidal components provide complementary information. The color stripe boundaries give sparse, but absolute depths, while the sinusoid gives dense relative depths. These points are used as known points in the sinusoid processing. The sinusoidal component is resistant to reflectance variations in the scene and only a few correctly sensed color transitions (also known as color edges) are needed as start points. A sinusoidal pattern also avoids the need to do any sub-pixel location estimation of pattern features (e.g. transitions). The depths are computed at pixel locations. A pattern of vertical colored stripes with a horizontal sinusoid overlaid (as shown in Figure 2.11) is projected by a standard LCD projector and the images are captured from a different viewpoint with a color camera.

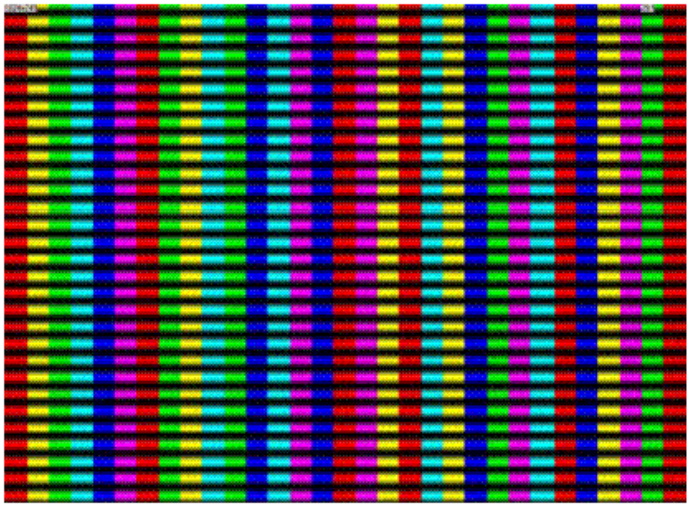


Figure 2.11 A pattern of vertical colored stripes with a horizontal sinusoid overlaid.

The sinusoid of intensity is embedded in the value channel and the colored stripes are embedded in the hue channel in HSV color space. Each frame can be processed independently to recover depth information. Pixels are given color labels and scores by a Bayesian classifier. The stripe transitions are identified, labeled, and scored according to the color labels of pixels adjacent to each boundary. The scene in segmented into regions using a snake technique [Kass 1988] and the luminance component in each region is processed to recover the dense depth map using the color transition depths as starting points. Processing each frame independently eliminates the need for temporal coherence. The scene can change significantly between frames without affecting the depth recovery. Using more colors leads to a greater number of possible transitions, but the colors are more difficult to distinguish. Since the sinusoid is used to get the dense range map, having a large number of transitions is not critical. In the experiment, six colors were used allowing 30 unique transitions. The depth at each pixel is computed from the apparent phase shift between the projected and observed sinusoids. The main assumption made in phase unwrapping is that between two adjacent pixels along the unwrapping path the phase changes by less than 2π. When the pattern is projected onto a speaker about 650 mm far away from the projector, the ridges and noise are less than ±1 mm from the surface. The root mean squared error is about 0.7 mm.

In the method proposed by Adan [Adan 2005], a 2D color pattern was designed for dynamic feature tracking. In order to design the features on M, seven colors (red, green, blue, magenta, cyan, yellow and white) associating one color to each point of M are used. The color that has been associated to the point c is denoted by C. Consequently a codeword corresponding to point c consists on the pair [C, $V(C)$] where the topology and properties for points remain for colors. The color has been inserted in M by applying an iterative algorithm that starts with a random assignment and that converges when no repeated codes are found. Therefore, for the feature point, [*b*, *bymymg*] put in the Figure 2.12 is equivalent to [*b*, *gbyymm*]. Each codeword has six neighboring codewords and six distances can be computed accordingly. Since 95% of the codes have a Hamming distance greater than one, most of misinterpreted and lost codewords can be corrected and the number of unlabeled spots decreases. Second, the code recovering and the correspondence process are simplified because of the codeword disorder property. For example, tasks such as computing the angle of $V(C)$ with respect to C do not need to be

carried out thus reducing the computational cost. This novel structured light sensor based on a disordered codeword strategy which solves easily and robustly the correspondence problem. This component yields medium spatial resolution of around 5 mm on a surface 1 m far away. The system has been tested in a real environment for 2 DOF soft movements. Tracking validation has been carried out introducing a 6 DOF manipulator robot inside the experimental setup which is able to measure the feature point location error.

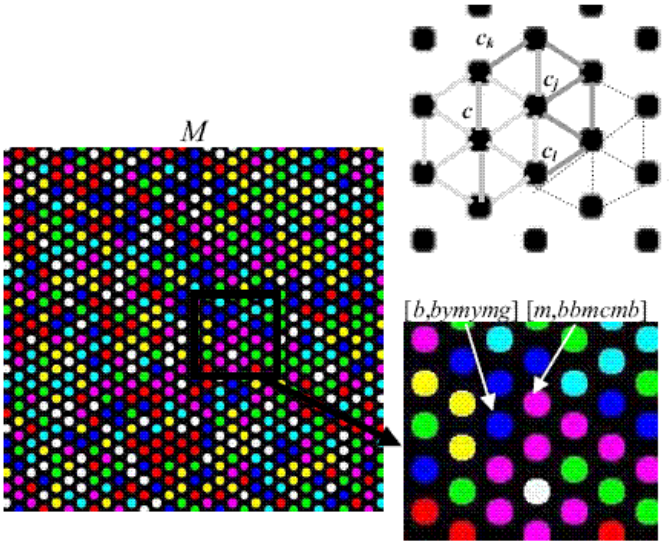


Figure 2.12 A 2D color pattern that designed for dynamic feature tracking task [Adan 2005].

In [Albitar 2007], a new monochromatic pattern for a robust structured light coding based on the spatial neighborhood scheme was proposed. The work aims to overcome classical problems such as shadows and occlusions. In such situations, a key factor is to elaborate a pattern with a significant Hamming distance between the codewords. Actually, the Hamming distance quantifies the differences between the codewords. If this distance is higher than 1, it quantifies the capability to correct the errors that can occur in the segmentation and decoding stages. It was shown in [Morano 1998] that better results can be achieved when the Hamming distance is higher than 3. However, most of referenced methods used such a distance but with a significant number of symbols were used. In order to simplify the decoding stage, it would be more efficient to handle fewer symbols. To represent the symbols, the author proposed to use geometrical primitives and generate

a monochromatic pattern. In practice, monochromatic patterns are more robust than colored ones. Firstly, it is much easier to deal with an image having only two intensity levels. Secondly, if the scene is very colored, the choice of colors to be used in the pattern can be very tricky. Moreover, the realization of compact projection devices with a monochromatic light is much simpler thanks to the technology innovations.

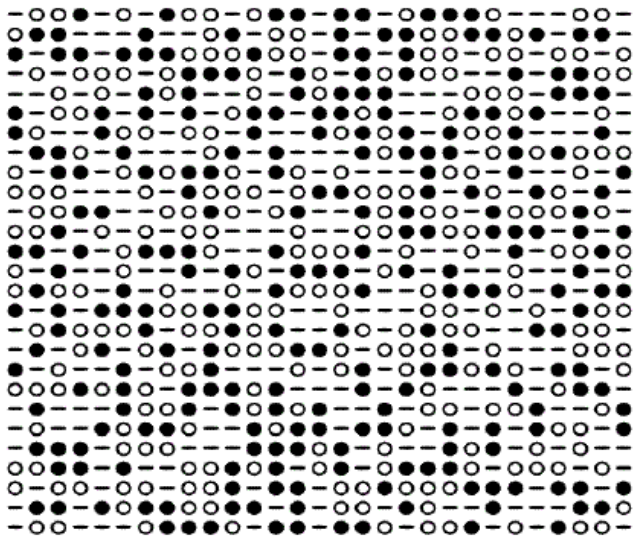


Figure 2.13 A 2D structured light pattern based M-array. The pattern element is defined to a small set of symbols associated to simple geometrical primitives [Albitar 2007].

Since the codeword of a primitive depends on its neighborhood, it is necessary to detect this neighborhood easily and without imposing any constraints on the acquisition device. To solve this problem, the primitives can be linked as it was proposed in [Yee 1994]. Unfortunately, this can cause new difficulties in the detection and segmentation stages. So, it would be more convenient to consider an alphabet with minimum number of symbols in building an M-array while being able to detect the neighborhood correctly. The proposed pattern is robust as it allows a high error rate characterized by an average Hamming distance higher than 6. The pattern element is defined to a small set of symbols associated to simple geometrical primitives as shown in Figure 2.13. One of these primitives embeds the local orientation of the pattern. This is helpful while performing the search for the relevant neighborhood during the decoding process. The aim of this work is to use this pattern for the real-time 3D reconstruction of dynamic scenes, particularly in endoscopic surgery, with fast and reliable detection and decoding stages.

Ongoing results are presented to assess both the capabilities of the proposed pattern and the reliable decoding algorithm with projections onto simple 3D scenes and onto internal structures of a pig abdomen. No metric measurement results are provided.

In [Desjardins 2007], the array construction starts by adding one 3×3 code in the upper left corner of an empty $n \times m$ array. Every element of this 3×3 code is chosen randomly. The next step is to fill out the first three rows and then the first three columns of the array. Every time three new color elements are added, a new code is created and compared with all previously introduced codes to assure its uniqueness. Different possible permutations of color elements are tested until the new 3×3 code created is unique or until all possible permutations have been exhausted. This iterative process continues until a full array of 53×38 is created as shown in Figure 2.14. Thresholding method was used to segment the pattern elements. A labeling algorithm locates the different regions on the binary images and the centroids are extracted and encoded consequently. To improve the feature density, a shifting process is proposed. The principle consists of collecting a sequence of pairs of images while the same pseudorandom color pattern is progressively shifted over the scene by realignment of the color grid over the projector LCD array. The accuracy is estimated to be within 2 mm for an object located about 1.5 m from the sensor.

In the method proposed by Chen and Li [Chen 2008], a grid patterns that can meet the practical requirements of uniquely indexing for solving uncertain occlusions and discontinuities in the scene is introduce. With a given set of color primitives, the constructed pattern is guaranteed to be a large matrix and desired shape with the restriction that each word in the pattern matrix must be unique. The centroid of the rectangular pattern element is defined as the feature point. And each feature point is encoded by the colors of its 4-adjacent (i.e. top-left-bottom-right) neighbors. A flood search algorithm is introduced to generate the grid-point net. In the experiment, a set with 7 colors is used, which can generate a matrix for a 38×212 rectangle or an 82×82 square as shown in Figure 2.14. The relative observed error is about 2% due to the system is not well calibrated. The author put the emphasis on the efficiency analysis. As reported in the paper, system can reach 30 fps with further optimization.

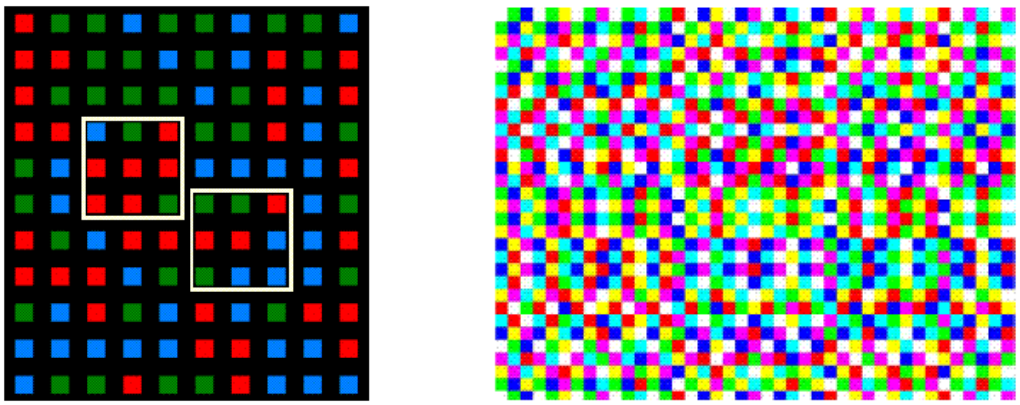


Figure 2.14 (Left) A color pattern designed (53×38) by [Desjardins 2007], it contains 3 colors and its window size is 3×3. (Right) A 38×212 tessellated pattern designed by [Chen 2008], 7 colors are used, and its codeword is formed by its 4-adjacents (i.e. top-left-bottom-right).

2.4 Surface Orientation Recovery

Approaches for visual recovery of surface orientation can be categorized into two classes: image-based and structured light projection-based. The first makes use of the natural features on the target surface and assumes certain ideal property (constant surface albedo, uniform textures etc.) of them. It is usually named shape from *X* techniques, where *X* can be shading, texture, polarization etc. The second class utilizes artificial textures or patterns that are projected onto the target surface.

Shape from shading is one representative method of the shape from X technique. The gradual variation of the image intensity is utilized to recover the surface shape. The Lambertian reflectance model is usually assumed, in which the grey level at any position in the image is determined by the surface orientation and the incident angle of light there [Lange 1999]. To a grey level image, the method aims to recover the light source direction and the surface orientation in 3D at each image position. The shape from shading problem can be formulated as that of finding the solution of a nonlinear first order partial differential equation called the brightness equation. It has been shown to be an ill-posed problem that generally has infinitely many solutions [Woodham 1978]. Additional heuristics on surface smoothness or intensity gradient are usually adopted to let surface orientations be determined in more precise form.

2.4.1 Shape from Shading

Shape from shading is one representative method of the shape from X technique. To a grey level image the method aims to recover the light source direction and the surface orientation in 3D at each image position. The gradual variation of the image intensity $p = \partial z/\partial x$, $q = \partial z/\partial y$ that represented by partial deviation of surface $z = f(x, y)$ in the x and y directions. Given an image, E, and a reflectance map, R, the shape from shading problem can formulated as a brightness equation as

$$E(x, y) = R(p, q) \tag{2.4}$$

Lambertian model is usually assumed to the surface in most shape from shading methods. In this model, the grey level at a pixel in the image depends on the light direction and the surface normal. Therefore, the Lambertian surface can be modeled as:

$$I_L = A\rho\cos\theta \tag{2.5}$$

where A is the light intensity, ρ is the surface albedo and θ is the incident angle.

Find a unique solution to shape from shading problem is difficult, since the equation has two unknowns. Additional constraints such as brightness constraint, smoothness constraint, and intensity gradient constraint are usually assumed [Zhang 1999a]. Based these assumptions, several methods have been proposed such as [Brooks 1985], [Pentland 1988], [Zheng 1991], and [Bichsel 1992] etc.

Ikeuchi and Horn first applied the calculus of variations to the shape-from-shading problem [Ikeuchi 1981]. They effectively solved a functional minimization problem in recovering object surface orientation. It is known that the occluding boundary provides important constraints on the solution of the shape-from-shading problem. The difficulty with using the gradient to specify surface orientation is that, as already mentioned, at least one of p and q is unbounded on the occluding boundary. This problem can be overcome by specifying surface orientation in another way. Consider the mapping from p-q space to f-g space specified by the equations:

$$f = \frac{2p}{1 + \sqrt{1 + p^2 + q^2}}$$

$$g = \frac{2q}{1+\sqrt{1+p^2+q^2}} \tag{2.6}$$

It is easy to verify that $f^2 + g^2 < 4$ for all visible parts of a surface. The orientation of a point on the occluding boundary corresponds to a point on a circle of radius 2 in f-g space. Thus, occluding boundaries present no difficulties now. This we may regard as a search for two functions, f and g, defined over Ω that corresponds to a smooth surface satisfying the image irradiance equation. The integral of the brightness error with following form is proposed to solve f and g.

$$\iint_{\Omega} (E(x,y) - R(f(x,y), g(x,y))^2 \tag{2.7}$$

The minimization of the above function was done through various calculi. In various calculus, the Euler equations, which are differential equations, are first computed, and then solved discretely to minimize the energy function. In [Zheng 1991], the Euler equation is simplified by taking the first order Taylor series of the reflectance map and representing the depth, gradient and their derivatives in discrete form.

A linear approximation of the reflectance map in p-q is adopted in [Pentland 1988]. The Linear approaches reduce the non-linear problem into a linear through the linearization of the reflectance map. The idea is based on the assumption that the lower order components in the reflectance map dominate. Therefore, these algorithms only work well under this assumption.

2.4.2 Photometric Stereo

Photometric stereo provides a method to estimate local surface orientation by using several images of the same surface taken from the same viewpoint but under illumination from different directions as shown in Figure 2.15. It was first proposed in [Woodham 1980]. The light sources are ideally point sources some distance away in different directions, so that in each case there is a well-defined light source direction from which to measure surface orientation. Therefore, the change of the intensities in the images depends on both local surface orientation and illumination direction. In the shape from shading problem, we want to infer two surface gradients, p and q, from intensity. In other words, we can only identify a curve of possible solutions. If we rotate the light source,

the reflectance function also rotates. A particular facet's position on the reflectance map does not change. Therefore, it should lie at the intersection of the two solution curves. For a surface that obeys Lambertian model, three images are required to resolve any possible ambiguities as shown in Figure 2.16.

Unlike single image-based shape from shading algorithms, photometric stereo makes no assumption of the smoothness of the surface (surface partial derivatives p and q vary continuously over the surface). Photometric stereo method requires additional lighting. However, it can be easily implemented at a reasonable computational cost. Each image brings along its own unique reflectance map, therefore each image will define a unique set of possible orientations for each point. Photometric stereo can recover not only surface orientation (p, q), but also surface albedo.

Consider three light sources with illumination vectors $\mathbf{L}_1$, $\mathbf{L}_{21}$ and $\mathbf{L}_3$, the equation can be rewritten in matrix form:

$$\mathbf{I} = \rho \mathbf{L} \mathbf{N} \tag{2.8}$$

Provided $\mathbf{L}_1$, $\mathbf{L}_{21}$ and $\mathbf{L}_3$ are non-coplanar, then the photometric illumination matrix $\mathbf{L}$ is non-singular and its inverse matrix. That means $\mathbf{L}^{-1}$ exists.

Define matrix $\mathbf{M}$ as:

$$\mathbf{M} = \begin{bmatrix} m_1 & m_2 & m_3 \end{bmatrix} = \mathbf{L}^{-1}\mathbf{I} = \rho \mathbf{N} \tag{2.9}$$

Then, the surface gradient components can be calculated as:

$$\begin{aligned} p &= -m_1 / m_3 \\ q &= -m_2 / m_3 \\ \rho &= \sqrt{m_1^2 + m_2^2 + m_3^2} \end{aligned} \tag{2.10}$$

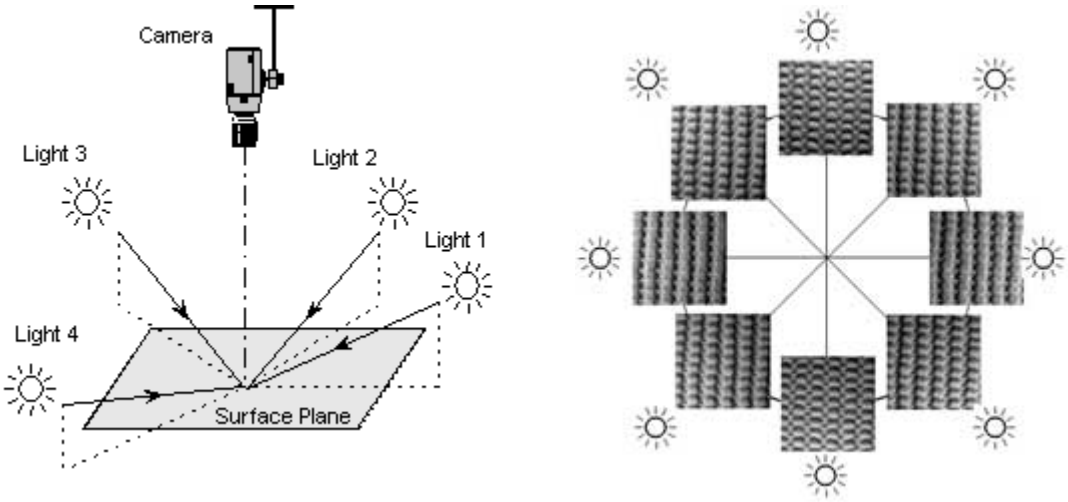


Figure 2.15 Photometric stereo method. (Left) Surface normal is estimated from multiple illuminations. (Right) Image captured from different illumination directions.

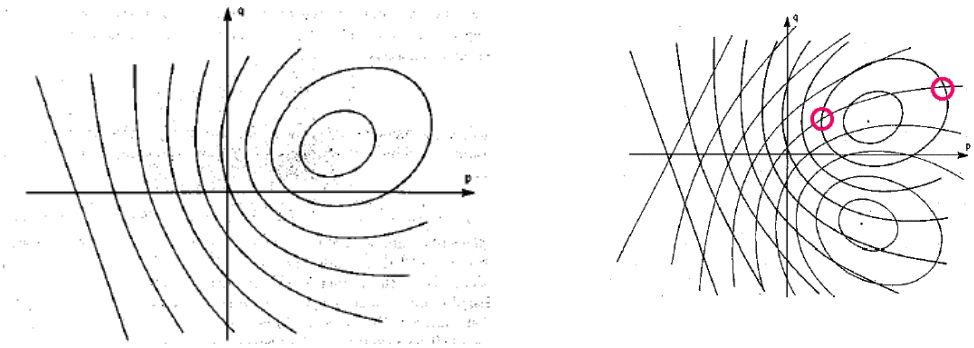


Figure 2.16 Reflectance map in *p-q* space. (Left) Single image provide single *p-q* curve in the shape from shading method. (Right) Multiple images captured from different illumination directions are used to solve the unique solution in photometric method.

2.4.3 Shape from Polarization

Polarization has proven to be a useful source of information in the analysis of light scattering from surfaces in computer vision. There are a number of ways in which polarization arises and can be used in surface analysis. One familiar example is when the incident light is polarized and the polarization of the scattered light is analyzed [Ragheb 2003]. Most research aimed at extracting and interpreting information from polarization

data involves placing a linear polarizer in front of a camera and taking images of an object with the polarizer oriented at different angles [Wolff 1991].

Existing work has demonstrated the usefulness of polarization in surface height recovery [Miyazaki 2002], overcoming the surface orientation ambiguity associated with photometric stereo; image segmentation; recognition and separation of reflection components; and distinguishing true laser stripes from inter-reflections for triangulation based laser scanning. An important contribution to shape from diffuse polarization was made by Miyazaki [Miyazaki 2003], who use a similar method to that described in this paper and estimate the illumination distribution from specularities. In [Atkinson 2006], the author presented a theoretical account of how diffuse polarization can be used to estimate surface normal directions and present results for a greater range of objects. One of our main aims here is to explore the extent to which surface orientation can be estimated from diffuse polarization alone. The work is based on the Fresnel theory for the reflection of electromagnetic waves from interfaces between media of different refractive indices. The Fresnel equations [Tan 2003], which are the principal results of this theory, give the ratio of reflected to incident light wave amplitudes for light that is linearly polarized either parallel to the plane of incidence (the plane that contains the incident and reflected light rays) or perpendicular to the plane of incidence. The equations provide a means by which the local surface normals of an object can be estimated. The zenith angle (i.e., the angle between the surface normal and the viewing direction) can be determined by the degree of polarization and may be recovered by solving a nonlinear equation. The azimuth angle of the surface normal (the angle of the projection of the normal onto the image plane relative to a reference) is determined by the orientation of the polarization. Once a field of surface normals for the object under study is to hand, then the surface height function may be recovered using standard surface integration techniques. To determine the required polarization, at least three images are needed.

2.4.4 Structured Light Projection Method

Structured light projection methods have also been proposed to recover local surface orientation. In the method proposed by Shrikhande and Stockman [Shrikhande 1989], a grid pattern is used. Surface orientations at the intersections are induced from the change

of the lengths of the grid edges in the image data. The result however indicates that there could be large errors at places where the grid-cell edges have large distortion, and such distortion is often induced by sharp edges in 3D. Sugihara [Sugihara 1984] proposed a method that infers surface orientation using the distortion of a known pattern projected on the surface from a structured light source. The projected textures on the object surface are detected and the distortion from the regular pattern is measured in order to estimate the surface orientation.

Winkelbach and Wahl [Winkelbach 2001] used a uni-direction strip pattern to compute the surface normal at each edge point as shown in Figure 2.17. A fringe pattern is projected onto the object twice from two different angles, each time with an image captured, so that two surface tangents are available for each image position. Surface normal at every image point can be interpolated from the data in the two images. A simplified method which determines surface normals from the slopes and intervals of the stripes in the image was also proposed [Winkelbach 2002]. It is based on the assumption that the surface patch between two strip edges is planar or very smooth. Thus, the tilt angle can be estimated from the deformed widths of the strip by comparing them with the strip width measured for a reference plane.

Compared with shape from X methods, structured light projection methods does not depend upon the existence of natural features, it makes less assumption on the target surface, and is more robust to imaging noise or illumination condition.

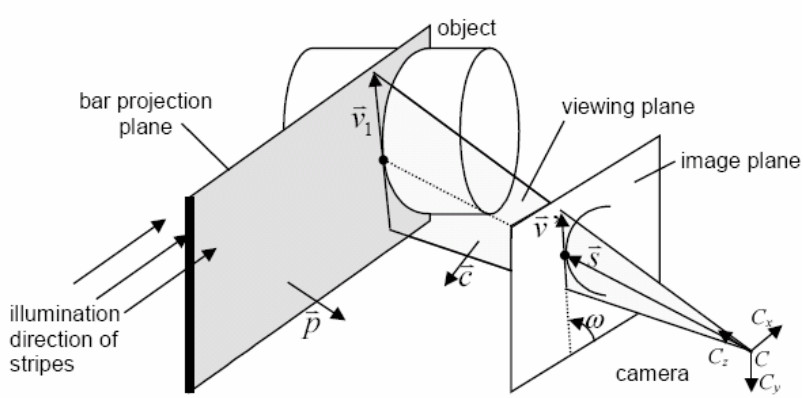


Figure 2.17 Projection of a stripe edge onto an object surface and reflected to the image plane.

2.5 Summary

3D position and orientation are both important in surface description. A comprehensive review is presented in this chapter about their recovery methods. In the 3D depth or position recovery techniques, we introduced time of flight, laser scanning, Moiré, shape from focus, stereo vision, and structured light-based systems. The emphasis is put to the structured light system including principle, triangulation, and coding schemes. When introduce the coding methods, we put our attention to the 2D coding methods. A number of previous coding strategies and related experimental results are provided.

As for the surface orientation recovery techniques, we classify them into image-based and structured light-based method. The first class contains shape from shading, photometric stereo, and shape from polarization etc. It utilizes one or multiple image intensities to estimate the illumination direction as well as the surface normal. By contrast, structured light-based method is much straightforward. It derives the surface normal from the geometry relations between the camera and projector. The ill-posed problem raised from image-based methods can be avoided. It also makes less assumption to the target surface and in turn more robust to illumination and image noises.

Chapter 3

Calibration of Projector-Camera System

Calibration is a crucial step in structured light-based range sensing device. It involves determination of both the intrinsic parameters of the camera and the projector that constitute the device, as well as the extrinsic parameters between the two instruments. Traditional solution requires the use of external calibration object with accurately measured pattern printed on it. This chapter presents a calibration design that makes use of an LCD panel as the calibration object. LCD panel's planarity is of industrial grade and is thus dependable. Pattern shown on the LCD panel is programmable and is thus convenient to produce in high precision. We show that with the design, the projector-camera system parameters can be calibrated with far less images in much higher accuracy. Extensive experiments are shown to illustrate the dramatic improvement in performance.

3.1 Introduction

Structured light system is one of the more convenient and accurate visual methods for 3D reconstruction. It is convenient because it requires only one camera and a projector to operate, and it needs no dense texture in the target surface to determine depth. It is accurate because it determines depth by the optical triangulation mechanism, and generally can achieve higher accuracy in depth determination than other visual methods like shape from shading [Bichsel 1992] etc.

Calibration is however a crucial step in the structured light system. Both the camera and the projector are required to be accurately calibrated for accurate depth recovery. Traditional calibration comprises two stages: camera calibration and projector calibration. Camera calibration has been widely researched. As for projector calibration, there is important difference from camera calibration. Since the projector cannot actually image the scene, methods for camera calibration cannot be directly adopted. The traditional calibration setting requires the use of a carefully constructed calibration object whose dimensions is precisely known or measured. One of the often-used calibration objects is a planar surface with accurately measured pattern printed on it. The calibration plane can be rotated and translated freely in the visual field of the camera, and it is imaged by the camera. From the image data of the pattern print on the calibration plane, the camera's intrinsic parameters as well as its position and orientation with respect to the calibration object can be determined. In the second stage, the calibrated information of the camera is used for projector calibration. With the spatial position of the calibration plane unchanged, a known pattern is projected from the projector onto the calibration plane whose reflection is imaged by the camera. The image information allows feature correspondences between the image plane of the camera and the image plane of the projector (we refer to the pattern generation panel of the projector as the image plane of the projector) be drawn, and in turn allows the projector's intrinsic parameters as well as its position and orientation with respect to the calibration plane to be also determined. Using the calibration object as the bridge between the camera end and the projector end, the positions of the camera and projector can be related and their relative geometry determined.

Obviously, calibration errors in the camera will be propagated to projector calibration and be enlarged there. Our experiments show that projector calibration could have error one order of magnitude larger than that of camera calibration. Hence, the accuracy of camera calibration determines the whole system's performance. To improve calibration accuracy, elaborate patterns need be used. In many of the existing systems these patterns are produced with precision as fine as 0.1 mm, and they are usually specially designed. This leads to inconvenience and high cost. On the other hand, if the widely adopted, less costly version of the calibration object - a pattern print glued onto a planar surface to form a calibration plane - is used, other problems arise. The pattern print has limited

accuracy in its metric figures. The planar surface has uncertain planarity. There is also inevitable folding of the print against the planar surface, leading to further compromise in the planarity of the calibration plane. All these would limit the calibration accuracy.

In this chapter, we present a calibration design that makes use of an LCD panel as the calibration plane. While patterns displayed on it are used for camera calibration, patterns projected onto and reflected by it when it is set to total dark are used for projector calibration. LCD panel's planarity is of industrial grade and is thus far more dependable; even consumer-level LCD panels have assurance that the substrate glass surface of the panel is of planarity deviation no more than 0.05 μm [Corning 2008]. Pattern shown on the LCD panel is programmable and is thus convenient to produce in high precision. We show that with the design, the projector-camera system parameters can be calibrated with far less images in much higher accuracy. Overall, the whole system can be built from off-the-shelf low-cost components, is convenient to use, and is of high accuracy.

The chapter is organized as follows. In Section 3.2, related works are briefly reviewed. In Section 3.3, the models for the camera and projector are introduced. In Section 3.4, the proposed calibration system that makes use of an LCD panel as the calibration plane is presented. The calibration steps are also outlined. In Section 3.5, extensive experimental results are shown that evaluate the performance of the proposed calibration design in a number of different measures, and with comparison with those of the traditional methods. A summary is given in Section 3.6.

3.2 Previous Work

Much work has addressed camera calibration. Among the most widely used ones are Tsai's method [Tsai 1987] and Zhang's method [Zhang 2000]. Tsai's method relies on accurate 3D measurement with respect to an external 3D calibration object to which a reference coordinate frame is defined. All the measurements of position and orientation are with respect to this reference coordinate frame. The method has been widely used in multi-camera applications. Zhang's work adopts the use of a planar calibration object instead. The advantages of the method are that planar calibration object is easier to construct than the 3D one, and both extrinsic parameters (the camera-to-world or camera-to-camera geometries) and intrinsic parameters (the internal parameters) of the cameras

can be calibrated in the same system. A similar method was also proposed by Sturm and Maybank [Sturm 1999], and there the singularities of the calibration results were discussed.

In contrast, fewer works have been devoted to the calibration of projector-camera system [Ribo 2005]. In the work of [Sansoni 2000], a relatively simple model for the optical geometry of the projector-camera system is proposed, which consists of three parameters: the baseline, the distance of the baseline from the reference surface, and the orientation of the projector with respect to the camera. The system requires the use of a specially designed calibration object - a block of parallepipedic shape. Its dimensions, measured by a Coordinate-Measurement Machine, must have an uncertainty one order of magnitude less than that of the targeted 3D resolution the structured light system is aimed at. The system comprises a specially developed LCD projector with VGA resolution, plus a video camera (of resolution 756×581 pixels) equipped with a CCTV manual iris zoom (of range 12.5-75 mm). Gray code and phase shifting codification are used for 3D reconstruction. Their experimental results show that at a working distance of 1500 mm, depth determination can reach a resolution of 70 μm, and the estimated accuracy can reach 0.1% of the depth range. Such figures are however only similar to those of the other calibration systems.

In the system proposed by Sadlo et al. [Sadlo 2005], a printed checkerboard pattern is used. The calibration plane is divided into two regions. One blank region is used as the projector screen and another is covered with a printed checkerboard pattern for camera calibration. Color markings on the four corners of the checkerboard pattern are used to ease pattern detection in the image data. With an image resolution of 1024×768 pixels, under the calibrated system parameters a spherical object could be reconstructed with a maximal error of 0.5 mm.

In the system developed by Legarda-Sáenz et al. [Sáenz 2004], a calibration object of size about 800×600 mm and with 140 physical target marks distributed uniformly over it was used. The target marks were measured with accuracy of 0.1 mm, and they were used first for camera calibration. Then a series of continuous periodic patterns that contain phase information (like the phase of a sinusoidal pattern) were projected from the projector to the calibration object, which was then pictured by the camera again. The phase

information in the patterns was so constructed that a one-to-one correspondence between the projector's image plane and the camera's image plane could be established. That was the main contribution of the work. Such one-to-one mapping between the projector's image plane and the cameras' image plane was then used to interpolate the image positions of the 140 target marks in the projector's image plane, thereby allowing those target marks be like pictured by the projector. Such image positions of the target marks in the projector were then used for projector calibration. In their experiments a camera of resolution 736×572 pixels and nominal focal length 12 mm, and an LCD projector of resolution 832×624 pixels and nominal focal length 50 mm, were used. The calibration was conducted using 19 positions of the calibration object. At a working distance of about 750 mm, the system could reconstruct a plane with standard deviation of planarity no more than 0.45 mm.

A similar idea was adopted by Song and Huang [Song 2006], but extended to a concept that allows the projector to be treated as if it can "capture" image. While in [Sáenz 2004] the image positions of 140 target marks were reconstructed in the projector's image plane, here the image of the entire calibration plane was reconstructed in the projector. With this concept, the calibration of the structured light system becomes essentially the same as the calibration of traditional stereo vision systems that consist of cameras only. In [Song 2006], a root mean square error of planarity from 0.10 mm to 0.22 mm was reported.

Methods such as [Fofia 2003] and [Li 2003] that require no use of calibration object gain convenience at the expense of reconstruction accuracy. In contrast, methods that require the use of calibration object can potentially achieve higher accuracy, but the accuracy hinges upon how accurate the metric figures of the calibration object and the markings on it are produced or measured.

Overall, we believe the approach of using a calibrated camera to calibrate the projector, even at the expense of having the accuracy of projector calibration partially influenced by that of camera calibration, has simplicity and convenience, and is a practical approach.

3.3 Projector-Camera System Model

The geometry between the camera, projector, and world coordinate frames is depicted in Figure 3.1. The perspective pinhole model is used to model the camera and projector, as illustrated by Equation (3.1) and (3.2). The 3D position $\mathbf{M}_W = [X, Y, Z]^T$ of any feature point in space, measured against the world coordinate frame, is related to the image position $\mathbf{m}_C = [u_C, v_C]$ or $\mathbf{m}_P = [u_P, v_P]$ by the following:

$$\begin{bmatrix} \mathbf{m}_{C/P} \\ 1 \end{bmatrix} \cong \begin{bmatrix} \alpha_u^{C/P} & \gamma^{C/P} & u_0^{C/P} \\ 0 & \alpha_v^{C/P} & v_0^{C/P} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{I}_3 & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} \end{bmatrix} \begin{bmatrix} \mathbf{M}_{C/P} \\ 1 \end{bmatrix} \tag{3.1}$$

$$\begin{bmatrix} \mathbf{M}_{C/P} \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{R}_{C/P} & \mathbf{t}_{C/P} \\ 0 & 1 \end{bmatrix}}_{\mathbf{E}_{C/P}} \begin{bmatrix} \mathbf{M}_W \\ 1 \end{bmatrix} \tag{3.2}$$

where the notation *C/P* refers to either the camera and the projector measurement or parameters, with *C* for camera and *P* for projector. There are five intrinsic parameters ($\alpha_u^C, \alpha_v^C, \gamma^C, u_0^C, v_0^C$ or $\alpha_u^P, \alpha_v^P, \gamma^P, u_0^P, v_0^P$) and 6 extrinsic parameters (($\mathbf{R}_C$, $\mathbf{t}_C$) or ($\mathbf{R}_P$, $\mathbf{t}_P$), collectively expressed as 4×4 matrix $\mathbf{E}_C$ or $\mathbf{E}_P$) for the camera and projector respectively. With respect to the same world coordinate system, extrinsic parameters $\mathbf{E}_C$ and $\mathbf{E}_P$ can be related by matrix $\mathbf{E}$, i.e., the relative transformation matrix between the projector and camera, as indicated by Equation (3.3). The calibration problem is to determine the intrinsic parameters of the camera and projector respectively and the relative transformation matrix $\mathbf{E}$:

$$\begin{bmatrix} \mathbf{M}_C \\ 1 \end{bmatrix} \cong \underbrace{\begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix}}_{\mathbf{E}} \begin{bmatrix} \mathbf{M}_P \\ 1 \end{bmatrix} \tag{3.3}$$

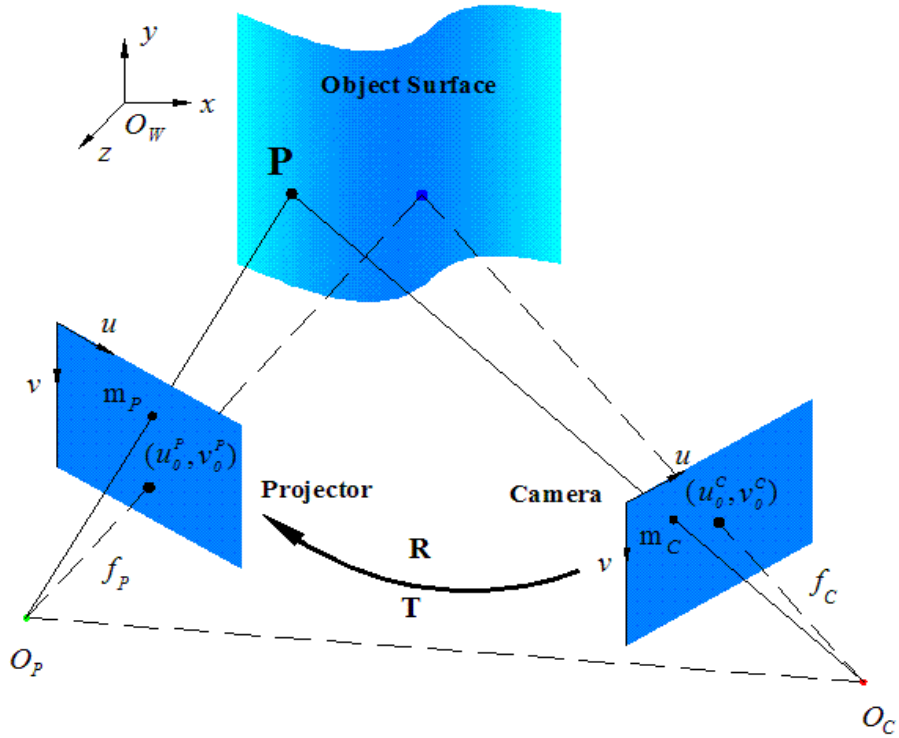


Figure 3.1 The geometry between the camera, projector, and the world coordinate frames. With the geometry calibrated, depth value of any world point P can be calculated from the two corresponding points: (u^c, v^c) and (u^p, v^p) via triangulation.

3.4 Calibration using LCD Panel

Accurate system model and solution for the model parameters are important. But another point often overlooked is the calibration pattern itself, especially when high accuracy is required. Once the system model is set up, the accuracy of the system relies heavily on the pattern selection and processing. Traditionally, a printed checkerboard paper which is attached onto a planar object is used as the calibration pattern. Another distinguishable pattern projected from the projector on the same planar object is used to calibrate the projector. Often the calibration algorithms assume that the observed plane and pattern are ideal; the print of the pattern, the manufacture of the calibration plane, and more precisely planarity errors are not considered. In reality, however, a standard copy paper has a thickness variation within itself from 0.08 to 0.11 mm. When the paper is attached to a planar board, the planarity of the final pattern could suffer from non-uniform folding

of the paper against the board. Even the planarity of the board itself could be limited. By experience, the flatness deviation of the pattern can well exceed 0.1 mm.

To overcome these drawbacks, in this work we propose an alternative calibration design. We use an LCD panel instead as the calibration plane. While patterns displayed on the panel are used for calibrating the camera, patterns projected from the projector onto the panel and reflected by it at a total-dark setting of its display are used for calibrating the projector. The advantages of this design include the following:

a. *Simple and Convenient Calibration Setup: A consumer LCD panel is all that is additionally required for the calibration task. Displaying pattern in and projecting pattern to an LCD panel are also much less tedious processes than acquiring precisely measured pattern print and gluing it to a planar board.*

b. *Ideal planarity: The flatness of LCD panel is of industrial grade; many off-the-shelf projectors have guarantee that planarity deviation of the panel is less than 0.05 μm. To the stated calibration task, it is almost as if the panel surface is an ideal plane.*

c. *Programmable Patterns: Different types of pattern in various combinations of colors can be generated at ease by computer, and displayed on the LCD panel as the calibration patterns.*

d. *High accuracy: There is no print error and the pattern dimensions are all measurable with ease in pixels of the LCD panel.*

e. *High image quality: Pattern images with high contrast can be achieved.*

3.4.1 Camera Calibration

A pattern with known dimensions on either the LCD panel in the proposed calibration setup or on an external calibration board in the traditional setup is shown to the camera and imaged. The method in [Zhang 2000] is then adopted for camera calibration. By introducing the homography constraint between the camera image plane Π_{CCD} and the calibration plane Π_{LCD}, the intrinsic parameters of the camera can be estimated. Once the camera intrinsic parameters $\mathbf{A}_C$ are known, the position of the camera with respect to the

calibration plane, or alternatively speaking the position of the calibration plane with respect to the camera - $\mathbf{R}_C$, $\mathbf{t}_C$ - can be determined as well. The procedure of the camera calibration can be depicted as following.

- *Intrinsic and Extrinsic Parameters*

Denote the rotation and homography matrix between the camera image plane and the calibration plane by $\mathbf{R}_C$ and $\mathbf{H}_C$. Write the two matrices in vector form as:

$$\begin{aligned} \mathbf{R}_C &= \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 \end{bmatrix} \\ \mathbf{H}_C &= \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3 \end{bmatrix} \end{aligned} \tag{3.4}$$

According to the definition of homography, we can get:

$$\begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3 \end{bmatrix} = \lambda \mathbf{A}_C \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t}_C \end{bmatrix} \tag{3.5}$$

where λ is an arbitrary scalar. Since $\mathbf{r}_1$ and $\mathbf{r}_2$ are orthonormal, we have following two constrains to $\mathbf{A}_C$:

$$\mathbf{h}_1^T \mathbf{A}_C^{-T} \mathbf{A}_C^{-1} \mathbf{h}_2 = 0 \tag{3.6}$$

$$\mathbf{h}_1^T \mathbf{A}_C^{-T} \mathbf{A}_C^{-1} \mathbf{h}_1 = \mathbf{h}_2^T \mathbf{A}_C^{-T} \mathbf{A}_C^{-1} \mathbf{h}_2 \tag{3.7}$$

Define matrix **B** by the following equation:

$$\mathbf{B} \cong \mathbf{A}_C^{-T} \cdot \mathbf{A}_C^{-1} = \begin{bmatrix} \frac{1}{(\alpha_u^C)^2} & -\frac{\gamma^C}{(\alpha_u^C)^2 \alpha_v^C} & \frac{v_0^C \gamma^C - u_0^C \alpha_v^C}{(\alpha_u^C)^2 \alpha_v^C} \\ -\frac{\gamma^C}{(\alpha_u^C)^2 \alpha_v^C} & \frac{(\gamma^C)^2}{(\alpha_u^C)^2 (\alpha_v^C)^2} + \frac{1}{(\alpha_v^C)^2} & -\frac{(v_0^C \gamma^C - u_0^C \alpha_v^C)\gamma^C}{(\alpha_u^C)^2 (\alpha_v^C)^2} - \frac{v_0^C}{(\alpha_v^C)^2} \\ \frac{v_0^C \gamma^C - u_0^C \alpha_v^C}{(\alpha_u^C)^2 \alpha_v^C} & -\frac{(v_0^C \gamma^C - u_0^C \alpha_v^C)\gamma^C}{(\alpha_u^C)^2 (\alpha_v^C)^2} - \frac{v_0^C}{(\alpha_v^C)^2} & \frac{(v_0^C \gamma^C - u_0^C \alpha_v^C)^2}{(\alpha_u^C)^2 (\alpha_v^C)^2} + \left(\frac{v_0^C}{\alpha_v^C}\right)^2 + 1 \end{bmatrix} \tag{3.8}$$

Since **B** is symmetry, define a 6-D vector **b** by:

$$\mathbf{b} = \begin{bmatrix} B_{11} & B_{12} & B_{22} & B_{13} & B_{23} & B_{33} \end{bmatrix}^T \tag{3.9}$$

Let the ith column vector of **H** be $\mathbf{h}_i = \begin{bmatrix} \mathbf{h}_{i1} & \mathbf{h}_{i2} & \mathbf{h}_{i3} \end{bmatrix}^T$, we have:

$$\mathbf{h}_i^T \mathbf{B} \mathbf{h}_j = \mathbf{v}_{ij}^T \mathbf{b} \tag{3.10}$$

where

$$\mathbf{v}_{ij} = [h_{i1}h_{j1} \quad h_{i1}h_{j2} + h_{i2}h_{j1} \quad h_{i2}h_{j2} \quad h_{i3}h_{j1} + h_{i1}h_{j3} \quad h_{i3}h_{j2} + h_{i2}h_{j3} \quad h_{i3}h_{j3}]$$

If we rewrite 2 homogeneous equations in **b**:

$$\begin{bmatrix} \mathbf{v}_{12}^T \\ (\mathbf{v}_{11} - \mathbf{v}_{22})^T \end{bmatrix} \mathbf{b} = 0 \tag{3.11}$$

If n images of the model plane are observed, by stacking n such equations as Equation (3.11), we have $\mathbf{Vb} = 0$, where **V** is a 2×6 matrix. If $n > 3$, we will have in general a unique solution **b** defined up to a scale factor.

The solution of Equation (3.11) is well known as the singular vector of **V** associated with the smallest singular value. Once **b** is estimated, we can compute the camera intrinsic matrix $\mathbf{A}_C$ from Equation (3.8). When $\mathbf{A}_C$ is known, the extrinsic parameters for each calibration image plane can be calculated by following equations:

$$\begin{aligned} \mathbf{r}_1 &= \lambda \mathbf{A}_C^{-1} \mathbf{h}_1 \\ \mathbf{r}_2 &= \lambda \mathbf{A}_C^{-1} \mathbf{h}_2 \\ \mathbf{r}_3 &= \mathbf{r}_1 \times \mathbf{r}_2 \\ \mathbf{t}_C &= \lambda \mathbf{A}_C^{-1} \mathbf{h}_3 \\ \lambda &= \frac{1}{\left\| \mathbf{A}_C^{-1} \mathbf{h}_1 \right\|} \end{aligned} \tag{3.12}$$

A *maximum likelihood estimation* can be used to solve following equation.

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \left\| m_{Cij} - \hat{m}_C(\mathbf{A}_C, \mathbf{R}_{Ci}, \mathbf{t}_{Ci}, \mathbf{M}_{Wj}) \right\|^2 \tag{3.13}$$

where $\hat{m}_C(\mathbf{A}_C, \mathbf{R}_{Ci}, \mathbf{t}_{Ci}, \mathbf{M}_{Wj})$ indicates the projector of point $\mathbf{M}_{Wj}$ in the ith image. To solve the non-linear minimization problem, *Levenberg-Marquardt Algorithm* (or *LMA*) is used in Zhang's method.

- *Camera Lens Distortion*

As reported in some literatures such as [Wei 1994], the distortion function is mainly dominated by the radial components in the camera distortion model. It has also been proved in our experimental results as shown in Section 3.5.1. In Zhang's method, only first k_1 and second order k_2 radial distortion coefficients are considered. Assume (u_C, v_C) is the distortion–free image coordinates, and (x_C, y_C) be its normalized image coordinates. The observed distorted image position $(\breve{u}_C, \breve{v}_C)$ can be expressed as:

$$\begin{aligned} \breve{v}_C &= v_C + (v_C - v_0^C)[k_1(x_C^2 + y_C^2) + k_2(x_C^2 + y_C^2)^2] \\ \breve{u}_C &= u_C + (u_C - u_0^C)[k_1(x_C^2 + y_C^2) + k_2(x_C^2 + y_C^2)^2] \end{aligned} \tag{3.14}$$

In matrix form:

$$\underbrace{\begin{bmatrix} \breve{u}_C - u_C \\ \breve{v}_C - v_C \end{bmatrix}}_{\mathbf{d}} = \underbrace{\begin{bmatrix} k_1 \\ k_2 \end{bmatrix}}_{\mathbf{k}} \underbrace{\begin{bmatrix} (u_C - u_0^C)(x_C^2 + y_C^2) & (u_C - u_0^C)(x_C^2 + y_C^2)^2 \\ (v_C - v_0^C)(x_C^2 + y_C^2) & (v_C - v_0^C)(x_C^2 + y_C^2)^2 \end{bmatrix}}_{\mathbf{D}} \tag{3.15}$$

The linear least-square solution of **k** can be expressed as:

$$\mathbf{k} = (\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T\mathbf{d} \tag{3.16}$$

It can be solved by using maximum likelihood estimation and *LMA* again.

In practice, a printed checkerboard with known dimension is usually used as the calibration target. The calibration plane is placed in front of the camera with different poses and then captured by the camera. Grid-points between checkerboard elements are extracted and then used to calculate parameters of camera sensor and lens. Several off-the-shelf tools or codes are available from the internet, for example, a Matlab Toolbox for Camera Calibration [Bouguet 2007], OpenCV Library [Intel 2008] based *C* language etc.

3.4.2 Projector Calibration

Once the camera is calibrated, with the spatial position and orientation of the LCD panel kept still we employ the projector to project a known pattern onto the LCD panel. The reflection from the panel is then imaged by the camera, and the image data is used for

calibrating the projector. In the process the LCD panel is instructed to have the total-dark setting on its display, so that confusion between the displayed pattern and reflected pattern on the calibration object in the image data - an inevitable issue in traditional calibration - is avoided. That is another advantage of using the LCD panel as the calibration object.

- *Intrinsic and Extrinsic Parameters*

Suppose we have a feature point $\mathbf{m}_p$ (i.e., (u_p, v_p)) on the projector's image plane, whose projection on the calibration plane Π_{LCD} is a certain $\mathbf{P}_L$ (i.e., (x_p, y_p, z_p)). Suppose also that $\mathbf{m}_c$ (i.e., (u_c, v_c)) is the corresponding image position on the camera's image plane that we can observe from the reflection image of Π_{LCD}. Since the camera parameters $\mathbf{A}_C$, $\mathbf{R}_C$, $\mathbf{t}_C$ have been made known from the prior camera calibration process, we have the following for $\mathbf{P}_L$:

$$\begin{bmatrix} \mathbf{m}_c \\ 1 \end{bmatrix} \cong \mathbf{A}_C \begin{bmatrix} \mathbf{I}_3 & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} \end{bmatrix} \begin{bmatrix} \mathbf{R}_C & \mathbf{t}_C \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{P}_L \\ 1 \end{bmatrix} \tag{3.17}$$

which represents two scalar constraints for $\mathbf{P}_L$. The fact that $\mathbf{P}_L$ lies on the calibration plane Π_{LCD} - which has also been made known in the camera calibration process - represents the third constraint for $\mathbf{P}_L$. The three constraints together allow $\mathbf{P}_L$ to be precisely determined.

Such a correspondence between $\mathbf{m}_p$ and $\mathbf{m}_c$ represents two scalar constraints for the projector's intrinsic parameters $\mathbf{A}_P$ and extrinsic parameters $\mathbf{R}_P$, $\mathbf{t}_P$, as expressed by:

$$\underbrace{\begin{bmatrix} u_p \\ v_p \\ 1 \end{bmatrix}}_{\mathbf{m}_p} \cong \underbrace{\begin{bmatrix} \alpha_u^P & \gamma^P & u_0^P \\ 0 & \alpha_v^P & v_0^P \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{A}_P} \begin{bmatrix} \mathbf{I}_3 & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} \end{bmatrix} \begin{bmatrix} \mathbf{R}_P & \mathbf{t}_P \\ 0 & 1 \end{bmatrix} \underbrace{\begin{bmatrix} x_p \\ y_p \\ z_p \\ 1 \end{bmatrix}}_{\mathbf{P}_L} \tag{3.18}$$

In the procedure, the 3D information $\mathbf{P}_L$ is first determined from image data $\mathbf{m}_c$ via Equation (3.17), and then the known $\mathbf{P}_L$ is used to determine the projector parameters $\mathbf{A}_P$, $\mathbf{R}_P$ and $\mathbf{t}_P$ via Equation (3.18). With this, projector calibration can be treated as traditional

camera calibration, which is to determine intrinsic and extrinsic parameters from known 3D information. On this, any state-of-the-art camera calibration method like [Lenz 1988], [Weng 1992], and [Heikkil 1997] can be used to tackle the problem.

As reported in the literature survey, bigger calibration errors can be found in projector calibration procedure. The error can be caused by the model of projector and lens or the calibration means. To improve the lens distortion estimation, we used a more complex model proposed in [Weng 1992]. The proposed calibration procedure consists of two steps. The first step employs a non-iterative algorithm to directly compute a closed-form solution of all the extrinsic parameters and some major intrinsic parameters based on a distortion-free projector model. In the second step, the parameters estimated in the first step are improved iteratively through a nonlinear optimization, taking into account camera distortions. According to minimum variance estimation, the objective function to be minimized is the mean-square discrepancy between the observed image points and their inferred image projections computed with the estimated calibration parameters. Details of this camera calibration-like step are available in [Weng 1992].

- *Projector Lens Distortion*

Only radial distortion is considered in the methods of [Zhang 2000] and [Tsai 1987], and it has shown to be accurate enough in most applications. Weng improved the model proposed by [Faugeras 1986] including three types of lens distortion: radial distortion, decentering distortion and prism distortion. The radial distortion is modeled in the same way Tsai did. The decentering distortion is due to that the optic centre of the lenses is not correctly aligned with the centre of the camera. This type of distortion introduces a radial and tangential distortion. Moreover, the thin prism distortion arises from imperfection in lens design and manufacturing as well camera assembly. This type of distortion can be modeled adding a thin prism in the optic system, causing radial and tangential distortions [Weng 1992]. Adding the three types of distortions, the following equations are obtained. They are represented by following equations respectively.

$$\begin{aligned} \delta_{ur} &= k_1 u(u^2 + v^2) + O[(u,v)]^5 \\ \delta_{vr} &= k_1 v(u^2 + v^2) + O[(u,v)]^5 \end{aligned} \tag{3.19}$$

$$\begin{aligned} \delta_{ud} &= p_1(3u^2 + v^2) + 2p_2uv + O[(u,v)]^4 \\ \delta_{vd} &= p_1(u^2 + 3v^2) + 2p_1uv + O[(u,v)]^4 \end{aligned} \tag{3.20}$$

$$\begin{aligned} \delta_{up} &= s_1(u^2 + v^2) + O[(u,v)]^4 \\ \delta_{vp} &= s_2(u^2 + v^2) + O[(u,v)]^4 \end{aligned} \tag{3.21}$$

Total amount of distortion along the u and v axes can be represented by the combination of the above six equations. Assuming that terms of order higher than three are negligible, we obtain:

$$\begin{aligned} \delta_u(u,v) &= k_1u(u^2 + v^2) + p_1(3u^2 + v^2) + 2p_2uv + s_1(u^2 + v^2) \\ \delta_v(u,v) &= k_1v(u^2 + v^2) + p_1(u^2 + 3v^2) + 2p_1uv + s_2(u^2 + v^2) \end{aligned} \tag{3.22}$$

Let

$$g_1 = s_1 + p_1, g_2 = s_2 + p_2, g_3 = 2p_1, g_4 = 2p_2$$

Then, the Equation (3.22) can be expressed as:

$$\begin{aligned} \delta_u(u,v) &= k_1u(u^2 + v^2) + (g_1 + g_3)u^2 + g_4uv + g_1v^2 \\ \delta_v(u,v) &= k_1v(u^2 + v^2) + g_2u^2 + g_3uv + (g_2 + g_4)v^2 \end{aligned} \tag{3.23}$$

3.4.3 Calibration Flow

After calibration of both the camera and the projector, $\mathbf{R}_C$, $\mathbf{t}_C$, $\mathbf{R}_P$, $\mathbf{t}_P$ can be used to determine the relative transformation matrix **E** from Equation (3.3). Based on this idea, we have the calibration procedure modified to the following steps:

a. *A checkerboard pattern with known dimensions is displayed on the LCD, and is imaged by the camera.*

b. *Keeping its position unchanged, set the LCD panel to a total dark and use it as a projection screen.*

c. *Project a known pattern from the projector onto the LCD panel, whose reflection is again imaged by the camera. The image of the displayed pattern and that of the projected pattern form an image pair.*

d. *Change the pose of LCD and repeat steps b-c to acquire enough image pairs.*

e. *Each image pair is associated with a particular pose of LCD panel. Use those images related to the LCD panel's pattern displays to calibrate the camera's intrinsic parameters. Also, use the image data to determine, for each pose of the LCD panel, the position Π_{LCD} of the LCD screen with respect to the camera.*

f. *Then use the image data related to the LCD panel's reflections of projected patterns to calibrate, through Equations (3.17) and (3.18), the projector's intrinsic parameters, as well as its position with respect to the camera.*

3.4.4 3D Triangulation

With the calibration result, depth can be recovered from the triangulation mechanism [Ribo 2005] as expressed by the following:

$$z_c = \frac{(\mathbf{R}\overline{\mathbf{m}}_c \cdot \overline{\mathbf{m}}_p)(\overline{\mathbf{m}}_p \cdot \mathbf{T}) - \left\|\overline{\mathbf{m}}_p\right\|^2 (\mathbf{R}\overline{\mathbf{m}}_c \cdot \mathbf{T})}{\left\|\mathbf{R}\overline{\mathbf{m}}_c\right\|^2 \left\|\overline{\mathbf{m}}_p\right\|^2 - (\mathbf{R}\overline{\mathbf{m}}_c \cdot \overline{\mathbf{m}}_p)^2} \tag{3.24}$$

where $\overline{\mathbf{m}}_c$ and $\overline{\mathbf{m}}_p$ refer to the two corresponding points on the camera's and projector's image planes respectively in homogeneous coordinates.

Notice that there exist more robust triangulation methods [Hartley 2004], which conduct 3D determination more globally like considering the minimization of the sum of squared differences between the corresponding points in the two images. Such methods are particularly useful in the case that large noise is present in the image positions of the feature points. However, in our system, the feature points on the projector side are generated data not observed ones, and they are known a priori without much error. Our experiments show that for the equipments we experimented with, direct triangulation already offers sufficiently accurate result. The experimental results in Section 3.5.1 show that the distortions in both the camera and the projector were relatively small in relation

to our task, and imposed almost no effect to the final reconstruction accuracy. For simplicity we thus in the triangulation process skip the use of the lens distortion parameters for 3D reconstruction.

In the procedure, the camera and projector use the same calibration plane but different patterns for calibration. While the camera is calibrated from patterns displayed on the LCD screen, the projector is calibrated from reflection of patterns on the screen. By using the LCD panel as the calibration plane, errors caused by imperfections of the patterns and the calibration plane, and confusion between displayed pattern and reflected pattern in the image data could be much reduced.

3.5 Experimental Results

In this experiment the calibration quality is evaluated based upon the re-projection error, which is the average Euclidean distance in the image space between two sets of image positions – the observed ones of certain image features, and those that are projected to the image plane from the corresponding features in 3D (on the calibration plane) using the calibrated camera parameters. We compared the results of our calibration design with those of the traditional setup.

The projector-camera system we used in our experiments consisted of a DLP projector of resolution 1024×768 (*NEC LT25* projector) and a camera of resolution 2400×1600 (*Pentax K100D* DSLR camera with *Pentax DA* 18-55 mm lens), both being off-the-shelf equipments as shown in Figure 3.2. The focal length of projector was of range 25-31 mm. The system was configured for a working distance about 750 mm, and an object thickness of about 50 mm (but not exceeding 100 mm).

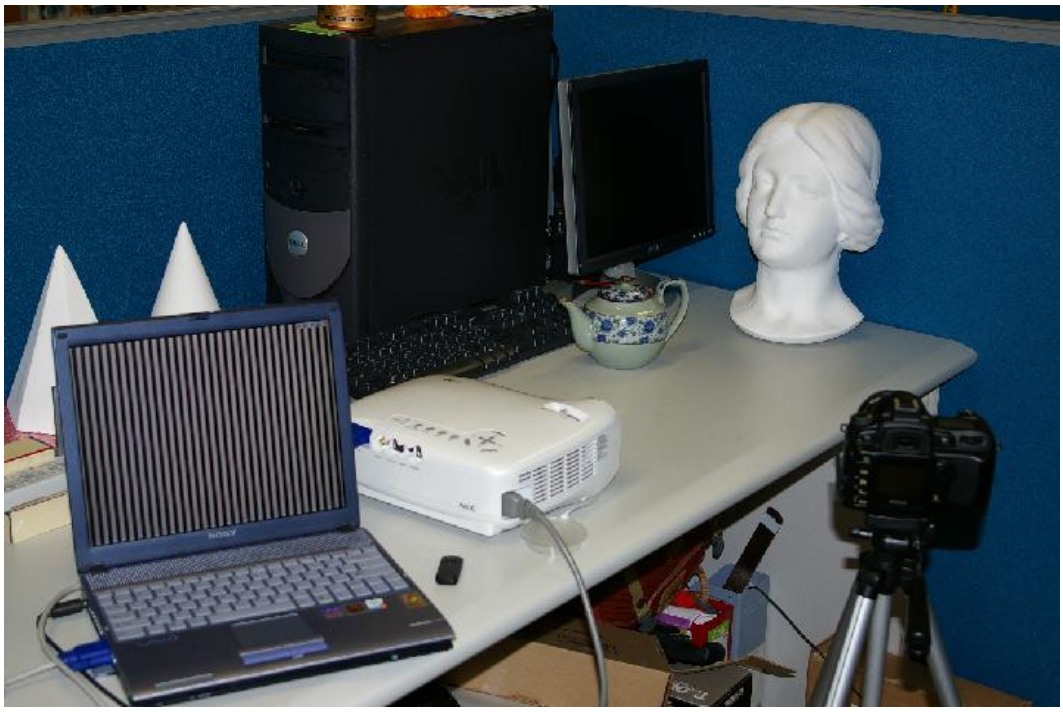


Figure 3.2 Experimental system, which consists of a DLP projector, a DSLR camera, and an LCD display for system calibration.

The calibration setup comprised a 15' LCD panel that was used as the calibration plane. At any new pose of the LCD panel, the orientation of the LCD panel was adjusted to avoid reflection glare. The calibration patterns we used are displayed in Figure 3.3.

The use of the LCD panel let the calibration patterns be displayed with high contrast, allowing the pattern image to be processed with ease and accuracy. We had tried as many as 20 images (under different positions and orientations of the LCD panel) for camera calibration, but the result showed that 6-8 images were generally enough to obtain stable calibration parameters. We therefore in all the experiments used only 8 images for calibration. In contrast, experiments show that the traditional method required 20-30 images to have enough calibration accuracy on the same projector-camera system. We believe the great reduction of the required image number owes to the many advantages of using the LCD panel as the calibration object.

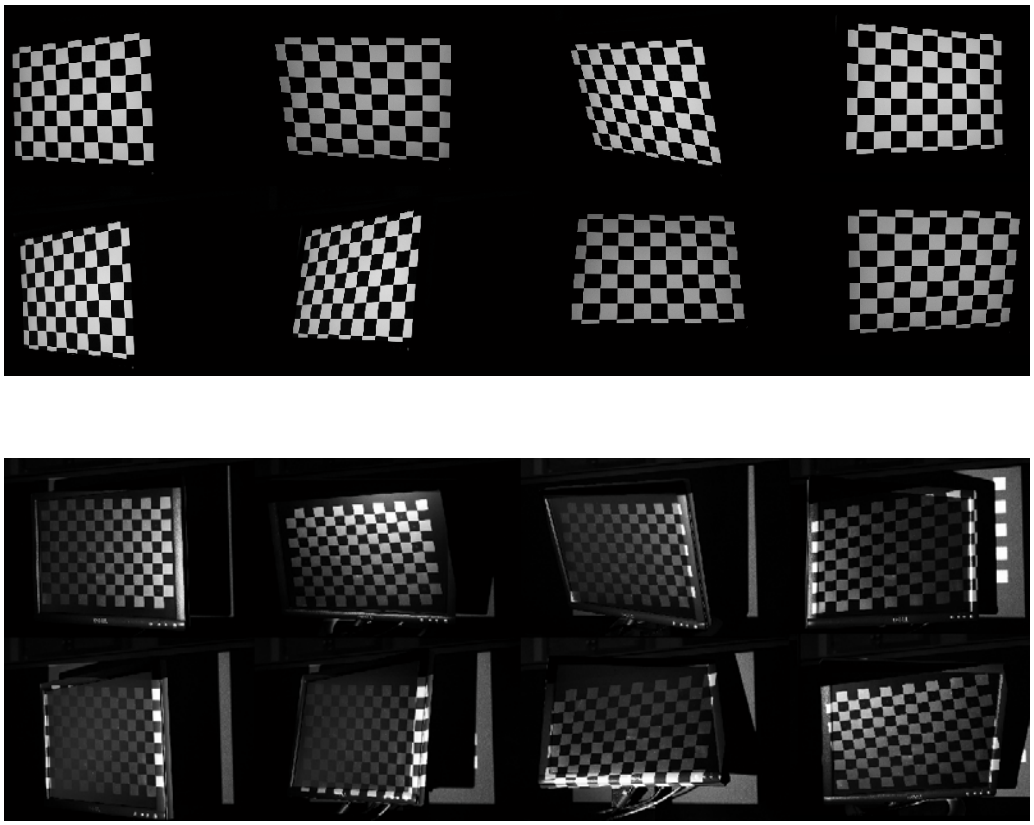


Figure 3.3 (Top) The checkerboard pattern displayed on the LCD screen for camera calibration in our setup. (Bottom) Patterns projected onto the LCD screen for projector calibration.

The traditional calibration setup we employed in our experiments used a printed checkerboard pattern of A4 size as the calibration object. In traditional calibration setup there is always the issue of confusion between the pattern print and the reflected pattern in the image data. We avoided the confusion by having the pattern print purposefully made of a rather light grey level - a grey level of 128. As shown in Figure 3.4, this allowed the pattern print and the reflected pattern is more distinguishable. In the experiments with the traditional calibration setup, a total of 20 pairs of images under different poses of the calibration object were used.

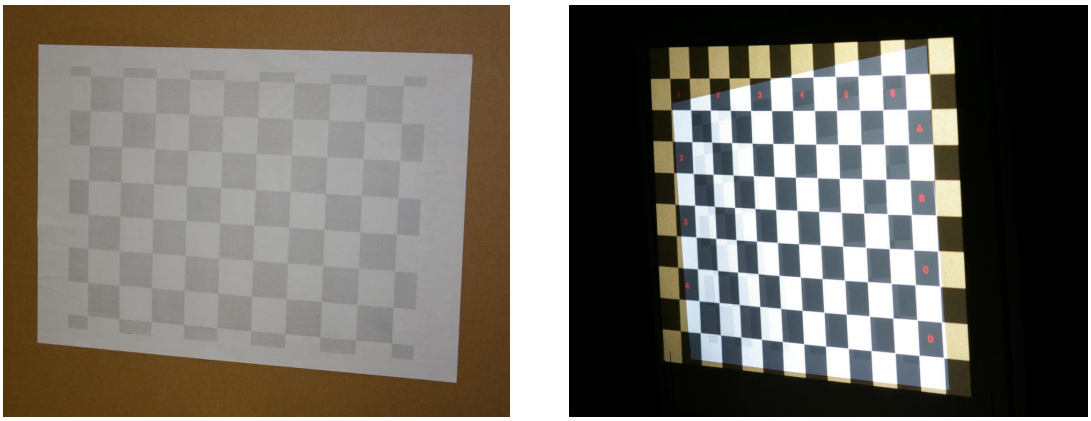


Figure 3.4 (Left) The checkerboard pattern print of grey level 128 that was for camera calibration under the traditional calibration design. (Right) The reflected pattern on top of the above pattern, that was for projector calibration.

3.5.1 Lens Distortion

We have incorporated lens distortion parameters in calibrating the camera and the projector, in the same way as described in Zhang and Weng's method. The result shows that tangential distortion of both camera and projector can be ignored. The lens distortion is mainly detonated by the radial distortion as shown in Figure 3.5. However, in most region of the sensor plane, the radial distortion can be controlled within 2 pixels. Compared with the size of the sensor, the distortion is relative very small. The experiments also show that the distortions in both the camera and the projector were relatively small in relation to our task, and imposed almost no effect to the final reconstruction accuracy. To simplify the calculation, both the radial and tangential distortions are not considered in the triangulation model as described in Section 3.4.4.

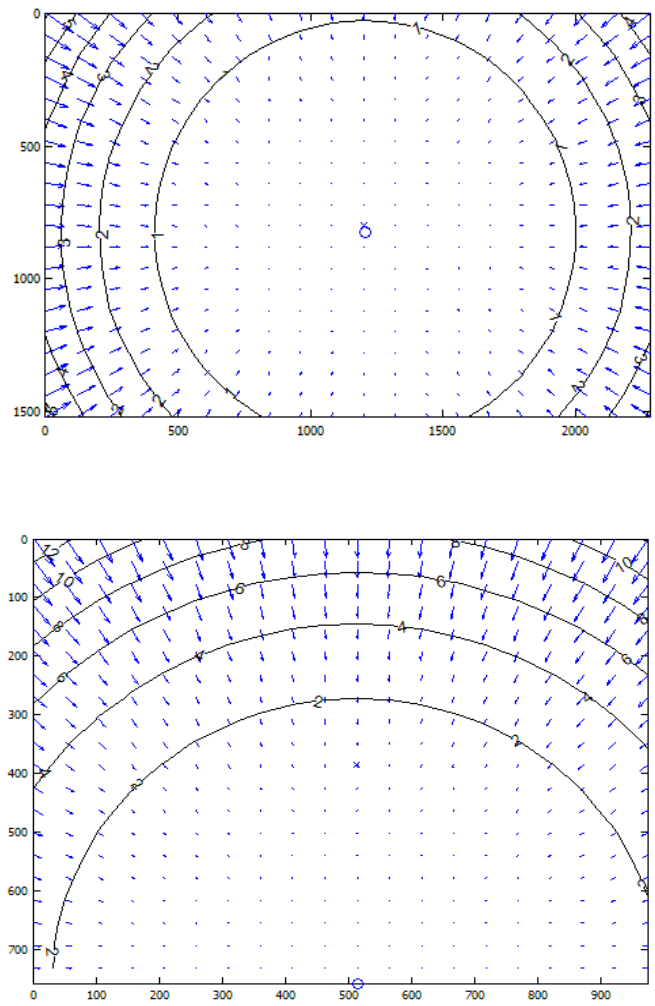


Figure 3.5 (Top) Radial distortion of camera. (Bottom) Radial distortion of projector.

3.5.2 Re-projection Error

Following the procedure described in Section 3.4, the system was calibrated with our and the traditional calibration setups respectively. The re-projection error distribution of the camera and projector's are shown in Figure 3.6. Some major calibration parameters are listed in Table 3.1: *RPE* is the re-projection error; f_C, f_P are the focal lengths; u_{C0}, v_{C0}, u_{P0}, v_{P0} are about the principal points of the camera and the projector respectively. The skew parameter and aspect ratio of both equipments were very close to 0 and 1, so we do not include them in Table 3.1.

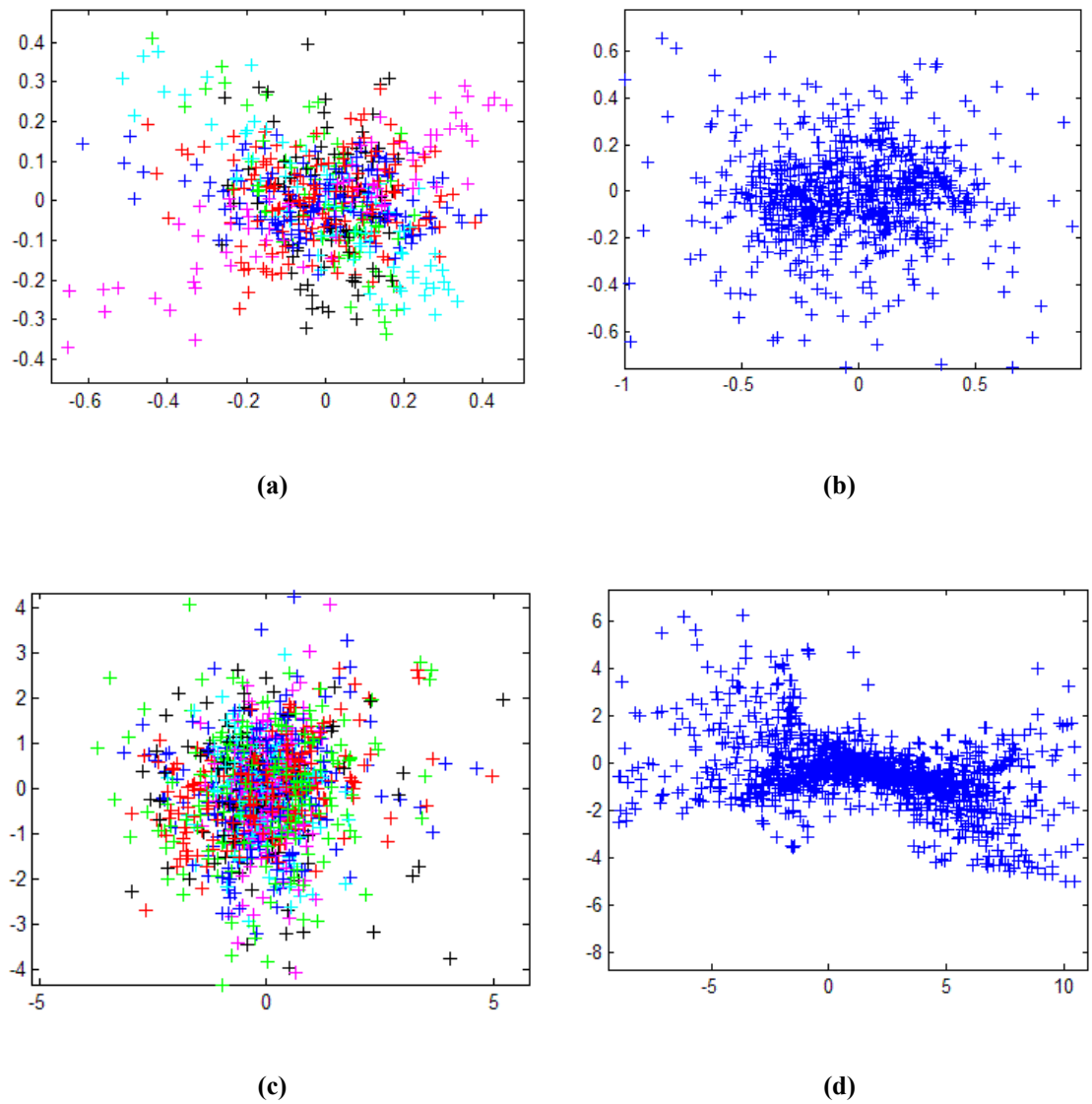


Figure 3.6 RPE distribution: (a) Camera calibration result. (b) Projector calibration result, both from the use of LCD panel as the calibration object. (c) Camera calibration result. (d) Projector calibration result, both from the use of printed pattern as the calibration object.

The result shows that with the use of the LCD panel, uncertainty can be reduced by 5-10 times, even with less than half of the images used (8 vs. 20). The *RPE* was about 0.1 pixel only under the use of 8 calibration images in the proposed calibration design. The *RPE* in projector calibration can be controlled to within 0.5 pixel. Compared with the *RPE* of 6 pixels in the use of the printed calibration pattern, the improvement was more than 10 times. Figure 3.6 (d) shows certain non-uniformity in the distribution, but the

non-uniformity does not always show up in the experiments, and when it shows up it does not appear to have any particular pattern. We believe the occasional non-uniformity of the distribution comes from the uneven non-planarity of the printed pattern and board, and features chosen from different parts of the calibration object cause different error distributions.

3.5.3 Planarity Check

In this experiment, we evaluate directly the improvement in 3D reconstruction under the proposed calibration design, as 3D reconstruction is the ultimate goal of the system. We compared the measured planarity of an ideal plane using the calibration parameters obtained respectively from our calibration system and from the traditional one (which used a printed pattern as the calibration object). We chose the LCD panel itself as the ideal plane, since its planarity is of an assured quality close enough to ideality. It was placed about 750 mm away from the camera in all related experiments.

In both calibration systems, a pattern was projected onto the LCD panel surface, some feature points were extracted, and their depth values calculated with the calibration parameters. A total of 80 points were extracted and their depth values calculated through triangulation with respect to the camera reference frame. A plane was then fitted to the reconstructed 3D points.

With calibration result from the proposed calibration system, the residual errors of the points with respect to the above the fitted plane were calculated and plotted in Figure 3.7. The error had a root mean square value of only 0.04 mm and a standard deviation of 0.05 mm.

When the calibration result obtained from the printed calibration pattern was used for 3D reconstruction, the error had a root mean square value of 0.405 mm and a standard deviation of 0.480 mm.

The result shows that 3D reconstruction accuracy can be improved by almost 10 times under the new calibration system.

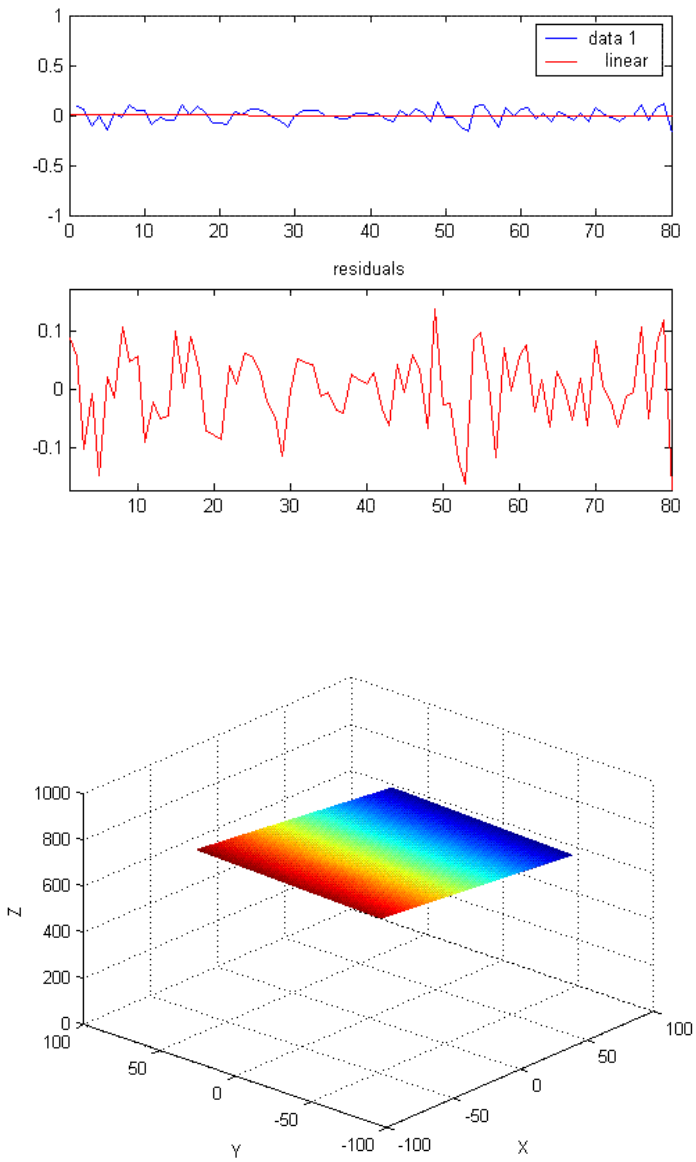


Figure 3.7 (Top) Residual errors with respect to the fitted plane under the proposed calibration setup. (Bottom) Reconstruction of the plane with respect to the camera coordinate frame. The reconstructed plane had an orientation [-0.4611, 0.0259, -0.8870] with respect to the camera reference frame.

3.5.4 Motion Measurement

In this experiment, the known motion of a plane is used to evaluate the calibration accuracy. A linear stage with motion accuracy of 1 μm was used to move an external plane. The plane was shifted 10 times in steps of 1 mm. After each motion, a pattern was projected on the plane for its reconstruction. Totally 11 planes were reconstructed, and the results under the proposed calibration are shown in Figure 3.8. The distances between every two adjacent planes were calculated and they are shown in Table 3.2. The absolute mean deviation of the measurement under the proposed calibration setup has a value of 0.077 mm or 0.01% of the depth range, while that from the traditional calibration setup had 0.168 mm. That was an accuracy improvement of more than two times.

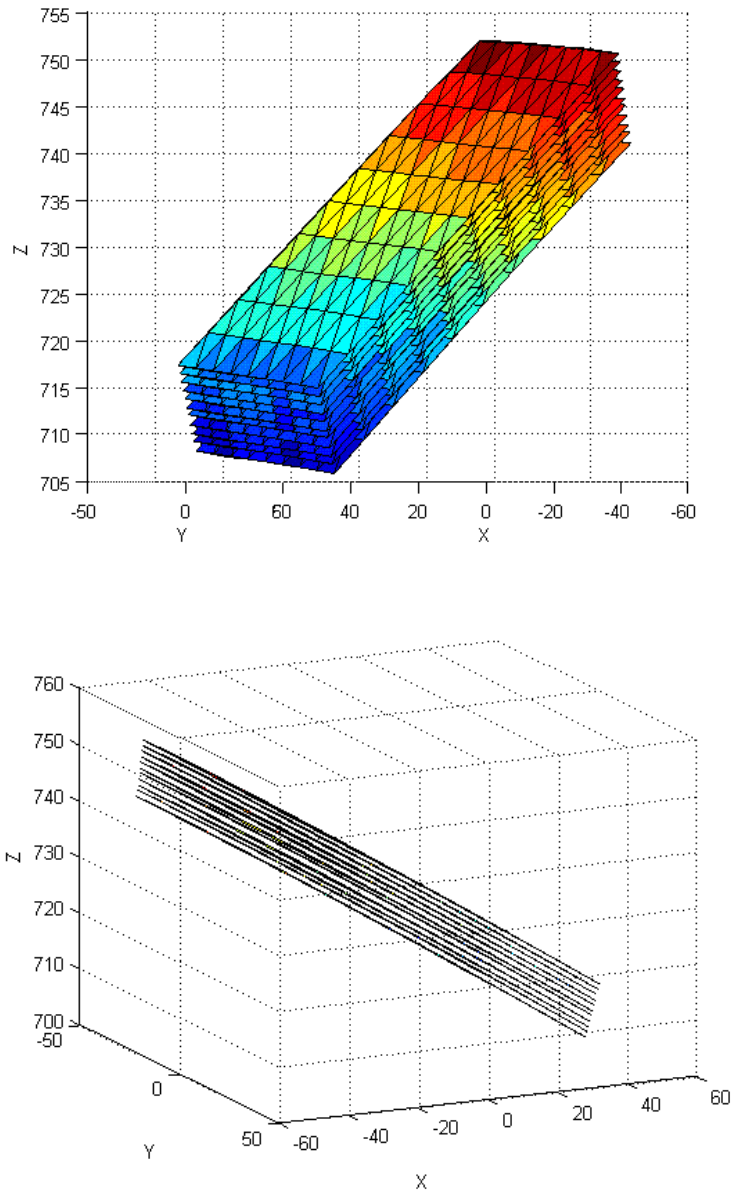


Figure 3.8 (Top) 11 reconstructed planes under the proposed calibration. (Bottom) Side view of the reconstructed planes.

Table 3.1 Calibration results from the use of LCD and printed pattern respectively (Unit: pixel).

Parameter	**LCD Pattern**	**Printed Pattern**
Camera RPE	[0.1308 0.0994]	[1.0446 1.3258]
f_C	[3255.9 3265.5] ± [3.9 3.5]	[3204.6 3197.8] ± [24.3 23.6]
u_{C0}, v_{C0}	[1199.8 835.0] ± [3.6 2.9]	[1380.4 837.5] ± [45.8 30.3]
Projector RPE	[0.4720 0.3281]	[6.1614 3.2826]
f_P	[2225.6 2223.8] ± [12.6 12.5]	[1962.3 1876.7] ± [47.1 45.1]
u_{P0}, v_{P0}	[511.7 747.8] ± [5.8 16.5]	[686.4 698.9] ± [78.4 98.8]

Table 3.2 Measurement results of the various motion steps of a moving plane. (Unit: mm)

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Error**
LCD Pattern	0.986	0.986	1.060	0.942	1.128	0.851	1.076	1.074	0.876	1.075	0.077
Printed Pattern	1.102	1.112	1.174	1.063	1.249	0.846	1.316	1.191	0.939	1.265	0.168

3.6 Summary

In this chapter, we have described an accurate, convenient, and efficient calibration design for the projector-camera structured light system. The calibration method makes use of an LCD panel as the calibration plane. In our experiments, re-projection error of 0.1 pixel was achieved in the camera calibration using only 8 images, which shows that uncertainty and re-projection error could be much reduced. The root mean square error in the planarity check was found to be only 0.04 mm, and an absolute mean deviation of only 0.077 mm or 0.01% of the depth range was obtained when measuring motion distance of 1 mm. Both indicate significant improvement over those of the traditional calibration setup. The result shows that by using the LCD panel as the calibration plane, the projector-camera system parameters can be calibrated in far less images with much higher accuracy. In fact, the programmability of the LCD panel in its pattern display makes the system more adaptable to other tasks and opens up many other applications.

Chapter 4

Image Feature Extraction: the Grid-points

On the image features for establishing correspondences between the projector's pattern panel and the camera's image plane, we propose to use rhombic pattern with binary (i.e., black and white) or colored elements as the projected pattern, and the grid-points between neighboring rhombic elements but not the centroids of the pattern elements themselves as the feature points. Grid-point in the pattern owns the two-fold rotation symmetry, or so-called cmm symmetry, which is largely preserved on the image side after perspective projection and imaging. We propose a grid-point detector that exploits such a symmetry. By avoiding the direct use of raw image intensity, the detector is less sensitive to image noise and surface texture. Comparison with traditional operators shows its promising robustness and accuracy.

4.1 Introduction

Feature extraction is a crucial step in any structured light systems. How precisely can the feature points be extracted from the image data determines directly the accuracy of 3D determination. In this thesis, we focused on a rhombic pattern consists of binary or colored elements. Moreover, the intersection points of each two adjacent pattern elements, which we refer to as the *grid-points*, are the feature points where 3D sampling is to take place. The 3D reconstruction accuracy of the system hinges much upon how precisely are the grid-points localized in the image domain.

As a classical problem in computer vision, a variety of feature detectors has been proposed, for example LoG [Marr 1980], Harris' [Harris 1988], SUSAN [Smith 1997], and SIFT [Lowe 2004] etc. However, most of them focused on the robustness rather than the precision of feature location.

In a projector-camera ranging system, the captured projection image is usually blurred and distorted. That may caused by the influence from the optical ability (MTF, depth of view, sensor noise etc.) and the surface property (texture, reflectance, and curvature etc.). As for the rhombic pattern we used, that will cause the adjacent rhombic elements disconnected in the image. It means there is often challenge in pinpointing where the grid-points are in the image data. Traditional corner detection algorithms such as LoG, Harris', or SUSAN cannot deal with this issue. Since the true object surface under operation can be a generic curved surface, the many corner extractors proposed for checkerboard pattern based calibration, which often assume ideal planarity of the pattern, are not applicable either. To extract the precise grid-point or corner locations in the image of a projected pattern, new mechanism is much needed.

In this chapter, we propose a new grid-point detector that exploits a certain symmetry the grid-points own – the two-fold rotation symmetry or *cmm* symmetry in the terminology of plane symmetry [Hermann 1989] – for boosting the localization precision. Such a symmetry in the pattern illumination is largely preserved in the image data after perspective projection [Gool 1996], and quasi-invariant with image noise, image blur, perspective distortion, and nonzero curvature of the illuminated surface in 3D. With a hypothesis step followed by a refinement step, the detection process can reach sub-pixel accuracy.

The chapter is organized as follows. Previous works on feature detectors are briefly reviewed in Section 4.2. In Section 4.3, the proposed grid-point detector is described. Robustness investigation and comparison between various feature detectors are given in Section 4.4. A summary is offered in Section 4.5.

4.2 Previous Work

Image feature detection is a classical problem for its importance to a large number of computer vision tasks including tracking, stereo vision, motion estimation, 3D

reconstruction, and object recognition etc. A good literature survey about image feature detectors can be found in [Basu 2002]. A feature point can be the point of intersection of two dominant edges of different orientations (i.e., a grid-point), a singular point with locally maximum or minimum intensity, the end of a line segment, or a point on a contour with locally maximum curvature. The quality of feature detector is often judged by its ability to detect the same feature in the images taken from different viewpoints and under different illumination condition [Schmid 2000]. Many feature extractors proposed in the literatures are based upon raw image intensities. While the use of direct intensities can minimize the computation effort, it also makes the extractors very sensitive to imaging noise, surface texture, reflectance, illuminations etc., and therefore compromises the accuracy in localizing the true feature positions in the image. A brief introduction about some popular used feature detectors is given in following section.

4.2.1 Laplacian of the Gaussian Operator

Laplacian of the Gaussian (or LoG) [Marr 1980] is one of the first and most common feature detectors. In the method, Gaussian function firstly is used to smooth the image. Then the second derivative of the smoothed image is taken; it has a zero crossing at the edge.

Given an input image $I(x, y)$, this image is convolved by a Gaussian kernel

$$L(x,y,\sigma) = I(x,y,\sigma) * g(x,y,\sigma) \tag{4.1}$$

where

$$g(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$$

and σ is the standard deviation of the Gaussian function. Then, the feature response can be defined as:

$$\nabla^2 L = L_{xx} + L_{yy} \tag{4.2}$$

LoG operator is an orientation-independent operator and its scale is given by σ. It breaks down at corners, curves, and at locations where image intensity function varies in a

nonlinear manner along an edge [Berzins 1984]. The using of second order derivative also makes it quite sensitive to image noise.

4.2.2 Harris' Corner Detector

It was proposed by Harris and Stephens [Harris 1988], which in turn was an improvement of a method proposed by Moravec [Moravec 1980]. Instead of using shifted windows, Harris and Stephens improved upon Moravec's corner detector by considering the differential of the corner score with respect to local image directions. In the detector, an auto-correlation matrix **A** is defined as:

$$\mathbf{A} = \begin{bmatrix} I_x I_x & I_x I_y \\ I_x I_y & I_y I_y \end{bmatrix} \tag{4.3}$$

where I_x, I_y indicate the intensity derivatives with respect to the two dimensions of the image domain respectively. A feature point or corner is detected at image position (x, y) if there the matrix **A** has two large, distinct positive eigenvalues (λ_1 and λ_2). The use of intensity derivatives however makes the operator sensitive to noise and compromises the location precision.

4.2.3 SUSAN Corner Detector

In the SUSAN (Smallest Univalue Segment Assimilating Nucleus) corner detector [Smith 1997], a circular mask M is selected over the nucleus. And a pixel in this mask is represented by $\vec{m} \in M$. The nucleus is at $\vec{m}_0$. Every pixel is compared to the nucleus using the comparison function $c(\vec{m})$ with form:

$$c(\vec{m}) = e^{\frac{(I(\vec{m}) - I(\vec{m}_0))^6}{t}} \tag{4.4}$$

where t determines the width, and the power of the exponent has been determined empirically. The area of the USAN is given by:

$$n(M) = \sum_{\vec{m} \in M} c(\vec{m}) \tag{4.5}$$

The number of pixels within the circular mask, which have similar brightness to that of the nucleus, is calculated.

If c is the rectangular function, then n is the number of pixels in the mask, which are within t of the nucleus. The response of the SUSAN operator is given by:

$$R(M) = \begin{cases} g - n(M) & if\,(n(M) < g) \\ 0 & otherwise \end{cases} \tag{4.6}$$

where g is named the `geometric threshold'. The value of g determines the minimum size of the univalue segment. If g is large enough, then this becomes an edge detector. For corner detection, two further steps are needed. Firstly, the centroid of the USAN is found. A proper corner will have the centroid far from the nucleus. The second step insists that all points on the line from the nucleus through the centroid out to the edge of the mask are in the USAN. As the operator is not based upon image derivatives, it is more robust to noise and rather fast.

4.2.4 Spline Fitting Detector

There are also some methods that make use of spline fitting [Medioni 1987] or image contours [Shi 1994]. Points on curve or contour with large curvatures are selected as the feature points. An earlier approach to corner detection is to detect points where the curvature of level curves and the gradient magnitude are simultaneously high. A differential way to detect such points is to compute the rescaled level curve curvature (the product of the level curve curvature and the gradient magnitude raised to the power of three) and to detect positive maxima and negative minima of this differential expression at some scale t in the scale-space representation L of the original image.

$$\widetilde{k}(x, y; t) = L_x^2 L_{yy} + L_y^2 L_{xx} - 2L_x L_y L_{xy} \tag{4.7}$$

A main problem with this approach, however, is that it is sensitive to noise and to the choice of the scale level. Moreover, the extraction of parametric entities like parametric curves etc. from the image is still a challenging issue.

4.2.5 Specialized Grid-point Detector

Grid-point detection has also been studied in camera calibration, which often requires the use of pattern like checkerboard patterns in the calibration object. Many approaches have been proposed. One popular approach mean is to find edges in the image and fit lines to them. Grid-points are then localized as the intersection points of these lines. In the method proposed by Shu [Shu 2003], the processing begins by finding corners in the image using Harris' corner detector. Then the grid-points are connected via *Delaunay* triangulation. Pairs of neighboring triangles are combined into quadrilaterals and then a topological filter is used to find regular grids. The camera calibration toolbox for *MATLAB* also provides the function to extract grid-points. Some input parameters and manual selections are required [Bouguet 2007].

A 2D random pattern named three-level checkerboard pattern was proposed in [Ito 1995]. The grid-point is encoded by the grey level of its adjacent pattern elements. Simple thresholding is used to extract the grid-points. That makes the system unstable and inaccurate. Maximal error of 2 mm was obtained at the working distance of 800 mm. In [Pagès 2003], some vertical and horizontal lines are encoded by *De Bruijn* sequence. The intersection points of these lines are defined as the feature points. Some simple templates are designed to detect these points in the grid. Similar methods are also reported in [Vuylsteke 1990]. A pseudorandom color pattern with circular elements is used in [Morano 1998]. Different from those of the regular random pattern, its codewords are generated by iterating between adding one new code word and checking it against all preceding ones. The feature points are defined as the centroids of the circular elements. No quantitative experimental results are provided yet.

4.3 *cmm*-Symmetry based Grid-point Detector

Under influence of the projector-camera system's optical ability (MTF, depth of view, sensor noise etc.) and of the surface property (texture, reflectance, curvature etc.), adjacent rhombic elements in the pattern are often disconnected in the image as shown in Figure 4.1. It means there is a challenge in pinpointing where the grid-points are in the image data. Traditional corner detection algorithms such as Harris' or SUSAN cannot deal with this issue. Since the true object surface under operation can be a generic curved

surface, the many corner extractors proposed for pattern-based calibration, which often assume ideal planarity of the patterns, are not applicable either. To extract the precise grid-point or corner locations in the image of a projected pattern, new mechanism is much needed.

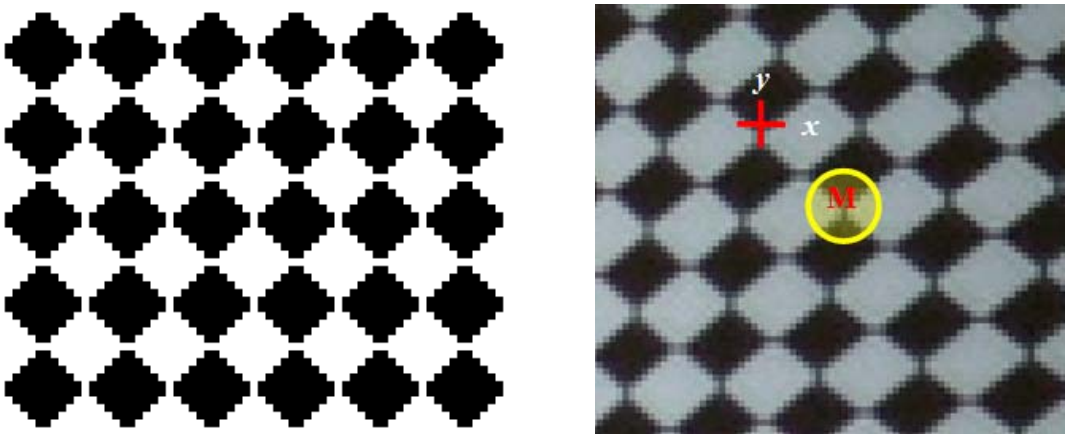


Figure 4.1 Pattern with rhombic elements. (Left) The grid-point can be perfect located in the original pattern image. (Right) After projection, the grid-points are invisible.

It can be observed from Figure 4.1 that any grid-point on the pattern generation plane of the projector has a certain local symmetry: a small circular disc of intensities centered at the grid-point resembles itself when rotated by 180°. We refer to the symmetry as the *cmm* symmetry. Such a symmetry on the illumination side is largely preserved in the image data, and is quasi-invariant with image noise, image blur, and perspective distortion for the reason that linearity of the segment that divides the circular disc into two symmetrical halves is preserved under the processes. If the circular disc is sufficiently small, even nonzero-curvature of the illuminated surface in 3D has little effect to the symmetry. This quasi-invariance of the symmetry is what we exploit for precise grid-point detection from the image.

To measure the *cmm* symmetry at any candidate position of grid-point in the image plane, we first crop out a local small circular disc from the image at the candidate position. The circular disc of colored intensities is then transformed to a binary disc with the description of each element reduced to either a foreground element or a background element. The correlation coefficient between such a disc and its 180° rotation is defined

as the measure of the *cmm* symmetry at the candidate position. Positions with this measure being strong and a local maximum are then recorded as the initial candidates of grid-points. A refinement step is then performed to determine the grid-point positions with sub-pixel accuracy.

However, before conducting the above process we have a coarse filtering step first that applies simple masking to the image to hypothesize candidate positions of the grid-points. This step is meant to speed up the entire process by letting the above process be applied not to every image position but only to positions with high chance to be the true grid-points. Below we elaborate in more details the two-step process of the grid-point detector.

4.3.1 Hypothesis of Grid-point Positions

A grid-point in the proposed pattern is characterized by the presence of, in two orthogonal directions about the point, neighboring regions of drastically different intensities. We use this property to hypothesize candidate positions of grid-points in the image. The color image is first transformed into a grey level one. A template in the shape of a cross as shown in Figure 4.1 is then used for convolution with the transformed image $I(x, y)$. The absolute difference d with following form is defined as the response value of the mask at position (x, y).

$$d = \left| \sum_{i=-\varepsilon}^{\varepsilon} I(x+i, y) - \sum_{j=-\varepsilon}^{\varepsilon} I(x, y+j) \right| \tag{4.8}$$

where ε indicates the size of template. The measures of d indicates the difference of image intensities accumulations along the x and y directions at the position. The image positions with high response values are kept as candidates of grid-points.

Such a template-based scheme is simple and efficient. However, the scheme merely considers raw image intensities, and that compromises its accuracy and robustness against image noise. It is used here only as a preliminary step to hypothesize possible grid-point positions for speeding up the process.

4.3.2 Confirmation of Grid-points and Position Refinement

Two-fold rotation symmetry is displayed at positions of true grid-points. This can be used for confirmation of the grid-point features as well as for their precise localization with sub-pixel accuracy.

To measure the strength of *cmm* symmetry at any candidate position, we use the correlation coefficient between a small circular window of the position and its 180° rotation. Since image intensities are in general normally distributed, the *Pearson's Product-Moment Correlation Coefficient* (or *PMCC*) [Cohen 1988] is adopted in our method.

Suppose M_C is a circular mask centered at any candidate image point *C*, and M_C' the mask created by rotating M_C by 180° around *C*. By the definition of *PMCC*, we can write the correlation coefficient ρ_C as:

$$\rho_C = \frac{\sum_{i=1}^{n}(M_{Ci} - \overline{M}_C)(M_{Ci}' - \overline{M}_C')}{\sqrt{\sum_{i=1}^{n}(M_{Ci} - \overline{M}_C)^2 \sum_{i=1}^{n}(M_{Ci}' - \overline{M}_C')^2}} \tag{4.9}$$

where *i* here refers to individual element of mask M_C or M_C', *n* is the size of the masks, M_{Ci}, M_{Ci}' indicate image intensities of the *i*th elements of the M_C and M_C' masks respectively, and $\overline{M}_C, \overline{M}_C'$ are the means of intensity distributions of the two masks. The above equation can be replaced by an equivalent formula, which avoids using the means and is therefore faster:

$$\rho_C = \frac{n\sum_{i=1}^{n} M_{Ci} M_{Ci}' - \sum_{i=1}^{n} M_{Ci} \sum_{i=1}^{n} M_{Ci}'}{\sqrt{n\sum_{i=1}^{n} M_{Ci}^{\ 2} - (\sum_{i=1}^{n} M_{Ci})^2}\sqrt{n\sum_{i=1}^{n} M_{Ci}'^2 - (\sum_{i=1}^{n} M_{Ci}')^2}} \tag{4.10}$$

The operator is only applied to image points *C* that pass the pre-processing step. The response values at these points are calculated, and a map named the response image ρ can be generated from the values as shown in Figure 4.2. A threshold *t* is used on the ρ-map to extract image points with the ρ value being a local maximum. A small region of size

say 3×3 pixels is selected around every such point. The final grid position (x_g, y_g) at the vicinity of every such point C is then computed with sub-pixel accuracy as the weighted average of all positions in this 3×3 region, with the weight being the ρ value ρ_i of each ith position (x_{Ci}, y_{Ci}) in the region:

$$(x_g, y_g) = \sum_{i=1}^{9} \rho_i . (x_{Ci}, y_{Ci}) \Big/ \sum_{i=1}^{9} \rho_i \tag{4.11}$$

Since structure rather than raw intensity mask of grid-point is used, the described method has higher robustness than intensity-based methods against image noise, blur, surface texture, curvature, and perspective distortion.

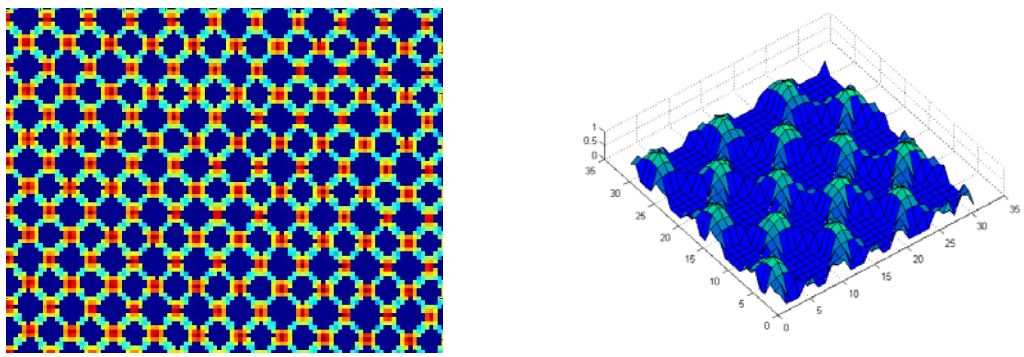


Figure 4.2 Response value of the proposed operator on a real image.

In summary, the grid-point detection procedure consists of the following steps:

a. *Pre-process the image by convolving it with the cross template to hypothesize candidate grid-point positions.*

b. *At every hypothesized grid-point position C, a circular window M_C of radius r is cropped out from the image for the subsequent refinement step.*

c. *M_C is rotated about C by 180° to form M_C', which is supposedly the counterpart of M_C in the cmm-symmetry at a true grid-point.*

d. *Compute correlation coefficient ρ_C between M_C and M_C' for every candidate grid-point position C.*

e. *Among all the candidate grid-point positions C, select initial grid-points from the image where ρ_C is locally maximum.*

f. *Apply the weighted averaging step to the above positions to extract the grid-point positions with sub-pixel accuracy.*

While the first hypothesis step suggests the more probable image positions for consideration using a template based scheme that is efficient, the second refinement step exploits the quasi-invariant *cmm*-symmetry property of the true grid-points for their confirmation, and further localize them with sub-pixel accuracy.

4.4 Experimental Results

A structured light system consisting of a DLP projector and a digital camera was used for experiments. The system was first calibrated using the method described in Chapter 3. The pattern shown in Figure 4.1 was projected through the DLP projector. The pattern element had a size of 11×11 pixels on the projector's display panel. The image resolution of the camera was set to 1500×1000 pixels. The size of the captured pattern element varied from 12 to 20 pixels in the image plane according to the surface property. Real image data with extra noise were experimented to test its anti-noise ability. The result is also compared with traditional methods such as template, Harris' and SUSAN methods. A rhombic color pattern is used in the second experiment. The grid-point detection result is also evaluated on a curved surface.

4.4.1 Robustness Check under Extra Gaussian Noise

To evaluate the robustness of the proposed grid-point detector, zero-mean Gaussian noise with $\sigma = 0.01$ was added to the image as shown in Figure 4.3 (b), (d) and (f). From the results, we can see that, even without extra noise, the Harris's and SUSAN detectors had trouble extracting all the grid-points correctly. Even the one successfully extracted had rather poor localization accuracy. With the extra noise added, the results were unbearable.

Using the proposed detector, all the grid-points could be exacted correctly with visually reasonable localization accuracy. Even with the extra noise added, the grid-points could still be correctly detected and the accuracy was only slightly affected. Experiments show that the new operator is effective, accurate, and robust against noise. Extensive experiments show that the new detector outperforms the traditional methods by a large margin.

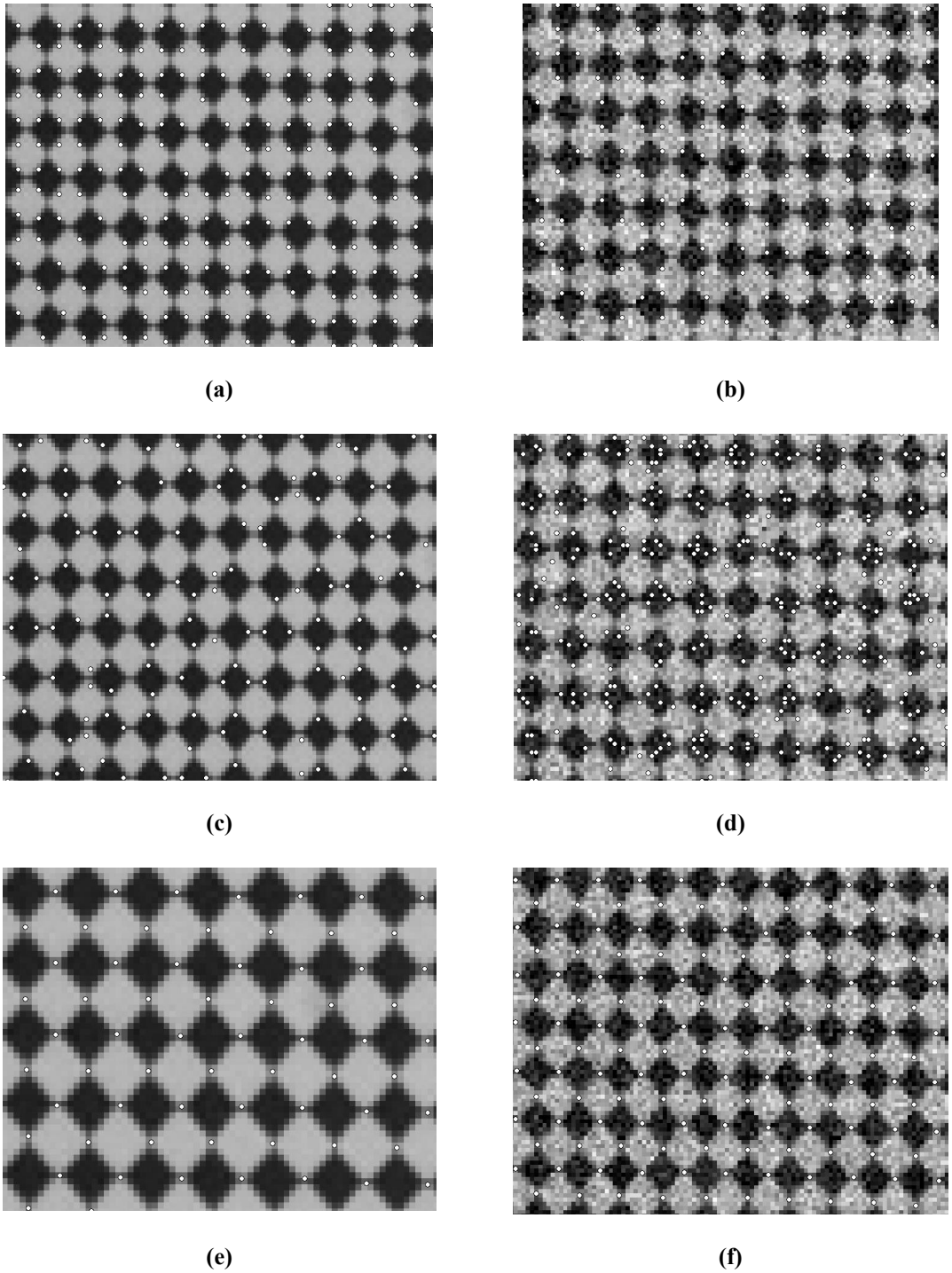


Figure 4.3 Grid-point extraction results of three operators. (a) Result from Harris' corner detector. (b) Harris' corner detector on noisy image. (c) Result from SUSAN corner detector. (d) SUSAN corner detector on noisy image. (e) Result from the proposed corner detector. (f) The proposed corner detector on noisy image.

4.4.2 Color Pattern under Extra Gaussian Noise

A rhombic color pattern is used in this experiment. The color pattern was projected onto a planar board and extra Gaussian noise with σ=0.03 is added to test its robustness when deal with color image. Figure 4.4 (a) shows grid-point detection result by the template method. Figure 4.4 (b) shows the result by the proposed method.

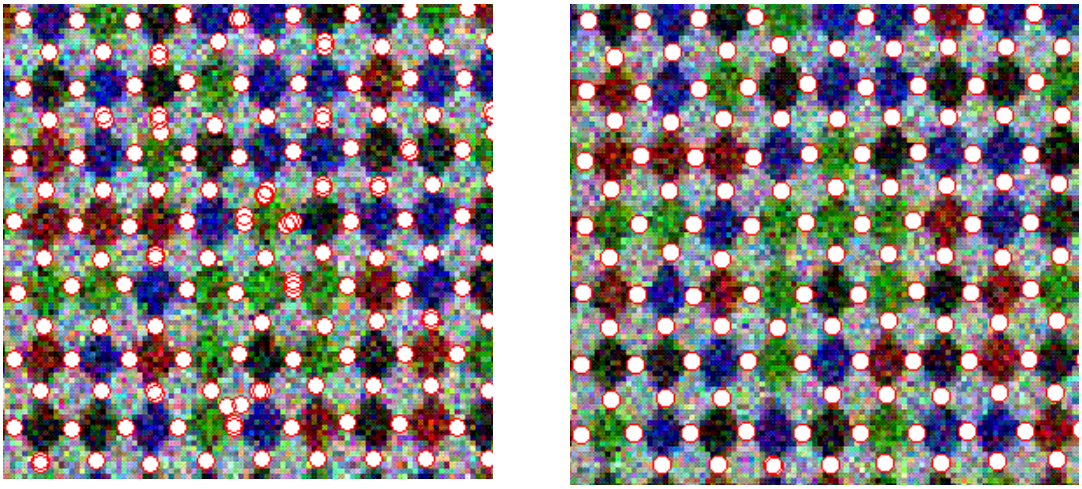
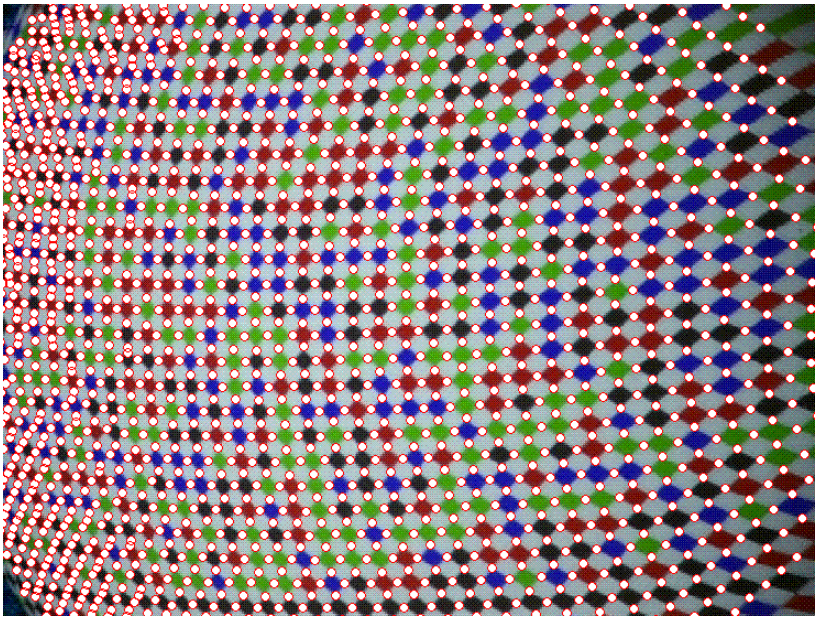


Figure 4.4 Grid-point detection results on a color pattern image with extra Gaussian noise (σ = 0.03). (a) Result from the template method. (b) Result by the proposed detector.

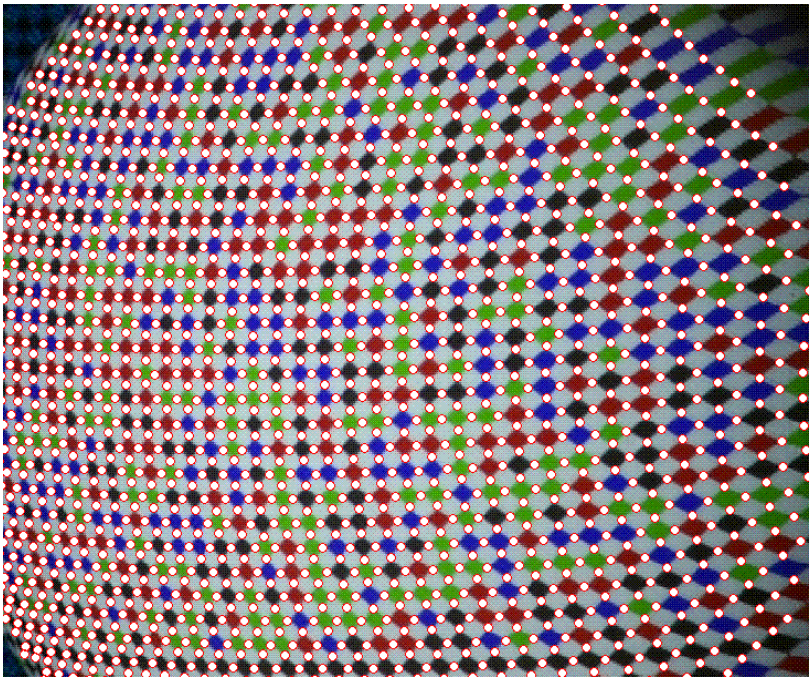
4.4.3 Experiment with Spherical Surface

A spherical object is used to test the performance of the proposed grid-point detector on curve surface. As shown in Figure 4.5, both the template and the proposed operator can detect all the grid-points correctly in the planar region. But the template method failed to extract the grid-points on the left side, where the density of grid-point is increasing. Many redundant feature points are detected as shown in Figure 4.5 (a).

With comparison, result from the proposed method is much better, even there is big distortion or the grid-points' density increased as shown in Figure 4.5 (b). Quantitative evaluation about its location accuracy can be found in Chapter 6.



(a)



(b)

Figure 4.5 Grid-point detection results on a spherical surface. (a) Result from the template method. (b) Result by the proposed grid-point detector.

4.5 Summary

Feature extraction is a crucial step in a structured light system. It directly decides the final accuracy of 3D reconstruction. In this chapter, an accurate and robust grid-point detection algorithm is proposed. The detector utilizes the *cmm* symmetry, which can be preserved in a local region around the grid-point in perspective projection. Correlation coefficient between the mask and its rotational mask is defined as the response value. A template method is first introduced to remove most false grid-points and in turn to speed up the whole procedure. Afterwards, a refining step is performed to obtain the final grid-point positions with sub-pixel accuracy. The operator uses structure not raw image intensities, and is thus more robust to image noise, surface texture, and perspective projection etc. than previous intensity-based methods. Experiment on noisy image shows that the operator has good localization accuracy and robustness. Experiments on colored patterns and curved surface show its good applicability.

Chapter 5

Coding over Time: Gray Code plus Strip Shifting Mechanism

This chapter describes a projector-camera system incorporates a strip shifting mechanism into the Gray code structured light system for encoding the illumination. The system is calibrated with the novel calibration means as described in Chapter 3. The results show that, with the proposed design of calibration and codification, even a structured light system made of household-quality projector and camera can achieve accurate system calibration in far less image-capture steps, as well as high-quality 3D reconstruction. Performance of the 3D reconstruction mechanisms is presented in comparison with those of the traditional systems.

5.1 Introduction

Structured light system is an effective 3D reconstruction means. In the most basic form it consists of only a projector and a camera - equipments that are today only household appliances. Its essence lies in the use of coded illumination from the projector to the target scene, which allows feature correspondences over the imaged scene to be readily established between the projector's pattern generation plane (i.e., the plane of the LCD array or pattern grating in the projector) and the camera's image plane, thereby making 3D reconstruction possible. With the coded illumination, not only is the demand of

texture on the target shape for 3D reconstruction relieved, the difficult correspondence problem often encountered in other visual methods like stereo vision is also avoided.

Temporal and spatial coding schemes are two major approaches of structuring the illumination, both with the same purpose of equipping each position in the projector's pattern generation plane with a unique code that is preserved under light projection from the projector, reflection by a scene point in 3D, and image captured by a camera. With the preservation and uniqueness of the codeword, correspondences can be readily established between positions in the projector's pattern generation plane and positions in the camera's image plane, and depth measures of the associated scene points can be reconstructed via the well-known triangulation process. More specifically, in the temporal coding scheme, a series of patterns are projected at different instants onto the target surface. The codeword for any given pixel on the camera's image plane is decided by the illumination from the projector for that pixel over time; in other words, the bits of the codeword are multiplexed in time. Structured light systems using this encoding scheme can adopt strong data density and achieve high accuracy in the measurements. Examples of temporal coding methods are Gray code [Valkenburg 1998], and Gray code combined with phase shifting [Krattenthaler 1994] or line shifting methods [Gühring 2000] etc. In the spatial coding scheme, codeword is constructed not over time but from neighborhood in the pattern. The codeword can be a window of grey level intensities, colors, or geometry primitives etc. The performance of the scheme is however sensitive to the variation of surface reflectance of the target shape and the presence of texture on it. Compared with the temporal coding methods, larger depth measurement errors could arise, especially if there are heavy occlusions and color distortions in the scene. However, the spatial coding scheme has the advantage that the encoding and decoding processes can be achieved with a single image capture. It is therefore particularly suitable for dynamic applications. *De Bruijn* sequences [Salvi 1998], M-arrays method [Vuylsteke 1990] are two spatial coding methods widely employed.

This chapter is organized as follows. In Section 5.2, a proposed Gray code encoding system with the strip shifting pattern incorporated is introduced. Section 5.3 presents extensive experimental results that evaluate the calibration and reconstruction performance. Concluding remarks and possible future work are offered in Section 5.4.

5.2 Strip Shifting Combined with Gray Code Pattern

Gray code is an encoding scheme widely used for its simplicity and robustness. If a binary Gray code pattern of codeword length n is to be used, an image sequence consisting of n+1 binary strip patterns need to be projected sequentially to encode the scene. With that, the scene image can be separated into 2^n distinguishable regions each with a unique different codeword, and the image positions are encoded with only pixel accuracy. Phase shifting and line shifting are two techniques often used with the Gray code scheme to achieve sub-pixel accuracy. In theses methods, a serial Gray code patterns are projected onto the target surface firstly. As a result, the image of the surface can be divided into some labeled sub-regions, as described above. Then sinusoidal intensity patterns or line patterns will be projected to encode different segments of each sub-region in the finer scale. The sinusoidal patterns are usually implemented as a number of grey levels. For example, a total of four Gray code patterns, with each of them offset in the phase from the previous one by $\pi/2$ of the period, are projected to the target surface. This way, for every given pixel (x, y) of the camera image, four intensity values: I_1, I_2, I_3, I_4 are obtained. Then, the phase value of pixel $I(x, y)$ can be determined as:

$$\phi(x,y) = \tan^{-1}\left[\frac{I_4(x,y) - I_2(x,y)}{I_1(x,y) - I_3(x,y)}\right] \tag{5.1}$$

For each image pixel, the phase value of it will add to the Gray codeword of the sub-region it belongs to, and form a global unique codeword for the pixel. This way, reconstruction finer than the sub-regions can be achieved. However, the phase shifting method as a method involving analog intensity values is more prone to the disturbance from imaging noise and surface texture. In contrast, the line shifting method encodes each sub-region with line features instead of intensity. That makes it more robust against image noise. However, the accuracy of line shifting method is constrained by the facility (like resolution and contrast) of the projector and the line detection accuracy in image processing. It would also encounter difficulty with target surfaces with reflectivity variation and texture.

5.2.1 Gray Code used with Strip Shifting Pattern

In our system, strip shifting is used in place of the line shifting strategy. Edges of binary strips (alternate dark and bright strips) are easier to detect in image data than lines. With edges more precisely located, higher reconstruction accuracy and stronger robustness can be attained.

A series of n Gray code patterns are first projected in order to divide the target surface into 2^n sub-regions, each with a unique n-bit long Gray codeword. Suppose that each of such sub-regions is m pixel wide on the projector's pattern generation plane. Then a strip pattern in half of the finest strip width of the Gray code sequence is shifted m-1 times, each with a separate image capture by the camera, to encode each sub-region with additional bits. Such a periodic pattern by itself has periodic ambiguity of m projector pixels in distinguishing different points of the target surface. However, by combining the two codewords together – one from the Gray code projection and the other from the strip shifting pattern – the m sub-divisions can be introduced to each of the 2^n sub-regions to achieve finer 3D reconstruction with sub-pixel accuracy. The procedure can be expressed as:

$$G = 2^n,\ S = m,;\ P = G + S \tag{5.2}$$

where G is the Gray codeword, S is the local codeword generated by the strip shifting pattern and P is the final unique codeword.

For example, suppose that the projector's pattern generation plane is of width $1024=2^{10}$ pixels. Gray codeword of 8 bit length (i.e., n=8) can separate the projector's pattern generation plane into 2^8 sub-regions. Then, the a strip pattern of width 4 pixels is used as the strip shifting pattern and shifted 3 times in a step length of 1 pixel as illustrated by Figure 5.1. Upon the shifting, each column of the projector's pattern generation plane will be attached with a unique 10-bit codeword. While the 10-bit long gray code pattern allows precision up to only pixels, this mechanism allows sub-pixel precision to be reached.

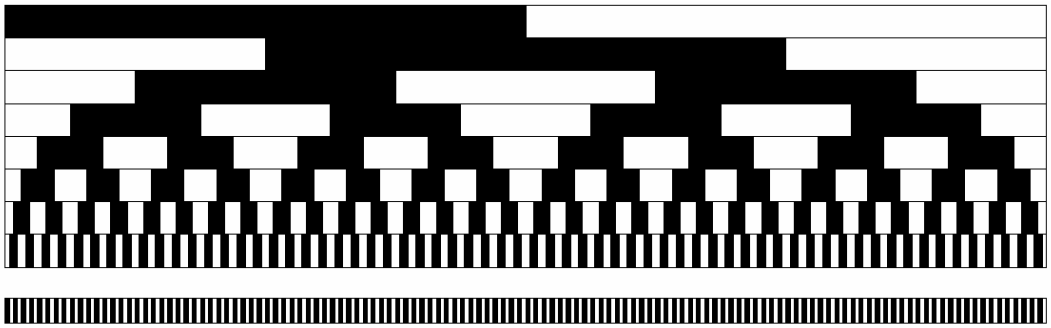


Figure 5.1 The use of Gray code together with strip shifting for 3D reconstruction. (Top) A series of 8 Gray code patterns are used to construct 2^8 sub-regions each with a unique 8-bit long codeword. (Bottom) A strip pattern of the width 4 pixels is shifted 3 times to encode positions within each sub-region.

5.2.2 Strip Edge Detection

Zero-crossing method, which based up on the second derivative of the image intensity is a widely used way of detecting edge location with sub-pixel accuracy. The principle is simple and can be reduced to a masking process. However, the method is sensitive to image noise and its performance could be affected by surface texture. An alternative method is to project both the regular and the reverse patterns (i.e. positive and negative patterns). The intersection of the positive and negative strip edge profiles could pinpoint strip's edge location with sub-pixel accuracy as shown in Figure 5.2. Experimental results show that a precision of 0.1 pixel in locating edge can be obtained with either method.

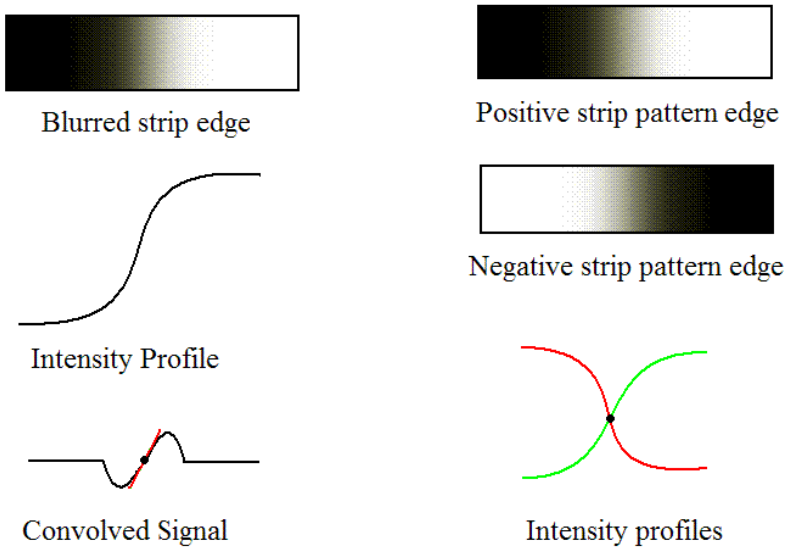


Figure 5.2 (Left) Principle of zero-crossing edge detection. (Right) Sub-pixel edge detection from two complementary patterns.

5.3 Experimental Results

Our experimental system consisted of a DLP projector of resolution 1024×768 (*NEC LT25* projector) and a camera of resolution 1500×1200 (*Pentax K100D* DSLR camera with *Pentax DA* 18-55 mm lens), both being only consumer-grade equipments. The focal length of the projector was of range 25-31mm. The system was configured for a working distance of about 850 mm. The total cost of the two instruments was less than *USD 1, 500*. A 15' LCD of consumer grade was used for system calibration.

In this section, we present the performance of such an affordable structured light system in three measures. The first measure is calibration accuracy, which is about how accurate the calibration of the system parameters under the proposed calibration scheme could be. Ground truth of the camera and projector parameters are generally inaccessible for direct accuracy evaluation, so we used the re-projection error for evaluation, which is about how close the projections of 3D data through the calibrated system parameters are to the observed image positions. The second measure is metric reconstruction accuracy, which is about how accurately metric reconstruction can be achieved under the described

ranging system. The third measure is 3D reconstruction quality for generic object, which is about a visual check if how much detail of free-form object can be extracted using the proposed ranging system.

5.3.1 Calibration Accuracy

A total of 8 pairs of images were used under the proposed calibration design in Chapter 3 to calibrate the projector-camera system. The result shows that with the use of the LCD panel, uncertainty could be reduced by 5-10 times even with less than half of the images (8 pairs vs. 20 pairs) used in traditional method. More specifically, the *RPE* was about 0.09 pixel only with the use of 8 calibration image pairs under the proposed method. The *RPE* in projector calibration could be controlled to within 0.5 pixel as displayed in Figure 5.3. Compared with the *RPE* of 5 pixels in the projector calibration by using printed calibration pattern, the improvement was more than 10 times.

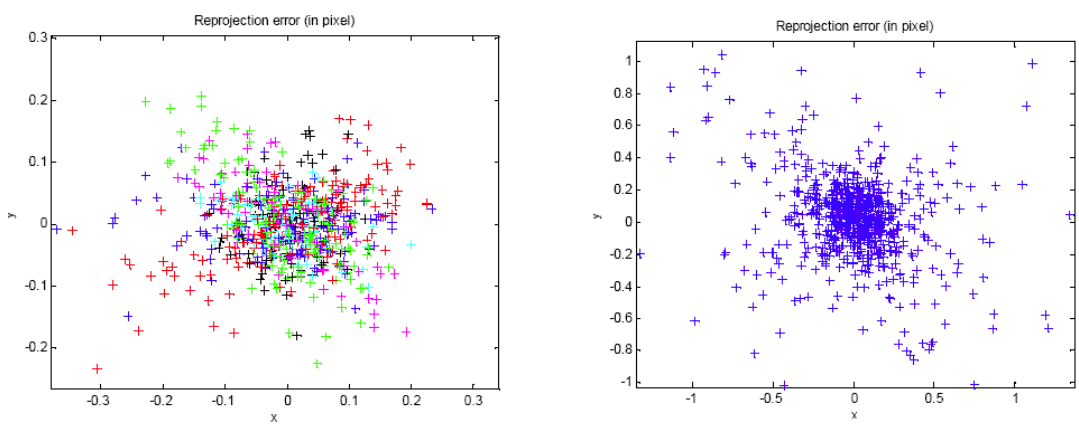


Figure 5.3 ***RPE*** **distribution of camera (left) and projector (right) under the calibration results.**

5.3.2 Metric Reconstruction Accuracy

In this experiment, a planar board and a plaster ball were reconstructed to evaluate the metric reconstruction accuracy of our ranging system. Gray code patterns were projected onto a planar board from a distance of about 850 mm. From the recovered depth map a region about 100×100 mm large on the plane was cropped out. A plane was fitted to the 3D data in the least square sense, and the planarity error in the fitting process was noted.

An absolute mean error of 0.11 mm, with a standard deviation of only 0.08 mm, was obtained.

Figure 5.4 shows the reconstruction of a plaster ball. The background points and noisy points were first removed by a simple filtering process. To measure the reconstruction accuracy, a sphere was fitted in the least square error sense to the reconstructed data points. The radius of the fitted sphere was 96.90 mm with an average residue fitting error of 0.15 mm or 0.02%. The accuracy is satisfactory despite the imperfect target ball. A phase shifting pattern was proposed in [Sadlo 2005]. The maximal error of 0.5 mm was reported while reconstructing a steel ball with diameter of 150±0.1 mm. The high accuracy of the proposed system owes both to the reconstruction method and to the accurate calibration means.

Without changing the configuration (working distance, image resolution etc.) of the system, a traditional Gray code pattern was also conducted for comparison. In the traditional Gray code system, a total of 11 pattern images are projected sequentially. As a result, each image point can be encoded with a codeword up to 2^{10}. In the sphere reconstruction, an absolute mean error of 0.32 mm was obtained. For reference, a planarity error of 0.3 mm was reported in [Rocchini 2001], which also utilized the traditional Gray code pattern. Figure 5.5 (c) shows its reconstruction result on a bust model. The ripples on the reconstructed surface are caused by the low resolution of the Gray code system. Since it can only reach accuracy with pixel level.

5.3.3 Reconstruction of Free-form Shapes

A bust was scanned with the proposed patterns firstly. The encoded image with and without periodic ambiguities between sub-regions were generated as shown in Figure 5.5 (a). In this experiment, more than 300,000 points were reconstructed. The data were processed and rendered by off-the-shelf CAD/CG software to get a 3D model as shown in Figure 5.5 (d). Though ground truth of the shape in 3D was not available, visual examination showed that even the tiny scratches (depth variations over those scratches were less than 0.1 mm approximately) on the bust could be reconstructed well. The whole process took only seconds on a Pentium-4 computer.

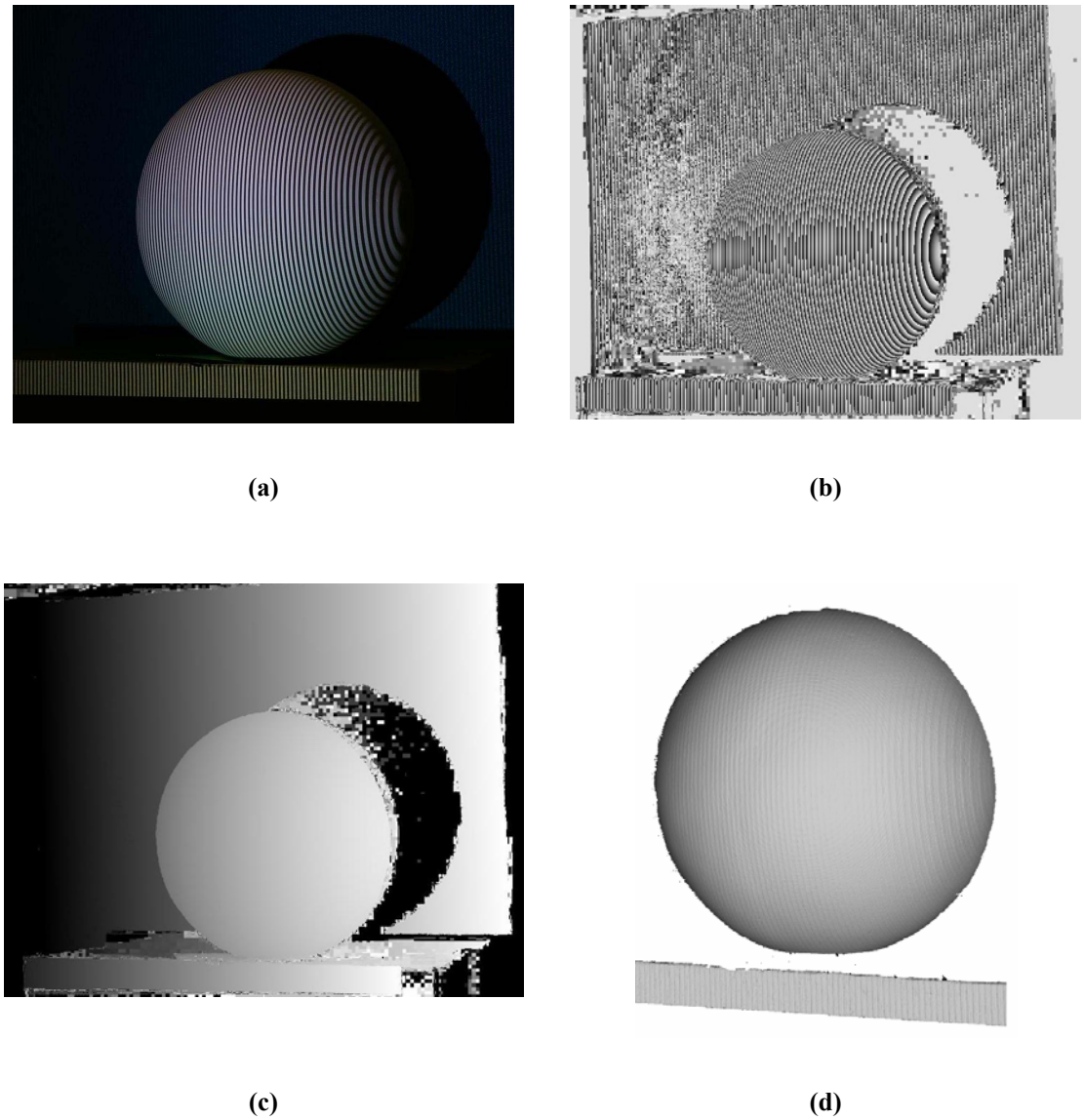


Figure 5.4 Metric reconstruction quality evaluation experiment with a spherical object. (a) A ball imaged under strip pattern projection. (b) Encoded image with a periodic ambiguity of 8 pixels. (c) Final encoded image after ambiguity removal. (d) Reconstructed 3D model with the proposed pattern design.

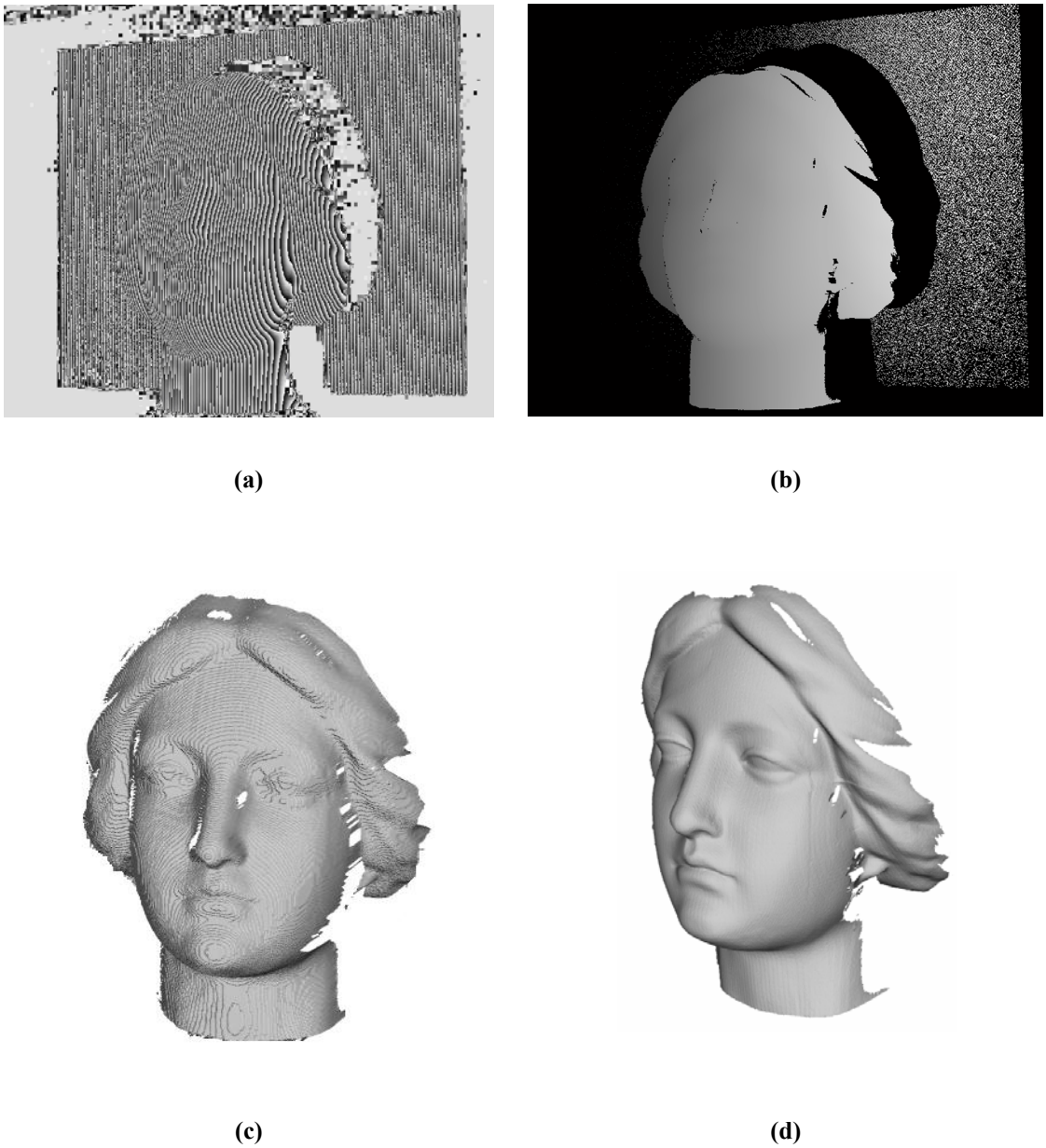


Figure 5.5 Quality evaluation of the reconstruction of a bust model. (a) Encoded image with a periodic ambiguity of 8 pixels. (b) Final encoded image after ambiguity removal. (c) Reconstructed 3D model without using strip pattern. (d) Reconstruction result with the improved pattern design.

5.3.4 Reconstruction of Texture and Reflective Object

A ceramic teapot with strong color and texture variations was also reconstructed, and the result is displayed in Figure 5.6 as well. Compared with the plaster model, the teapot had plentiful texture and color, and was reflective. In this particular experiment, the method of projecting both positive and negative patterns was used to improve the edge detection accuracy. As shown in Figure 5.6 (d), even minute details on the teapot surface were reconstructed clearly.

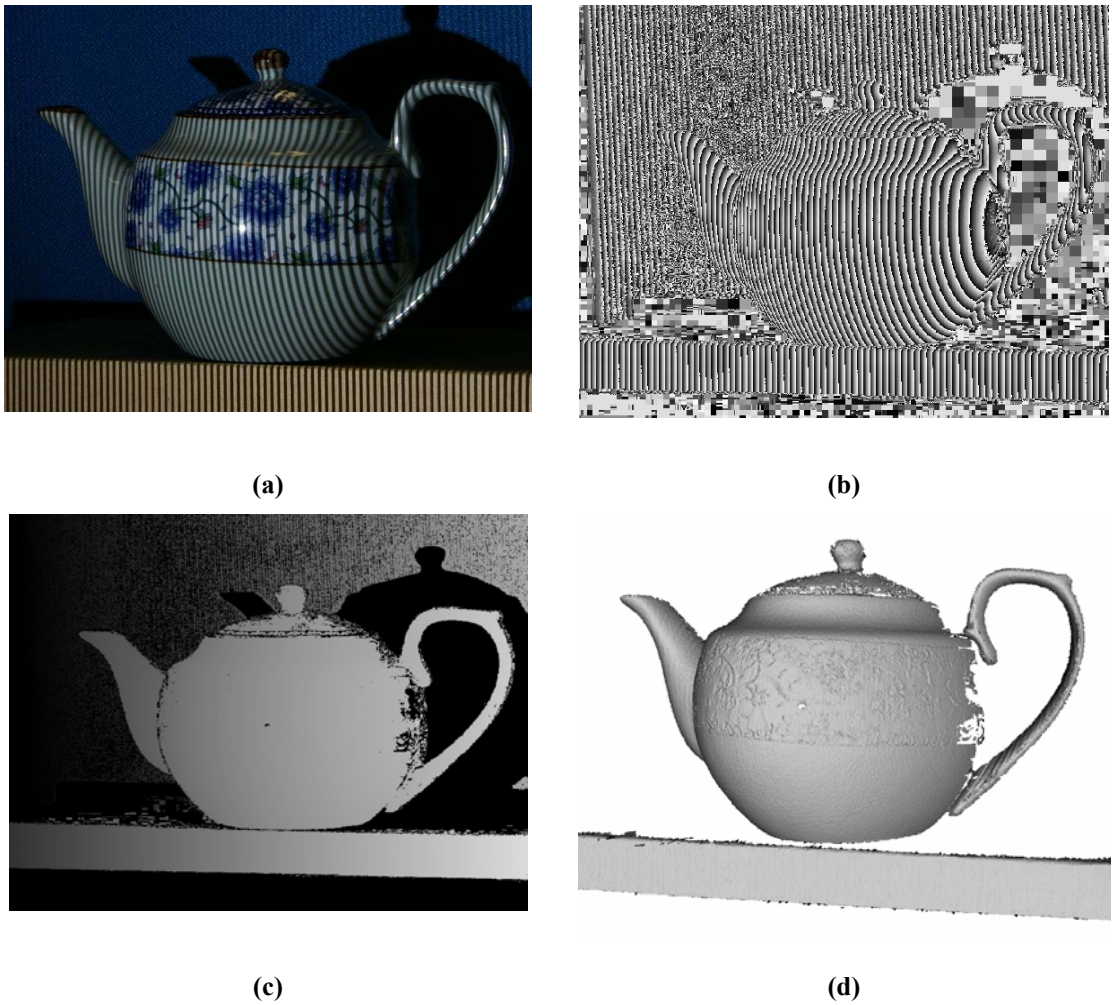


Figure 5.6 Quality evaluation of a ceramic teapot which have plentiful texture and reflective. (a) Object imaged under strip pattern projection. (b) Encoded image with a periodic ambiguity of 8 pixels. (c) Final encoded image after ambiguity removal. (d) Reconstructed 3D model with the improved pattern design.

5.4 Summary

In this chapter, we have described an affordable, convenient, yet accurate structured light system composed of only consumer-grade projector and camera. In terms of operations, there are two major ones: system calibration and 3D reconstruction. We have shown that with proposed LCD pattern, the system can be calibrated with far less images than the traditional calibration scheme, yet with more accurate calibration result. The 3D reconstruction operation is the traditional Gray code pattern based, but augmented by the use of strip shifting to enhance the resolution of 3D reconstruction. Specifically, while strip shifting allows high-resolution pattern information to be projected to the target scene or object, the Gray code projection removes the periodic ambiguity in the pattern. We also provide a way of allowing the strip edges to be located with sub-pixel accuracy, thereby raising the reconstruction quality. All these collectively contribute to the attractive features of the system. Extensive experimentation on calibration quality alone, on the reconstruction of some known 3D shapes, and on shape determination of some free-form shapes show that the system is of high quality though low construction cost.

Chapter 6

Coding over Spatial Domain: Color Coding plus the use of Grid-points as Image Features

Coding can also be established over the spatial domain. This corresponds to the use of color coded structured light in the literature. The advantage of such a system is that only one single image is necessary for 3D reconstruction. That makes it particularly suitable for real-time operation on objects that could move fast. We adjust a pseudorandom color pattern to one with rhombic pattern elements, and we use grid-points between the pattern elements, rather than the centroids of the pattern elements, as the image features. We show that for the same projected pattern, the use of grid-points rather than pattern element centroids as used traditionally can double the number of feature points with higher location precision. In addition, we show that not only are the grid-points useful, the grid-lines among the pattern elements can also be used to derive surface orientation information about the imaged object. The 3D position information as determined from the grid-points, and the 3D orientation as determined from the grid-lines, are information independently recovered, and they together would allow the 3D shape of the imaged object to be revealed in more complete form. Extensive experiments are provided to show the capability and accuracy of the mechanism.

6.1 Introduction

Temporal and spatial coding schemes are two main approaches of structuring the illumination. Compared with temporal coding scheme, fewer images are needed in the spatial coding scheme. It is therefore particularly suitable for use in dynamic applications. In spatial coding scheme, only the depth value of feature points, which have been encoded, can be reconstructed. Its application is often limited by the lower reconstruction data density. Surface orientation or surface normal is information complementary to 3D position in describing shape. Yet in the literature, orientation information is often generated from surface points, but then that way its accuracy is determined by the determination accuracy and density of the surface points. Although the absolute accuracy of reconstructed model can be high, even a small amount of noise in measured positions can cause large errors when surface normals are computed. More important, it does not provide information beyond what is already contained in the position data.

In this chapter, we propose a mechanism of determining both surface position and orientation independently. The mechanism involves projection of a specially designed pseudorandom color pattern. Rhombic color elements are adopted for the pattern. Grid-points between pattern elements are defined as the feature points and encoded by the colors of their neighboring pattern elements. A grid-point detector is proposed to extract the grid-points in sub-pixel accuracy. The number of feature points can be doubled with respect to previous patterns, since each pattern element possesses only one centroid but can be attributed with two grid-points. The grid-point also with much higher location precision since it is insensitive to image noise and perspective projection. Furthermore, with consideration of the geometric relation and calibration parameters of the projector-camera system, as we proposed in Chapter 6, the surface normal at each grid-point can be determined also. Surface normals so determined provide information beyond what is in the position data. All the above can be achieved with only a single pattern illumination.

This chapter is organized as follows. In Section 6.2, previous work on 2D pseudorandom structured light is briefly reviewed. The principle of proposed pattern design is presented in Section 6.3. In Section 6.4, extensive experiments on various surfaces are shown. A summary of this chapter can be found in Section 6.5.

6.2 Determine Surface Normal from Image Tangents

Suppose we program a pattern on the display panel of the projector, which consists of a 2D array of alternate dark and bright rectangular blocks, and project it to the object surface. The edge of the blocks forms the grid-lines, and the intersections of every two grid-lines form the grid-points. The grid-lines and grid-points are accessible in both the projector panel and the image data.

6.2.1 The Underlying Principle

Consider any grid-point $\mathbf{p}_p$ in the pattern panel of the projector, and the two accompanying grid-lines that compose it, as illustrated by Figure 6.1 (a). Suppose the grid-point and grid-lines induce an imaged grid-point $\mathbf{p}_c$ and two imaged grid-lines (or more correctly grid-curves, as the original grid-lines are generally modulated by the curvature of the surface in 3D and appear not as lines in the image data) on the image plane via a 3D point $\mathbf{P}$ on the object surface in space. Suppose the tangents to the grid-lines at point $\mathbf{p}_p$ in the pattern projection panel are $\mathbf{t}_{p1}$ and $\mathbf{t}_{p2}$ respectively, and the tangents to the grid-lines at point $\mathbf{p}_c$ in the image plane are $\mathbf{t}_{c1}$ and $\mathbf{t}_{c2}$.

Tangent $\mathbf{t}_{p1}$ and the grid-point $\mathbf{p}_p$ in the pattern projection panel together form a plane $\Pi(\mathbf{p}_p, \mathbf{t}_{p1})$ of illumination from the light source, which is reflected by the object surface at point $\mathbf{P}$ and becomes the plane of projection $\Pi(\mathbf{p}_c, \mathbf{t}_{c1})$ to the image plane. The intersection of the two light planes $\Pi(\mathbf{p}_p, \mathbf{t}_{p1})$ and $\Pi(\mathbf{p}_c, \mathbf{t}_{c1})$ actually defines a tangent $\mathbf{t}_1$ in 3D to the object surface at point $\mathbf{P}$. $\mathbf{p}_p$ and $\mathbf{t}_{p1}$ are fully accessible as they are entities under system design, and so are $\mathbf{p}_c$ and $\mathbf{t}_{c1}$ as they are entities observable from the image data. Thus the two light planes $\Pi(\mathbf{p}_p, \mathbf{t}_{p1})$ and $\Pi(\mathbf{p}_c, \mathbf{t}_{c1})$ are both constructible, and their intersection $\mathbf{t}_1$ can be determined. In fact the tangent $\mathbf{t}_1$ to the object surface at point $\mathbf{P}$ is merely the cross-product of the surface normals of the two light planes.

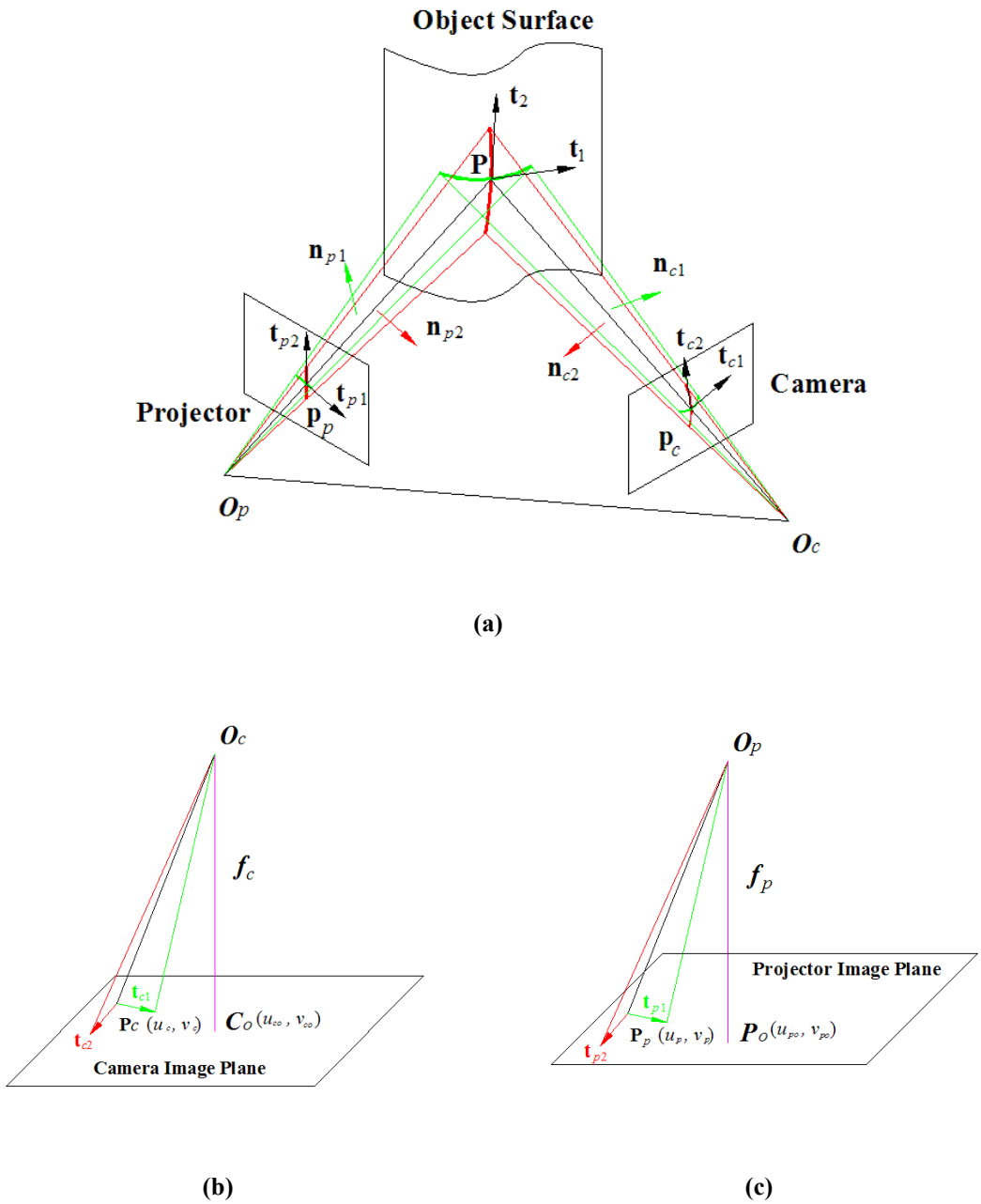


Figure 6.1 Determining surface orientation from two image tangents, for object surface under pattern illumination.

Similarly, another tangent $\mathbf{t}_2$ to the object surface at point **P** can be determined as the cross-product of the surface normals of two other light planes: $\Pi(\mathbf{p}_p, \mathbf{t}_{p2})$ and $\Pi(\mathbf{p}_c, \mathbf{t}_{c2})$, which are both accessible from design and image observations. In other words, by simply taking one image of the object surface that is under projection of a proper pattern, for any imaged grid-point $\mathbf{p}_p$ at position (x,y) on the image plane, the surface orientation $\mathbf{n}(x,y)$ of the object surface at the associated 3D point can be determined simply as $\mathbf{n}(x,y)=\mathbf{t}_1\times\mathbf{t}_2$ from image observations $\{\mathbf{p}_c, \mathbf{t}_{c1}, \mathbf{t}_{c2}\}$ and pattern data $\{\mathbf{p}_p, \mathbf{t}_{p1}, \mathbf{t}_{p2}\}$. This is the underlying principle of this work.

6.2.2 The Path from Image Tangents to 3D Orientation

Suppose $\mathbf{n}_{c1}$ and $\mathbf{n}_{c2}$ are the surface normals of the light planes $\Pi(\mathbf{p}_c, \mathbf{t}_{c1})$ and $\Pi(\mathbf{p}_c, \mathbf{t}_{c2})$ on the camera side, and $\mathbf{n}_{p1}$ and $\mathbf{n}_{p2}$ the surface normals of the light planes $\Pi(\mathbf{p}_p, \mathbf{t}_{p1})$ and $\Pi(\mathbf{p}_p, \mathbf{t}_{p2})$ on the projector side. Suppose the intrinsic parameters of the camera and projector (which is modeled as another perspective camera, except that light is coming up out of it instead of going into it) have been calibrated, which are focal lengths f_c, f_p, and principal points $\boldsymbol{C}_c\ (u_{c0}, v_{c0})$, $\boldsymbol{C}_p(u_{p0}, v_{p0})$. Then $\mathbf{n}_{c1}$, $\mathbf{n}_{c2}$, $\mathbf{n}_{p1}$, $\mathbf{n}_{p2}$ can be determined as:

$$\vec{\mathbf{n}}_{ci} = \mathbf{p}_{ci} \times \mathbf{l}_{ci}\,,\ \vec{\mathbf{n}}_{pi} = \mathbf{p}_{pi} \times \mathbf{l}_{pi}\,,\ i = 1, 2 \tag{6.1}$$

where $\mathbf{p}_{ci} = [u_{ci}, v_{ci}, -f_c]$, $\mathbf{p}_{pi} = [u_{pi}, v_{pi}, -f_p]$ represent the projective points on image and projector plane respectively, $\mathbf{l}_{ci}$ and $\mathbf{l}_{pi}$ indicate the projective lines which defined by the image tangents $\mathbf{t}_{c1}$, $\mathbf{t}_{p1}$ and projective points $\mathbf{p}_{ci}$ and $\mathbf{p}_{pi}$.

Assume the tangential angle of $\mathbf{t}_{c1}$ is θ_{c1} in camera image plane as shown in Figure 6.1 (b). Numerically, the surface normal of incident light plane can be formulated as:

Parallel model:
$$\mathbf{n}_{c1} = \{-f_c \sin\theta_{c1}, f_c \cos\theta_{c1}, 0\} \tag{6.2}$$

Perspective model:
$$\mathbf{n}_{c1} = \{-f_c \sin\theta_{c1}, f_c \cos\theta_{c1}, (u_c - u_{c0})\sin\theta_{c1} - (v_c - v_{c0})\cos\theta_{c1}\} \tag{6.3}$$

Similarly, $\mathbf{n}_{c2}$, $\mathbf{n}_{p1}$ and $\mathbf{n}_{p2}$ can be determined also.

So the two 3D tangents $\mathbf{t}_1$, $\mathbf{t}_2$ at surface point **P** that are associated with image point $\mathbf{p}_p=(x, y)$, if with reference to the camera coordinate system, can be obtained as:

$$\mathbf{t}_1 = \mathbf{n}_{c1} \times (\mathbf{R}\mathbf{n}_{p1}) \tag{6.4}$$

$$\mathbf{t}_2 = \mathbf{n}_{c2} \times (\mathbf{R}\mathbf{n}_{p2}) \tag{6.5}$$

where **R** represents the rotational relationship between the camera and projector coordinate frames. Finally, the surface orientation to the object surface point **P** that is associated with the image position (*x, y*) can be determined as:

$$\mathbf{n}(x, y) = \mathbf{t}_1 \times \mathbf{t}_2 \tag{6.6}$$

Notice that the determination of local surface orientation as expressed by Equation (6.6) is a deterministic process that requires only image information local to the specific point to operate. Unlike shape from shading and many other methods, it does not require assumption about how local surface orientations at neighboring points are related. More specifically, it requires no process of iterations to determine local surface orientations.

6.2.3 The Need of Correspondence

Given a projective point $\mathbf{p}_c$ and its image tangent $\mathbf{t}_c$, the surface normal $\mathbf{n}_c$ of its projective plane $\prod(\mathbf{p}_c, \mathbf{t}_c)$ can be fully decided. Under the parallel projection, $\mathbf{n}_c$ can be solely decided by $\mathbf{t}_c$. The simplification can fast the computation, but do harm to the accuracy. To improve that, the perspective projection is suggested in this paper. The effect from $\mathbf{p}_c$ is considered in the calculation. In other words, even $\mathbf{t}_c$ is a constant in the image plane, $\mathbf{n}_c$ still change slightly according to the image position. The same improvement has done to the projector. So that it is required to setup the correspondence of between $\mathbf{p}_c$ and $\mathbf{p}_p$.

6.3 2D Rhombic Pseudorandom Pattern Design

Pseudorandom Binary Sequence (*PRBS*, maximal length shift register sequences, or *m*-sequences) are certain binary sequences of length $n=2^m-1$. These sequences have been popular used in range finding, scrambling, fault detection, modulation, and synchronizing [Ball 1975] etc. It was firstly introduced into structured light domain by Morita [Morita 1988] for its window property. Since each code only appear once in the whole sequence or array, the corresponding problem can be avoided. This section gives a brief

introduction to the construction of a pseudorandom sequence and array and the feature definition in the final pattern image.

6.3.1 Background

Many strategies have been proposed to generate a random pattern with unique window property. The main difference between the techniques is the way in which the elements of the array are represented in the pattern. In Chapter 2, we have given a comprehensive survey of previous 2D pseudorandom patterns, such as [Morano 1998], [Salvi 1998], [Adan 2005], [Albitar 2007], [Desjardins 2007], and [Chen 2008]. Some authors prefer to define the pattern as an array of color spots, where each color represents one of the symbols of the coding alphabet. Other authors prefer to define different shapes for each symbol. When perceiving the projected pattern, an algorithm to recover the maximum number of visible windows must be fulfilled. This is the crucial step of these systems. Since spatial neighborhood is used, not all the pattern elements will be visible from the camera's point of view, due to the shadows and occlusions. The robustness of these methods depends on the correct decodification of the visible parts, taking advantage of the properties of the pseudorandom array [Salvi 2004].

To evaluate a random pattern, we should focus on two aspects: (1) how many feature points can be encoded; and (2) how many pattern primitives are needed. Actually, they are two conflicting problems. We have to use more colors or pattern primitives to encode more features when the window size is predestined. Presently, the color coding technique tends to use more pattern primitives. Using more pattern primitives can improve the feature density, hamming distance and decrease the window size. But it will make the pattern segmentation and feature detection harder. Compared with pattern design, relative few works have been denoted to the feature extraction, which determines the final reconstruction accuracy directly. Only some simple processing such as thresholding, template methods are used. That makes their results with low accuracy as reported in Chapter 2.

Generally, to make a successful structured light system, we should consider all the three key issues as we mentioned before. In this chapter, we will focus on the pattern design. In the pattern design, we have to consider following issues: (1) mathematical principle, e.g. how to construct the pattern matrix with small window size as possible; and (2) how to

design the pattern element and so that more feature points can be encoded; and (3) how to design the pattern structure to make it not only derive 3D depth but also other 3D information such as surface normal; and (4) respective feature detector should also be designed accordingly. Previous work only focused on one or two aspects. In this chapter, we fill focus all the aspects in the pattern design.

6.3.2 Pseudorandom Sequence

Generally, a *PRBS* can be generated by a given primitive polynomial through *LFSR* (*Linear Shift Feedback Register*). It follows that if $h(x)$ is a primitive polynomial of degree m, the shift register goes through all 2^m-1 distinct nonzero states before repeating, and produces an output sequence $a_0a_1a_2\ldots$ of period 2^m-1. We call any segment $a_i a_{i+1}\cdots a_{i+2^m-2}$ of length 2^m-1 a pseudorandom sequence.

Let $h(x)$ be a fixed primitive polynomial of degree m, and let $\boldsymbol{\delta}_m$ be the set consisting of the pseudo random sequences obtained from $h(x)$. These pseudo random sequences are the 2^m-1 different segment:

$$a_i a_{i+1}\cdots a_{i+2^m-2},\ i=0,1,\cdots,2^m-2 \tag{6.7}$$

of length 2^m-1 from the output of the shift register specified by $h(x)$. The autocorrelation function $\rho(i)$ of a sequence $s_0, s_1, \ldots, s_{n-1}$, of length n is defined by:

$$\rho(i)=\frac{1}{n}\sum_{j=0}^{n-1} s_j \bar{s}_{i+j}, i=0,\pm1,\pm2,\cdots \tag{6.8}$$

Autocorrelation function of a pseudorandom sequence of length 2^m-1 can be given by:

$$\rho(0)=1,\ \rho(i)=-1/n,\ \text{for } 1\le i\le 2^m-2 \tag{6.9}$$

It can be proved that this is the best possible autocorrelation function of any binary sequence of length 2^m-1. If a window of width m is sliding along a pseudorandom sequence in $\boldsymbol{\delta}_m$, each of the 2^m-1 nonzero binary m-tuples is seen exactly once-see, since $h(x)$ is a primitive polynomial.

A binary linear code of length n, dimension k, and minimum distance d consists of 2^k binary vectors $u, \ldots, u_n$, $u_i=0$ or 1, called codeword, which: i) form a linear space (i.e. the

modulo 2 sum of two codeword is a codeword), and ii) are such that any two codeword differ in at least *d* places [MacWilliams 1976].

6.3.3 Pseudorandom Array

Pseudorandom arrays can be constructed by folding the pseudorandom sequences. Take a number of the form $n=2^{k1k2}-1$ such that $n_1=2^{k1}-1$ and $n_2= n/n_1$ are relatively prime and greater than 1. The starting point is a pseudo random sequence $a = a_0, \ldots, a_{n-1}$, in $\boldsymbol{\delta}_{k1k2}$ obtained from a primitive polynomial $h(x)$ of degree $m = k_1k_2$. We use a to fill up an $n_1 \times n_2$, by writing a down the main diagonal and continuing from the opposite side whenever an edge is reached. An $n_1 \times n_2$ array formed in this way will be called a pseudorandom array. There are n different arrays, one from each pseudorandom sequence in $\boldsymbol{\delta}_{k1k2}$. The pseudorandom sequences in at least look random to the eye, whereas the corresponding arrays have a conspicuous nonrandom feature, being symmetric about a column of zeros. If the pseudorandom array is denoted by:

$$b = \begin{bmatrix} b_{00} & b_{01} & \cdots & b_{0,n_2-1} \\ b_{10} & b_{11} & \cdots & b_{1,,n_2-1} \\ \cdots & \cdots & \cdots & \cdots \\ b_{n_1-1,0} & \cdots & \cdots & b_{n_1-1,n_2-1} \end{bmatrix}$$

$$a_i = b_{i_1 i_2} \tag{6.10}$$

$$i_1 \equiv i(\mathrm{mod}\, n_1), 0 \le i_1 < n_1$$

$$i_2 \equiv i(\mathrm{mod}\, n_2), 0 \le i_2 < n_2$$

Conversely, given i_1 and i_2, there is a unique value of i in the range $0 \le i < n = n_1 n_2$, such that Equation (6.10) holds, by the Chinese Remainder Theorem (since n_1 and n_2 are relatively prime). The arrays are described by polynomials with following form:

$$b(x, y) = \sum_{i=0}^{n_1-1} \sum_{j=0}^{n_2-1} b_{ij} x^i y^j \tag{6.11}$$

With n, n_1, n_2, k_1, k_2 as defined above, let $h(z)$ be a fixed primitive polynomial of degree $m=k_1k_2$. Each of the pseudorandom sequences in $\boldsymbol{\delta}_m$, folds into a distinct $n_1 \times n_2$

pseudorandom array. Let $\boldsymbol{\beta}_{\mathrm{m}}$, be the set of 2^m arrays so formed, together with the zero arrays. Let b be any $n_1 \times n_2$ array of area $n = n_1 \times n_2$. The 2D autocorrelation function of b is defined as follows:

$$\rho(i,j) = \frac{A-D}{n}$$
$$i = 0, \pm 1, \pm 2, \cdots \tag{6.12}$$

A is the number of positions in which b and b shifted i places down and j to the right agree, and D is the number in which they disagree, then if a $k_1 \times k_2$ window is sliding over a pseudo random array in $\boldsymbol{\delta}_{\mathrm{m}}$, each of the 2^{k1k2}-1 nonzero binary $k_1 \times k_2$ arrays only appear once.

The same constructions can be used to obtain pseudorandom sequences and arrays with entries which, instead of being 0's and l's, are taken from an alphabet of q symbols, for any q which is a prime or a power of a prime. Let q be any prime power and let $GF(q)$ be the *Galois Field* with q elements. We need a polynomial $h(x)$ of degree m with coefficients from $GF(q)$ which: (1) is not the product of two such polynomials of lower degree, and (2) has as a zero a primitive element of $GF(q^{\mathrm{m}})$, say E. We call such a polynomial primitive over $GF(q)$. The primitive polynomials $h(x)$ that used in binary pseudorandom array are primitive over $GF(2)$ as:

$$h(x) = x^m + h_{m-1}x^{m-1} + \cdots + h_1 x + h_0 \tag{6.13}$$

with $h_i \in GF(q), h_0 \neq 0$. The feedback path then forms

$$a_{i+m} = -h_{m-1}a_{i+m-1} - h_{m-2}a_{i+m-2} - \cdots - h_i a_{i+1} - h_0 a_i \tag{6.14}$$

The equation is the recurrence describing the output sequence. This is an infinite sequence of period q^m-1(if the starting point is not zero), and each nonzero state appears once in a period. A segment of the output sequence of length q^m-1 is called a pseudorandom sequence over $GF(q)$.

Let $\boldsymbol{\delta}_{\mathrm{m}}(q)$ be the set of such sequences, together with 0. The autocorrelation function is given by:

$$\rho(0) = 1, \rho(i) = \frac{1}{q^m - 1}$$
$$1 \le i \le q^m - 2 \tag{6.15}$$

More details about the proof and deduction of the pseudorandom array can be found in [MacWilliams 1976] and [Etzion 1988].

6.3.4 Pattern Image Construction

A pattern in matrix form has been generated. The number in the matrix represents different pattern primitive. In order to project the pattern through a LCD or DMD projector, we need to transform it into image form. To generate an appropriate pattern image, there are some points we should pay attention:

a. *The pattern is expected to have square shape, since the inspected region is usually square.*

b. *Considering the possible occlusions, the unique window is expected to be compact as possible.*

c. *The selected colors should be easy distinguished.*

d. *The pattern structure should facilitate surface normal recovery.*

e. *Pattern size and resolution is confined to the resolution of LCD and camera.*

In order to find a pattern can meet above requirements, we setup two tables based *GF*(2) and *GF*(4) as shown in Table 6.1 and Table 6.2

Table 6.1 Pseudorandom sequence and array size in *GF*(2), where 4≤m≤16: L represent the length of sequence, k1 and k2 indicate the unique window size, n1 and n2 indicate the array's dimension.

m	L	k1	k2	n1	n2
4	2^4-1 = 15	2	2	3	5
5	2^5-1 = 31	5	1	31	1
6	2^6-1 = 63	3	2	7	9
7	2^7-1 = 127	7	1	127	1
8	2^8-1 = 255	4	2	15	17
9	2^9-1 = 511	3	3	7	73
10	2^{10}-1 = 1023	5	2	31	33
11	2^{11}-1 = 2047	11	1	2047	1
12	2^{12}-1 = 4095	6	2	63	65
13	2^{13}-1 = 8191	13	1	8191	1
14	2^{14}-1 = 16383	7	2	127	129
15	2^{15}-1 = 32767	5	3	31	1057
16	2^{16}-1 = 65535	8	2	255	257

Table 6.2 Pseudorandom sequence and array size in *GF*(4), where 4≤*m*≤9: L represent the length of sequence, k1 and k2 indicate the unique window size, n1 and n2 indicate the array's dimension.

m	L	k1	k2	n1	n2
4	4^4-1 = 255	2	2	15	17
5	4^5-1 = 1023	5	1	1023	1
6	4^6-1 = 4095	3	2	63	65
7	4^7-1 = 16383	7	1	16383	1
8	4^8-1 = 65535	4	2	255	257
9	4^9-1 = 262143	3	3	63	4161

Obviously, when $n_1 = 63$ and $n_2 = 65$, the pattern is approximate to square. The LCD or DMD panel in the projector has a resolution of 1024×768 pixels.

If $n_1 = 63$ and $n_2 = 65$, the size of each pattern element is about 16×16 pixels. If we use two elements in the pattern, the unique window has a dimension of 6×2 as shown in Table 6.1. It's too long to decode especially occlusion arises.

In Table 6.2, the window size can be shortened to 3×2, while 4 elements are used.

When the parameters of the pseudorandom array are decided, (i.e. q=4, n_1=63, n_2=65, k_1=3, k_2=2), a primitive polynomial as follow is used to generate the pseudorandom sequence in $GF(4)$.

$$2x^6 + 2x^5 + x^4 + 3x^3 + 2x^2 + 2x + 1 \tag{6.16}$$

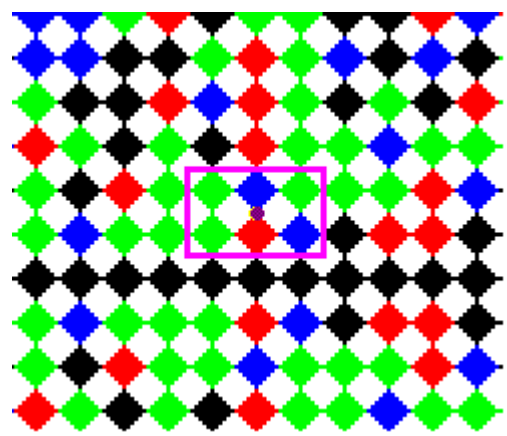
With the given primitive polynomial, the LFSR can output a pseudorandom sequence with length 4095. A part of the sequence is as shown in Table 6.3.

After a folding process [MacWilliams 1976], the sequence can be transformed into an array with size of 63×65. After a rotation of 90˚, we can get a pattern with size of 65×63, and the window size is changed to 2×3.

To produce the pattern image, the four elements *1,2,3,4* in the array are replaced with four different color elements respectively. We used *red*, *green*, *blue* and *black*. The background is set to white. One example can be found in Figure 6.2: (a) shows a codeword (*2-3-2-2-1-3*) in pattern matrix and (b) shows its corresponding image form. The codeword is represented by a color sequence *G-B-G-G-R-B*. Figure 6.3 shows the global view of the rhombic color pattern.

3	2	1	4	2	2	4	4
3	4	4	2	1	2	3	4
4	4	1	3	1	2	4	2
2	4	2	4	1	2	2	3
4	1	2	2	3	2	2	2
3	2	2	2	1	3	4	1
4	4	4	4	4	4	4	4
3	2	2	2	1	3	4	1
4	1	2	2	3	2	2	2
2	4	2	4	1	2	2	3
4	4	1	3	1	2	4	2

(a)



(b)

Figure 6.2 A partial view of pseudorandom array with size of 65×63. (a) A part of the pseudorandom array, *2-3-2-2-1-3* indicates one unique codeword. (b) In the pattern of image form, the codeword *2-3-2-2-1-3* is replaced with a color sequence: *G-B-G-G-R-B*.

Table 6.3 A part of the pseudorandom sequence with length 4095 and the window size is 1×6.

... 3 4 2 3 4 2 2 4 4 2 2 4 1 1 1 3 2 2 3 4 1 4 1 3 3 1 2 2 3 2 2 1 3 2 1 2 2 2 2 2
2 3 1 4 2 3 1 1 2 2 4 4 3 2 2 1 3 1 4 1 3 1 3 2 4 3 4 4 4 4 1 3 3 3 3 3 2 3 3 2 2 3
2 1 3 1 4 3 4 3 4 4 3 3 1 4 3 4 4 1 3 2 4 4 4 1 4 1 1 1 1 4 2 1 4 3 2 4 4 4 3 3 3 2
3 2 4 3 1 3 3 3 2 4 1 4 4 1 4 4 1 3 1 2 2 2 3 2 4 3 3 4 1 4 4 3 1 1 4 2 4 1 3 3 4 1
1 4 1 1 3 4 1 3 4 1 4 3 1 3 3 2 4 1 3 1 2 1 3 4 3 3 2 3 2 1 1 4 1 2 1 2 3 1 2 4 2 1
4 2 1 3 3 3 4 1 4 1 1 4 3 1 4 4 2 3 1 4 2 4 4 2 3 1 1 4 1 2 3 1 1 2 4 3 2 4 4 1 1 2
1 3 1 2 3 3 3 3 4 2 2 3 3 2 ...

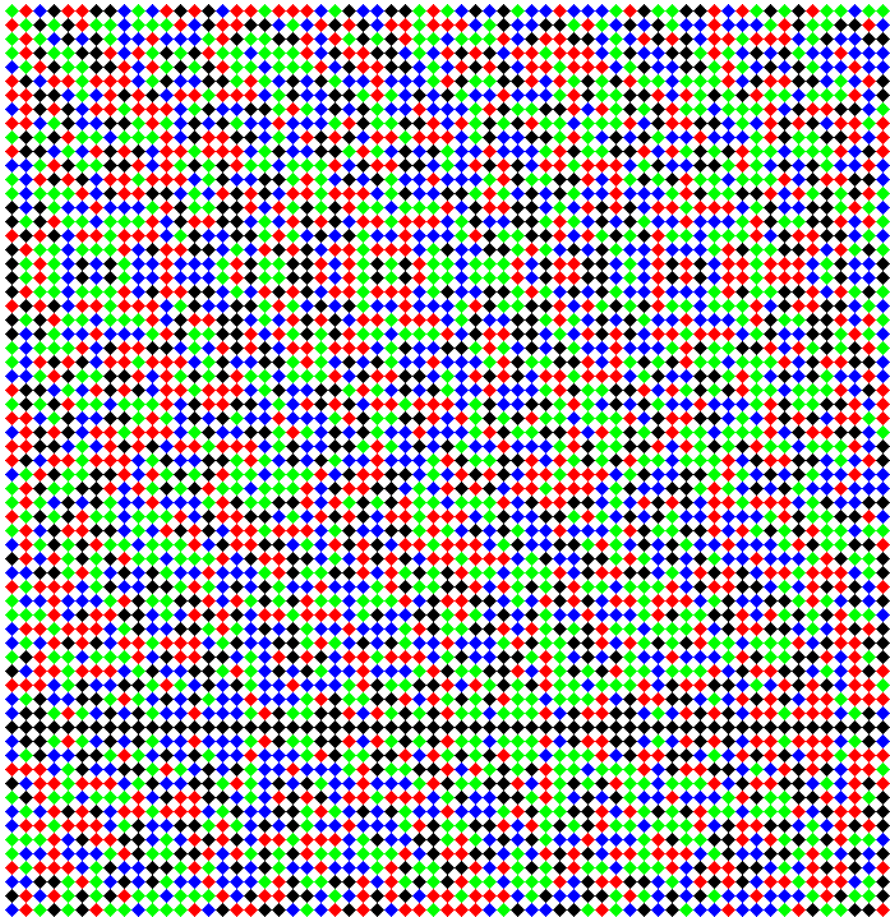


Figure 6.3 A 2D pseudorandom color pattern with rhombic elements: the diagram of pattern composed of 65×63 elements, which belong to red, green, blue, and black color.

6.3.5 Feature Definition

In previous works, the pattern element is usually designed to the rectangle or circle with different color or grey levels. Moreover, the centroids of these pattern elements are extracted by specially designed feature detectors and then are defined as the feature points. With this design, the number of feature points cannot exceed the number of elements in the pattern. Suppose a 2D pattern, which is defined by a pseudorandom matrix with $m \times n$ elements and has a window property of $a \times b$. With traditional feature point definition, the maximum number of feature points, which can be encoded, is:

$$N_{traditional} = (m-a+1) \times (n-b+1) \tag{6.17}$$

Centroid of pattern element is usually defined as the feature points. However, the centroid position cannot be accurate located in the distorted pattern elements under perspective projection.

In this work, we propose to use a rhombic pattern element. And the grid-points between each neighboring rhombic elements is defined as the feature point. With the proposed *cmm*-symmetry based grid-point detector (see Chapter 4), the grid-point can be accurately detected under perspective projection and image noise.

According to the colors of their neighboring elements on the left and right sides, the grid-points can be classified into two types: $\mathbf{P}_1$ and $\mathbf{P}_2$ as shown in Figure 6.4. When all the grid-points are detected, they can be divided easily since $\mathbf{P}_1$ has two background elements on the left-right sides. Because the baseline between the camera and the projector is almost parallel to the scanning line of the camera sensor in our system, the color identification of the left-right elements can be much simplified. When all the grid-points belong to $\mathbf{P}_1$ have been identified, the others can be categorized to $\mathbf{P}_2$.

After extraction and separation of the grid-points, each grid-point can be encoded by the color of the 2×3 rhombic elements surrounding it, in the form of:

$$c1–c2–c3–c4–c5–c6$$

where ci (i =1…6) indicates the color of the corresponding element for any grid-point in Figure 6.3.

For instance, in Figure 6.4, the codeword of point $\mathbf{P}_1$ and $\mathbf{P}_2$ have the same codeword:

R-K-B-G-R-B *or* *1–4–3–2–1–3*

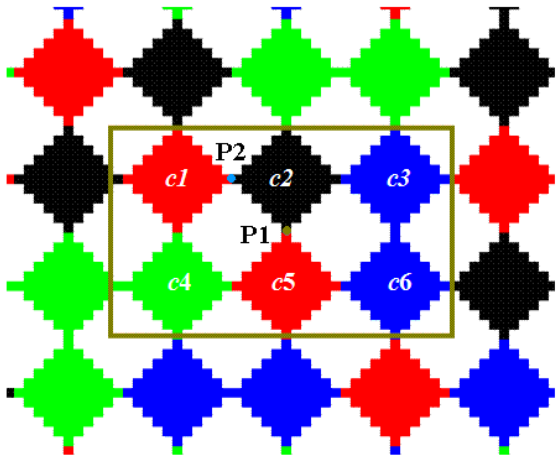


Figure 6.4 Grid-point P_1 and P_2 have the same codeword: *c1–c2–c3–c4–c5–c6*.

Figure 6.5 shows the coding procedure. When $\mathbf{P}_1$ is encoded by its surrounding 2×3 color window, $\mathbf{P}_2$ which formed by elements *c*1 and *c*2 is determined also. In practice, we need not to search the 2×3 color window around $\mathbf{P}_2$. We just need to search for $\mathbf{P}_1$'s nearest grid-point, which belongs to $\mathbf{P}_2$ class in the upper left. When the point is found, the same codeword (i.e. *c1–c2–c3–c4–c5–c6*) as $\mathbf{P}_1$'s will be assigned. To facilitate the searching process, $\mathbf{P}_1$ and $\mathbf{P}_2$ are saved in two separate data sets respectively. Experiment shows that the encoding and searching procedures are practical and efficient.

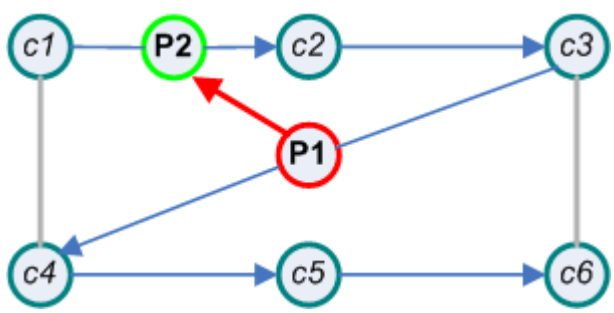


Figure 6.5 The coding process of grid-point P_1. When P_1 is encoded, its nearest grid-point P_2 will share the same codeword (i.e. c1–c2–c3–c4–c5–c6) with P_1.

By this new feature definition, the maximum number of the encoded feature points can reach 7,808 in comparison of 3,904 by using the centroids. The density of feature point can be doubled. If the centroids of the pattern elements are also encoded, the feature points' density can be further improved.

6.3.6 Grid-point vs. Centroid

As discussed, determining the centroids of pattern elements requires segmentation of the input image into pattern elements in the first place. That sets tight dependence between the individual centroid results. Unless the singular case is reached in which all pattern segments are segmented correctly, any error in segmenting a single pattern element will directly affect the segmentation results of the neighboring elements, and propagate to all centroid determinations. The accuracy in determining centroids is thus limited. Segmentation, a process directly related to raw intensities, is also generally sensitive to albedo variation, uneven illumination, and image noise.

In contrast, by exploiting two-fold symmetry of the features in the proposed method, grid-points can be determined independently of one another, and determined precisely in sub-pixel accuracy.

We have conducted experiments to investigate the above on a human face whose image is shown in Figure 6.6 (a). To extract the centroids of the pattern elements, we needed to conduct segmentation first. We used a morphological method (available in MATLAB) for it that is widely used. The process can be outlined as below:

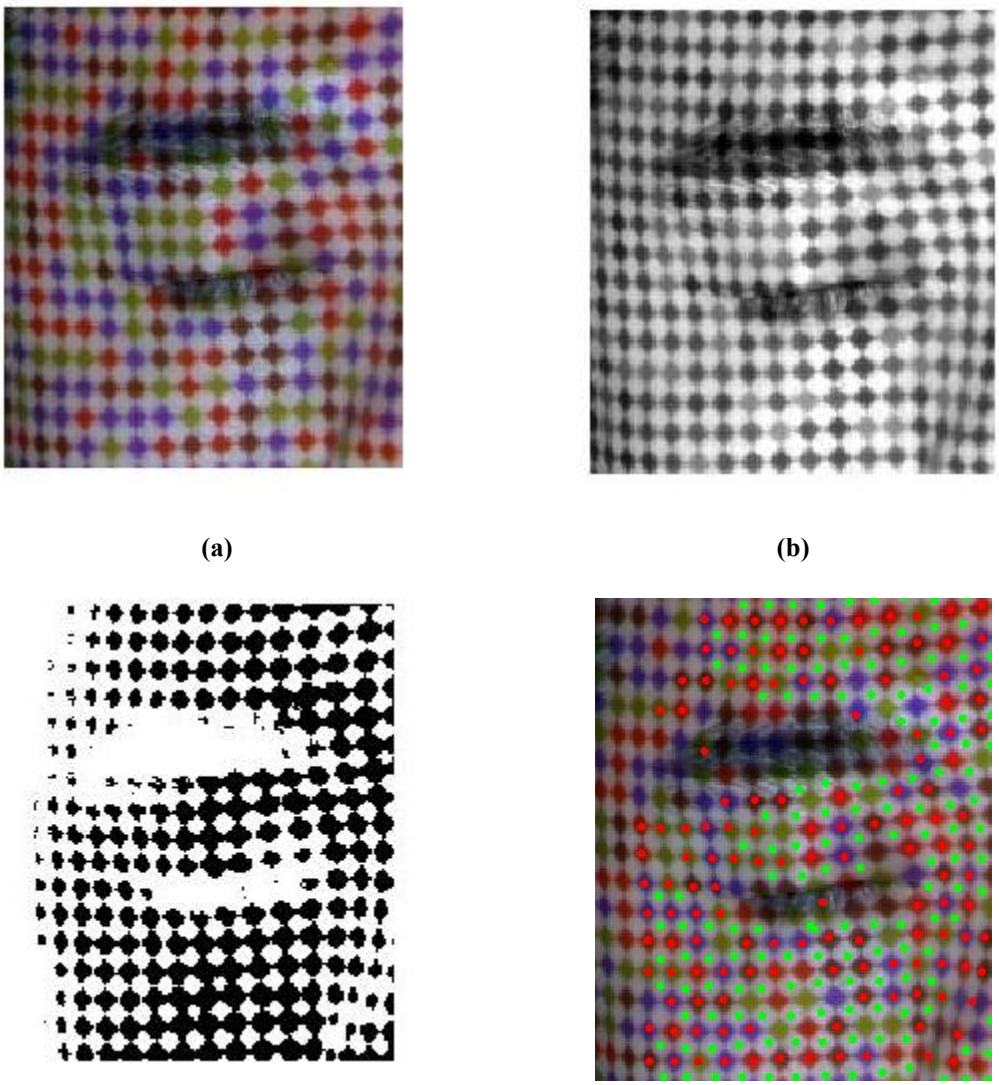
a. *The image was converted from a color one to a gray level one, which is shown in Figure 6.6 (b). A median filter was used to smooth the image;*

b. *The image was further transformed into a binary one using carefully adjusted thresholds.*

c. *Morphological operators including 'erode', 'hbreak', 'close' provided by MATLAB were used to separate the pattern elements, and the results are shown in Figure 6.6 (c).*

d. *The centroids of the individual rhombic elements were then calculated.*

The detected centroids are marked as color dots in Figure 6.6 (d). The centroids belong to either foreground or background, which are marked in red and green respectively. As can be observed from the result, the segmentation process had difficulty at places where the pattern elements were close to one another, where there was illumination change (e.g., the illumination degradation over the left side of the face), and where there was surface texture or albedo change (e.g. over the eyebrow). The segmentation error propagated to centroids determination.



(a) **(b)**

(c) **(d)**

Figure 6.6 Detection of centroids and grid-points. (a) Human face under projection of rhombic color pattern. (b) The grey level image transformed from the color image. (c) Segmentation of pattern elements by a morphological method. (d) Result of centroid detection..

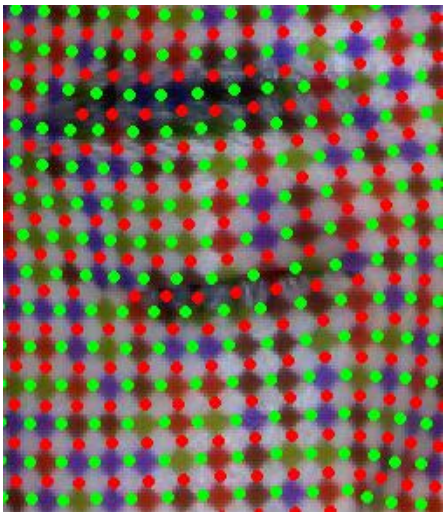


Figure 6.7 Result of grid-point detection by the proposed detector, which has the grid-points divided into two types P1 and P2 (marked as red and green dots respectively).

Figure 6.7 shows the result from the proposed grid-point detector. The grid-points belong to either type P_1 or type P_2, which are marked in red and green respectively. The grid-points were detected successfully with clearly much better localization accuracy.

Our experience on a number of image sets shows that, by the use of grid-points as the feature points and the proposed grid-point detector, more feature points can be extracted in higher accuracy.

6.3.7 Explicit 3D Reconstruction at Grid-points

When a grid-point has been extracted from the image, its codeword can be produced by judging the colors of its neighboring elements. Its corresponding point on the projector image plane is known as a priori. Therefore, its 3D coordinate on the object surface can be easily determined via triangulation method. Since more grid-points can be encoded in our pattern, a denser depth map is available.

Combined with surface orientation recovery mechanism we have proposed, to each grid-point, it surface normal can be decided. We also show that, the recovered surface normal has better accuracy than that estimated from surface points directly. Generally, with single illumination and image capture, both the position and surface normal (i.e. explicit 3D) at the grid-point can be determined independently.

6.4 Experimental Results

We have implemented the system using a DLP projector of XGA resolution and a camera of resolution 1.5M pixels (i.e., image resolution: 1500×1000 pixels). The working distance of system was about 850 mm. The system was first calibrated using the method described in Chapter 3. Extensive experiments have been conducted to examine how accurately the proposed mechanism can determine 3D position and orientation.

6.4.1 3D Depth Recovery

The color pattern was projected onto a planar board and its planarity was measured. To measure the planarity, a plane was fitted in the least square error sense to the discrete 3D points recovered. An absolute mean error of 0.131 mm or 0.015%, with a standard deviation of only 0.1 mm or 0.011%, was obtained. A relative deviation of about 0.2% was reported in previous works [Ito 1995] & [Kiyasu 1995], which used 2D pseudorandom patterns also.

Table 6.4 shows the accuracy of a number of structured light systems. Since different test objects and error evaluation criterions are used in these methods, the results in the table are only for reference.

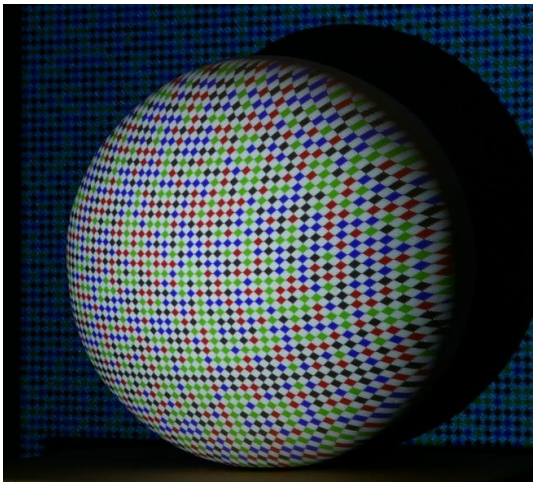
From the table, we can see that, our result far beyond previous structured light systems that used 2D pattern also. The result we had is even more favorable than that of some of the existing temporal methods like those in [Sadlo 2005] & [Gühring 2000]. The good accuracy can be contributed to our accurate calibration design and the proposed grid-point detector.

We had also conducted the experiment on a sphere with the radius about 97 mm. The position reconstruction result is shown in Figure 6.8. A sphere was fitted in the least square error sense to the reconstructed data points. The absolute mean fitting residue was 0.202 mm, with a standard deviation of 0.067 mm.

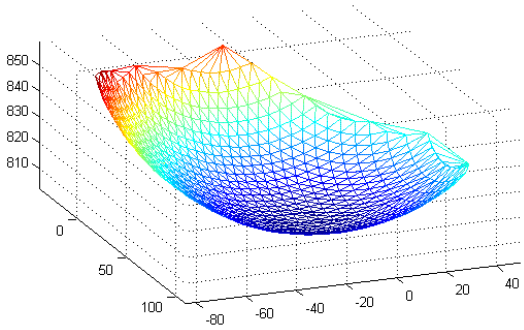
As shown in Figure 6.8 (c) and (d), a ribbon like surface is also reconstructed.

Table 6.4 Accuracy comparison between some structured light systems.

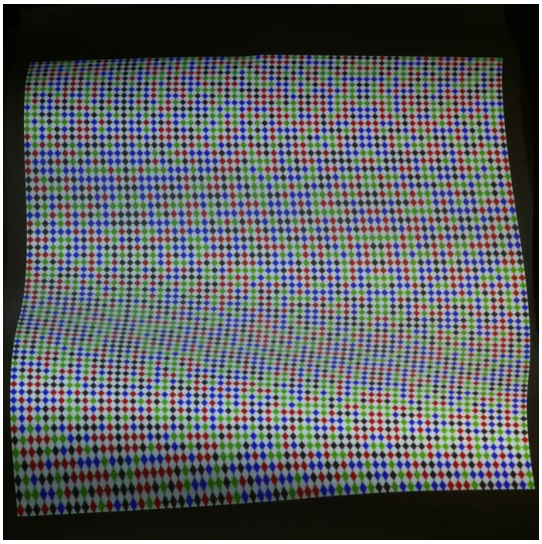
Reference	Method	Accuracy (working distance)
[Kiyasu 1995]	2D color pattern	1 mm (500 mm) or 0.2%
[Fong 2005]	2D color pattern	0.7 mm (650 mm) or 0.11%
[Chen & Li 2008]	2D color pattern	about 2%
[Desjardins 2007]	2D color pattern	2 mm (1.5 m) or 0.11%
[Sadlo 2005]	Gray code & phase shifting	about 0.1- 0.5 mm
[Gühring 2000]	Gray code & line shifting	Std. ~ 0.17-0.28 mm
[Sohmura 2004]	Laser	0.2 mm
[Demeyere 2007]	Laser	1%
[Harding 1998]	Moiré	0.5 ~ 1 mm
[Subbarao 1995]	Shape from focus	12.4 mm (0.6 m) or 2.07%
Our method	*2D Pattern*	*0.11-0.13 mm (850 mm) or 0.015%* *Std. 0.1 mm or 0.011%*



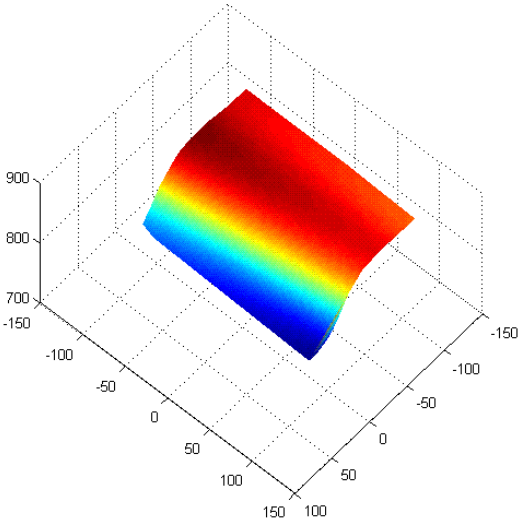
(a)



(b)



(c)



(d)

Figure 6.8 Reconstruction of a sphere and a ribbon-like surface. (a) A sphere with color pattern projected on it. (b) Reconstructed 3D surface. (c) A ribbon-like surface with color pattern projected on it. (d) 3D reconstruction result.

6.4.2 Surface Orientation Recovery

Experiments have also been conducted to examine the accuracy of the recovered surface normals. Similar to the planarity experiment, the surface normals of the grid-points on the inspected plane about 100×100 mm were reconstructed. The result shows only an average discrepancy of 0.69˚ and a standard deviation of 0.13˚ in the distribution of local surface orientations.

When the parallel projection models are used, an average deviation of 1.65˚ and standard deviation of 0.48˚ were obtained. Note that in methods like [Sugihara 1984], [Winkelbach 2001] and [Winkelbach 2002] other earlier works a discrepancy in the range of 2˚ to 8˚ were reported. The higher accuracy in our method can owe to the accurate system model, feature detector, and precise correspondences.

We have also conducted experiment on the above spherical object and ribbon like surface. The surface normal at each grid-point was recovered using the proposed method. The result was compared with that on the fitting sphere. There was a mean discrepancy of only 1.13˚. Figure 6.9 shows the recovered orientation map.

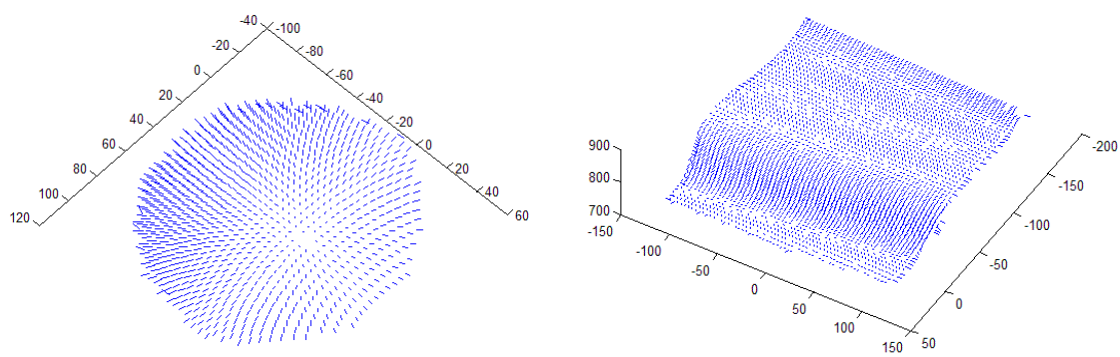


Figure 6.9 Recovered surface orientation map of a sphere (left) and ribbon-like surface (right).

We can evaluate the quality of surface orientation recovery in a way alternative to the above. Triangular mesh was generated from the reconstructed 3D points on the sphere as shown in Figure 6.8 (b). Surface normal at each grid-point was calculated from those of the surrounding triangle patches as the weighted average, with the weight being the area of each of the triangular patches. The mean discrepancy between this weighted average

and the surface normals on the fitting sphere was 5.41˚. In compared with the result of 1.13˚ that obtained with our method, the improvement is dramatically.

6.4.3 Results on Bust and Human Face

This experiment aims to test the performance on surface with generic shape and texture. A bust and the human face are used. The inspected surfaces under pattern illumination are shown in Figure 6.10 (a) and Figure 6.11 (a).

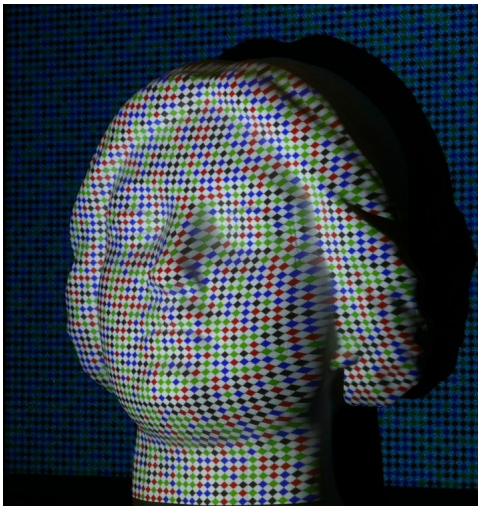
Figure 6.10 (b) shows that almost all the grid-points can be extracted correctly with the proposed gird point detector.

Compared with the bust, the grid-points that projected on the human face are more difficult to recognize. However, most of them can still be detected as shown in Figure 6.11 (b). That demonstrates that the proposed detector is reliable and robust.

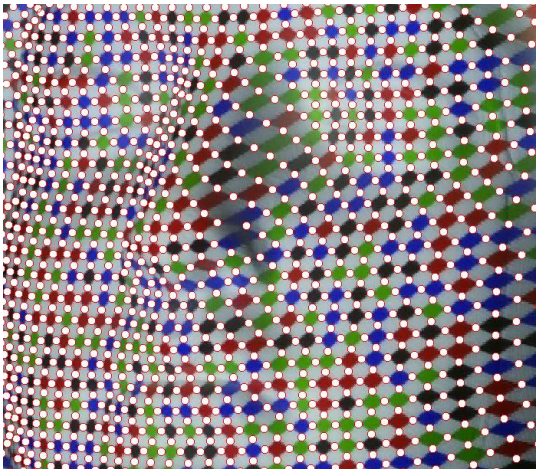
The interpolated 3D shapes based upon the recovered 3D points are shown in Figure 6.10 (c) and Figure 6.11 (c). The recovered surface orientation map is as shown in Figure 6.10 (d) and Figure 6.11 (d) as needle diagrams.

The human face is more challenge than the bust surface for its higher color variance and presence of texture over its surface; it also contained more occlusions in the captured image data. There were grid-points missing or encoded incorrectly. We used median filtering to remove outliers of the recovered 3D points, and we used simple interpolation over local neighborhood to fill in the occasional omissions of 3D reconstruction. The reconstructed 3D model is as shown in Figure 6.12. The top figure shows the 3D model plotted by the MATLAB function: surf(). The bottom figure shows the 3D model with texture mapping. The texture mapping procedure was completed by another MATLAB function: warp(), which needs two input parameters: surface points and original image. To align the 3D model with the 2D image, manual operation is needed.

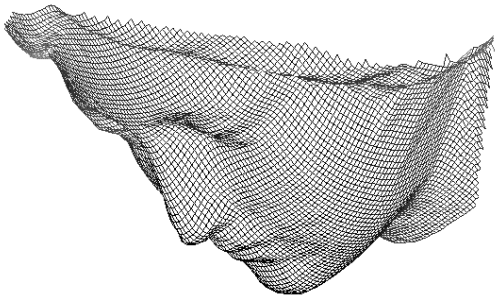
Although, ground truths of these surfaces were not available for performance evaluation, but visual examination clearly shows that the reconstructions were of reasonable quality.



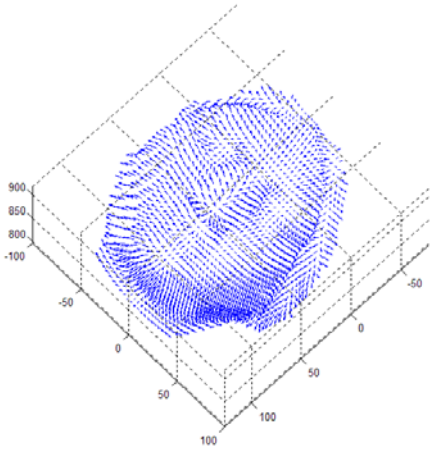
(a)



(b)

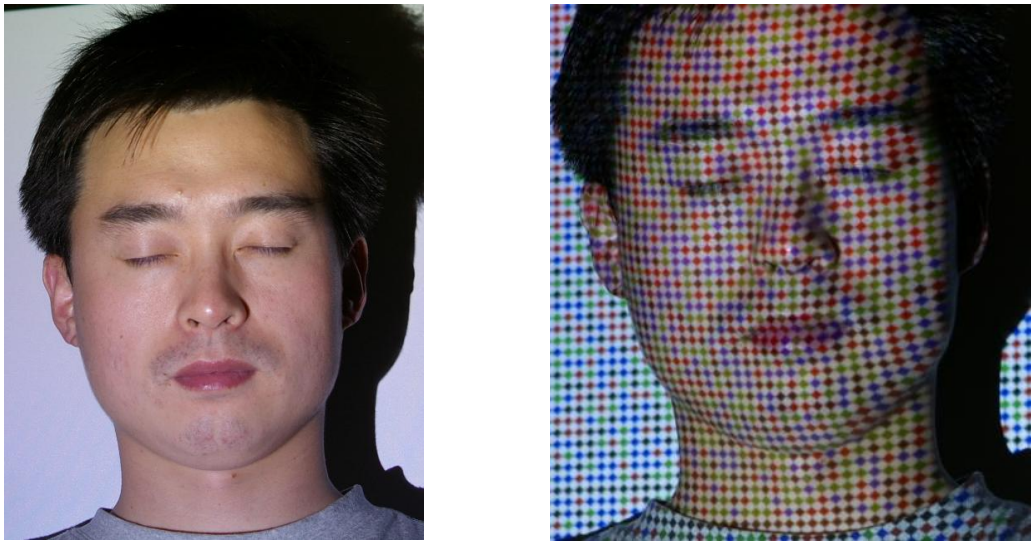


(c)

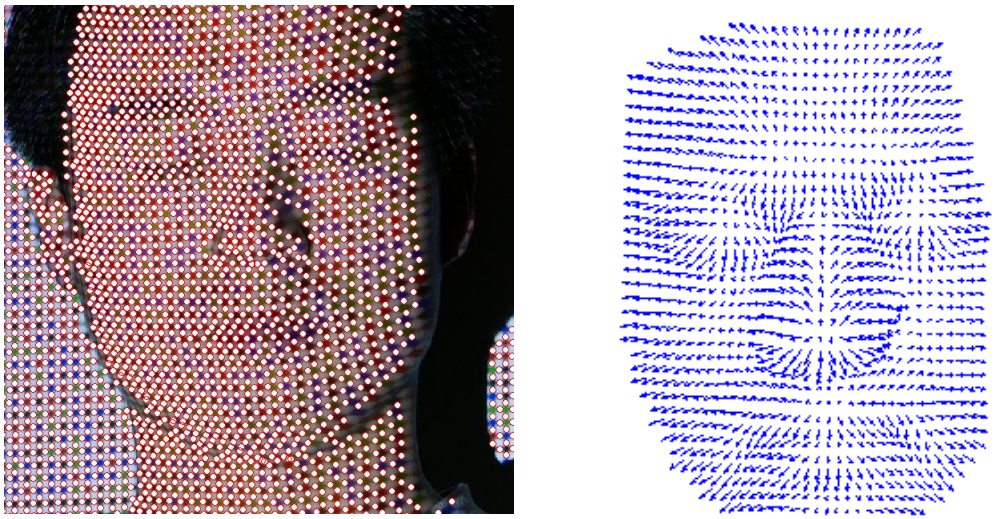


(d)

Figure 6.10 Reconstruction of a bust. (a) A bust under pattern illumination. (b) Grid-point detection result. (c) Interpolated surface only using the recovered 3D points. (d) Interpolated surface orientation map only using the recovered surface normals at the grid-points.

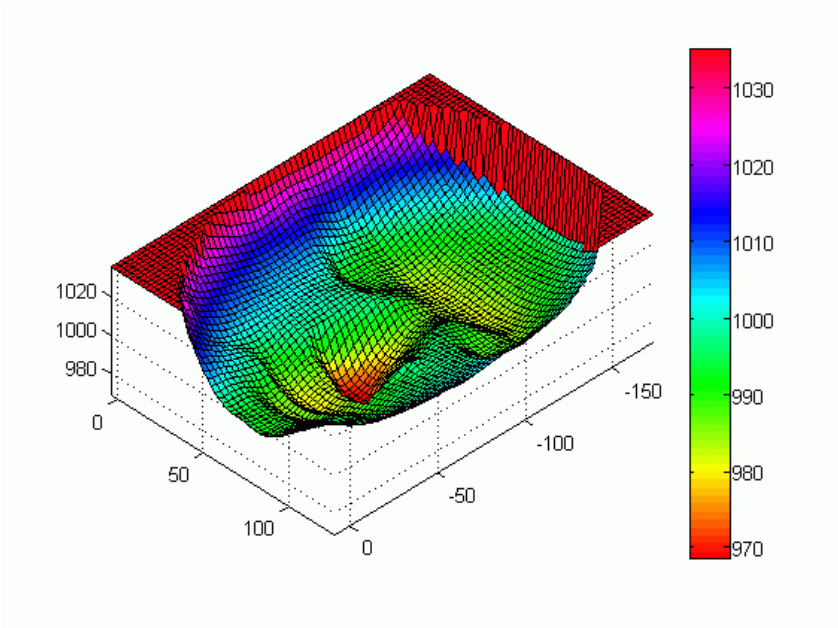


(a) **(b)**



(c) **(d)**

Figure 6.11 Reconstruction of human face. (a) Human face without pattern illumination. (b) Human face with pattern projection. (c) Grid-point detection result. (d) Interpolated surface orientation map only using the recovered surface normals at the grid-points.



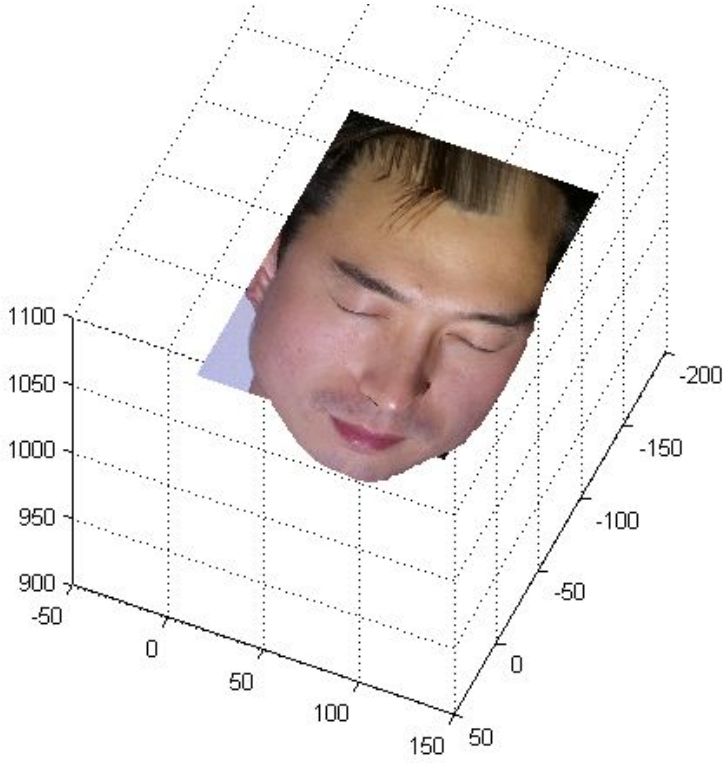


Figure 6.12 Reconstructed 3D model of a human face. (Top) Interpolated 3D model only using the reconstructed depth at the grid-points. (Bottom) 3D model with texture mapping.

6.5 Summary

In this chapter, we have proposed a 2D pseudorandom rhombic color pattern which can determine both surface position and orientation at the grid-point independently. As opposed to the use of the pattern element centroids, the grid-points between pattern elements are defined as the feature points and they are extracted via the proposed feature detector as we have introduced in Chapter 4. With this design, the density of feature points can be doubled with respect to previous patterns that used the centroid as the feature point. The grid-point also with much higher location precision since it is insensitive to image noise and perspective projection. In addition, we show that not only are the grid-points useful, the grid-lines among the pattern elements can also be used to derive surface orientation information about the imaged object. The 3D depth information as determined from the grid-points, and the 3D orientation as determined from the grid-lines, are information independently recovered, and they together would allow the 3D shape of the imaged object to be revealed in more complete form. Perspective projection models and precise correspondence are used in the orientation recovery mechanism, which makes it with high accuracy far beyond previous methods.

Experiment of surface position recovery on a plane shows an absolute mean error of 0.131 mm or 0.015%. An average discrepancy of 0.69˚ and a standard deviation of 0.13˚ are reported in the surface normal recovery experiment. Both are more accurate than previous work. Extensive experiments have confirmed the feasibility of the mechanism and probed the reconstruction quality of different objects and human face.

Chapter 7

Coding-Free Mechanism: Projection of Merely a Regular Pattern

In this chapter, we explore how coding in structured light mechanism can be made even unnecessary. We adopt the above concept of recovering surface orientation from grid-lines, and show that by the use of a regular pattern, like a rhombic binary pattern, an orientation map about the imaged object can be recovered. Specifically, we show that here the correspondences over grid-lines between the projector and the camera's image plane can be approximated with a linear mapping and in turn boost the accuracy of surface normal estimation. We go on and show that, the above orientation map can be converted to an absolute depth map via a 2-step transformation. Firstly, the image plane x-y is transformed into the object plane X-Y where weak perspective projection is assumed. Secondly, the surface integration is conducted on the new reference plane X-Y. To improve the accuracy, three colored elements are embedded into the regular binary pattern. The absolute 3D depths at their centroids are reconstructed via triangulation. Then used to compute the transformation scales s_x and s_y. A weighted procedure is also proposed to recover the absolute depth at any point (X, Y). Experiments on a variety of object surfaces and human face show that, 3D shape can be recovered with ease and promising precision.

7.1 Introduction

According to metric property of the models that reconstructed, the 3D recovery approaches can be classified into two classes: triangulation-based and integration-base. The first one provides absolute distance between the camera and objects as well as its real size, such as stereo vision [Taylor 2003], structured light-based approach [Salvi 2004], and laser strip scanning [Kawasaki 2006] etc. The second class only provides relative depth or shape without absolute scale. The correspondence problem is avoided in approaches, such as shape from shading [Horn 1989], shape from texture [Garding 1993], shape from specularity [Healey 1987], shape from polarization [Miyazaki 2002], and photometric stereo methods [Clark 1992] etc. In these approaches, surface normals are estimated from local intensity or geometry features, and then the relative 3D model can be integrated from the image gradients that inferred from surface normals. Relative depth description has the scale ambiguity in comparison with the absolute depth description, but is a shape description sufficient for many applications including model registration, object recognition, pose estimation, and segmentation [Bronstein 2005] etc.

In Chapter 2, we have given a review about previous work on surface orientation recovery methods. Compared to the image-based approaches, structured light-based approaches do not depend upon the existence of natural features, it makes less assumption on the target surface, and therefore is more robust to imaging noise or illumination condition. However, previous structured light-based methods assumed a parallel projection and image capture a parallel imaging as well. In other words, the intrinsic parameters of both the camera and projector are not considered, and errors due to the much simplified projection and imaging models are inevitable.

In this chapter, we propose a method of recovering 3D surface with metric measure or absolute scale from a single image that is captured under projection of a regular rhombic pattern. We adopt the above concept of recovering surface orientation from grid-lines. In comparison with the crude parallel models in previous work, the more accurate perspective projection model is adopted for both devices. A linear mapping is proposed to overcome the correspondence problem and to improve the accuracy of orientation determination. With this method, surface normal at each grid-point can be determined also with promising accuracy. Based on the recovered sparse and discrete orientation

map, a dense and regular orientation map can be generated for 3D integration with a proposed procedure. Furthermore, we will show that, the integrated 3D model can be transformed into the real metric space with a 2-step procedure. Firstly, the image plane *x*-*y* is transformed into the object plane *X*-*Y* where weak perspective projection is assumed. Secondly, the surface integration is conducted on the new reference plane *X*-*Y*. To improve the accuracy, three colored elements are embedded into the regular binary pattern, and the 3D depths at the centroids are reconstructed and used as the reference points.

The chapter is organized as follows. The principle of correction for discrepancy from perspective projection is presented in Section 7.2. Section 7.3 describes the transformation from surface orientation map to absolute 3D model. Experimental results on a variety of surfaces are given in Section 7.4. A summary of this chapter is offered in Section 7.5.

7.2 Correction for Discrepancy from Perspective Projection

As we have described in Chapter 6, the precise correspondence between camera and projector can boost the accuracy of surface normal determination. Moreover, in that design, a 2D color pattern is used, and the correspondence problem can be solved by the unique codeword. In contrary, the pattern used in this chapter is merely a regular pattern without codeword inside. That makes the precise correspondences unavailable.

Compared with the perspective projection model, the projector illumination process cannot be exactly captured by the parallel projection model. And discrepancy from it will cause bigger error to the determination of $\mathbf{n}_{p1}$ and $\mathbf{n}_{p2}$ at positions farther away from the principal point of the projector panel. In this work, a simple correction mechanism is proposed to fix the problem. For each grid-point $\mathbf{p}_c$ (=(u_c, v_c)) in the image plane, we adopt a linear mapping, as described below, to estimate more finely the associated grid-point $\mathbf{p}_p$ (=(u_p, v_p)) in the projector panel:

$$[u_p, v_p] = [\frac{u_c - u_{\min}}{u_{\max} - u_{\min}} W_p, \frac{v_c - v_{\min}}{v_{\max} - v_{\min}} H_p] \tag{7.1}$$

where ($u_{\min}$, $v_{\min}$), ($u_{\max}$, $v_{\max}$) represent the top-left and bottom-right grid-point positions in the image plane, W_p and H_p are the width and height of the pattern on the projector panel counted in projector panel's pixels. This way, a rough estimation of the correspondence under the perspective projection model is used to determine $\mathbf{n}_{p1}$ and $\mathbf{n}_{p2}$. And the improvement of determination accuracy can be obtained as shown in the experimental section.

7.3 Absolute 3D Determination from Normal Integration

Let $\boldsymbol{I}(x, y)$ be the observed image intensity and $z(x, y)$ be the unknown surface height above the x-y plane. The shape from shading problem can be formulated as follow:

$$I(x,y) = \Re(p,q,\beta,l,\rho) \tag{7.2}$$

where

$$p = \partial z/\partial x, q = \partial z/\partial y \tag{7.3}$$

are the surface tangents in x and y directions respectively. And β is the illumination direction vector, l is the vector from camera to object surface, and ρ is the albedo of the surface material or intrinsic reflectivity.

Surface normal at given point (x, y) can be represented by:

$$n(x,y) = (p \quad q \quad -1) \tag{7.4}$$

It has been observed that shape from shading can be expressed as a problem of solving a first-order nonlinear partial differential equation in x and y. In deriving iterative solutions to Equation (7.2) by calculus of variations, it appears to be much more straightforward to solve for surface orientation than to solve directly for z. A reasonable consistency constrain to place on the surface slopes is that they be integrable, where integrability can be defined as:

$$z_{xy}(x,y) = z_{yx}(x,y) \tag{7.5}$$

They correspond to a surface with second partial derivatives independent of the order of differentiation. As a result, the surface height at any particular point is independent of

the path of integration. This can also be thought of as smoothness constrain adopted in many shape from shading calculus, such as [Brooks 1985], [Frankot 1988], [Horn 1989], and [Zhang 1999a] etc.

Such constrains in the iterative process are avoided in our method. Since the absolute surface orientation can be determined directly from the image tangents and system geometry parameters as we have shown in Chapter 6. The main difference is that, only the surface normal at the grid-point can be determined. As a result, only a sparse and discrete orientation map can be recovered. To facilitate the integration, a median filtering is introduced to remove the outliers among them that mostly exist around the image boundary. Then the discrete data go through bilinear interpolation to form regular and dense surface orientation map that can be used by previous 3D integration methods.

7.3.1 From Image Plane *x-y* to Object Plane *X-Y*

Traditional surface integration is conducted on the image plane *x-y*. So decided integrated depth value *z* also referenced to *x-y* plane. To makes the integrated depth with real metric measure, we have to move the reference plane from the image plane *x-y* to the real object plane or *X-Y* plane. Figure 7.1 shows three different projection models: perspective, orthographic, and weak perspective (or scaled orthographic) projection model. Perspective projection model is more reasonable and accurate, but it needs accurate *Z* value to decide the object plane. In the model of orthographic projection, its field of view is usually very limited. Moreover, both the camera and the projector in our structured light system are based perspective projection models. It not suitable either. As a good approximation to perspective model, weak perspective model can be considered as a combination of perspective and orthographic models. It based on two assumptions: (1) depth range ΔZ is small, with respective to the average distance $\overline{Z}$ from the camera; and (2) the object is close to the optical axis, with respective to the average distance $\overline{Z}$ from the camera. Both can be satisfied in our system.

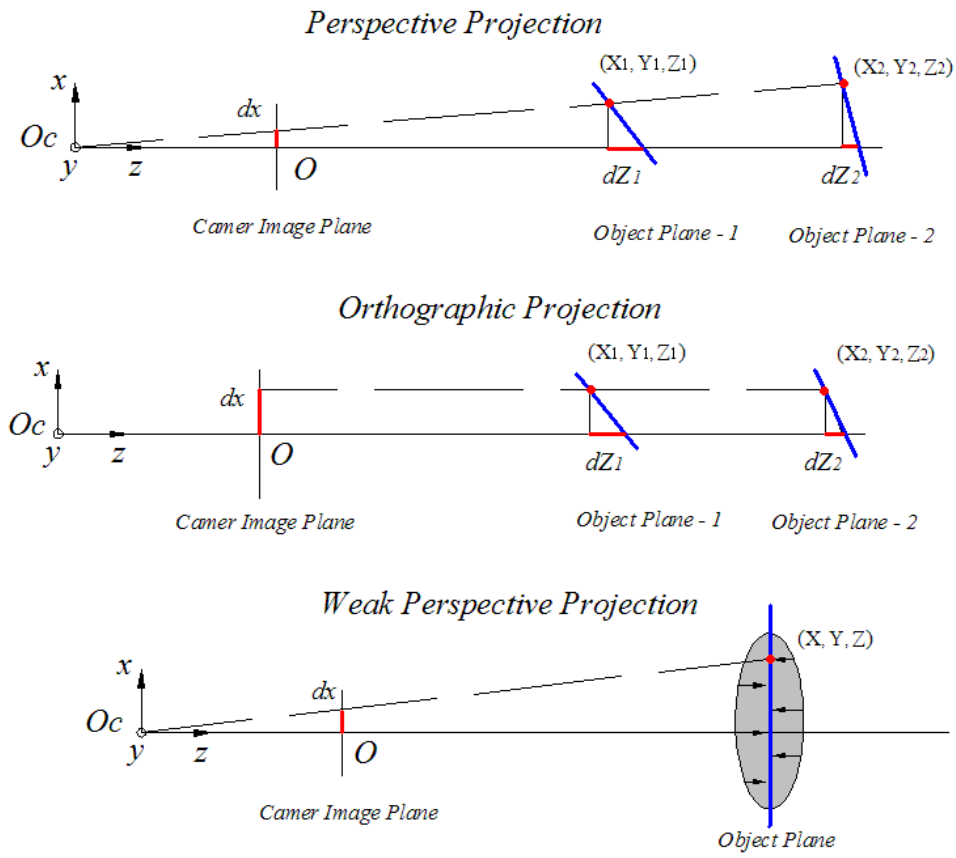


Figure 7.1 Illustrations of three projection models: perspective, orthographic and weak perspective.

The transformation from image plane *x*-*y* to object plane *X*-*Y* under weak perspective projection can be formulated as:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{7.6}$$

We can get:

$$\begin{cases} X = \dfrac{x}{s_x} \\ Y = \dfrac{y}{s_y} \end{cases} \tag{7.7}$$

Since image point (x, y) on the image plane x-y and object point (X, Y) have the same reference point – O, the principle point of camera, that makes the determination of s_x and s_y simple. In theory, only one correspondence between x-y and X-Y is enough to calculate s_x and s_y. However, to boost the accuracy, multiple corresponding points are needed. Suppose there are m correspondences, e.g.

$$\{(x_i, y_i) \quad (X_i, Y_i, Z_i)\} \;\; i = 0,1,\ldots, m-1$$

Using simple average measure, s_x and s_y can be solved by:

$$\begin{cases} s_x = \dfrac{1}{m}\displaystyle\sum_{i=0}^{m-1} \dfrac{x_i}{X_i} \\ s_y = \dfrac{1}{m}\displaystyle\sum_{i=0}^{m-1} \dfrac{y_i}{V_i} \end{cases} \tag{7.8}$$

When we get the transformation scales s_x and s_y, image plane x-y can be scaled to the object plane X-Y, where an orthographic projection is assumed and then are used as the reference plane for surface integration.

7.3.2 3D Recovery from Surface Orientation Map

The interpolated regular surface orientation map is not guarantied to be integrable. To generate an integrable gradient map $(\tilde{p}, \tilde{q})$ from the original possibly un-integrable orientation map $(\hat{p}, \hat{q})$, where

$$\partial\tilde{p}/\partial Y = \partial\tilde{q}/\partial X \tag{7.9}$$

It is also known as the integrability constrain. A minimization of distance measure with form:

$$E = \iint (|\tilde{p} - p|^2 + |\tilde{q} - q|^2) dXdY + \lambda \iint (p_X^2 + p_Y^2 + q_X^2 + q_Y^2) dXdY \tag{7.10}$$

can be used to estimate $(\tilde{p}, \tilde{q})$, where the second term indicates a smooth constrain with a scale λ. The minimum distance property makes this an orthogonal projection. The task of finding such a projection is simplified if we can represent the surface slopes by a finite set of integrable basis functions that satisfy Equation (7.9).

Suppose that we represent the surface $Z(X, Y)$ by the functions $\phi(X,Y,\boldsymbol{\omega})$ as:

$$Z(X,Y) = \sum_{\omega \in \Omega} C(\boldsymbol{\omega})\phi(X,Y,\boldsymbol{\omega}) \tag{7.11}$$

where $\boldsymbol{\omega} = (\omega_X, \omega_Y)$ is a 2D index, Ω is a finite set of indexes, and the members of $\{\phi(X,Y,\boldsymbol{\omega})\}$ are not necessarily mutually orthogonal. If each $\phi(\boldsymbol{\omega})$ satisfies Equation (7.9), then it follows that Z can also be expressed in terms of this expansion, giving

$$\begin{aligned} p &= \sum_{\boldsymbol{\omega} \in \Omega} C(\boldsymbol{\omega})\phi_X(X,Y,\boldsymbol{\omega}) \\ q &= \sum_{\boldsymbol{\omega} \in \Omega} C(\boldsymbol{\omega})\phi_Y(X,Y,\boldsymbol{\omega}) \end{aligned} \tag{7.12}$$

Suppose that the members of $\phi_X(X,Y,\boldsymbol{\omega})$ and $\phi_Y(X,Y,\boldsymbol{\omega})$ are mutually orthogonal. Then the coefficients for the expansions of $\hat{p}$ and $\hat{q}$ such that

$$\begin{aligned} \hat{p} &= \sum_{\boldsymbol{\omega} \in \Omega} \hat{C}_1(\boldsymbol{\omega})\phi_X(X,Y,\boldsymbol{\omega}) \\ \hat{q} &= \sum_{\boldsymbol{\omega} \in \Omega} \hat{C}_2(\boldsymbol{\omega})\phi_Y(X,Y,\boldsymbol{\omega}) \end{aligned} \tag{7.13}$$

The expansion coefficients $C(\boldsymbol{\omega})$ that minimize Equation (7.11) given a possibly non-integrable estimate of surface slopes $(\tilde{p},\tilde{q})$ are given by:

$$\tilde{C}(\boldsymbol{\omega}) = \frac{P_X(\boldsymbol{\omega})\hat{C}_1(\boldsymbol{\omega}) + P_Y(\boldsymbol{\omega})\hat{C}_2(\boldsymbol{\omega})}{P_X(\boldsymbol{\omega}) + P_Y(\boldsymbol{\omega})} \tag{7.14}$$

for each $\boldsymbol{\omega} \in \Omega$, where

$$\begin{aligned} P_X(\boldsymbol{\omega}) &= \iint \left|\phi_X(X,Y,\boldsymbol{\omega})\right|^2 dXdY \\ P_Y(\boldsymbol{\omega}) &= \iint \left|\phi_Y(X,Y,\boldsymbol{\omega})\right|^2 dXdY \end{aligned} \tag{7.15}$$

The integrated surface $\tilde{Z}(X,Y)$ and integrable surface slopes $(\tilde{p},\tilde{q})$ are then obtained by substituting $\tilde{C}(\boldsymbol{\omega})$ into Equation (7.11) and (7.12).

The usefulness of the expansion depends on the selections of basis functions. Frankot and Chellappa proposed to use Fourier basis functions as:

$$\phi(X,Y,\boldsymbol{\omega}) = \exp(j\omega_X X + j\omega_Y Y) \tag{7.16}$$

Since it is particular computationally and forms a complete orthornormal basis for *Z*. Using this basis, the surface can be represented by:

$$\widetilde{Z}(X,Y) = \sum_{\boldsymbol{\omega}\in\Omega} \widetilde{C}(\boldsymbol{\omega}) \exp\{j\boldsymbol{\omega}.(X,Y)\} \tag{7.17}$$

where $\widetilde{C}$ are the coefficient of the Fourier series expansion of $\widetilde{Z}$. For images of size $N\times N$, a reasonable choice for Ω would be $2\pi n \times 2\pi m$, with $n,m \in \{0,1,...,N-1\}$.

Let $\hat{C}_X, \hat{C}_Y, \widetilde{C}_X, \widetilde{C}_Y$ be the Fourier coefficients for $\hat{p}, \hat{q}, \widetilde{p}, \widetilde{q}$, respectively.

Then $\widetilde{C}$ can be represented by:

$$\widetilde{C}(\boldsymbol{\omega}) = \frac{-j\omega_X \hat{C}_X(\boldsymbol{\omega}) - j\omega_Y \hat{C}_Y(\boldsymbol{\omega})}{\omega_X^2 + \omega_Y^2} \tag{7.18}$$

with the Fourier coefficients of the constrained surface slopes as:

$$\begin{aligned} \widetilde{C}_X(\boldsymbol{\omega}) &= j\omega_X \widetilde{C}(\boldsymbol{\omega}) \\ \widetilde{C}_Y(\boldsymbol{\omega}) &= j\omega_Y \widetilde{C}(\boldsymbol{\omega}) \end{aligned} \tag{7.19}$$

In general, the Fourier coefficients of the raw slope estimates are combined in proportion to their relative frequencies.

For computer implementation, *Discrete Fourier Transform* (*DFT*) can be used. Fourier coefficients $\widetilde{C}$ in Equation (7.18) can be reformulated to:

$$\widetilde{C}(\boldsymbol{\omega}) = \frac{\alpha_X(\omega_X)\hat{C}_X(\boldsymbol{\omega}) + \alpha_Y(\omega_Y)\hat{C}_Y(\boldsymbol{\omega})}{\alpha_X^2 + \alpha_Y^2} \tag{7.20}$$

with

$$\widetilde{C}_X(\boldsymbol{\omega}) = \alpha_X(\omega_X)\widetilde{C}(\boldsymbol{\omega})$$
$$\widetilde{C}_Y(\boldsymbol{\omega}) = \alpha_Y(\omega_Y)\widetilde{C}(\boldsymbol{\omega}) \tag{7.21}$$

where α_X and α_Y are the Fourier coefficients of the appropriate discrete differentiation operator in *X*-*Y*. Then, the estimated surface $\widetilde{Z}(X,Y)$ can be constructed by performing the inverse *DFT* of $\widetilde{C}(\boldsymbol{\omega})$. More details about the proof and deduction can be found in [Frankot 1988]. A method based on solving the Eikonal equation by the use of Fast Marching Method is also proposed by Jeffrey et al. recently [Ho 2006].

Since the reference plane has been moved to the object plane *X*-*Y*, the integrated depth also with metric meanings. However, too boost the accuracy of 3D recovery, a weighted procedure is proposed in following section.

7.3.3 A Weighted Procedure for Absolute Depth Recovery

Since the surface has been assumed integrable, we have:

$$dZ(X,Y) = pdX + qdY \tag{7.22}$$

where *Z* represents the absolute depth in object plane *X*-*Y*.

The integration from any reference point (X_0, Y_0) to any point (X, Y) can be formulated as:

$$Z(X,Y) = \underbrace{\int_{(X_0,Y_0)}^{(X,Y)} (pdX + qdY)}_{\overline{Z}(X,Y)} + Z(X_0,Y_0) \tag{7.23}$$

It shows the case where only one reference point (X_0,Y_0) is used.

From Equation (7.23), we can see that any absolute depth value $Z(X,Y)$ can be represented by an integrated term $\overline{Z}(X,Y)$ and the depth $Z(X_0,Y_0)$ at a given reference point (X_0,Y_0). In other words, if we get $Z(X_0,Y_0)$, we can calculate the absolute depth $Z(X,Y)$ at any object point (X,Y).

Given a number of reference points (X_i,Y_i) with known absolute depth $Z_i(X_i,Y_i)$, where $i = 0,1,\ldots,m-1$, we have the following equation:

$$Z(X,Y) = \int_{(X_i,Y_i)}^{(X,Y)} (pdX + qdY) + Z(X_i,Y_i) \tag{7.24}$$

The above equation can be expanded as follow:

$$Z(X,Y) = \underbrace{\int_{(X_0,Y_0)}^{(X,Y)} (pdX + qdY)}_{\overline{Z}(X,Y)} - \underbrace{\int_{(X_0,Y_0)}^{(X_i,Y_i)} (pdX + qdY)}_{\overline{Z}(X_i,Y_i)} + Z(X_i,Y_i) \tag{7.25}$$

Obviously, $\forall i \neq 0$, $\overline{Z}(X_i,Y_i)$ can be determined without referring to any point (X,Y).

Suppose *m* reference points are available, we can get the following equation groups:

$$\forall (X,Y) \qquad \begin{aligned} Z(X,Y) &= \overline{Z}(X,Y) - \overline{Z}(X_0,Y_0) + Z(X_0,Y_0) \\ Z(X,Y) &= \overline{Z}(X,Y) - \overline{Z}(X_1,Y_1) + Z(X_1,Y_1) \\ &\vdots \\ Z(X,Y) &= \overline{Z}(X,Y) - \overline{Z}(X_{m-1},Y_{m-1}) + Z(X_{m-1},Y_{m-1}) \end{aligned} \tag{7.26}$$

Add a distance weight w_i to $\overline{Z}(X_i,Y_i)$ in Equation (7.25), we can get:

$$Z(X,Y) = \overline{Z}(X,Y) + Z(X_i,Y_i) - w_i\overline{Z}(X_i,Y_i)$$

$$w_i = \frac{1/d_i(X,Y)}{\sum_{i=0}^{m-1} 1/d_i(X,Y)} \tag{7.27}$$

where $d_i(X,Y)$ indicates the distance between any point (X,Y) and the reference point (X_i,Y_i).

Substitute Equation (7.27) into Equation (7.26) and accumulate the all equations in Equation (7.26), we have:

$$Z(X,Y) = \overline{Z}(X,Y) + \frac{1}{m}\sum_{i=0}^{m-1}\left[Z(x_i,y_i) - w_i\overline{Z}(x_i,y_i)\right] \tag{7.28}$$

In practice, we can embed three colored pattern elements (for easy identify, we used red, green, and blue) into the binary pattern as shown in Figure 7.2.

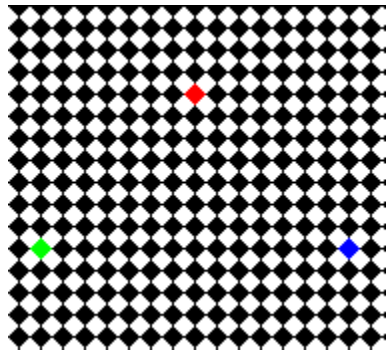


Figure 7.2 A regular rhombic pattern with three colored elements embedded.

The absolute depth at their centroids can be determined via triangulation and indicated by:

$$\{Z(X_r,Y_r) \quad Z(X_g,Y_g) \quad Z(X_b,Y_b)\}$$

Consequently, Equation (7.28) can be expanded as:

$$\begin{aligned} Z(X,Y) = \bar{Z}(X,Y) &+ \frac{1}{3}\left[Z(X_r,Y_r) + Z(X_g,Y_g) + Z(X_b,Y_b)\right] \\ &- \frac{1}{3}\left[w_r\bar{Z}(X_r,Y_r) + w_g\bar{Z}(X_g,Y_g) + w_b\bar{Z}(X_b,Y_b)\right] \end{aligned} \tag{7.29}$$

$$\begin{aligned} w_r &= \frac{1/d_r(X,Y)}{\left[1/d_r(X,Y) + 1/d_g(X,Y) + 1/d_b(X,Y)\right]} \\ w_g &= \frac{1/d_g(X,Y)}{\left[1/d_r(X,Y) + 1/d_g(X,Y) + 1/d_b(X,Y)\right]} \\ w_b &= \frac{1/d_b(X,Y)}{\left[1/d_r(X,Y) + 1/d_g(X,Y) + 1/d_b(X,Y)\right]} \end{aligned} \tag{7.30}$$

7.4 Experimental Results

In this section, the accuracy of the proposed method is evaluated by projecting on a planar surface and spherical object. A bust model and a real human face are used as the last experiment. The system was calibrated using the LCD pattern as described in Chapter 3.

7.4.1 Accuracy of Surface Orientation Recovery

A planar surface with a distance of about 850 mm was used to evaluate the accuracy of the proposed surface orientation recovery method. Grid-points in a region about 100×100 mm were reconstructed. For all the measured points, only an average discrepancy of 0.82˚ and standard deviation of 0.15˚ in the distribution of local surface orientations are obtained. Compared with the results in Section 6, which reports an average discrepancy of 0.69˚ and a standard deviation of 0.13˚, the deviation is slightly. It shows that the proposed linear mapping is accurate enough to overcome the discrepancy from perspective projection without precise correspondence.

7.4.2 Spherical Reconstruction

We have also conducted experiment on a spherical object of the radius about 97 mm. Before integration, the scale value s_x and s_y, which transform the image plane to object plane are calculated. As shown in Figure 7.3, $(x_r, y_r),(x_g, y_g),(x_b, y_b)$ are the centroids of red, green, and blue pattern element. And their corresponding object points are denoted by $(X_r, Y_r, Z_r),(X_g, Y_g, Z_g),(X_b, Y_b, Z_b)$ respectively.

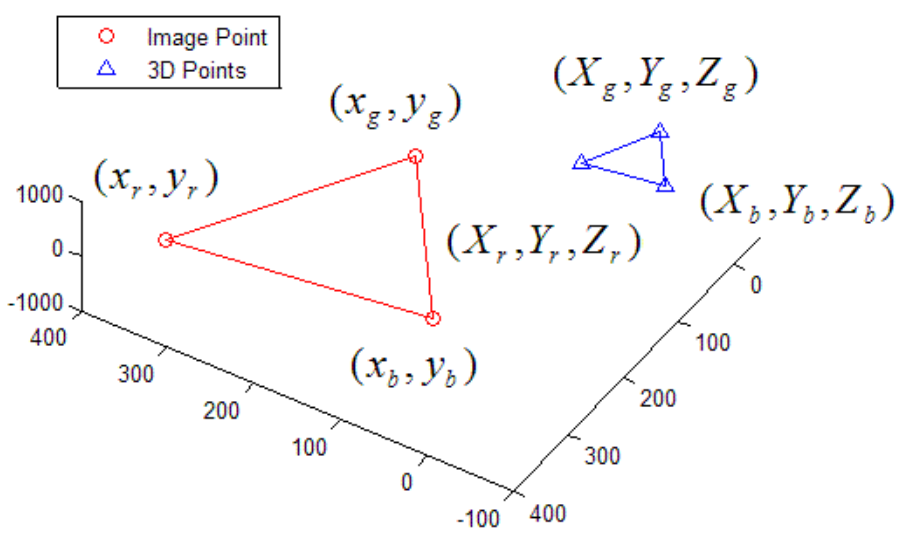


Figure 7.3 Calculation of s_x and s_y, which transforms image plane *x-y* to object plane *X-Y*.

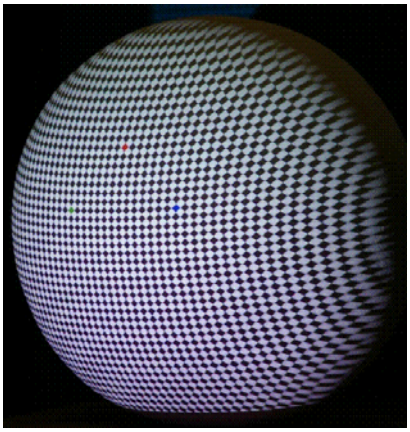
Using Equation (7.8), three values of s_x and s_y can be calculated. The results are:

$$s_x = [0.316, 0.318, 0.311],\ s_y = [0.324, 0.327, 0.321]$$

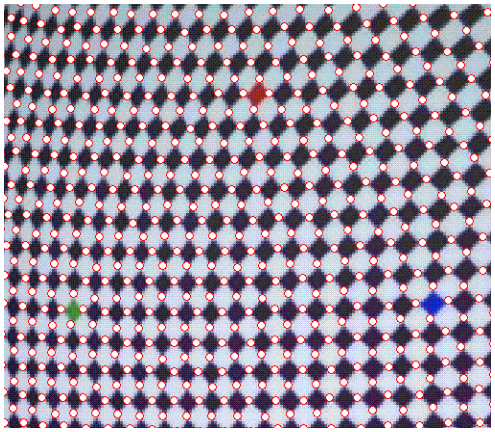
We can see that the deviations in each group are relative small. That shows the transformation is accurate and robust.

The shape reconstruction result by the proposed method is as shown in Figure 7.4: Figure 7.4 (a) shows the object under illumination of the rhombic pattern, Figure 7.4 (b) shows the appearance of grid-points detected in the image data and Figure 7.4 (c) shows the determined local surface orientations map at these grid-points. The regular data was generated by interpolating the recovered discrete orientation map and then integrated for 3D shape reconstruction as shown in Figure 7.4 (d).

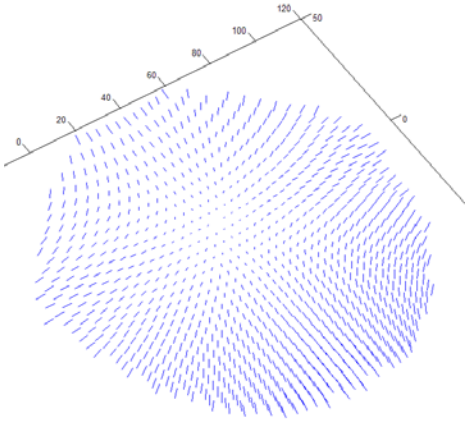
Centroids of the three colored elements are extracted and matched to reconstruct their absolute 3D coordinates. The generated 3D model with absolute scale is as shown in Figure 7.5. A sphere was fitted in the least square error sense. The radius of the fitted sphere is 94.95 mm with a standard deviation of 1.82 mm or 0.23% with respect to the working distance. The estimated sphere center position is [-20.9 35.6 897.9] mm with respect to the camera reference frame.



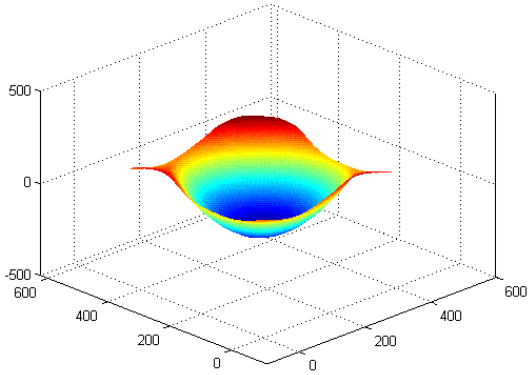
(a)



(b)



(c)



(d)

Figure 7.4 Reconstruction of a spherical surface by the proposed method. (a) Object surface under pattern illumination. (b) Grid-points detection result. (c) Reconstructed local surface orientations at the grid-points. (d) Integrated 3D surface without using reference points.

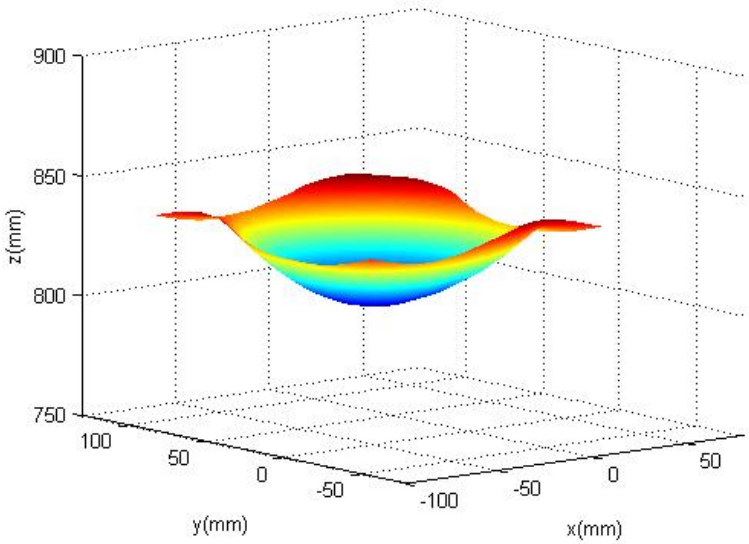
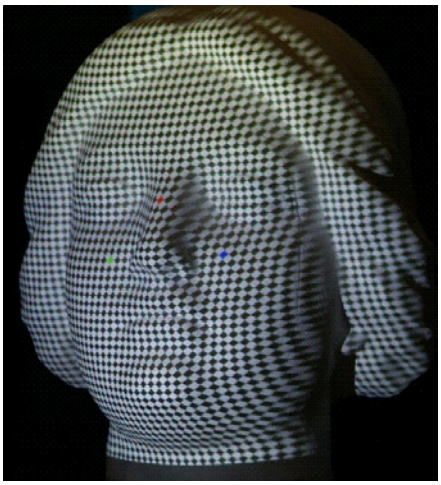


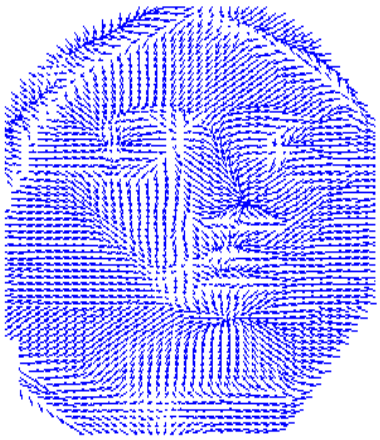
Figure 7.5 Reconstructed 3D model with absolute scale by the proposed method.

7.4.3 Reconstructions of Bust and Human Face

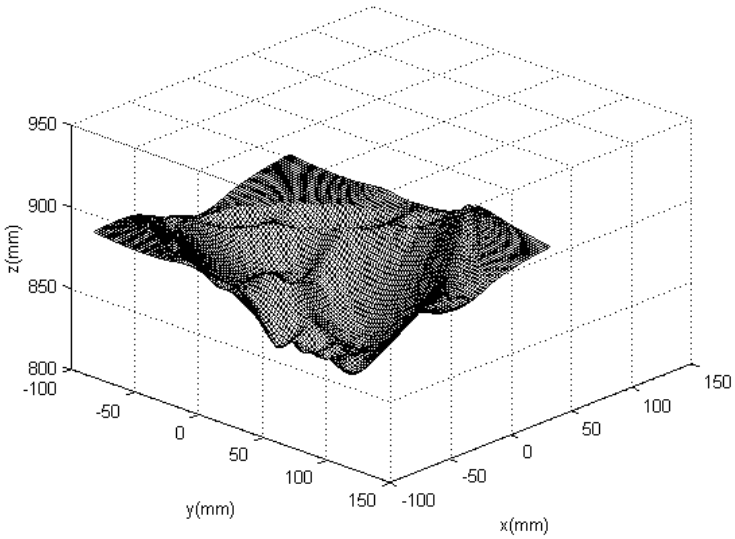
A bust was reconstructed and the results are shown in Figure 7.6. A more challenge experiment is conducted with a real human face where texture and occlusion exists. Due to the robust feature detector and surface normal determination mechanism, the surface orientation map can be constructed successfully as shown in Figure 7.7. Though ground truths of these surfaces were not available for evaluating in precise terms the reconstruction quality, visual check showed that the reconstruction was reasonable.



(a)

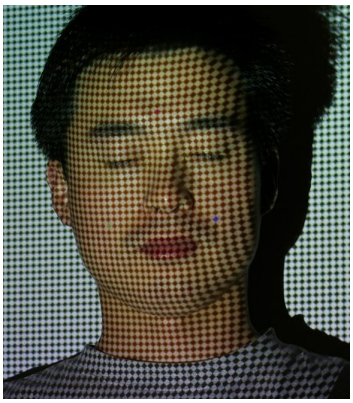


(b)

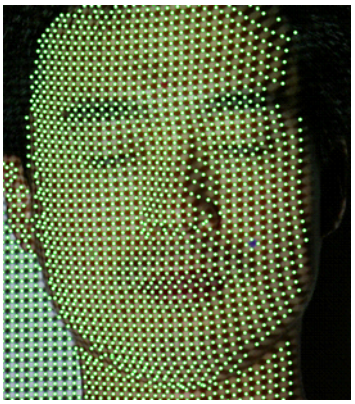


(c)

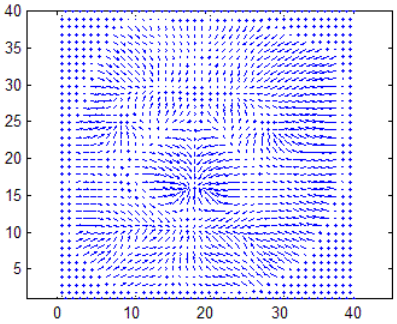
Figure 7.6 Reconstruction of a bust model by the proposed method. (a) Bust surface under pattern illumination. (b) Interpolated orientation map. (c) Reconstruction result with in absolute scale.



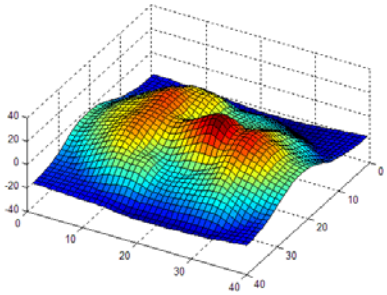
(a)



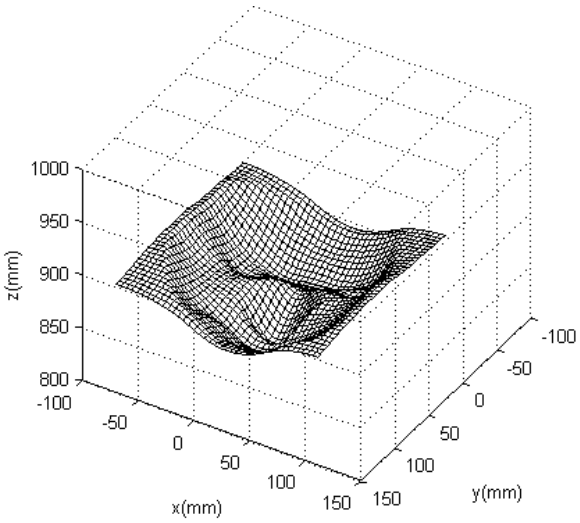
(b)



(c)



(d)



(e)

Figure 7.7 Reconstruction of a human face by the proposed method: (a) human face under pattern illumination; (b) grid-point detection results; (c) interpolated surface orientation map; (d) integrated 3D model; (e) reconstructed human face with metric measure.

7.5 Summary

We adopt the concept of recovering surface orientation from grid-lines, and show that by the use of a rhombic binary pattern, an orientation map about the imaged object can be recovered. Specifically, we show that here the correspondences problem can be avoided by making one specific design and one approximation. We go on and show that, as long as no less than one reference point on the imaged object is available where absolute 3D is known, the above orientation map can even be converted to an absolute depth map via a 2-step algorithm. First, the image plane *x*-*y* is transformed to the object plane *X*-*Y* where weak perspective projection is assumed. Secondly, the surface integration is conducted on the new reference plane *X*-*Y*. To boost the reconstruction accuracy, several colored elements are embedded into the binary pattern, and their centroids are reconstructed and used as the reference points. In the planar experiment, an average discrepancy of only 0.82˚ and a standard deviation of 0.15˚ are reported. A sphere with radius of about 97 mm is used for its accuracy evaluation. The radius of the fitted sphere is 94.95 mm with a standard deviation of 1.82 mm or 0.23% with respect to the working distance.

Certainly, with only one image capture as the input, with no coding injected into the projected pattern, the accuracy of 3D reconstruction is behind triangulation-based mechanisms. However, the need of a single image and the avoidance of the use of color make it attractive for certain applications that demand only approximate shape of the imaged object.

Chapter 8

Conclusion and Future Work

This chapter presents the conclusion and some perspectives opened by this work. The contributions of the thesis are first discussed. The list of publications related to this work is then presented. Finally, future possible works beyond this research are pointed out.

8.1 Conclusion and Contributions

This thesis has focused on a structured light system that can recover both surface position and orientation of the inspected object surface via single pattern projection. Three key issues in a structured light system: calibration, feature detection, and codification have been involved in this work. All the contributions in this work can be enumerated as follows.

The first contribution in this work is a comprehensive survey of previous work on 3D recovery as we have addressed in Chapter 2. The survey is conducted from two aspects: 3D position recovery and 3D orientation recovery. A number of 3D position recovery methods are introduced, and the emphasis is put on structured light-based systems. The coding scheme is classified into temporal and spatial. The main coding strategies and recent progress are reported. In the review of surface orientation recovery, we classified them into two classes: image-based and structured light-based methods. In the first class, we introduced the methods like shape from shading, polarization, and photometric stereo. Several structured light-based methods are also presented for comparison.

The second contribution is that we have proposed a novel calibration means for the projector-camera structured light system. The calibration system makes use of an LCD panel as the calibration plane (or named LCD pattern). The patterns shown on the panel are used for camera calibration. When we turn it off, it can be used as a projector screen for the calibration of projector. The result shows that, with the new calibration design, the projector-camera system can be calibrated in far less images with much higher accuracy. In our experiments, re-projection error of only 0.1 pixel was achieved in the camera calibration using only 8 images. The root mean square error in the planarity check was found to be only 0.04 mm, and an absolute mean deviation of only 0.077 mm or 0.01% of the depth range was obtained when measuring motion distance of 1 mm. Both indicate significant improvement over those of the traditional calibration setups.

The third contribution is that we have proposed a novel grid-point detector. Grid-point in the pattern owns the two-fold rotation symmetry, or so-called *cmm* symmetry, which is largely preserved on the image side after perspective projection and imaging. We propose a grid-point detector that exploits such a symmetry. By avoiding the direct use of raw image intensity, the detector is less sensitive to image noise and surface texture. A template is first conducted to select potential grid-points and speed up the whole process. Then correlation coefficient between the mask and its rotational mask is defined as the response value. At last, a refining step is performed to obtain the grid-points with sub-pixel accuracy. Experiment on noisy image shows that the operator has good localization accuracy and is insensitive to image noise. Experiments on color patterns and generic surfaces show its good applicability.

The fourth contribution is that we have developed a low-cost and high-accuracy dense 3D scanning system. The 3D reconstruction operation is the traditional Gray code pattern based, but augmented by the use of strip shifting to enhance the resolution of 3D reconstruction. Specifically, while strip shifting allows high resolution pattern information to be projected to the target scene or object, the Gray code projection removes the periodic ambiguity in the pattern. We also provide a way of allowing the strip edges to be located with sub-pixel accuracy, thereby raising the reconstruction quality. Combined with the novel calibration method, a high accuracy ranging system can be composed with consumer level projector and camera.

The fifth contribution is the proposal of a pseudorandom rhombic color pattern, which can determine both surface position and orientation at the grid-point independently. As opposed to the use of the pattern element centroids, the grid-points between pattern elements are defined as the feature points and they are extracted via the proposed feature detector as we have introduced in Chapter 4. With this design, the density of feature points can be doubled with respect to previous patterns that used the centroid as the feature point. The grid-point also with much higher location precision since it is insensitive to image noise and perspective projection. In addition, we show that not only are the grid-points useful, the grid-lines among the pattern elements can also be used to derive surface orientation information about the imaged object. The 3D position information as determined from the grid-points, and the 3D orientation as determined from the grid-lines, are information independently recovered, and they together would allow the 3D shape of the imaged object to be revealed in more complete form. Perspective projection models and precise correspondence are used in the orientation recovery mechanism, which makes it with high accuracy far beyond previous methods. Surface position recovery experiment on a planar surface shows an absolute mean error of only 0.131 mm or 0.015%. An average discrepancy of 0.69˚ and a standard deviation of only 0.13˚ are reported in the surface normal recovery experiment. Both are more accurate than previous work. Extensive experimental results have confirmed the feasibility of the mechanism and probed the reconstruction quality of different objects and human face.

The last contribution is that we have proposed coding-free structured light pattern for absolute 3D depth reconstruction. We adopt the concept of recovering surface orientation from grid-lines, and show that by the use of a regular pattern, like a binary rhombic pattern, an orientation map about the imaged object can be recovered. Specifically, we show that here the correspondences problem can be avoided by making one specific design and one approximation. We go on and show that, as long as no less than one reference point on the imaged object is available where absolute 3D is known, the above orientation map can even be converted to an absolute depth map by a simple integration process. To boost the reconstruction accuracy, several colored elements are embedded into the binary pattern, and their centroids are reconstructed and used as the reference points. In the planar experiment, an average discrepancy of only 0.82˚ and a standard

deviation of 0.15˚ are reported. A sphere with radius of about 97 mm is used for its accuracy evaluation. The radius of the fitted sphere is 94.95 mm with a standard deviation of 1.82 mm or 0.23% with respect to the working distance. Certainly with only one image capture as the input, with no coding injected into the projected pattern, the accuracy of 3D reconstruction is behind those of the time-coded (as proposed in Chapter 5) and spatially coded (as proposed in Chapter 6) mechanisms. However, the need of a single image and the avoidance of the use of color make it attractive for certain applications that demand only approximate shape of the object.

8.2 Related Publications

[1] Z. Song and R. Chung et al. Surface Orientation Recovery of Specular Micro-surface via Binary Pattern Projection. *Proceedings of the SPIE, Machine Vision Applications in Industrial Inspection* XIV, Vol. 6070, pp. 254-262. 2006.

[2] Z. Song and R. Chung. An Accurate Camera-Projector System Calibration by the use of an LCD Panel. *Proceedings of the 2007 International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV),* pp. 164-170. 2007.

[3] Z. Song and R. Chung. 3D Shape Recovery by the use of Single Image plus Simple Pattern Illumination. *Proceedings of 3rd International Symposium on Visual Computing (ISVC)*, pp. 268-277. 2007.

[4] Z. Song and R. Chung. Grid-point Extraction Exploiting Point Symmetry in a Pseudorandom Color Pattern. *International Conference on Image Processing (ICIP)*. 2008.

[5] Z. Song and R. Chung. Determining both 3D Position and Orientation via the Projection of a Rhombic Pseudorandom Pattern. *British Machine Vision Conference (BMVC)*. 2008.

[6] Z. Song and R. Chung. Use of LCD Panel for Calibrating Structured Light-based Range Sensing System. *IEEE Trans. on Instrumentation and Measurement.* In press. 2008.

[7] Z. Song and R. Chung. Off-the-shelf Structured Light-based System for Accurate 3D Reconstruction. *HKIE Trans.* Accepted. 2008.

[8] Z. Song and R. Chung. Use of Grid-point Features in Pseudorandom Color Pattern Based Structured Light System. *IEEE Trans. on Image Processing*. Submitted.

[9] Z. Song and R. Chung. Absolute 3D Recovery via Single Illumination of a Rhombic Pattern. For submission.

[10] Z. Song and R. Chung. Explicit 3D Recovery via Single Illumination of Pseudorandom Color Pattern. For submission.

8.3 Future Work

The extension of this work focused on four aspects: (1) further improvement of the proposed calibration design by considering the thickness of the LCD's substrate glass; and (2) optimization and automation of the grid-point extraction algorithm; and (3) how to fully utilize the independent surface position and orientation that have been recovered; and (4) system efficiency improvement including synchronization of a video camera and projector, and algorithm optimization.

In the proposed calibration mechanism, the effect from the substrate glass (with a thickness around 0.5 mm) of the LCD screen was not considered. The transparency and refraction induced by the glass can make slight deviation to the co-planarity of the displayed pattern and the projected pattern. Future work can address on this issue by adding a compensation value to the calibrated LCD plane. Further improvement of the calibration result to the projector is expected.

The *cmm*-symmetry based grid-point detector has been proven effective and accurate even noise and texture exist. However, some parameters in the operator need to be adjusted by experience. For example, the size of mask and template are required to be similar with the size of pattern elements in the image. If the elements are greatly distorted in the image, e.g. they are too separated or close. That may cause false detections. To overcome this disadvantage, future work should focus on adaptive selection of the mask size and some thresholds in the operator.

We have shown that, the recovered surface normal at grid-point with higher accuracy than that directly estimated from the 3D points. The future work should also focus on how to fully utilize this information. Two potential applications can make use of it. The

first one is to use the independent surface normals for precise 3D modeling. As reported in [Nehab 2005], shape from shading or photometric stereo is used to generate an independent normal field instead of computing it from measured positions. The author proved that by the use of these independently measured normal maps allows for high quality renderings from certain viewpoints. The second potential application is to generate a more accurate surface description model such as the spin image [Johnson 1999]. The spin image representation comprises a set of descriptive images associated with the oriented points on the surface of an object. These images are created by constructing a pose-invariant 2D coordinate system at an oriented point (3D point with surface normal vector) on the surface, and accumulating the coordinates of other points in a 2D histogram called a spin image [Correa 2001]. The representation assumes that surfaces are approximated by regular polygonal surface meshes, of a given mesh resolution, defined as the median length of all edges in the mesh. The spin image representation is robust to scene clutter and occlusion and therefore suitable for object recognition and surface matching problems.

The current image processing algorithms are developed under MATLAB. We plan to rewrite and optimize some kernel functions with *C* language to make it faster and more efficiency. To make a real time ranging system, the still camera we are using now should also be replaced by a video camera that is synchronized to the projector. With these improvements, a real-time structured light system can be established.

Bibliography

[Adan 2005] A. Adan, F. Molina, et al. 3D Feature Tracking using a Dynamic Structured Light System. Second Canadian Conference on Computer and Robot Vision (CRV 2005), pp.168-175. 2005.

[Albitar 2007] I. C. Albitar, P. Graebling, C. Doignon. Robust Structured Light Coding for 3D Reconstruction Computer Vision. International Conference on Computer Vision, pp. 1-6. 2007.

[Alessandro 2006] C. Alessandro, C. Claudio, S. Raimondo. 3D Face Detection Using Curvature Analysis. Pattern Recognition, Vol. 39, pp. 444-455. 2006.

[Alexander 2005] M. Alexander, M. Bronstein et al. Three-Dimensional Face Recognition. International Journal of Computer Vision, Vol. 64(1), pp. 5-30. 2005.

[Astrom 1999] K. Astrom, R. Cipolla, P. Giblin. Generalised Epipolar Constraints. International Journal of Computer Vision, Vol. 33(1), pp. 51-72. 1999.

[Asundi 1993] A. Asundi. Computer Aided Moiré Methods. Optical Lasers Engineering. Vol. 17, pp. 107-116. 1993.

[Atkinson 2006] G. Atkinson, E. R. Hancock. Recovery of Surface Orientation from Diffuse Polarization. IEEE Transactions on Image Processing, Vol. 15(6), pp.1653-1664. 2006.

[Ball 1975] J. R. Ball, A. H. Spittle and H. T. Liu. High Speed M-sequence Generation: A Further Note. Electronic Letter, Vol. 11, pp.107-108. 1975.

[Basu 2002] M. Basu. Gaussian-Based Edge-Detection Methods - A Survey. IEEE Transactions on Systems, Man, and Cybernetics (C), Vol. 32 (3), pp. 252-260. 2002.

[Beardsley 1994] P. A. Beardsley, A. Zisserman and D. W. Murray. Navigation using Affine Structure from Motion. European Conference on Computer Vision, Vol. 2, pp. 85-96. 1994.

[Bergmann 1995] D. Bergmann. New Approach for Automatic Surface Reconstruction with Coded Light. Proceedings of Remote Sensing and Reconstruction for Three-Dimensional Objects and Scenes, Vol. 2572, pp. 2-9. 1995.

[Berzins 1984] V. Berzins. Accuracy of Laplacian Edge-detectors. Computer Vision Graphic Image Process, Vol. 27, pp. 195-210, 1984.

[Bichsel 1992] M. Bichsel and A. P. Pentland. A Simple Algorithm for Shape from Shading. Computer Vision and Pattern Recognition, pp. 459-465. 1992.

[Blais 1986] F. Blais and M. Rioux. Real-Time Numerical Peak Detector. Proceeding on Signal Processing, Vol. 11, pp. 145-155. 1986.

[Blais 2004] F. Blais. Review of 20 Years of Range Sensor Development. Journal of Electronic Imaging, Vol. 13(1), pp. 231-240. 2004.

[Bouguet 2007] J. Bouguet. www.vision.caltech.edu/bouguetj/calib_doc/index.html. Matlab Camera Calibration Toolbox.

[Bronstein 2005] A. M. Bronstein, M.M. Bronstein and R. Kimmel. Three-Dimensional Face Recognition. International Journal of Computer Vision, Vol. 64(1), pp. 5-30. 2005.

[Brooks 1985] M. J. Brooks and B. K. P. Horn. Shape and Source from Shading. International Joint Conference on Artificial Intelligence, pp. 932-936. 1985.

[Brooks 1985] M. J. Brooks and B. K. P. Horn. Shape and Source from Shading. International Joint Conference on Artificial Intelligence, pp. 932-936. 1985.

[Campbell 2001] R. J. Campbell, P. J. Flynn. A Survey of Free-form Object Representation and Recognition Techniques. Computer Vision and Image Understanding, Vol. 81, pp. 166-210. 2001.

[Chen 2004] Y. F. Chen, H. T. Wang, Y. H. Wang. Pyramid Surface Reconstruction from Normal. Third International Conference on Image and Graphics, pp. 464-467. 2004.

[Chen 2000] F. Chen, G. M. Brown, and M. Song. Overview of Three-Dimensional Shape Measurement using Optical Methods. Optical Engineering, Vol. 39(1), pp. 10-22. 2000.

[Chen 2008] S. Y. Chen, Y. F. Li. Vision Processing for Realtime 3-D Data Acquisition Based on Coded Structured Light. IEEE Transactions on Image Processing, Vol. 17(2), pp. 167-176. 2008.

[Clark 1992] J. J. Clark. Active Photometric Stereo. Computer Vision and Pattern Recognition, pp. 29-34. 1992.

[Cohen 1988] J. Cohen. Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Lawrence Erlbaum Associates, Hillsdale, NJ. ISBN 0-8058-0283-5. 1988.

[Corning 2008] http://www.corning.com/

[Correa 2001] S. R. Correa, L. G. Shapiro, M. Melia. A New Signature-based Method for Efficient 3-D Object Recognition. Computer Vision and Pattern Recognition, Vol. 1, pp. 769-776. 2001.

[Demeyere 2007] M. Demeyere, D. Rurimunzu and C. Eugène. Diameter Measurement of Spherical Objects by Laser Triangulation in an Ambulatory Context. IEEE Transactions on Instrumentation and Measurement, Vol. 56(3), pp. 867-872. 2007.

[Desjardins 2007] D. Desjardins and P. Payeur. Dense Stereo Range Sensing with Marching Pseudorandom Patterns. Canadian Conference on Computer and Robot Vision, pp. 216-226, 2007.

[Etzion 1988] T. Etzion. Construction for Perfect Maps and Pseudorandom Arrays. IEEE Transaction on Information Theory, Vol. 34(5), pp. 1308-1317. 1988.

[Fancourt 2007] C. Fancourt. Registration and Matching of Perspective Surface Normal Maps. IEEE International Conference on Image Processing, Vol. 6(4), pp. 552-555. 2007.

[Faugeras 1986] O. D. Faugeras and G. Toscani. The Calibration Problem for Stereo. Computer Vision and Pattern Recognition, pp. 15-20. 1986.

[Faugeras 1993] O. Faugeras. Three-Dimensional Computer Vision, A Geometric Viewpoint. Cambridge, MA, MIT Press. 1993.

[Fofia 2003] D. Fofia, J. Salvi and E. M. Mouaddib. Uncalibrated Reconstruction: An Adaptation to Structured Light Vision. Pattern Recognition, Vol. 36(7), pp. 1631-1644. 2003.

[Fong 2005] P. Fong and F. Buron. High-Resolution Three Dimensional Sensing of Fast Deforming Objects. International Conference on Intelligent Robotics and Systems, pp. 1606-1611. 2005.

[Forsyth 2003] A. D. Forsyth and J. Ponce. Computer Vision: A Modern Approach. Prentice Hall. 2003.

[Frankot 1988] R. T. Frankot and R. Chellappa. A Method for Enforcing Integrability in Shape from Shading Algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 10, pp. 439-451. 1988.

[FRVT 2006] http://www.frvt.org/FRVT2006/Results.aspx

[Garding 1993] J. Garding. Direct Estimation of Shape from Texture. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 15, pp. 1202-1208. 1993.

[Gool 1996] L.Van. Gool, T. Moons, and M. Proesmans. Mirror and Point Symmetry under Perspective Skewing. IEEE Conference on Computer Vision and Pattern Recognition, pp. 285-292. 1996.

[Griffin 1992] P. Griffin, L. Narasimhan and S. Yee. Generation of Uniquely Encoded Light Patterns for Range Data Acquisition. Pattern Recognition, Vol. 25 (6), pp. 609-616. 1992.

[Gühring 2000] J. Gühring. Dense 3-D Surface Acquisition by Structured Light using Off-the-Shelf Components. Proceedings of SPIE, Vol. 4309, pp. 220-231. 2000.

[Haitjema 2001] H. Haitjema, W. Pril, P.H.J. Schellekens. A Silicon-Etched Probe for 3D Coordinate Measurements with an Uncertainty below 0.1 μm. IEEE Transactions on Instrumentation and Measurement, Vol. 50(6), pp. 1519-1523. 2001.

[Harding 1998] K. Harding and L. Bieman. High Speed Moiré Contouring Methods Analysis. Proceeding of SPIE on Three-Dimensional Imaging, Optical Metrology, and Inspection IV, Vol. 3520, pp. 27-35. 1998.

[Harris 1988] C. Harris and M.J. Stephens. A Combined Corner and Edge Detector. 4th Alvey Vision Conference. pp. 147-152. 1988.

[Hartley 1997] R. Hartley. Triangulation. Computer Vision and Image Understanding, Vol. 68(2), pp. 146-157. 1997.

[Hartley 2004] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge, U.K.: Cambridge University Press. 2004.

[Healey 1987] G. Healey and T. O. Binford. Local Shape from Specularity, International Conference on Computer Vision, pp. 151-160. 1987.

[Heikkil 1997] J. Heikkil, O. Silvén. A Four-step Camera Calibration Procedure with Implicit Image Correction. Computer Vision and Pattern Recognition, pp. 1106-1112. 1997.

[Hermann 1989] W. Hermann. *Symmetry*. Princeton University Press, ISBN 0-691-02374-3. 1989.

[Ho 2006] J. Ho, J. W. Lim, M. H. Yang and D. J. Kriegman. Integrating Surface Normal Vectors Using Fast Marching Method, European Conference on Computer Vision, Vol. 3, pp. 239-250. 2006.

[Horn 1989] B. K. P. Horn. Height and Gradient from Shading. International Journal of Computer Vision, pp. 37-75. 1989.

[Horn 1999] E. Horn, N. Kiryati. Toward Optimal Structured Light Patterns. Image and Vision Computing, Vol. 17(2), pp. 87-97. 1999.

[Ikeuchi 1981] K. Ikeuchi and B.K.P. Horn. Numerical Shape from Shading and Occluding Boundaries. Artificial Intelligence, Vol. 17(3), pp. 141-184. 1981.

[Inokuchi 1984] S. Inokuchi, K. Sato, F. Matsuda. Range Imaging System for 3-D Object Recognition. International Conference on Pattern Recognition, pp. 806-808. 1984.

[Intel 2008] http://www.intel.com/technology/computing/opencv

[Ishii 2007] I. Ishii, K. Yamamoto, et al. High-Speed 3D Image Acquisition using Coded Structured Light Projection. International Conference on Intelligent Robots and Systems, pp. 925-930. 2007.

[Ito 1995] M. Ito and A. Ishii. A Three-Level Checkerboard Pattern (TCP) Projection Method for Curved Surface Measurement. Pattern Recognition, Vol. 28, pp. 27-40. 1995.

[Johnson 1999] A. E. Johnson and M. Hebert. Using Spin Images for Efficient Object Recognition in Cluttered 3D Scenes. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 21(5), pp. 433-449. 1999.

[Kass 1988] M. Kass, A. Witkin, D. Terzopoulos. Snakes, Active Contour Models. International Journal of Computer Vision, Vol. 1, pp. 321-331. 1988.

[Kawasaki 2006] H. Kawasaki and R. Furukawa. 3D Acquisition System using Uncalibrated Line-Laser Projector. International Conference on Pattern Recognition, Vol. 1, pp. 975-979. 2006.

[Kiyasu 1995] S. Kiyasu, H. Hoshino, K. Yano and S. Fujimura. Measurement of the 3-D Shape of Specular Polyhedrons using an M-array Coded Light Source. IEEE Transactions on Instrumentation and Measurement, Vol. 44(3), pp. 775-778. 1995.

[Koninckx 2006] T. P. Koninckx, L. J. van Gool. Real-Time Range Acquisition by Adaptive Structured Light. IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 28(3), pp. 432-445. 2006.

[Krattenthaler 1994] W. Krattenthaler, K. J. Mayer and H. P. Duwe. 3D Surface Measurement with Coded Light Approach. Proceedings of the 17th meeting of the Austrian Association for Pattern Recognition on Image Analysis and Synthesis, Vol. 12, pp. 103-114. 1994.

[Lange 1999] H. Lange. Advances in the Cooperation of Shape from Shading and Stereo Vision. Second International Conference on 3-D Imaging and Modeling, pp. 46-58, 1999.

[Lange 2000] R. Lange and P. Seitz. Seeing Distances - A Fast Time-of-Flight 3D Camera. Sensor Review, Vol. 20 (3), pp. 212-217(6). 2000.

[Lee 1985] C. H. Lee and A. Rosenfeld. Improved Methods of Estimating Shape from Shading using the Light Source Coordinate System. Artificial Intelligence, Vol. 26. pp. 125-143. 1985.

[Lenz 1988] R. K. Lenz, and R. Y. Tsai,: Techniques for Calibration of the Scale Factor and Image Center for High Accuracy 3-D Machine Vision Metrology. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 10 (5), pp. 713-720. 1988.

[Levin 2004] A. Levin, A. Zomet, S. Peleg and Y. Weiss. Seamless Image Stitching in the Gradient Domain. European Conference on Computer Vision, Vol. 4, pp. 377-389. 2004.

[Li 2003] Y. F. Li and S. Y. Chen. Automatic Recalibration of an Active Structured Light Vision System. IEEE Transactions on Robotics and Automation, Vol. 19(2), pp. 259-268. 2003.

[Liere 2002] R. van Liere, W. de Leeuw, J. Mulder, et al. Virtual Reality in Biological Microscopic Imaging. IEEE International Symposium on Biomedical Imaging, pp. 879-882. 2002.

[Loaiza 1999] H. Loaiza, J. Triboulet, S. Lelandais, et al. A Multi-Configuration Stereoscopic Vision System for Domestic Mobile Robot Localization. Proceedings of the First Robot Motion and Control, pp. 207-212. 1999.

[Lowe 2004] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, Vol. 60(2), pp. 91-110. 2004.

[MacWilliams 1976] F. J. MacWilliams and N. J. A. Sloane. Pseudorandom Sequences and Arrays. Proceedings of the IEEE, Vol. 64(12), pp. 1715-1729. 1976.

[Marr 1976] D. Marr and T. Poggio. Cooperative Computation of Stereo Disparity. Science, Vo. 194, pp. 283-287. 1976.

[Marr 1980] D. Marr and E. Hildreth. Theory of Edge Detection. Proceedings of the Royal Society London, Vol. 207, pp. 187-217. 1980.

[Marr 1982] D. Marr. Vision: *A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W. H. Freeman. 1982.

[Max 1999] N. Max. Weights for Computing Vertex Normals from Facet Normals. Journal of Graph Tools, Vol. 4(2), pp.1-6. 1999.

[Medioni 1987] G.Medioni and Y. Yasumoto. Corner Detection and Curve Representation using Cubic B-Spline. Computer Vision, Graphics and Image Processing, Vol. 39, pp. 267-278. 1987.

[Miyazaki 2002] D. Miyazaki, M. Saito, Y. Sato, and K. Ikeuchi. Determining Surface Orientations of Transparent Objects Based on Polarization Degrees in Visible and Infrared Wavelengths. Journal of the Optical Society of America A, Vol. 19, pp. 687-694. 2002.

[Miyazaki 2003] D. Miyazaki, R. T. Tan, et al. Polarization-Based Inverse Rendering from a Single View. International Conference on Computer Vision, Vol. 2, pp. 982-987. 2003.

[Moore 2001] T. D. Moore, D. Vanderstraeten and P. Forsell. Determination of BGA Structural Defects and Solder Joint Defects by 3D x-ray Laminography. IEEE Proceeding of 8th Physical and Failure Analysis of Integrated Circuits, pp. 146-150. Singapore. 2001.

[Morano 1998] R. A. Morano, C. Ozturk, R. Conn, et al. Structured Light using Pseudorandom Codes. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20(3), pp. 322-327. 1998.

[Moravec 1980] H. Moravec. Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover. Tech Report CMU-RI-TR-3 Carnegie-Mellon University, Robotics Institute. 1980.

[More 1977] J. More. The Levenberg-Marquardt Algorithm, Implementation and Theory. Numerical Analysis, Lecture Notes in Mathematics 630. Springer-Verlag, 1977.

[Morita 1988] H. Morita, K. Yajima, S. Sakata. Reconstruction of Surfaces of 3-D Objects by M-array Pattern Projection Method. International Conference on Computer Vision, pp. 468-473, 1988.

[Nayar 1994] S. K. Nayar and Y. Nakagawa. Shape from Focus. IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 16(8), pp. 824-831. 1994.

[Nayar 1996] S. K. Nayar, M. Watanabe and M. Noguchi. Real-time Focus Range Sensor. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18(12), pp. 1186-1198. 1996.

[Nehab 2005] D. Nehab, R. Szymon et al. Efficiently Combining Positions and Normals for Precise 3D Geometry. ACM Transactions on Graphic, Vol. 24(3), pp. 536-543. 2005.

[Pagès 2003] J. Pagès, J. Salvi and C. Matabosch. Robust Segmentation and Decoding of a Grid Pattern for Structured Light. Pattern Recognition and Image Analysis, pp. 689-696. 2003.

[Pagès 2003a] J. Pagès, J. Salvi, R. García and C. Matabosch. Overview of Coded Light Projection Techniques for Automatic 3D Profiling. International Conference on Robotics and Automation, pp. 133-138. 2003.

[Pentland 1988] A. Pentland. Shape Information from Shading: a Theory about Human Perception. International Conference on Computer Vision, pp. 404-413. 1988.

[Petriu 1992] E. M. Petriu, T. Bieseman et al. Visual Object Recognition using Pseudorandom Grid Encoding. International Conference on Intelligent Robots and Systems, pp. 1617-1624. 1992.

[Ragheb 2003] H. Ragheb and E. R. Hancock. A Probabilistic Framework for Specular Shape From Shading. Pattern Recognition, Vol. 36, pp. 407-427. 2003.

[Ribo 2005] M. Ribo, M. Brandner. State of the Art on Vision-Based Structured Light Systems for 3D Measurements. International Workshop on Robotic and Sensors Environments Ottawa, Canada, 2005.

[Rocchini 2001] C. Rocchini, Paulo Cignoni, et al. A Low Cost 3D Scanner based on Structured Light. Proceedings of EUROGRAPHICS, Vol. 20(3), pp. 299-308. 2001.

[Sadlo 2005] F. Sadlo and T. Weyrich. A Practical Structured Light Acquisition System for Point-Based Geometry and Texture. Proceedings of the Eurographics Symposium on Point-Based Graphics, pp.89-98. 2005.

[Sáenz 2004] R. Legarda-Sáenz, T. Bothe, and W. P. Jüptner. Accurate Procedure for the Calibration of a Structured Light System. Optical Engineering, Vol. 43(2), pp. 464-471. 2004.

[Salvi 1998] J. Salvi, J. Batlle, E. Mouaddib. A Robust-Coded Pattern Projection for Dynamic 3D Scene Measurement. Pattern Recognition Letters, Vol. 19, pp. 1055-1065. 1998.

[Salvi 2004] J. Salvi, J. Pagès and J. Batlle. Pattern Codification Strategies in Structured Light Systems. Pattern Recognition, Vol. 37(4), pp. 827-849. 2004.

[Sansoni 2000] G. Sansoni, M. Carocci, and R. Rodella. Calibration and Performance Evaluation of a 3-D Imaging Sensor Based on the Projection of Structured Light. IEEE Transactions on Instrumentation and Measurement, Vol. 49(3), pp. 628-636. 2000.

[Schmid 2000] C. Schmid, R. Mohr and C. Bauckhage. Evaluation of Interest Point Detectors. International Journal of Computer Vision, Vol. 37(2), pp.151-172. 2000.

[Shi 1994] J. Shi and C. Tomasi. Good Features to Track. Computer Vision and Pattern Recognition, pp. 593-600, 1994.

[Shrikhande 1989] N. Shrikhande and G. Stockman. Surface Orientation from a Projected Grid. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 2(6). 1989.

[Shu 2003] C. Shu, A. Brunton, and M. Fiala. Automatic Grid Finding in Calibration Patterns using Delaunay Triangulation. Technical Report NRC-46497/ERB-1104, Nat. Research Council. 2003.

[Smith 1997] S. M. Smith and M. Brady. SUSAN - A New Approach to Low Level Image Processing. International Journal of Computer Vision, Vol. 23(1), pp. 45-78. 1997.

[Sohmura 2004] T. Sohmura, N. Mitsumasa. High-Resolution 3-D Shape Integration of Dentition and Face Measured by New Laser Scanner. IEEE Transactions on Medical Imaging, Vol. 23(5), pp. 633-638. 2004.

[Song 2006] Z. Song, P. S. Huang. Novel Method for Structured Light System Calibration. Optical Engineering, Vol. 45(8). 2006.

[Song 2007] Z. Song, R. Chung and S. WL. Yeung. An Accurate Camera-Projector System Calibration by the use of an LCD Panel. Proceedings of the 2007 International Conference on Image Processing, Computer Vision, and Pattern Recognition.Vol.1, pp. 164-170. 2007.

[Spoelder 2000] H. J.W. Spoelder, F. M. Vos et al. Some Aspects of Pseudo Random Binary Array-based Surface Characterization, IEEE Transactions on Instrumentation and Measurement, Vol. 49(6), pp. 1331-1336. 2000.

[Spoelder 1998] H. J. W. Spoelder, F. M. Vos, E. Petriu and F. C. A. Groen. A Study of the Robustness of Pseudorandom Binary Array Based Surface Characterization. IEEE Transaction on Instrument Measurement, Vol. 47(4), pp. 833-838. 1998.

[Steinman 2000] S. Steinman, B. Steinman, R. Garzia. Foundations of Binocular Vision: A Clinical Perspective. McGraw-Hill Medical, New York. 2000.

[Sturm 1999] P. Sturm and S. Maybank: On Plane-based Camera Calibration: A General Algorithm, Singularities, Applications. Computer Vision and Pattern Recognitions, pp. 432-437. 1999.

[Subbarao 1995] M. Subbarao and T.S. Choi. Accurate Recovery of Three Dimensional Shape from Image Focus. IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 17(3), pp. 266-274. 1995.

[Sugihara 1984] K. Sugihara, K. Okazaki, F. Kaihua and N. Sugie. Regular Pattern Projection for Surface Measurement. 2nd International Symp. Robotics Research, pp. 17-24. 1984.

[Tan 2003] R. T. Tan and K. Ikeuchi. Separating Reflection Components of Textured Surfaces using a Single Image. International Conference on Computer Vision, Vol. 2, pp. 870-877. 2003.

[Taylor 2003] C. J. Taylor. Surface Reconstruction from Feature Based Stereo. 9th IEEE International Conference on Computer Vision, pp. 184-190. 2003.

[Treleaven 2007] P. Treleaven, J. Wells. 3D Body Scanning and Healthcare Applications. Computer. Vol. 40(7), pp. 28-34. 2007.

[Tsai 1987] R. Y. Tsai. A Versatile Camera Calibration Technique for High Accuracy 3D Machine Vision Metrology using off-the-shelf TV Cameras and Lenses. IEEE J. Robotics and Automation, Vol. 3(4), pp. 323-344. 1987.

[Valkenburg 1998] R. J. Valkenburg and A. M. McIvor. Accurate 3D Measurement using a Structured Light System. Image and Vision Computing, Vol. 16(2), pp. 99-110. 1998.

[Vassilev 2000] T. Vassilev and S. Bernhard. Efficient Cloth Model for Dressing Animated Virtual People. Proceeding of Learning to Behave Workshop, pp. 89-100, 2000.

[Vuylsteke 1990] P. Vuylsteke and A. Oosterlinck. Range Image Acquisition with a Single Binary-Encoded Light Pattern. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12(2), pp. 148-164. 1990.

[Wei 1994] G. Wei and S. Ma. Implicit and Explicit Camera Calibration: Theory and experiments. *IEEE* Transactions on Pattern Analysis and Machine Intelligence, Vol. 16(5), pp. 469-480. 1994.

[Weng 1992] J. Weng, P. Cohen and M. Herniou. Camera Calibration with Distortion Models and Accuracy Evaluation. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14(10), pp. 965-980. 1992.

[Winkelbach 2001] S. Winkelbach, F. M. Wahl. Shape from 2D Edge Gradients. Pattern Recognition, Lecture Notes in Computer Science, Vol. 2191, pp.377-384. 2001.

[Winkelbach 2002] S. Winkelbach, F. M. Wahl. Shape from Single Stripe Pattern Illumination. Pattern Recognition, Lecture Notes in Computer Science, Vol. 2449, pp. 240-247 2002.

[Wolff 1991] L. B.Wolff and T. E. Boult. Constraining Object Features using a Polarization Reflectance Model. IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 13(7), pp. 635-657. 1991.

[Woodham 1978] R. J. Woodham. Photometric Stereo: A Reflectance Map Technique for Determining Surface Orientation from Image Intensity. SPIE, Vol. 155, pp. 136-143. 1978.

[Woodham 1980] R.J. Woodham. Photometric Method for Determining Surface Orientation from Multiple Images. Optical Engineering, Vol. 19(1), pp. 139-144. 1980.

[Yee 1994] S. Yee and P. Griffin. Three-Dimensional Imaging System. Optical Engineering, Vol. 33(6), pp. 2070-2075. 1994.

[Zhang 1999] R. Zhang, P. Tsai, J. E. Cryer, and M. Shah. Shape from Shading: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 21(8), pp. 690-705. 1999.

[Zhang 2000] Z. Zhang. A Flexible New Technique for Camera Calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22(11), pp.1330-1334. 2000.

[Zheng 1991] Q. Zheng and R. Chellappa. Estimation of Illuminant Direction, Albedo, and Shape from Shading. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 13(7), pp. 680-702. 1991.