

**3D machine vision systems for robotic weeding and plant phenotyping**

by

**Ji Li**

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Co-majors: Agricultural Engineering; Human Computer Interaction

Program of Study Committee:

Lie Tang, Major Professor

Stuart Birrell

Matthew Darr

Phillip Jones

Brian Steward

Iowa State University

Ames, Iowa

2014

Copyright © Ji Li, 2014. All rights reserved.

## DEDICATION

I dedicate this work to my wife.

## TABLE OF CONTENTS

|   | Page |
|---|------|
| DEDICATION.....   | ii   |
| LIST OF FIGURES .....   | vi   |
| LIST OF TABLES .....  | xi   |
| ACKNOWLEDGEMENTS.....   | xii  |
| ABSTRACT.....   | xiv  |
| CHAPTER 1. GENERAL INTRODUCTION .....                                       | 1    |
| 1.1 Introduction.....   | 1    |
| 1.2 Plant Phenotype and Phenotyping .....                                   | 2    |
| 1.2.1 Background .....  | 2    |
| 1.2.2 Ideal Phenotyping System.....   | 4    |
| 1.2.3 Imaging Based Phenotyping .....                                       | 7    |
| 1.3 Machine Vision System for Automated Precision Weed Control .....        | 13   |
| 1.3.1 Background .....  | 13   |
| 1.3.2 Machine Vision Based Plant Species Identification System .....        | 15   |
| 1.4 Research Objective .....  | 17   |
| 1.5 Dissertation Overview .....   | 17   |
| 1.6 Reference .....   | 18   |
| CHAPTER 2. 3D IMAGING BASED CROP PLANT AND WEED<br>RECOGNITION SYSTEM ..... | 25   |
| 2.1 Abstract .....  | 25   |
| 2.2 Introduction.....   | 25   |
| 2.3 Sensor and Data Collection System.....                                  | 29   |
| 2.4 Experiment Design.....  | 32   |
| 2.5 Algorithm Design.....   | 35   |
| 2.5.1 Initial Noise Filtering.....  | 35   |
| 2.5.2 Neighbor Points Table Development.....                                | 36   |
| 2.5.3 Curvature and Normal Estimation of Surface.....                       | 38   |
| 2.5.4 Advanced Noise Clearance Algorithm .....                              | 44   |
| 2.5.5 Gradient of Amplitude and Depth Image.....                            | 48   |
| 2.5.6 Percentile Index of Amplitude Image .....                             | 49   |
| 2.5.7 Statistics of Extracted Features .....                                | 54   |
| 2.5.8 Crop Plant Segmentation Algorithm.....                                | 68   |

|   |         |
|---|---------|
| 2.6 Results & Discussion .....  | 82      |
| 2.6.1 Broccoli Recognition Result .....   | 82      |
| 2.6.2 Soybean Recognition Result .....  | 90      |
| 2.6.3 Processing Speed .....  | 95      |
| 2.7 Conclusions .....   | 95      |
| 2.8 Reference .....   | 96      |
| <br>CHAPTER 3. MACHINE VISION BASED REAL-TIME MOTION AND<br>ATTITUDE ESTIMATION .....     | <br>99  |
| 3.1 Abstract .....  | 99      |
| 3.2 Introduction .....  | 100     |
| 3.3 Sensor .....  | 104     |
| 3.4 Target Detection Algorithm .....  | 105     |
| 3.4.1 Challenges of Target Detection .....  | 105     |
| 3.4.2 Background Identification Algorithm for Target Detection<br>Acceleration .....      | 106     |
| 3.4.3 Optimization of Adaptive Threshold .....  | 112     |
| 3.5 Position and Attitude Estimation .....  | 114     |
| 3.5.1 Positioning Theory .....  | 114     |
| 3.5.2 Attitude Evaluation .....   | 118     |
| 3.6 Target Identification .....   | 118     |
| 3.7 Experiments and Results .....   | 119     |
| 3.7.1 Target Detection and ID Recognition Performance Experiment<br>and Result .....      | 119     |
| 3.7.2 Localization Experiment and Result .....  | 129     |
| 3.7.3 Attitude Estimation Experiment and Result .....                                     | 138     |
| 3.8 Conclusions .....   | 143     |
| 3.9 Reference .....   | 144     |
| <br>CHAPTER 4. DEVELOPMENT OF A 3D RECONSTRUCTION BASED<br>PLANT PHENOTYPING SYSTEM ..... | <br>146 |
| 4.1 Abstract .....  | 146     |
| 4.2 Introduction .....  | 147     |
| 4.3 Experimental Design .....   | 149     |
| 4.4 Sensors .....   | 151     |
| 4.5 Camera Posture Tracking Infrastructure .....  | 155     |
| 4.6 Position and Attitude Estimation of 3D Camera .....                                   | 159     |
| 4.6.1 Position and Attitude Estimation of 2D Camera .....                                 | 159     |
| 4.6.2 Calibration between 2D and 3D Cameras .....   | 161     |
| 4.7 3D Reconstruction .....   | 169     |
| 4.7.1 Preprocessing .....   | 170     |
| 4.7.2 3D Registration .....   | 171     |
| 4.8 Leaf and Stem Segmentation .....  | 174     |
| 4.8.1 Stem Segmentation .....   | 174     |

|  |         |
|--|---------|
| 4.8.2 Leaf Segmentation .....                                      | 176     |
| 4.9 Leaf Parameter Estimation.....                                 | 178     |
| 4.9.1 Leaf Points Regression .....                                 | 178     |
| 4.9.2 Leaf Parameter Estimation.....                               | 179     |
| 4.10 Results and Discussion .....                                  | 181     |
| 4.11 Conclusions.....  | 190     |
| 4.12 Reference .....   | 191     |
| <br>CHAPTER 5. GENERAL CONCLUSIONS.....                            | <br>194 |
| 5.1 Conclusions.....   | 194     |
| 5.1.1 3D Imaging Based Crop Plant and Weed Recognition System..... | 194     |
| 5.1.2 3D Reconstruction Based Plant Phenotyping System .....       | 195     |
| 5.2 Recommendations.....   | 196     |

## LIST OF FIGURES

|   | Page |
|---|------|
| Figure 2.1. Swissranger SR4000.....   | 30   |
| Figure 2.2. (a) Picture of data collection system and (b) drawing of 3D ToF camera setup.....   | 31   |
| Figure 2.3. Broccoli field of this study .....  | 34   |
| Figure 2.4. Soybean field of this study .....   | 34   |
| Figure 2.5. Amplitude image of broccoli with noise at left corner.....  | 36   |
| Figure 2.6. Normal and curvature estimation of soybean: (a) 2D color image, (b) amplitude image from ToF camera, (c) depth image, (d) neighbor count image Nc, (e) curvature estimation result, and (f) normal direction represented by color image.....                              | 44   |
| Figure 2.7. 3D point cloud data after initial noise clearance: (a) front view of 3D point cloud data and (b) side view with sparse point noise marked by red circle .....   | 46   |
| Figure 2.8. 3D point cloud data after advanced noise clearance: (a) front view of 3D point cloud data and (b) side view .....   | 47   |
| Figure 2.9. Gradient images: (a) amplitude image, (b) gradient of amplitude image , (c) depth image, and (d) gradient of depth image .....  | 49   |
| Figure 2.10. Broccoli amplitude image and histograms: (a) broccoli amplitude image, (b) broccoli segmented from amplitude image, (c) histogram of whole amplitude image, (d) histogram of background, and (e) histogram of broccoli .....   | 52   |
| Figure 2.11. Amplitude percentile image of broccoli .....   | 53   |
| Figure 2.12. Amplitude percentile image of soybean .....  | 54   |
| Figure 2.13. Sample image sets of soybean for feature distribution analysis: (a) depth image, (b) amplitude image, (c) manual segmentation image of soybean leaves, (d) API image, (e) curvature image, (f–g) depth and amplitude gradient images, and (h) neighbor count image.....  | 56   |
| Figure 2.14. Sample image sets of broccoli for feature distribution analysis: (a) depth image, (b) amplitude image, (c) manual segmentation image of soybean leaves, (d) API image, (e) curvature image, (f–g) depth and amplitude gradient images, and (h) neighbor count image..... | 57   |

|  |     |
|--|-----|
| Figure 2.15. Histograms and cumulative histograms of various features of soybean and background: (a–b) histogram and cumulative histogram of API feature of soybean and background; (c–d) histogram and cumulative histogram of curvature feature of soybean and background, (e–f) histogram and cumulative histogram of depth gradient of soybean and background, (g–h) histogram and cumulative histogram of intensity gradient of soybean and background, and (i–j) histogram and cumulative histogram of neighbor count feature of soybean and background.....       | 62  |
| Figure 2.16. Histograms and cumulative histograms of various features of broccoli and background: (a–b) histogram and cumulative histogram of API feature of broccoli and background, (c–d) histogram and cumulative histogram of curvature feature of broccoli and background, (e–f) histogram and cumulative histogram of depth gradient of broccoli and background, (g–h) histogram and cumulative histogram of intensity gradient of broccoli and background, and (i–j) histogram and cumulative histogram of neighbor count feature of broccoli and background..... | 67  |
| Figure 2.17. Images of the soybean segmentation procedure: (a) amplitude image, (b) depth image, (c) amplitude gradient image GI, (d) depth gradient image Gz, (e) neighbor count image Nc, (f) curvature image $\delta$ , (g) amplitude percentile image API, (h–i) intermediate image S1 and S2, (j–k) image S11 and S21, (l) image S12, (m) Image S13, and (n) final soybean segmentation result image.....   | 74  |
| Figure 2.18. Images of broccoli segmentation procedure: (a) amplitude image, (b) depth image, (c) amplitude gradient image GI, (d) depth gradient image Gz, (e) neighbor count image Nc, (f) curvature image $\delta$ , (g) amplitude percentile image API, (h–i) intermediate images S1 and S2, (j) image S11, (k) image S12, (l) image S13, (m) image S14, (n) image S15, (o) result image of flood fill, and (p) final broccoli segmentation result image.....  | 81  |
| Figure 2.19. Broccoli segmentation result: (a) 2D color image, (b) amplitude image from ToF camera, and (c) broccoli segmentation result .....   | 84  |
| Figure 2.20. Small plant missed by system: (a) amplitude image captured by ToF camera and (b) 2D color image .....   | 85  |
| Figure 3.1. Raw sample image captured from camera.....   | 110 |
| Figure 3.2. Edge extraction result .....   | 111 |
| Figure 3.3. Result of area analysis: The gray part is the area eliminated for later target searching; the white lines are the edges kept by the edge selection algorithm; and the color areas are kept for later target searching.....   | 111 |

|   |     |
|---|-----|
| Figure 3.4. Target detection result .....   | 112 |
| Figure 3.5. The world coordinate system defined by the target and camera coordinate system.....   | 116 |
| Figure 3.6. Two target samples .....  | 118 |
| Figure 3.7. Video frame of moving target captured indoors.....  | 121 |
| Figure 3.8. Frames of the videos of a moving target captured outdoors:<br>(a) a frame of the video whose camera direction is around $135^\circ$ to the sun and<br>(b) a frame of the video whose camera direction is around $45^\circ$ to the sun ..... | 122 |
| Figure 3.9. A frame of the video without a target in view, captured indoors .....   | 123 |
| Figure 3.10. A frame of the video without a target in view captured outdoors .....  | 123 |
| Figure 3.11. Target detection time cost when target is available.....   | 125 |
| Figure 3.12. Response time of target detection algorithm when no target is available .....  | 127 |
| Figure 3.13. Percentage of image areas eliminated for the later target searching .....  | 127 |
| Figure 3.14. The setup of localization experiment .....   | 131 |
| Figure 3.15. Location error map .....   | 134 |
| Figure 3.16. Standard deviation of positioning result.....  | 136 |
| Figure 3.17. X3M multi-axis absolute MEMS inclinometer and its three axis .....   | 138 |
| Figure 3.18. Method for mounting the inclinometer on the target.....  | 139 |
| Figure 3.19. The setup of the angle measurement experiment with the Z axis of the target parallel to gravity.....   | 139 |
| Figure 3.20. The angle measurement trajectory of two methods when the Z axis of the target is parallel to gravity.....  | 141 |
| Figure 3.21. The setup of the angle measurement experiment with the X axis of the target parallel to gravity.....   | 142 |
| Figure 3.22. The angle measurement trajectory of two methods when the X axis of the target is parallel to gravity.....  | 143 |



|   |     |
|---|-----|
| Figure 4.1. Corn plant used in this phenotyping research.....   | 151 |
| Figure 4.2. Data collection system built for this 3D reconstruction based phenotyping research: (a) picture of system and (b) translation relationship between sensors .....  | 154 |
| Figure 4.3. World coordinate system and chessboard pattern array used as the camera's pose estimation infrastructure.....   | 157 |
| Figure 4.4. Infrastructure setup .....  | 158 |
| Figure 4.5. Images collected by 2D and 3D camera: (a) chessboard pattern beacon detected by 2D camera, (b) depth image collected by 3D ToF camera, and (c) intensity image captured by 3D ToF camera.....   | 160 |
| Figure 4.6. 2D and 3D camera calibration images: (a) images of two chessboard targets captured by 2D camera at a farther viewpoint A, (b) image of right chessboard target captured by 2D camera at close viewpoint B, (c–e) intensity and depth image and point cloud data of left chessboard target captured by 3D camera at close viewpoint B..... | 166 |
| Figure 4.7. Coordinate system CL defined by left target board.....  | 167 |
| Figure 4.8. Inner corner points P1 and P2 on the X axis of the coordinate system defined by the left target board .....   | 167 |
| Figure 4.9. Translation relationship between cameras of the data collection system with calibration result .....  | 169 |
| Figure 4.10. (a) The 3D model's achievement without sparse point noise clearance and (b) the 3D model's achievement with sparse noise clearance .....   | 171 |
| Figure 4.11. Complete 3D model of corn plant achieved through 3D registration: (a) side view and (b) top view of 3D model, (c) the intensity and (d) depth image of the point cloud data view corresponding to the green points in (a) and (b).....   | 173 |
| Figure 4.12. The stem detected in different side views at different viewpoints .....  | 176 |
| Figure 4.13. The top view image of the point cloud of the 3D model; the stem is marked with red color .....   | 177 |
| Figure 4.14. Leaves and stem segmentation result .....  | 178 |
| Figure 4.15. Regression result of leaf 3 of plant 1 .....   | 179 |

|   |     |
|---|-----|
| Figure 4.16. Corn plant 1 and its 3D reconstruction result: (a) 2D color picture and (b) 3D reconstruction result ..... | 183 |
| Figure 4.17. Corn plant 2 and its 3D reconstruction result: (a) 2D color picture and (b) 3D reconstruction result ..... | 184 |
| Figure 4.18. Corn plant 3 and its 3D reconstruction result: (a) 2D color picture and (b) 3D reconstruction result ..... | 185 |
| Figure 4.19. Leaf and stem segmentation result: (a) plant 1, (b) plant 2, and (c) plant 3 .....                         | 186 |

## LIST OF TABLES

|  | Page |
|--|------|
| Table 2.1. Height of crop plant corresponding to each image capturing date ..... | 33   |
| Table 2.2. Feature distribution table for soybean .....                          | 68   |
| Table 2.3. Feature distribution table for broccoli .....                         | 68   |
| Table 2.4. Broccoli plant recognition accuracy rate .....                        | 84   |
| Table 2.5. Broccoli segmentation result .....                                    | 86   |
| Table 2.6. Soybean segmentation result .....                                     | 91   |
| Table 3.1. Response time cost of different target detection algorithm .....      | 128  |
| Table 3.2. Localization error (mm).....  | 133  |
| Table 3.3. Standard deviation of localization result (mm).....                   | 135  |
| Table 3.4. Corner positioning error (pixel).....                                 | 137  |
| Table 4.1. Measurement result of corn plant 1 .....                              | 187  |
| Table 4.2. Measurement result of corn plant 2 .....                              | 188  |
| Table 4.3. Measurement result of corn plant 3 .....                              | 189  |
| Table 4.4. Statistics of leaf parameter's measurement error rate.....            | 190  |

## ACKNOWLEDGEMENTS

First and foremost, I would like to thank Dr. Lie Tang for supporting and supervising me throughout my Ph.D.'s work. Without his patient discussion, encouragement and help in both the lab and the farm field, this work could not be possible. During the four and half years, Dr. Tang has helped me a lot in all aspects. His academic knowledge and experience inspired me when I encountered difficulties and guided my research projects to the right direction. Also, Dr. Tang helped me to refine my career goal and achieved the goal step by step. He not only encouraged me to collaborate with industries, but also provided great platform for collaboration. Moreover, Dr. Tang and his family are always there as my friends in Ames. Their care and supports for my life in Ames have greatly helped me finish my research work.

I also would like to thank my committee members, Dr. Brian Steward, Dr. Matthew Darr, Dr. Phillip Jones, and Dr. Stuart Birrell, for their support throughout the course of this research. Their guidance not only inspired me a lot for my research, but also taught me how to think creatively. The precious knowledge about various aspects I learned from them opened lots of doors to the fantastic engineering world, and prepare me well for any challenges in my later career life.

Additionally, I thank Iowa State University for all the resources and support. I greatly appreciate all the faculty members and staffs who have taught me or helped me. I enjoyed the pleasant life to study and live here. And I will be proud to be a cyclone forever.

I also would like to thank my lab colleagues: Mohd Taufik Bin Ahmad and Yin Bao. They helped me a lot for both farm and lab experiments. Their support also made my personal life more pleasant.

Finally, I would like give my special thanks to my family. Their love and support is the great power to keep me moving forward.

## ABSTRACT

The need for chemical free food is increasing and so is the demand for a larger supply to feed the growing global population. An autonomous weeding system should be capable of differentiating crop plants and weeds to avoid contaminating crops with herbicide or damaging them with mechanical tools. For the plant genetics industry, automated high-throughput phenotyping technology is critical to profiling seedlings at a large scale to facilitate genomic research. This research applied 2D and 3D imaging techniques to develop an innovative crop plant recognition system and a 3D holographic plant phenotyping system.

A 3D time-of-flight (ToF) camera was used to develop a crop plant recognition system for broccoli and soybean plants. The developed system overcame the previously unsolved problems caused by occluded canopy and illumination variation. Both 2D and 3D features were extracted and utilized for the plant recognition task. Broccoli and soybean recognition algorithms were developed based on the characteristics of the plants. At field experiments, detection rates of over 88.3% and 91.2% were achieved for broccoli and soybean plants, respectively. The detection algorithm also reached a speed over 30 frame per second (fps), making it applicable for robotic weeding operations.

Apart from applying 3D vision for plant recognition, a 3D reconstruction based phenotyping system was also developed for holographic 3D reconstruction and physical trait parameter estimation for corn plants. In this application, precise alignment of multiple 3D views is critical to the 3D reconstruction of a plant. Previously published research highlighted the need for high-throughput, high-accuracy, and low-cost 3D phenotyping systems capable of holographic plant reconstruction and plant morphology related trait

characterization. This research contributed to the realization of such a system by integrating a low-cost 2D camera, a low-cost 3D ToF camera, and a chessboard-pattern beacon array to track the 3D camera's position and attitude, thus accomplishing precise 3D point cloud registration from multiple views. Specifically, algorithms of beacon target detection, camera pose tracking, and spatial relationship calibration between 2D and 3D cameras were developed. The phenotypic data obtained by this novel 3D reconstruction based phenotyping system were validated by the experimental data generated by the instrument and manual measurements, showing that the system has achieved measurement accuracy of more than 90% for most cases under an average of less than five seconds processing time per plant.

## CHAPTER 1. GENERAL INTRODUCTION

### 1.1 Introduction

Cereal production must be increased by 70% before 2050 to meet the demand of increasing global population, whereas the rice supply has already become insufficient (Furbank et al., 2009). Additionally, the demand for biofuels will increase in the next decade (Sticklen, 2007), and the competition for the limited arable land between food and biofuel crop plants makes food security challenges even worse. Moreover, the abiotic stress resulting from climate change on temperature and rainfall patterns may lead to yield reduction (Sticklen, 2007). A phenotyping system that helps scientists explore the linkage between plant genotype and phenotype could equip humans to meet the challenges in the future.

Despite the potential future challenges in food production, the public currently calls for healthy food products that are free of chemicals and pollution. Various chemicals, including herbicides and pesticides, are widely used in modern agriculture production systems. These chemicals contribute to low-cost pest and weed control approaches, and they are indispensable in ensuring sufficient yield of food production. For example, in 2010, Earthbound Farms, the largest producer of organic products in North America, reported costs of up to \$1,000 per acre for weed control, mainly due to the labor cost of manual weeding. In contrast, conventional farmers spend only about \$50 per acre on herbicides (EFO). Because of the low cost and high efficacy of chemical weed control, herbicides are widely adopted to replace manual operation. Additionally, labor regulations also make manual work impractical. One example is the ban on manual weeding by the California Industry Safety Board in 2004. Even though chemical based weed control can reduce cost and increase food



yield, there are rising concerns about herbicide resistant weeds, pesticide resistant pests, the environmental impact of chemical runoff, and potential health issues caused by chemicals.

Currently there is no clean, effective, and affordable alternative approach to replace the usage of chemicals without losing yield. Research to address these problems is thus greatly needed.

## 1.2 Plant Phenotype and Phenotyping

### 1.2.1 Background

A plant's phenotype is the result of the dynamic interaction of its genotype and environment (Dhondt et al., 2013). Phenotypic parameters, such as leaf size, crop height, cereal yield, photosynthesis rate, nutrient intake rate, resistance to disease and drought, etc. are important for breeders (Foundation and Mcb, 2011). Understanding the linkage between a particular genotype and a specific phenotypic parameter is a core goal of modern biology. However, it is generally difficult due to the large number of genes and the interaction with complex and changeable environmental influences (Foundation and Mcb, 2011).

The fast development of technology has enabled rapid genome sequencing at a steadily declining cost, at rapidly increasing speed. The next generation sequencing method will capture the complete genotype information of a many representatives of a population (Foundation and Mcb, 2011). Scientists have collected abundant information of plant genotype due to the recent revolution of genomic technologies. The genomic information could not be fully capitalized without correct linkage between genotype and phenotype (Cobb et al., 2013; Foundation and Mcb, 2011; Furbank and Tester, 2011). Sophisticated phenotyping system could serve as the bridge (Furbank and Tester, 2011).

*Phenomics* is the science of large-scale phenotypic data collection and analysis (Foundation and Mcb, 2011). It typically employs large numbers of genetic variants for sampling variation in genetic diversity, and each genotype is assayed for a large number of traits to maximize the accuracy. The experiment condition is well monitored and controlled. In addition, the phenotypic data and metadata description of experiment conditions are well captured for the data analysis to reveal the relationship between the phenotypic feature and the genotype (Allen et al., 2010; Heffner et al., 2011; Lu et al., 2011; Nichols et al., 2011; Speliotes et al., 2010; Winzeler et al., 1999).

Challenges to extracting and quantifying sophisticated phenotypic features on a large scale are the “phenotyping bottleneck” to being able to interpret the relationship between genotype and phenotype masterfully (Foundation and Mcb, 2011). Traditional phenotyping is labor intensive, expensive, and destructive (Furbank and Tester, 2011).

Adapting advanced imaging techniques is clearing the phenotyping bottleneck (Cobb et al., 2013; Fiorani and Schurr, 2013; Foundation and Mcb, 2011; Furbank and Tester, 2011). Imaging technique, such as 2D color imaging, chlorophyll fluorescence imaging, thermography, etc. have been successful in enabling remote and noninvasive methods for capturing not only the morphological phenotype data but also the physiological status for plants with relatively simple canopy structures. However, the problems associated with the automated phenotyping of complex plant architecture still require great efforts to solve (Furbank and Tester, 2011).

Phenotyping in three dimensions has the potential for better comprehension of phenotype (Dhondt et al., 2013). Sensors such as stereo cameras, laser range sensors, time-

of-flight (ToF) cameras, and 3D light field cameras have been applied to improve the usability of the automatic phenotyping system.

How to process the images captured by the above mentioned techniques is also important in phenotyping. Lack of algorithms that extract, quantify, and summarize plant phenotypes could significantly hinder the application of image techniques in phenotyping (Foundation and Mcb, 2011). Image processing algorithms also play remarkable part in providing standardized digital data that facilitates the data storage, data transaction, and data analysis in a phenotyping system (Foundation and Mcb, 2011).

### **1.2.2 Ideal Phenotyping System**

Plant phenomics is of great importance for trait based physiological breeding. The genotype is the genetic constitution of an organism, and the phenotype is the composite of an organism's observable traits. The study of the genotype–phenotype relationship through the use of genomic data and the analysis of multigenic functions is called *phenomics* (Schilling et al., 1999). The main idea of physiological breeding is to identify superior crop varieties with important traits, and, consequently, to identify the corresponding genetic traits by looking for those candidates commonly encountered in the superior group. The discovery of superior genes can benefit superior plant species by transplanting desired genes to them. Plant phenomics requires precise, comprehensive, and quantitative analysis tools to facilitate the process of discovering the genes contributing to the traits.

The agricultural application of plant phenomics still has great challenges. First, the linkage between plant genotype and phenotype is often illusive (Miyao et al., 2007). In a real world environment, the linkage of a specific gene to the phenotype must be examined

carefully and comprehensively. Second, the capacity to extract a specific subtle phenotypic effect with the influence of other factors and measurement error is still low. The breeding industry is calling for systematic high-throughput phenotyping capacity to mine candidate germplasm for genes with agricultural importance. Plant phenotyping is the assessment of plant's traits, such as plant height, projected leaf area, leaf length and rolling, color, yield, resistance to biotic and abiotic stress, etc.

The desired agricultural traits of plants are high yield, good quality, and high biotic and abiotic stress tolerance across multiple environments. Screening plants with valuable traits and replicating the trials to phenotype a large population is laborious and time consuming. There are three main problems related to plant phenotyping that need to be solved for crop breeding.

First, the current commonly used phenotyping tools require destructive harvest at particular phenotyping stages, which is slow and costly. In addition, it is almost impossible to monitor and track the phenotype changes of a specific plant during its whole life cycle. Under the constraint of labor and cost, crop breeding programs using conventional phenotyping methods only measure the final yield for replicated trials over different environments and multiple seasons. These conventional methods are problematic, as yield itself is a poorly inherited trait for crop breeding (Richards et al., 2010).

Second, human factors may lead to measurement errors in phenotyping. A subtle trait improvement may indicate a promising plant candidate in crop breeding. Nevertheless, subtle trait improvement benefited by a gene is very sensitive to measurement variation and error. Currently, under most circumstances, the plant's physical characteristics are measured manually. The measurements conducted by different experts lead to measurement variation.

In particular, the harsh in field environment conditions, such as high temperature, strong sun light, and long working hours, would exhaust people's energy and thus decrease the accuracy and objectivity of the measurement. Manually measured plant candidates may be undervalued or overvalued. In other words, human factors may hurt the confidence of the measurement results and eliminate a superior candidate.

Third, a precise and automated phenotyping system will enable development of a searchable database that links gene sequences to plant structure, development, composition, and performance for genetic research. To establish such a database, all phenotyping measurements should be accurate and objective; all the measurement results should be digitalized and recorded in a searchable format. For example, the information of experiment design, plant material, and growth condition need to be saved in a standardized metadata format, so that complete and objective information can be provided. A fast, precise, and cost effective phenotyping system is indispensable to collecting trait information of numerous genes for crop breeding.

To solve the three problems, a desirable system with the following characteristics should be established:

- (1) High-throughput capacity
- (2) Noninvasive or nondestructive, so that complete lifecycle measurements are possible
- (3) Adequate accuracy for subtle trait improvement detection
- (4) Low-cost to promote large-scale implementation, and
- (5) Controllable environment facilities to eliminate nongenomic factors that lead to phenotyping improvement.

### **1.2.3 Imaging Based Phenotyping**

In the past decade, scientists have made great efforts to achieve such an ideal phenotyping system. Different imaging sensors, image processing techniques, robotics, and high performance computing have been applied for this ultimate goal.

#### **1.2.3.1 2D Imaging Based Phenotyping**

2D imaging based phenotyping systems were most widely used in research and commercial products, such as LemnaTec Scanalyser (LemnaTec GmbH, Germany).

2D color imaging can extract some relatively simple features of plants for crop breeding. For example, using 2D color images features, such as projected leaf area, can be extracted and used to estimate the growth rate and drought tolerance for rosette plants such as *Arabidopsis* (Granier et al., 2006; Walter et al., 2007). Some other relatively simple morphological information, such as plant height and canopy's diameter, roundness, circumference, and eccentricity, can also be estimated based on top view or side view 2D images (LemnaTec Scanalyser, LemnaTec GmbH, Germany).

An advanced 2D imaging system, called modulated chlorophyll fluorescence imaging, has been developed for phenotyping. It is a promising tool to score the resistance performance of plants under both abiotic and biotic stresses. This system estimates the photosynthetic response by analyzing the 2D chlorophyll fluorescence image stimulated by a modulated light source. It has been applied in the studies of photosynthetic responses in various conditions such as drought stress, cold, heat, and ultraviolet light (Jansen et al., 2009). Modulated chlorophyll fluorescence imaging has also been used to detect pathogen

infections that affect photosynthesis of a plant. For example, it successfully detected the infection symptom of the foliar and stem fungal pathogens that reduce plant photosynthesis by damaging transpiration (Chaerle et al., 2009; Scholes and Rolfe, 2009; Swarbrick et al., 2006).

Another advanced 2D imaging technique, thermography, has also been developed. Thermography could distinguish leaf surfaces from the backgrounds by detecting the temperature difference between leaf and backgrounds caused by transpiration. For example, it has been used to phenotype plant traits and disease related transpiration such as root fungal pathogen infection (Sirault et al., 2009).

The above mentioned 2D imaging techniques can provide projected images of objects instead of complete 3D plant structural information. Therefore, they can only obtain proxies for relatively simple shape and texture properties. Their inherent limitations make 2D imaging based phenotyping systems less useful and accurate for studying plants that have complicated and self-occluded canopies.

#### 1.2.3.2 3D Imaging Based Phenotyping

Based on 2D imaging methods, quasi 3D systems have been developed. Being different from real 3D systems, quasi 3D systems still apply 2D imaging sensors instead of 3D sensors. They do not have the plant's 3D architecture information. Instead, they estimate the properties of a plant by analyzing multiple 2D images acquired at different viewing angles based on the relationship between the projections at different viewpoints. Normally, one top view and two side view images would be used for a quasi 3D system ([www.plantphenomics.com](http://www.plantphenomics.com)). For example, a quasi 3D system based on 2D color images has

been applied to estimate biomass and leaf area (Rajendran et al., 2009). Another quasi 3D system based on the mathematical relationship of multiple views of chlorophyll fluorescence images has been used to study plant photosynthetic response (Konishi et al., 2009). These quasi 3D systems have high measurement accuracy when the target plants are at early growth stages; the accuracy would decrease as the target plants enter more advanced growth stages to form complex vegetative structures. Overall, quasi 3D systems function well only when the plants' vegetative structures do not have occlusions (Furbank and Tester, 2011).

Real 3D imaging techniques based on a 3D imaging sensor have started to show advantages in building a phenotyping system that has more desirable characteristics. These techniques aim at providing the in-depth 3D information of the plant. Phenotyping systems based on 3D imaging techniques may enable the development of 3D reconstruction and 3D architecture measurement throughout the growing cycle of the plant. These 3D models could be used to extract more detailed and more specific information, even for the life history of each leaf or each tiller. Moreover, compositing 2D information, including texture, color, chlorophyll fluorescence image, and thermography, to a 3D model of the plant will enable researchers to study the illness, pathogen infection and tolerance, transpiration, and photosynthetic response of any specific part of the plant (Furbank and Tester, 2011). Ideally, a 3D phenotyping system could automatically quantify the properties of particular interest and empower breeders to improve the corresponding plant traits.

An advanced 3D imaging sensor is key to the realization of 3D model reconstruction for in-depth phenotyping. There are four types of 3D imaging sensors that could be used for phenotyping, including stereo cameras, laser range sensors, ToF cameras, and 3D light field cameras. In the next section, each type of sensor is discussed in detail.



There are two types of stereo vision: passive and active type (Chen and Zheng, 1995). Passive stereo vision senses the same object by two or more 2D cameras that are placed at distinct locations. Passive stereo vision estimates the depth information by searching the correspondence between the images taken by the cameras and then doing triangulation. Researchers have applied this technique to analyze the transplant variables of sweet potatoes (He et al., 2003) and to reconstruct a 3D model of the corn plant canopy for phenotyping (Ivanov et al., 1995; Ivanov et al., 1994). Different from passive stereo vision, active stereo vision projects structured light, such as parallel lines and grids, on the object surface. The structured light is distorted by the object 3D geometry, and it forms new distorted patterns. An object's 3D shape can be recovered by analyzing the images of the distorted light pattern. Veeken et al. (2006) reported about a corn plant sensing system that estimated the 3D structure of a corn plant by projecting parallel lines onto it. The major challenge of applying passive stereo vision to agriculture is the mismatch or unavailability of correspondence due to the plant's lack of texture on leaves, the complexity of its canopy structure, and the occlusion problem. Additionally, changing lighting conditions, which is common for outdoor agricultural applications, can influence passive stereo vision performance (Omasa et al., 2007; Weiss et al., 2010). Active stereo cameras have been used to achieve great success for indoor applications recently. The David structured light scanner (David Structured Light Scanner SLS-2, David Vision System GmbH, Germany) can generate high-quality 3D images with precision up to 0.1% of the scan size (down to 0.06 mm). However, they are vulnerable to strong ambient light, and can hardly work outdoors except in the evening.

Light detection and ranging (Lidar) is a more powerful direct, 3D measurement sensor based on a laser. As it is an active sensor with a powerful laser light source, it is not

susceptible to ambient sunlight and can work reliably outdoors. Some researchers have positioned the Lidar several meters away from relatively large-scale sensing objects to estimate the height (Hopkinson et al., 2004; Hosoi et al., 2005) and canopy structure (Hosoi and Omasa, 2006; Van der Zande et al., 2006) of trees and the carbon stock of forest. The structural parameters of trees and forest, including tree height, stem count, biomass estimation, material distribution, etc., are provided by these systems. Based on the 3D points' density distribution of multiple image views captured by Lidar sensors, the dry weight of the stem, leaves, and ears of a corn field could also be estimated (Hosoi and Omasa, 2009). The recent Lidar technology development has enabled a high-resolution 3D sensor based on the optical-probe method with a range accuracy of 0.5mm at a distance of 3.5m. Omasa et al. (2007) reported attaining high-density and clean 3D data of aubergine (*Solanum melongena* L.) and sunflowers using this type of sensor, and have demonstrated the 3D point cloud composited with textures such as natural color, chlorophyll fluorescence, and leaf temperature information. One limitation of Lidar system is that it does not capture the 3D points of one frame simultaneously, and, therefore, it requires that the sensor and the plant be kept relatively still during the process, making the system difficult to deploy on a mobile platform or when conducting in-field phenotyping tasks on windy days.

A ToF 3D sensor is a good alternative technology for 3D reconstruction. It measures an environment's depth information based on the ToF of emitted modulated light reflected by the objects in the field of view. Two commonly used ToF cameras, PMD Camcube (PDM Tec, Germany) and IFM Efector 3D (IFM, USA), were reported to have a depth resolution of around 1 cm and to be usable with an autonomous phenotyping robot if there is no heavy dust (Klose et al., 2009). 3D plant modeling research was reported by combining a 2D color

camera and a 3D ToF camera, whose position and attitude was controlled and measured using a robotic arm (Alenya et al., 2011). In their research, instead of segmenting the plant from the background in 3D space, they combined 3D spatial data with 2D color information and performed segmentation based on color in 2D space. Their segmentation program, which relied on a fused depth/color image, took approximately 1 minute to run. This approach may get high accuracy and reliability if the robotic arm is of high positioning accuracy and precision, but the equipment cost is high. On the contrary, Rusu et al. (2008) developed a robust multiple-viewpoint cloud alignment approach based on 3D feature histograms. Their algorithm calculated the feature histogram of each point of every view and looked for the key feature points whose feature histogram is significantly different than others. By searching for the correspondence of key feature points between different views, the relationship between different views can be found and used for alignment. The advantage of this approach is that it is purely based on 3D image processing, and it does not depend on any other device such as high-precision robot arms, which lowers the system cost. However, it is calculation intensive and thus requires a long processing time.

A 3D light field camera is another newly available 3D sensor (Raytrix, Germany). It has a micro-lens array, and each micro-lens captures a tiny image. Therefore, the resulting raw image is a composition of tiny images. Software is used to extract the depth information based on the correspondence of these images. This technique features high spatial resolution and accuracy, and it can work reliably outdoors under sunlight. However, its depth calculation cannot be done in real-time and therefore is post processed. Additionally, its price is significantly higher than that of the ToF camera.

The 3D imaging sensor based phenotyping system is still at its infancy stage (Furbank and Tester, 2011). Although the development of advanced 3D sensors, including active stereo vision (Lidar and 3D ToF camera), makes 3D spatial data available with relatively reasonable resolution, accuracy, and noise, 3D reconstruction algorithms for crop plant phenotyping are still primitive. This method for 3D reconstruction of plants does not achieve adequate accuracy with real-time processing speed or a more affordable system cost. The 3D features extracted are still limited to those simple ones that are primarily based on the statistics of point cloud distribution. Efforts still need to be made to achieve a phenotyping system that can perform 3D reconstruction and extract in-depth 3D structural features with the accuracy, processing speed, and reliability desired by plant genomics researchers and the plant genetics industry.

### 1.3 Machine Vision System for Automated Precision Weed Control

#### 1.3.1 Background

An alternative weed control approach to herbicide weed control that is clean, effective, and affordable is important to reduce the chemical usage for agricultural production systems, as it will alleviate the concerns of herbicide-resistant weeds, environmental pollution, and human health issues.

Manual weeding becomes impractical because of labor costs and regulation. In contrast, automated weeding systems have great potential for reducing the economic and environmental cost while providing high-level weed control (Thompson et al., 1991). There are two research areas for robotic weed control, including weed removal mechanism for robotic actuation and the control of robotic actuator. Four weed control mechanisms,

including a mechanical system, a precision chemical spraying system, a thermal method, and a high-voltage electrical discharge method, have been reported, and the first three of them have been developed into successful commercial products (Slaughter et al., 2008a).

There are three methods to guide and control a weed actuator. First, guidance systems were developed to locate the crop row and accordingly control tractor steering and weeding actuator position. Machine vision based (Åstrand and Baerveldt, 2005; Kise et al., 2005; Søggaard and Olsen, 2003) and real-time kinematic global positioning system (RTK GPS)–based guidance systems were developed, and both types of systems have shown centimeter-level accuracy (Jørgensen et al., 2002; Nagasaka et al., 2004). Although the guidance system knows the crop row path, it is limited to identify and locate the crop plant, even for the machine vision based system. Therefore, it is good for interrow (between crop rows) weeding but poor at intrarow (within or close to crop rows) weed control (Slaughter et al., 2008a).

Second, GPS mapping systems have been applied for weed control. Some researchers equipped planters with RTK GPS to generate the crop seed map during planting. With this method, the plant detected at places other than those of the recorded crop seed planting position were recognized as weed. Ehsani et al. (2004) tested this kind of system for maize seed mapping, and reported that the average error between seed map and actual plant position after germination was about 34 mm. The error resulted from a RTK GPS's error, the motion of the planting device related to GPS antenna, seed bounce in the furrows, and different soil conditions (Griepentrog et al., 2005).

Third, the machine vision based plant species identification system is another approach to localize the crop plant and weed for weed actuator control. Although guidance systems and GPS mapping systems have been developed and have achieved some

commercial success for nonrobotic weeding, the realization of a practical plant species identification system with reliable performance still has not been achieved (Slaughter et al., 2008b). Further efforts to address the roadblocks to machine vision based plant identification system are needed.

### **1.3.2 Machine Vision Based Plant Species Identification System**

For plant species recognition, 2D cameras and corresponding image processing algorithms were applied. A number of plant identification studies have extracted and analyzed morphological features of a leaf or a whole plant, such as length, width, perimeter dimensions, roundness, circularity, convexity, moment, etc. Slaughter et al. (2008a) reviewed these types of systems and concluded that they generally demonstrated high recognition accuracy under ideal condition when there was no occlusion problem to leaf or plant and the leaf was undamaged. However, they are not robust to the occlusion problem caused by visual defects of the plant caused by insect damage or wind twist, which are common in the field.

Spectral imaging was reported to be effective in segmenting vegetation from soil (Kumar et al., 2002; Noh et al., 2005) and discriminating crop plants from weed (Vrindts and Baerdemaeker, 1997). Zwiggelaar (1998) reviewed several studies that applied spectral features to discriminate crop plants from weeds. He found that although using spectral features to discriminate certain set of weed and crop plants is effective, the spectral wavebands selected for classification are frequently different for different weed and crop pairs. In an actual field, there are various weed species, leading to complexity of wavebands selection and algorithm design to discriminate crop plants from different weed species. Moreover, as a passive sensor that is susceptible to environmental factors, including variation

of sunlight, it is not believed to be reliable for plant species classification (Du et al., 2007; Furbank and Tester, 2011; Jones and Vaughan, 2010).

Stereo vision was reported to be applied for corn plant recognition (Jin and Tang, 2009) and structure analysis (Veeken et al., 2006). The major challenge in using stereo vision for practical agricultural systems is the correspondence searching problems caused by the lack of leaf texture, the complexity of the canopy structure, occlusion, and variation in sunlight conditions (Weiss et al., 2010).

As active 3D imaging sensors, Lidar and ToF cameras are less susceptible to ambient sunlight than passive sensing systems are. Although they can provide 3D spatial data with reasonable resolution and accuracy, the current 3D image processing algorithms for plant recognition reported in the literature are still primitive. Weiss et al. (2010) reported nearly 99% accuracy in classifying six plant species using Lidar in laboratory experiments. They used basic extracted 3D features, and their experiment design included the factors encountered in real field conditions, such as surrounding weed and plant occlusion. In some other research, a corn plant spacing system based on a ToF camera was developed (Nakarmi and Tang, 2012). It achieved a 98% detection rate of corn plants, where crop plants were significantly taller than weeds and other objects. However, this system was not developed and tested for the conditions where crop plants and weeds are of comparable heights. In addition, its processing speed was not fast enough to meet the speed requirements of an automated weeding system.

#### 1.4 Research Objective

The overall objective of this research was to develop 3D machine vision technologies for high-throughput plant phenotyping and robotic weed control.

To address the ultimate goal of this research, the major work could be divided into three main objectives:

(1) To develop a 3D machine vision based crop plant recognition system for automated precision weed control.

(2) To develop a machine vision based motion and attitude estimation system. 3D reconstruction is used as the foundation to develop a sophisticated 3D plant phenotyping system to extract the complex 3D features of plants. Its main challenge is to track precisely the posture of a 3D camera with high accuracy in order to align different 3D image views together into a complete 3D model.

(3) To develop a 3D machine vision based plant phenotyping system for plant phenomics research.

#### 1.5 Dissertation Overview

This dissertation contains three main parts: a 3D imaging based crop plant recognition system (chapter 2), a 2D imaging based motion and attitude estimation system (chapter 3), and a 3D reconstruction based phenotyping system (chapter 4).

In chapter 2, a 3D imaging based crop plant recognition system is introduced, and broccoli and soybean plants were used as the study objects. The sparse noise filtering algorithm of 3D point cloud data is introduced first. Then, the fast speed 3D geometry feature



estimation algorithm is discussed. Further, based on the extracted 3D features, the crop plant segmentation algorithm is proposed along with the experimental results listed at last.

In chapter 3, a 2D camera and chessboard pattern beacon based motion and attitude estimation system is introduced. This part of the research presents a real-time approach to measure the position and attitude of a 2D camera in relation to stationary beacons. The work in this chapter is the foundation of the 3D reconstruction system presented in chapter 3, in which the relationship of different 3D point cloud view data for alignment is indirectly given by a 2D camera and beacons. At first, a beacon detection algorithm is introduced. Then, the theory of the motion and attitude calculation of a 2D camera based on the detected beacon images is described. The experimental results of the beacon detection algorithm and the motion and attitude measurement accuracy are provided.

In chapter 4, a 3D reconstruction based plant phenotyping system using a 2D camera, a 3D ToF camera, and chessboard pattern beacons is introduced. Specifically, algorithms for camera calibration, noise filtering, 3D reconstruction are presented and discussed. Then, based on the reconstructed 3D models, plant trait features in 3D space are extracted, and the performance of the system is evaluated.

Chapter 5 contains general conclusions drawn from the research as well as suggestions for future work.

## 1.6 Reference

- Alenya, G., B. Dellen, and C. Torras. 2011. 3D modelling of leaves from color and tof data for robotized plant measuring. In 2011 IEEE International Conference on Robotics and Automation. Shanghai, China: IEEE.
- Allen, H. L., K. Estrada, G. Lettre, S. I. Berndt, M. N. Weedon, F. Rivadeneira, C. J. Willer, A. U. Jackson, S. Vedantam, and S. Raychaudhuri. 2010. Hundreds of variants

- clustered in genomic loci and biological pathways affect human height. *Nature* 467(7317):832-838.
- Åstrand, B., and A.-J. Baerveldt. 2005. A vision based row-following system for agricultural field machinery. *Mechatronics* 15(2):251-269.
- Chaerle, L., S. Lenk, I. Leinonen, H. G. Jones, D. Van Der Straeten, and C. Buschmann. 2009. Multi-sensor plant imaging: Towards the development of a stress-catalogue. *Biotechnology Journal* 4(8):1152-1167.
- Chen, C., and Y. F. Zheng. 1995. Passive and active stereo vision for smooth surface detection of deformed plates. *IEEE Transactions on Industrial Electronics* 42(3):300-306.
- Cobb, J. N., G. DeClerck, A. Greenberg, R. Clark, and S. McCouch. 2013. Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. *Theoretical and Applied Genetics* 126(4):867-887.
- Dhondt, S., N. Wuyts, and D. Inzé. 2013. Cell to whole-plant phenotyping: the best is yet to come. *Trends in plant science* 18(8):428-439.
- Du, J.-X., X.-F. Wang, and G.-J. Zhang. 2007. Leaf shape based plant species recognition. *Applied Mathematics and Computation* 185(2):883-893.
- EFO. Organic farming 101: Weed control. Earthbound Farm Organic. Earthbound Farm Organic. Available at: <http://www.ebfarm.com/learn/organic-101/controlling-weeds>.
- Ehsani, M., S. K. Upadhyaya, and M. L. Mattson. 2004. Seed location mapping using RTK GPS. *Transactions of the ASAE* 47(3):909-914.
- Fiorani, F., and U. Schurr. 2013. Future scenarios for plant phenotyping. *Annual review of plant biology* 64:267-291.
- Foundation, N. S., and G. Mcb. 2011. Phenomics : Genotype to Phenotype A report of the Phenomics workshop sponsored by the USDA and NSF.

- Furbank, R. T., and M. Tester. 2011. Phenomics - technologies to relieve the phenotyping bottleneck. *Trends in Plant Science* 16(12):635-644.
- Furbank, R. T., S. von Caemmerer, J. Sheehy, and G. Edwards. 2009. C-4 rice: a challenge for plant phenomics. *Functional Plant Biology* 36(10-11):845-856.
- Granier, C., L. Aguirrezabal, K. Chenu, S. J. Cookson, M. Dauzat, P. Hamard, J. J. Thioux, G. Rolland, S. Bouchier-Combaud, A. Lebaudy, B. Muller, T. Simonneau, and F. Tardieu. 2006. PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in *Arabidopsis thaliana* permitted the identification of an accession with low sensitivity to soil water deficit. *New Phytologist* 169(3):623-635.
- Griepentrog, H.-W., M. Nørremark, H. Nielsen, and B. Blackmore. 2005. Seed mapping of sugar beet. *Precision Agriculture* 6(2):157-165.
- He, D. X., Y. Matsuura, T. Kozai, and K. C. Ting. 2003. A binocular stereovision system for transplant growth variables analysis. *Applied Engineering in Agriculture* 19(5):611-617.
- Heffner, E. L., J.-L. Jannink, and M. E. Sorrells. 2011. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *The Plant Genome* 4(1):65-75.
- Hopkinson, C., L. Chasmer, C. Young-Pow, and P. Treitz. 2004. Assessing forest metrics with a ground-based scanning lidar. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere* 34(3):573-583.
- Hosoi, F., and K. Omasa. 2006. Voxel-based 3-D modeling of individual trees for estimating leaf area density using high-resolution portable scanning lidar. *IEEE Transactions on Geoscience and Remote Sensing* 44(12):3610-3618.
- Hosoi, F., and K. Omasa. 2009. Estimating vertical plant area density profile and growth parameters of a wheat canopy at different growth stages using three-dimensional portable lidar imaging. *Isprs Journal of Photogrammetry and Remote Sensing* 64(2):151-158.
- Hosoi, F., K. Yoshimi, Y. Shimizu, and K. Omasa. 2005. 3-D measurement of trees using a portable scanning Lidar. *Phyton-Annales Rei Botanicae* 45(4):497-500.

- Ivanov, N., P. Boissard, M. Chapron, and B. Andrieu. 1995. Computer stereo plotting for 3-D reconstruction of a maize canopy. *Agricultural and Forest Meteorology* 75(1-3):85-102.
- Ivanov, N., P. Boissard, M. Chapron, and P. Valery. 1994. Estimation of the height and angles of orientation of the upper leaves in the maize canopy using stereovision. *Agronomie* 14(3):183-194.
- Jansen, M., F. Gilmer, B. Biskup, K. A. Nagel, U. Rascher, A. Fischbach, S. Briem, G. Dreissen, S. Tittmann, S. Braun, I. De Jaeger, M. Metzlaff, U. Schurr, H. Scharr, and A. Walter. 2009. Simultaneous phenotyping of leaf growth and chlorophyll fluorescence via GROWSCREEN FLUORO allows detection of stress tolerance in *Arabidopsis thaliana* and other rosette plants. *Functional Plant Biology* 36(10-11):902-914.
- Jin, J., and L. Tang. 2009. Corn Plant Sensing Using Real-Time Stereo Vision. *Journal of Field Robotics* 26(6-7):591-608.
- Jones, H. G., and R. A. Vaughan. 2010. Remote sensing of vegetation: principles, techniques, and applications. Oxford University Press, Oxford, UK.
- Jørgensen, M., H. T. Søgaaard, and P. S. Nielsen. 2002. Præcision ved automatisk styring af radsensere. *Grøn Viden-Markbrug*(268).
- Kise, M., Q. Zhang, and F. Rovira Más. 2005. A stereovision-based crop row detection method for tractor-automated guidance. *Biosystems Engineering* 90(4):357-367.
- Klose, R., J. Penlington, and A. Ruckelshausen. 2009. Usability study of 3D Time-of-Flight cameras for automatic plant phenotyping. *Bornimer Agrartechnische Berichte* 69:93-105.
- Konishi, A., A. Eguchi, F. Hosoi, and K. Omasa. 2009. 3D monitoring spatio-temporal effects of herbicide on a whole plant using combined range and chlorophyll a fluorescence imaging. *Functional Plant Biology* 36(10-11):874-879.
- Kumar, L., K. Schmidt, S. Dury, and A. Skidmore. 2002. Imaging spectrometry and vegetation science. In *Imaging spectrometry*, 111-155. Springer.

- Lu, Y., L. J. Savage, M. D. Larson, C. G. Wilkerson, and R. L. Last. 2011. Chloroplast 2010: a database for large-scale phenotypic screening of *Arabidopsis* mutants. *Plant physiology* 155(4):1589-1600.
- Miyao, A., Y. Iwasaki, H. Kitano, J.-I. Itoh, M. Maekawa, K. Murata, O. Yatou, Y. Nagato, and H. Hirochika. 2007. A large-scale collection of phenotypic data describing an insertional mutant population to facilitate functional analysis of rice genes. *Plant Molecular Biology* 63(5):625-635.
- Nagasaka, Y., N. Umeda, Y. Kanetai, K. Taniwaki, and Y. Sasaki. 2004. Autonomous guidance for rice transplanting using global positioning and gyroscopes. *Computers and Electronics in Agriculture* 43(3):223-234.
- Nakarmi, A. D., and L. Tang. 2012. Automatic inter-plant spacing sensing at early growth stages using a 3D vision sensor. *Computers and Electronics in Agriculture* 82:23-31.
- Nichols, R. J., S. Sen, Y. J. Choo, P. Beltrao, M. Zietek, R. Chaba, S. Lee, K. M. Kazmierczak, K. J. Lee, and A. Wong. 2011. Phenotypic landscape of a bacterial cell. *Cell* 144(1):143-156.
- Noh, H., Q. Zhang, S. Han, B. Shin, and D. Reum. 2005. Dynamic calibration and image segmentation methods for multispectral imaging crop nitrogen deficiency sensors. *Transactions of the ASAE* 48(1):393-401.
- Omasa, K., F. Hosoi, and A. Konishi. 2007. 3D lidar imaging for detecting and understanding plant responses and canopy structure. *Journal of Experimental Botany* 58(4):881-898.
- Rajendran, K., M. Tester, and S. J. Roy. 2009. Quantifying the three main components of salinity tolerance in cereals. *Plant Cell and Environment* 32(3):237-249.
- Richards, R. A., G. J. Rebetzke, M. Watt, A. G. Condon, W. Spielmeyer, and R. Dolferus. 2010. Breeding for improved water productivity in temperate cereals: phenotyping, quantitative trait loci, markers and the selection environment. *Functional Plant Biology* 37(2):85-97.
- Rusu, R. B., N. Blodow, Z. C. Marton, and M. Beetz. 2008. Aligning point cloud views using persistent feature histograms. In *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*. IEEE.

- Schilling, C. H., J. S. Edwards, and B. O. Palsson. 1999. Toward metabolic phenomics: analysis of genomic data using flux balances. *Biotechnology Progress* 15(3):288-295.
- Scholes, J. D., and S. A. Rolfe. 2009. Chlorophyll fluorescence imaging as tool for understanding the impact of fungal diseases on plant performance: a phenomics perspective. *Functional Plant Biology* 36(10-11):880-892.
- Sirault, X. R. R., R. A. James, and R. T. Furbank. 2009. A new screening method for osmotic component of salinity tolerance in cereals using infrared thermography. *Functional Plant Biology* 36(10-11):970-977.
- Slaughter, D., D. Giles, and D. Downey. 2008a. Autonomous robotic weed control systems: A review. *Computers and electronics in agriculture* 61(1):63-78.
- Slaughter, D. C., D. K. Giles, and D. Downey. 2008b. Autonomous robotic weed control systems: A review. *Computers and Electronics in Agriculture* 61(1):63-78.
- Søgaard, H. T., and H. J. Olsen. 2003. Determination of crop rows by image analysis without segmentation. *Computers and electronics in agriculture* 38(2):141-158.
- Speliotes, E. K., C. J. Willer, S. I. Berndt, K. L. Monda, G. Thorleifsson, A. U. Jackson, H. L. Allen, C. M. Lindgren, J. a. Luan, and R. Mägi. 2010. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature genetics* 42(11):937-948.
- Sticklen, M. B. 2007. Feedstock crop genetic engineering for alcohol fuels. *Crop Science* 47(6):2238-2248.
- Swarbrick, P. J., P. Schulze-Lefert, and J. D. Scholes. 2006. Metabolic consequences of susceptibility and resistance (race-specific and broad-spectrum) in barley leaves challenged with powdery mildew. *Plant Cell and Environment* 29(6):1061-1076.
- Thompson, J., J. Stafford, and P. Miller. 1991. Potential for automatic weed detection and selective herbicide application. *Crop protection* 10(4):254-259.
- Van der Zande, D., W. Hoet, L. Jonckheere, J. van Aardt, and P. Coppin. 2006. Influence of measurement set-up of ground-based LiDAR for derivation of tree structure. *Agricultural and Forest Meteorology* 141(2-4):147-160.

- Veeken, M. v. d., L. Tang, and J. W. Hofstee. 2006. Automated corn plant spacing measurement at early growth stages using active computer vision. In ASABE 2006 Annual International Meeting. Portland.
- Vrindts, E., and J. d. Baerdemaeker. 1997. Optical discrimination of crop, weed and soil for on-line weed detection. Precision agriculture '97. Volume II. Technology, IT and management. Papers presented at the First European Conference on Precision Agriculture, Warwick University, UK, 7-10 September 1997.
- Walter, A., H. Scharr, F. Gilmer, R. Zierer, K. A. Nagel, M. Ernst, A. Wiese, O. Virnich, M. M. Christ, B. Uhlig, S. Juenger, and U. Schurr. 2007. Dynamics of seedling growth acclimation towards altered light conditions can be quantified via GROWSCREEN: a setup and procedure designed for rapid optical phenotyping of different plant species. *New Phytologist* 174(2):447-455.
- Weiss, U., P. Biber, S. Laible, K. Bohlmann, and A. Zell. 2010. Plant species classification using a 3d lidar sensor and machine learning. In Ninth International Conference on Machine Learning and Applications (ICMLA). IEEE.
- Winzeler, E. A., D. D. Shoemaker, A. Astromoff, H. Liang, K. Anderson, B. Andre, R. Bangham, R. Benito, J. D. Boeke, and H. Bussey. 1999. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* 285(5429):901-906.
- Zwiggelaar, R. 1998. A review of spectral properties of plants and their potential use for crop/weed discrimination in row-crops. *Crop Protection* 17(3):189-206.

## CHAPTER 2. 3D IMAGING BASED CROP PLANT AND WEED RECOGNITION SYSTEM

### 2.1 Abstract

A 3D time-of-flight (ToF) camera was applied to develop a crop plant recognition system for broccoli and soybean plants. The developed system overcame the previously unsolved problems caused by occluded canopy and illumination variation. An efficient noise filter was developed to filter the sparse noise points in 3D space. Both 2D and 3D features, including gradient of amplitude and depth image, surface curvature, amplitude percentile index (API), normal direction, and neighbor point count in 3D space, were extracted and found effective for recognizing these two types of plants. According to the 3D geometry and 2D amplitude characteristics of broccoli and soybean, one segmentation algorithm was developed for each of them. For the experimental condition in which the crop plants were heavily surrounded with various types of weed plants, detection rates over 88.3% and 91.2% were achieved for broccoli plant and soybean leaves, respectively. Additionally, the crop plants were segmented out with relatively complete shape. Moreover, the program is optimized, and over 30 fps image processing speed was achieved by this system.

### 2.2 Introduction

An alternative approach to herbicide weed control that is clean and effective at an affordable cost is important to reduce the chemical usage for agricultural production systems, as it will alleviate the concerns of herbicide-resistant weeds, environmental pollution, and human health issues.



Manual weeding becomes impractical because of labor costs and regulation. In contrast, automated weeding systems have great potential for reducing the economic and environmental cost while providing high-level weed control (Thompson et al., 1991). There are two research areas for robotic weed control, including weed removal mechanisms for robotic actuation and the control of the robotic actuator. Four weed control mechanisms, including a mechanical system, a precision chemical spraying system, a thermal method, and a high-voltage electrical discharge system, have been reported, and the first three of them have been developed into successful commercial products (Slaughter et al., 2008a).

There are three methods to guide and control weed actuator. First, guidance systems were developed to locate the crop row and accordingly control tractor steering and the weeding actuators' position. Guidance systems based on machine vision (Åstrand and Baerveldt, 2005; Kise et al., 2005; Søgaaard and Olsen, 2003) and real-time kinematic global positioning system (RTK GPS) were developed, and both types of systems have shown centimeter-level accuracy (Jørgensen et al., 2002; Nagasaka et al., 2004). Although the guidance system knows the crop row path, it is limited in its ability to identify the crop plant, even for a machine vision based system. Therefore, it is good for interrow (between crop rows) weeding but poor at intrarow (within or close to crop rows) weed control (Slaughter et al., 2008a).

Second, GPS mapping systems were applied for weed control. Some researchers have equipped planters with RTK GPS to generate crop seed maps during planting. With this method, the plant detected at the places other than those of the recorded crop seed planting position were recognized as weeds. Ehsani et al. (2004) tested this kind of system for maize seed mapping, and reported that the average error between the seed map and the actual plant

position after germination is about 34 mm. The error resulted from the RTK GPS's error, the motion of the planting device related to GPS antenna, seed bounce in the furrows, and different soil conditions (Griepentrog et al., 2005).

Third, the machine vision based plant species identification system is an important approach to localize the crop plant for robotic weeding. Whereas guidance systems and GPS mapping systems have achieved development and some commercial success for non-robotic weeding, practical plant species identification system with reliable performance have still not been achieved yet (Slaughter et al., 2008b). The effort to address the roadblock of machine vision based plant identification system is desired.

For plant species recognition, two dimensional (2D) cameras and corresponding image processing algorithms were applied. A number of plant identification researchers have extracted and analyzed morphological features of a leaf or a whole plant, such as length, width, perimeter dimensions, roundness, circularity, convexity, moment, etc. Slaughter et al. (2008a) reviewed these types of systems and concluded that they generally demonstrate high recognition accuracy under ideal conditions, when there is no occlusion problem to the leaf or plant and the leaf is undamaged. However, they are not robust to the occlusion problem, and the visual defects of a plant caused by insect damage and wind twist are common in the field.

Spectral imaging was reported to be effective in segmenting vegetation from soil (Kumar et al., 2002; Noh et al., 2005) as well as discriminating crop plants from weed (Vrindts and Baerdemaeker, 1997). Zwiggelaar (1998) reviewed several studies that applied spectral features to discriminate crop plants from weeds. He found that although spectral features are effective to discriminate certain sets of weed and crop plants, the spectral

wavebands selected for classification are frequently different for different weed and crop pairs. In the practical field, there are various weed species, leading to complexity of wavebands selection and algorithm design to discriminate crop plants from different weed species. Moreover, as a passive sensor, which is susceptible to environment factors including variation of sunlight, it is not believed to be reliable for plant species classification (Du et al., 2007; Furbank and Tester, 2011; Jones and Vaughan, 2010).

Stereo vision was reported to be applied for corn plant recognition (Jin and Tang, 2009) and structural analysis (Veeken et al., 2006). The major challenge in using stereo vision for practical agricultural systems is the correspondence searching problems caused by the lack of leaf texture, the complexity of the canopy structure, occlusion, and variation in sunlight conditions (Weiss et al., 2010).

As active 3D imaging sensors, Lidar and ToF cameras are less susceptible to ambient sunlight than passive sensing systems are. Although they can provide 3D spatial data with reasonable resolution and accuracy, the current 3D image processing algorithms for plant recognition reported in the literature are still primitive. Weiss et al. (2010) reported nearly 99% accuracy in classifying six plant species using Lidar in laboratory experiments. They used basic extracted 3D features, and their experiment design included the factors encountered in real field conditions, such as surrounding weed and plant occlusion. In some other research, a corn plant spacing system based on a ToF camera was developed (Nakarmi and Tang, 2012). It achieved more than 98% corn plant detection accuracy with about 2–3 cm root mean square error (RMSE) plant stem localization error based on morphological analysis of 2D intensity and 3D spatial data, where crop plants were significantly taller than weed and other objects were. However, running the algorithm on an i7 quad core 2.8 Hz

CPU, it could only achieve a processing speed that was equivalent to 0.8 km/h travel speed, when acquiring sequential images along a crop row, which is too slow for use in an automated weeding system.

This research aims to apply a 3D imaging sensor and develop a corresponding machine vision algorithm to discriminate crop plants from weed plants that can work in practical field conditions. Ultimately, this research aims to provide a machine vision solution with satisfactory accuracy, reliability, and fast processing speed for fully automated robotic weeding operation.

### 2.3 Sensor and Data Collection System

A 3D ToF camera (Swissranger SR4000, MESA Imaging AG, Zurich, Switzerland) was applied as the sensor in this study. It has an illumination source, and it measures the distance of objects based on the ToF of the reflected light. Thanks to the powerful light source, this active 3D sensor has a degree of robustness to ambient sunlight, and it can work outdoors when the sunlight is not too strong. Otherwise, a cover or umbrella can be adopted to make this camera work by reducing the sunlight intensity. This camera can estimate x, y, z coordinate locations for each pixel relative to its sensor based coordinate system. It also can capture the amplitude image, which represents the intensity of the reflected signal emitted from its light source, as Figure 2.5 shows. Its resolution is  $144 \times 176$  pixels.

Figure 2.1 shows the SR4000 sensor. A data collection system was built by mounting the 3D sensor and a laptop on a modified golf cart, as Figure 2.2 (a) indicates, to collect the top view image of vegetation. The 3D sensor is marked by a red rectangle in Figure 2.2 (a). Figure 2.2 (b) shows the 3D camera installation on the data collection system. The height of

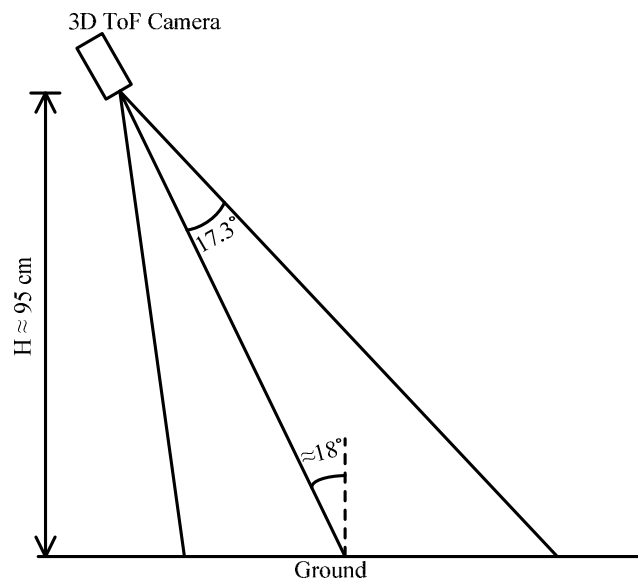
the 3D camera was around 95 cm, and the angle between its view direction and vertical direction was around  $18^\circ$ . In this case, the distance between camera and plant canopy was about 90 cm, and the camera's spatial resolution was around  $3.8 \text{ mm} \times 3.8 \text{ mm}$  per pixel. By pushing the golf cart and running the data capturing program, the continuous amplitude image and point cloud data with x, y, z value were collected and stored for the crop plant recognition study.



Figure 2.1. Swissranger SR4000



(a)



(b)

Figure 2.2. (a) Picture of data collection system and (b) drawing of 3D ToF camera setup

## 2.4 Experiment Design

The study objects of this crop plant recognition research were broccoli and soybean plants in the field with complex field conditions. Image data were collected in the broccoli and soybean fields shown by Figure 2.3 and Figure 2.4, respectively. It is clear to see that the broccoli field was full of various types of weeds that are common in Iowa, including brome grass (*Bromus inermis* Leyss), Waterhemp (*Amaranthus rudis*), pigweed (*Amaranthus*), lambsquarters (*Chenopodium album*), cockspur grass (*Echinochloa crus-galli*), purslane (*Portulaca oleracea*), bindweed (*Convolvulus arvensis*), and clover (*Trifolium*). The broccoli plants were heavily infested by weeds with similar height. The soybean field was also full of weeds, mainly including brome grass (*Bromus inermis* Leyss), purslane (*Portulaca oleracea*), waterhemp (*Amaranthus rudis*), cockspur (*Echinochloa crus-galli*), and bindweed (*Convolvulus arvensis*). The majority of weeds plants were brome grass plants. Some weed plants were higher than the soybean plants, and they partially occluded the soybean canopies (Figure 2.4).

The image data were collected between 5 and 8 p.m. on sunny days (June 26, July 6, and July 18, 2012). The 3D ToF camera works the best when ambient light is weak or when sun light is diffused. During the image acquisition process, we used an umbrella to block direct sunlight from getting into the field of view of the camera.

In this study, the data collection system was pushed through the path between the crop rows at the speed of around 0.2 m/s, to collect consecutive 3D images and amplitude images of 206 broccoli plants and three rows of soybean plants. The length of one row of soybean plant was about 20 m. The interrow distance of both the broccoli and the soybean field is around 30 cm. The intrarow distance of the broccoli field was approximately 46 cm.

In total, 48,179 and 18,793 image sets were collected for broccoli and soybean plants, respectively, over the three data collection times. Each plant has around 200 images, which were captured at different viewpoints when the system was pushed to pass by during the data collection process. Moreover, this approach allowed a more comprehensive testing for the crop plant recognition at different viewpoints.

Table 2.1 gives the heights of broccoli plants and soybean plants corresponding to each data collection date. There was little maturity difference observed in either the broccoli plants between June 26th and July 5th or the soybean plants between June 26 and July 5, which was probably because of the competition between weeds and crop plants.

Table 2.1. Height of crop plant corresponding to each image capturing date

| Data collection date | Height of broccoli plant (cm) | Height of soybean plant (cm) |
|----------------------|-------------------------------|------------------------------|
| June 26, 2012        | 18–23                         | Approx. 20                   |
| July 6, 2012         | 18–23                         | Approx. 28                   |
| July 18, 2012        | 18–30                         | N/A                          |





Figure 2.3. Broccoli field of this study



Figure 2.4. Soybean field of this study

## 2.5 Algorithm Design

The ToF camera used in this research collected 2D amplitude images and 3D information. This crop plant recognition research developed 2D and 3D image processing algorithms to process the amplitude and 3D images.

### 2.5.1 Initial Noise Filtering

Because of the strong ambient sunlight, some points of the data collected by the ToF camera did not have correct amplitude and 3D coordinate information. Therefore, it was important to clear those noise points to start.

The amplitude signal provided by the sensor is 16 bits wide with the range between 0 and 65,535, and a typical noise point has significant big amplitude value, shown as the white area marked by a red rectangle at the bottom left corner of Figure 2.5. In this study, it was effective to recognize the point with the amplitude value above 2000 as noise data, based on experience and observation. Additionally, the pixels of the 3D image with a depth value over 1.5 m were also judged as noise because the distance between the 3D sensor and the ground was less than approximately 1 m in this study. These two criteria were applied for initial noise clearance. Relatively clean data were achieved for later feature estimation, based on which a more advanced noise clearance algorithm would be adopted.

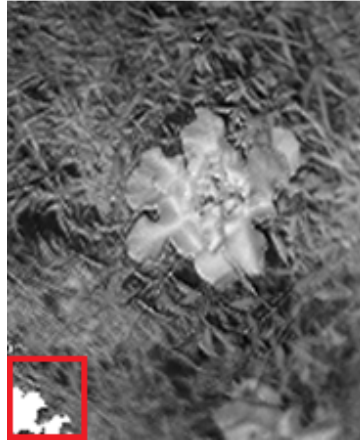


Figure 2.5. Amplitude image of broccoli with noise at left corner

### 2.5.2 Neighbor Points Table Development

Searching for the neighbor points of each point is the foundation of 3D image processing. Due to measurement error, the 3D image has some outliers, namely sparse point noise, shown as the areas marked by red circles in Figure 2.7 (b). Those outliers had relatively big distance from other points and were filtered out by analyzing each point's neighborhood pixels. The crop plant recognition study estimated the local 3D geometry feature, including the curvature and normal direction of the point cloud. Both tasks were based on the analysis of each point and its neighbors.

A neighbor point search algorithm calculation is intensive for nonorganized 3D point cloud data. Its efficiency and speed are critical for the processing speed of the whole system. To develop fast searching algorithms, various data structures, such as Octree, kd-tree, and bd-tree (Arya et al., 1998) have been shown to organize the 3D point cloud data. Even though these algorithms greatly improved the speed, the CPU cost to search for the neighbor points of every point of a 3D image is still a heavy burden for practical application.

The 3D point cloud data from the 3D ToF camera is organized information, which is stored inside a  $144 \times 176$  (camera's resolution) buffer array like 2D images. In this research, the point stored in the  $i_{th}$  row and the  $j_{th}$  column is represented as  $p_{i,j}$ . For each pixel in the buffer, only the pixels stored in its neighbor columns and rows in the buffer are the neighbor point candidates in 3D space, because only those points have similar x and y values.

Therefore, for point  $p_{i,j}$ , only the points  $P \in \{p_{x,y} | i - w \leq x \leq i + w, j - w \leq y \leq j + w\}$  in the square window whose side was  $2w + 1$  were taken, to calculate their distance to the query point  $p_{i,j}$  for neighbor points judgment. This research concerned the neighborhood within only 10 mm for each 3D data point, and the spatial resolution of the image was approximately  $4 \times 4 \text{ mm}^2$ ; a window with four sides, each nine pixels long, was big enough, according to the observation. Therefore,  $w$  was chosen to be four in this research.

There are two ways to define the neighbor point set, which is represented by  $P^k$ , of a query point  $p_q$ , including:

- 1) The closest  $k$  neighbors of the query point
- 2) All of the  $k$  neighbors whose distance to the query point is not bigger than a specific radius  $r$ .

This study adopted the second method because of the nature of this application and data. The neighbor points of the query data were collected to estimate the local surface geometry and clear the sparse noise. As the 3D image of this study contained crop plants, weed plants, and soil residue, the discontinuity of 3D geometry and occlusion happened all over the image. In this condition, the point cloud of some places was dense, whereas that of other areas was relatively sparse in 3D space. While the goal of the neighbor points searching algorithm was to find all of the local data for 3D geometry analysis, and the first

neighborhood definition, i.e. 1), may mistakenly select the points far away from the query point for the place where the point intensity is low. Therefore, the second definition, i.e. 2), is more appropriate for local geometry analysis. In this study, if the distance between two points was not bigger than 10 mm, they were recognized as neighbor points in 3D space, and they were referred to as being connected in 3D space. The selection of 10 mm for the radius of the neighborhood is discussed in section 2.5.3.

Instead of calculating the distance  $d$  between two points for neighbor points judgment in 3D space, the distance square  $d^2$  was calculated by comparing with the threshold  $100 \text{ mm}^2$  in order to avoid time consuming square root calculation.

$$d^2 = (x_p - x)^2 + (y_p - y)^2 + (z_p - z)^2 \quad (2.1)$$

For every 3D data point, its  $k$  neighbors in 3D space was found and stored for later image processing, and they are represented with  $P_{10}^k$ . Its number of neighbors was also stored inside of the neighbor count image,  $N_C$ , which is shown by Figure 2.6 (d). The maximum  $N_C$  value of Figure 2.6 (d) is 32, and that of the whole data set collected in this research is 40. As  $N_C$  indicates, the pixels of broccoli and soybean leaves have significantly higher number of neighbor points in 3D space than most pixels of soil and weed background because of their relatively smooth 3D geometry. This observation was helpful for this crop plant recognition study.

### 2.5.3 Curvature and Normal Estimation of Surface

Curvature and normal direction are important local features that capture the 3D geometry of the local surface around a query point  $p_q$ . These features are critical for this plant recognition study. One example is that a query point  $p_q$  and its neighbor points

belonging to the same soybean or broccoli leaf should have similar normal direction and small curvature because the local surface of the plant leaves is relatively smooth and continuous.

To estimate the curvature and normal direction of the local surface around a query point  $p_q$ , the first step is to search all of its neighbor points within a specific distance, and the selection of the radius for the neighborhood definition is important and application dependent. While the smaller maximum distance value will reduce the number of neighbor points and make the calculation result more susceptible to the noise of the 3D image, a larger value is more likely to mistakenly select the points of other objects' surfaces as the neighbor points to do surface normal and curvature estimation, leading to calculation error. In this study, the maximum distance of neighbor points for the local surface feature analysis was set as 10 mm, for several reasons:

- 1) Because the ToF camera was less than 0.9 m away for the plants, the point cloud was so dense that the distance between two closest points on a smooth surface was around 4 mm. Therefore, inside the sphere with a radius of 10 mm and centered at the query point  $p_q$ , normally over 10 neighbor points can be found. The number of points was enough to perform principal component analysis (PCA) to process the 3D coordinates of the neighborhood.
- 2) As the size of soybean leaves was small and the broccoli leaves were not flat, a small radius would make the local surface estimation algorithm more accurate and would help to avoid taking the relatively far away points for analysis.

- 3) The leaves of the crop plant and the weeds were close to each other, and occlusion of the canopy was common. A smaller radius can help to avoid taking the points of other leaves for local surface analysis.
- 4) The standard deviation of the 3D measurements of SR4000 used is around 4 mm, according to the datasheet. If the reflectance of the object is small, the standard deviation of measurement further increases. In this case, if the neighborhood definition radius is too small, the 3D measurement error can greatly affect the normal and curvature estimation accuracy of the local surface.

Considering the problems, the radius of 10 mm for neighbor points searching was chosen for the local surface feature estimation, based on the experiment. In addition, the corresponding neighbor points set  $P_{10}^k$  of each point was achieved by the previous step as introduced previously.

To estimate the normal direction of the local surface around a query point  $p_q$ , there are many methods to process the query point  $p_q$  and its neighbor points  $P_{10}^k$ . Klasing et al. (2009) introduced and compared seven surface normal estimation methods, including *PlaneSVD*, *PlanePCA*, *VectorSVD*, *QuadSVD*, *QuadTransSVD*, *AreaWeighted*, and *AngleWeighted*. Although the performance of these methods on agriculture applications have not been reported, these studies concluded that while the choice of best algorithm depends mostly on the 3D graph structure, *PlanePCA* has better performance in terms of both quality and speed for all cases, compared to the others.

*PlanePCA* stands for principal component analysis–based normal–estimation method, and the calculation process is introduced by Equation 2.2-2.6.

For the normal estimation of a query point, the set that contains it and its neighbor points  $P_{10}^k$  is represented using  $P$ , as Equation 2.2 shows. Moreover, set  $P$  has a total of  $k+1$  elements.

$$P = p_q \cup P_{10}^k \quad (2.2)$$

The  $i_{th}$  element of  $P$  is represented by  $p_i$ , which is a 3D feature vector consisting of the x, y, and z coordinate value of the point in the sensor's coordinate system, as Equation 2.3 indicates.

$$p_i = [x_i, y_i, z_i]^T \quad (2.3)$$

The mean vector  $\bar{p}$  and covariance matrix of  $P$  can be calculated using Equation 2.4 and 2.5, respectively.

$$\bar{p} = \frac{\sum_{i=1}^{k+1} p_i}{k+1} \quad (2.4)$$

$$C = \frac{1}{k} \sum_{i=1}^{k+1} (p_i - \bar{p})(p_i - \bar{p})^T \quad (2.5)$$

The covariance matrix  $C$  has three eigenvalue and eigenvector sets, which can be achieved by solving Equation 2.6.  $\vec{v}_j$ , and  $\lambda_j$  represents the  $j_{th}$  eigenvector and the corresponding eigenvalue, respectively.

$$C\vec{v}_j = \lambda_j\vec{v}_j, \quad j \in \{1, 2, 3\} \quad (2.6)$$

For *PlanePCA* based normal estimation method, the eigenvector corresponding to the lowest eigenvalue represents the normal direction at the query point.

In addition to normal calculation of local surface, PCA was also used for surface curvature estimation (Pauly et al., 2002). As Equation 2.7 shows, this study applied their method to represent surface curvature value  $\delta$  with the ratio of the minimum eigenvalue to the sum of three eigenvalues achieved by Equation 2.6. The curvature value  $\delta$  is positively



related to surface curvature. The bigger  $\delta$  indicates big curvature, and the maximum possible value of  $\delta$  is  $1/3$ . If  $\delta$  is equal to zero, it means that all the points are on a plane.

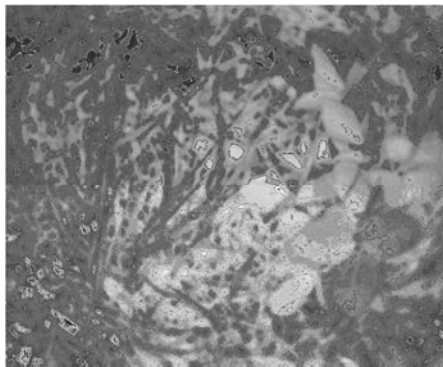
$$\delta = \frac{\min(\lambda_j)}{\lambda_1 + \lambda_2 + \lambda_3}, \quad j \in \{1, 2, 3\} \quad (2.7)$$

Moreover,  $\delta$  is not only related to curvature, it is also related to the noise level around the query point. The larger the noise level is, the higher  $\delta$  becomes.

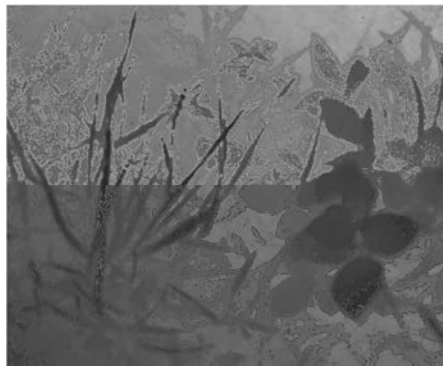
Figure 2.6 (e) is the curvature image. Figure 2.6 (f) is the color-coded image of normal direction. In Figure 2.6 (f), the normalized normal direction  $[u, v, w]$  is represented by encoded color  $[R, G, B]$ , which was calculated as  $[255(u+1), 255(v+1), 255(w+1)]$ . The black area of Figure 2.6 (f) represents the unavailability of a data point there because of noise filtering. Figure 2.6 also provides the corresponding depth image, amplitude image, and the color image captured by the 2D camera. As Figure 2.6 (e) indicates, the flat surface of scenery has low  $\delta$  value, whereas the weeds and soil have relatively high  $\delta$  value because of the discontinuity caused by the narrowness of the weed leaf and the roughness of earth surface. This indicates that  $\delta$  value may be an effective feature to differentiate the crop plant from soil and weeds. Figure 2.6 (f) uses color to represent the different normal direction of the surface. It shows that the normal direction around every point of a soybean leaf is almost consistent because of its 3D geometry characteristics, whereas the different leaves have different color because of their various normal directions. The weed and soil area is a mix of different colors because of the inconsistent normal direction caused by the discontinuity of their 3D geometry.



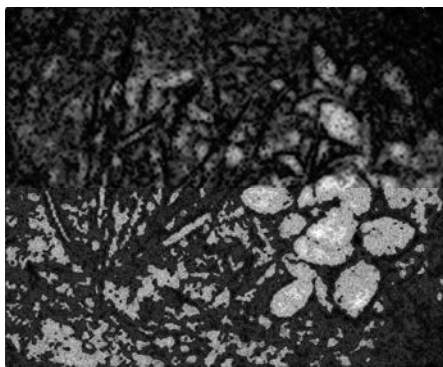
(a)



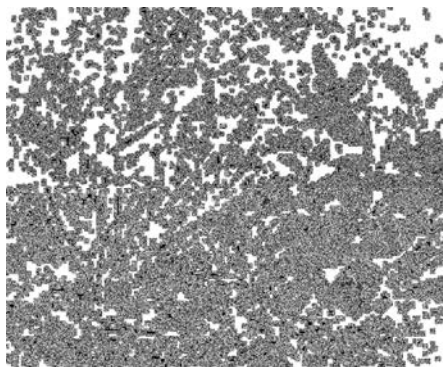
(b)



(c)

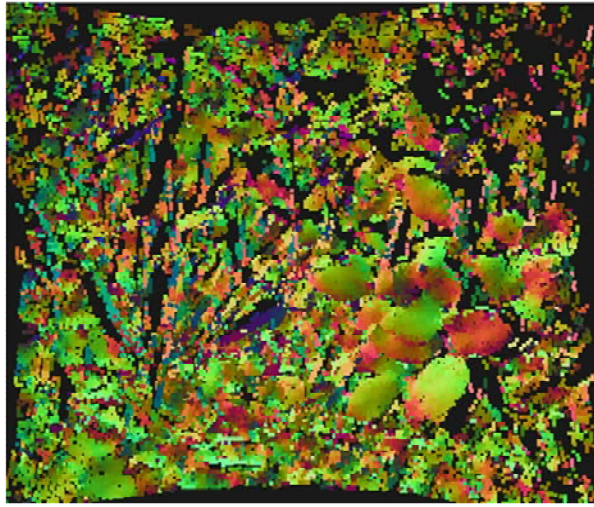


(d)



(e)

Figure 2.6 continued



(f)

Figure 2.6. Normal and curvature estimation of soybean: (a) 2D color image, (b) amplitude image from ToF camera, (c) depth image, (d) neighbor count image  $N_c$ , (e) curvature estimation result, and (f) normal direction represented by color image

#### 2.5.4 Advanced Noise Clearance Algorithm

Section 2.5.1 introduced the simple noise clearance method based on the amplitude intensity value and simple depth threshold. However, it is far from sufficient and can only be used as a preprocessing step. As Figure 2.7 indicates, even though the front view of 3D data after the initial noise seems clear, its side view indicates that it still has sparse point noise, as the area marked by a red circle in Figure 2.7 (b) shows.

After the neighbor points table, normal direction, and curvature values are achieved for the 3D point cloud data, further noise cleaning was conducted based on these results.

The first step of this advanced noise clearance algorithm was to organize all of the data points into different regions in 3D space based on neighborhood analysis. If the distance of two points was no bigger than 10 mm and the angle between their normal directions was

smaller than  $15^\circ$ , they were recognized as connected and placed inside the same region. To accomplish the task, the system analyzed every point of the 3D point cloud data except for the noise excluded at an earlier step as query point  $p_q$ . For each query point  $p_q$ , the angle  $\theta$  between its normal direction and that of all the neighbor points within 10 mm, which is  $P_{10}^k$ , achieved in previous step, was examined to check whether it is over the threshold.

Equation 2.8 gives the equation to calculate the angle between the normal directions of the query point  $p_q$  and the neighbor point.

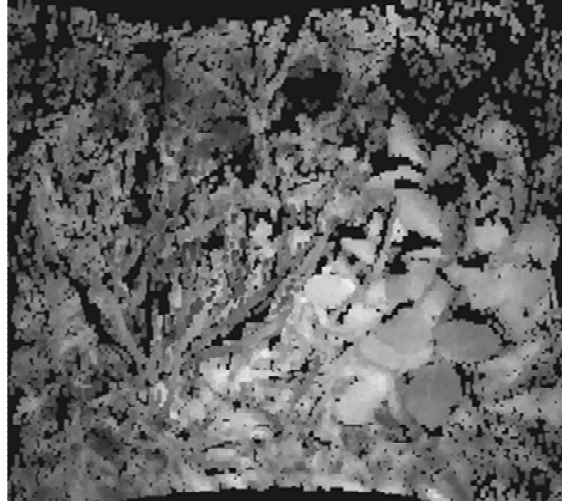
$$\theta = \arccos\left(\frac{\vec{n}_q \cdot \vec{n}_n}{|\vec{n}_q| |\vec{n}_n|}\right), \quad (2.8)$$

where  $\vec{n}_q$  is the normal direction at query point  $p_q$ , and the  $\vec{n}_n$  is that of the neighbor point  $p_n$ . Our system is only interested to know whether the angle between these two points is smaller than  $15^\circ$ , instead of the exact angle value; therefore, the problem can be simplified to check whether Equation 2.9 is satisfied in order to accelerate the program speed. Because  $\cos(15^\circ)$  in Equation 2.9 is a constant value (the equation avoids the time consuming inverse cosine function) the efficiency is improved further, as compared to Equation 2.8.

$$\vec{n}_q \cdot \vec{n}_n \geq |\vec{n}_q| |\vec{n}_n| \cos(15^\circ) \quad (2.9)$$

If Equation 2.9 was satisfied, the query point  $p_q$  and that specific neighbor point within 10 mm were classified into the same region as connected points. Moreover, this was iterated to process every point of the 3D point cloud data as the query point to check the relationship between it and each of its neighbor points. Based on that, it could be decided whether there was a link of neighbor points satisfying Equation 2.9 to connect any two points of the 3D image in 3D space. If so, these two points and the points of their link were classified into the same region in 3D space. Otherwise, these two points were split into different regions. In this way, all points were grouped into different regions in 3D space.

Then, points in the regions with a member count less than four are removed as sparse point noise.

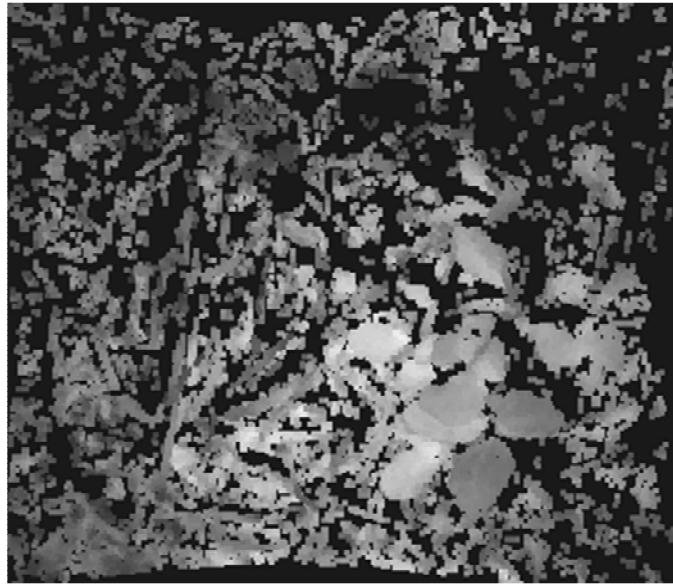


(a)

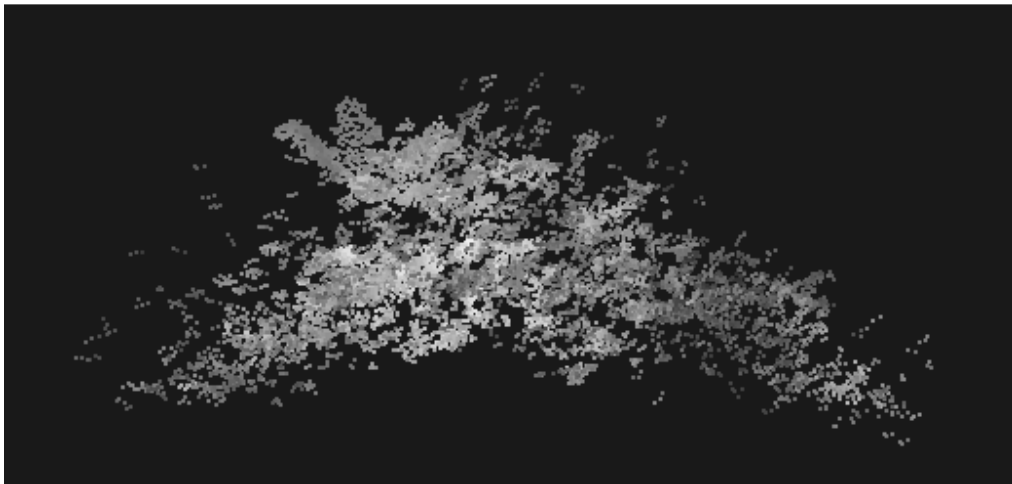


(b)

Figure 2.7. 3D point cloud data after initial noise clearance: (a) front view of 3D point cloud data and (b) side view with sparse point noise marked by red circle



(a)



(b)

Figure 2.8. 3D point cloud data after advanced noise clearance: (a) front view of 3D point cloud data and (b) side view

The advanced noise clearance result of the data shown in Figure 2.7 is given by Figure 2.8. By comparing Figure 2.7 (a) and Figure 2.8 (a), many points were removed by the advanced noise filter. These filtered points were sparse point noise; most of them were weed plants and soil residue.

### 2.5.5 Gradient of Amplitude and Depth Image

The gradient of amplitude image and depth image are important features for this plant recognition research. The Sobel operator was used to calculate the gradient for both images, and it applies two  $3 \times 3$  kernels to convolve with the original image to calculate the horizontal and vertical derivatives. For the original image represented with  $A$ , and the images of the horizontal and vertical derivatives represented with  $G_x$  and  $G_y$ , respectively, their relationship was expressed by Equation 2.10 and 2.11. In addition, the approximate derivatives value was calculated by summing the absolute value of the vertical and horizontal derivatives as Equation 12 indicates.

$$G_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} * A \quad (2.10)$$

$$G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * A \quad (2.11)$$

$$G = |G_x| + |G_y| \quad (2.12)$$

By using Equation 2.10–2.12 to process the amplitude image and depth image from the ToF camera, their gradient images were calculated respectively as Figure 2.9 shows. The depth gradient image and amplitude gradient image are represented using  $G_z$  and  $G_I$ , respectively, in this study.

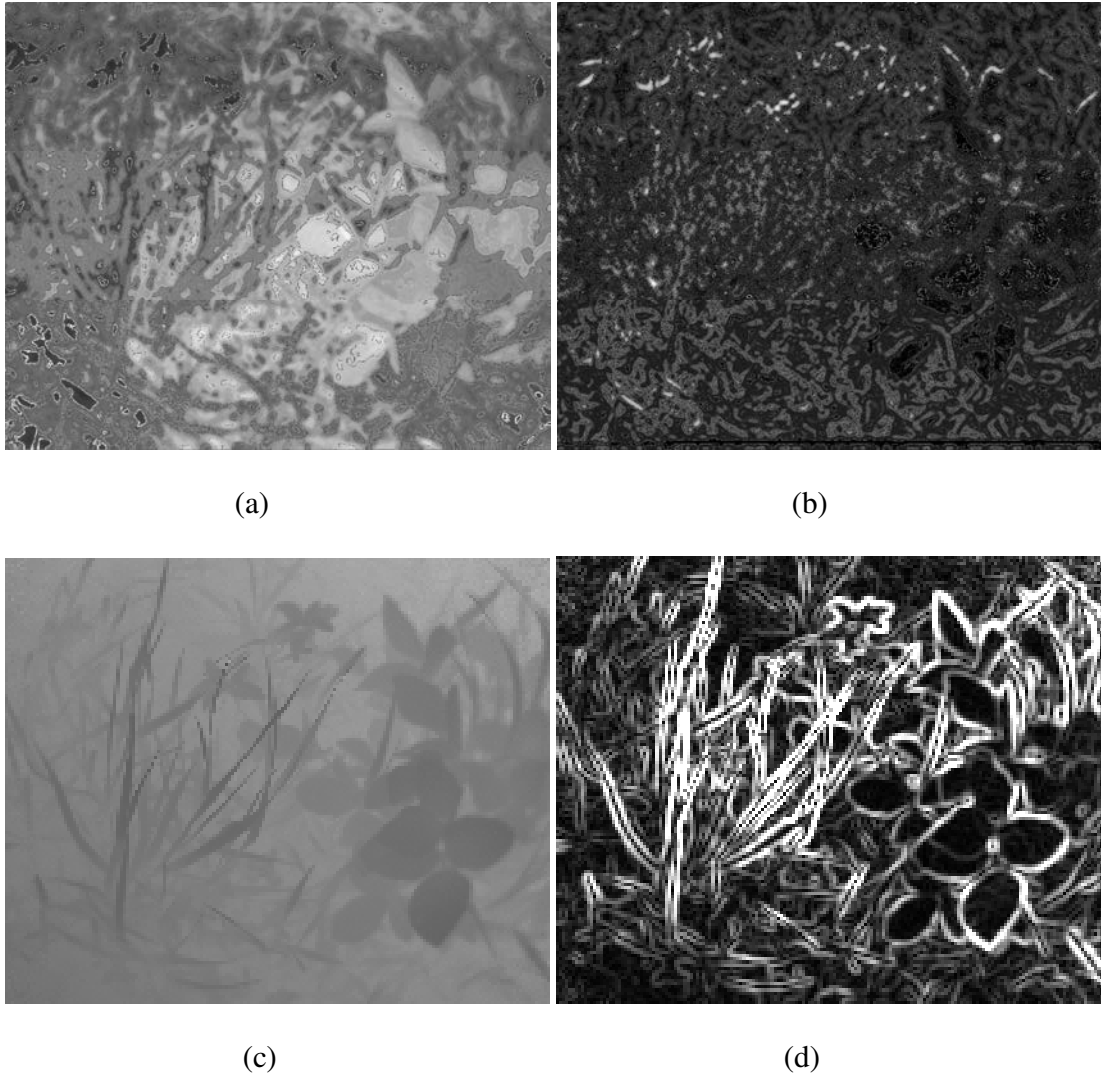


Figure 2.9. Gradient images: (a) amplitude image, (b) gradient of amplitude image, (c) depth image, and (d) gradient of depth image

As Figure 2.9 indicates, the amplitude and depth gradient is big for the place where the change of amplitude value is high and the discontinuity of 3D geometry happens respectively. They are critical features for the segmentation task of this research.

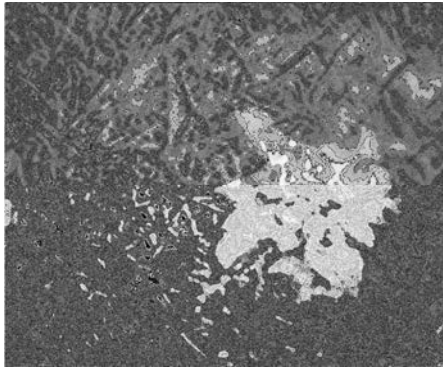
### 2.5.6 Percentile Index of Amplitude Image

As Figure 2.10 indicates, the broccoli plant has significantly higher intensity value than the weeds and soil in the amplitude image due to its higher reflectance for the 850 nm

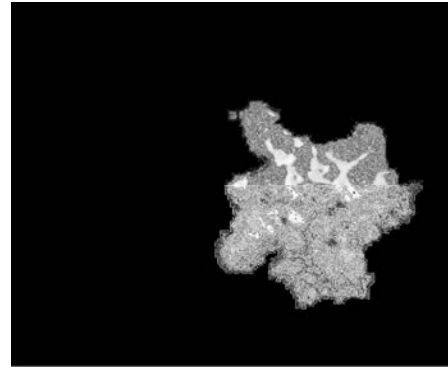


wavelength light emitted from the ToF camera. Figure 2.10 (a) is a broccoli amplitude image, and Figure 2.10 (b) is the broccoli area segmented out from the background. Figure 2.10 (d) and (e), which are the histograms of the background area and the broccoli area of amplitude image, respectively, indicate that the amplitude range of broccoli is significantly higher than the soil and weed background. For this data set, whereas the majority of the background's amplitude value is smaller than 150, the major range of broccoli's amplitude value is between 125 and 255. Therefore, the difference of their amplitude range can contribute to differentiating broccoli from the background.

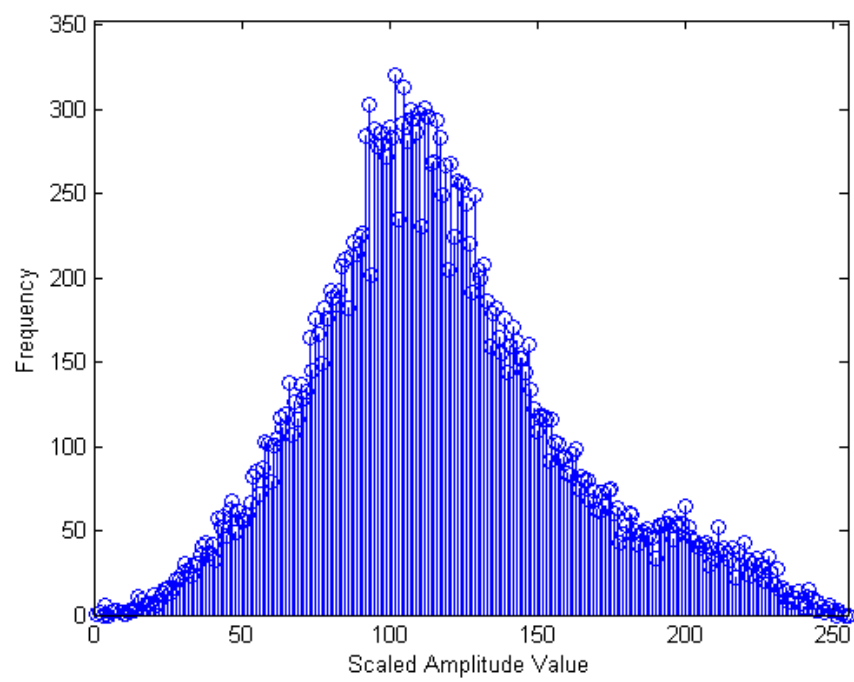
However, the previous histogram based threshold algorithm did not work well to segment the broccoli out from the amplitude image. Although the histograms of the background and the broccoli have significantly different ranges and peak values, the histogram of the whole amplitude image does not have two obvious peaks separated by a trough, as Figure 10 (c) shows. Because of this problem, the previous histogram based thresholding algorithms including the 2D maximum entropy thresholding algorithm (Kapur et al., 1985), minimum error thresholding algorithm (Kittler and Illingworth, 1986), and Otsu's method (Otsu, 1975) mistakenly recognized the whole image as a single area and therefore could not differentiate the broccoli from the background in the test conducted in this study.



(a)

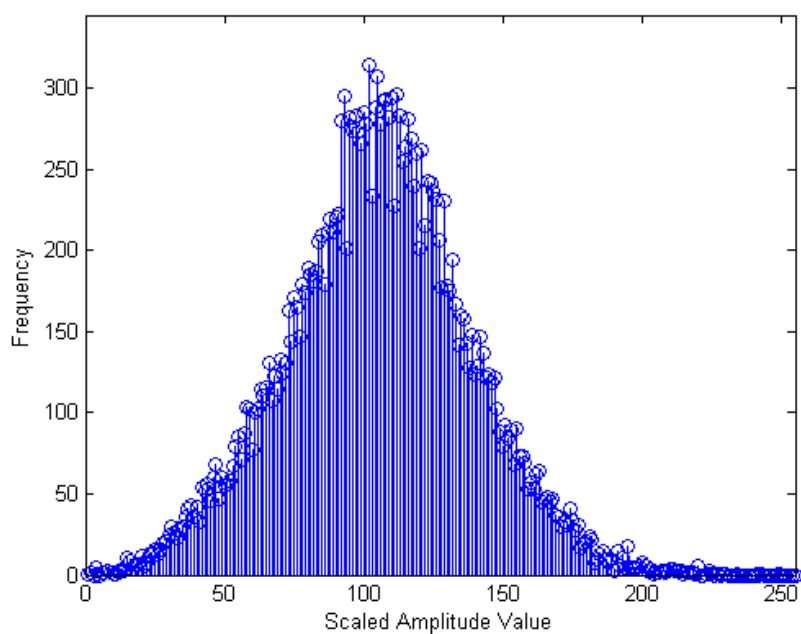


(b)

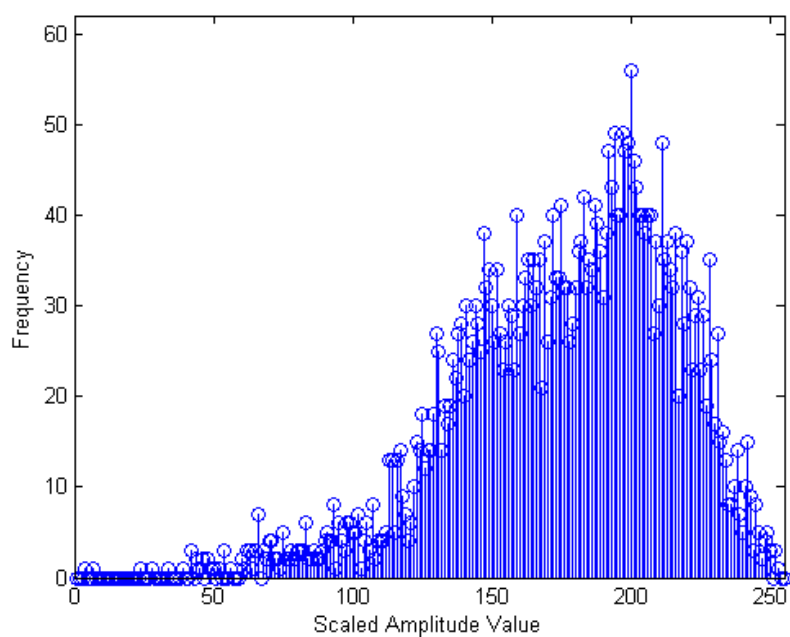


(c)

Figure 2.10 continued



(d)



(e)

Figure 2.10. Broccoli amplitude image and histograms: (a) broccoli amplitude image, (b) broccoli segmented from amplitude image, (c) histogram of whole amplitude image, (d) histogram of background, and (e) histogram of broccoli

To take the advantage of the fact that broccoli has a higher amplitude value than that of the soil and the weed background for crop plant recognition, this research created a feature called the API. While the data collection system was pushed through the field, the broccoli and soybean plants came into view from the right boarder, moved to the left side, and disappeared from the left side of view. Depending on the position of the crop plant, it partially or fully appeared in the image view. Therefore, the crop plant can occupy different area sizes in the image, and it can greatly influence the histogram of the amplitude image. In order to solve this problem, for the amplitude image with a dimension of  $144 \times 176$  pixels, each column was analyzed separately. The amplitude values of every column were ranked from low to high, and the percentile of each pixel within its corresponding column was its API value, which was stored in an API image. In this case, the API value of a pixel of broccoli and soybean was only related to its column, and was not susceptible to their position no matter whether the crop plant partially appeared at the left or the right boarder of the image or fully appeared in the center. Figure 2.11 is the API image of the amplitude image shown by Figure 2.10 (a). It is clear that the pixels of broccoli have significantly higher API value compared to the background pixels in the same column in the image.

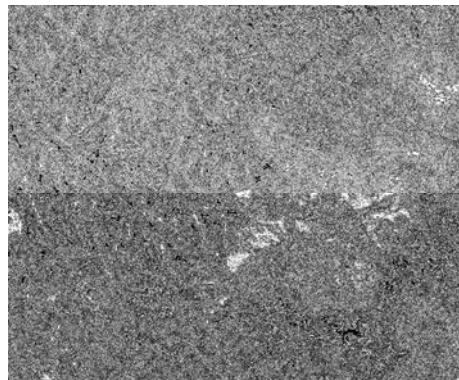


Figure 2.11. Amplitude percentile image of broccoli

The API feature extraction method was also applied to process the soybean's amplitude image. Even though the soybean's amplitude value was not significantly higher than that of the weeds in this study, API also can help to clear the soil from the image. Figure 2.12 shows the API calculation result of the soybean amplitude image given by Figure 2.9 (a). In Figure 2.12, the soil area has a low API value and therefore is dark, whereas the weeds and soybean leaves do not show significant difference.

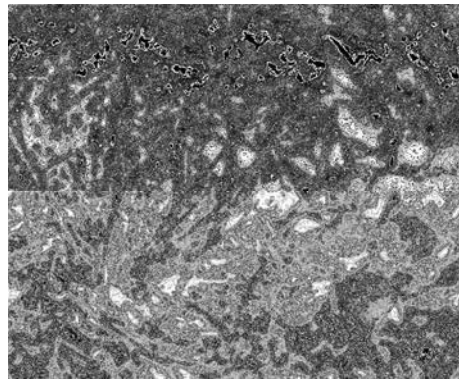


Figure 2.12. Amplitude percentile image of soybean

### 2.5.7 Statistics of Extracted Features

To develop a crop plant segmentation algorithm based on extracted 2D and 3D features, statistics analysis was done to investigate the distribution of each feature for broccoli, soybean, and background. Five randomly selected image sets of broccoli and soybean were used respectively for feature distribution analysis, and they were processed with manual segmentation to specify the area of crop plant and background. Figure 2.13 and 2.14 show one example set of depth and amplitude image, manual segmentation image, and five extracted features used for statistics analysis for soybean and broccoli, respectively.

For broccoli and soybean plants, histogram and cumulative histogram of each extracted feature were generated at the crop plant and background areas of five sample image

sets, respectively, as Figure 2.15 and 2.16 show. As Figure 2.15 (d) and 2.16 (d) indicate, the maximum cumulative histogram value of curvature feature is smaller than 1, this is because part of the crop plant and background area does not have effective curvature value because it does not have enough neighbor points for curvature estimation. For soybean image sets, 80% of the broccoli and only 22.41% of the background area had valid curvature value. For broccoli image sets, 85.85% of the soybean area and only 36.25% of the background had valid curvature value. Therefore, this is helpful to differentiate crop plant and background.

As the histogram and cumulative histogram indicate, the API and neighbor count features of broccoli and soybean leaves are mainly distributed in the higher value range, compared to the background, and the amplitude and depth gradient and curvature of these two crop plants are relatively smaller than the soil and weed background. Based on the feature distribution, this research carefully picked the threshold value for each feature in order to effectively remove the background area while keeping most of the crop plant area. The picked threshold and percentage of crop plant and background area within the range are given in Table 2.2 and 2.3. As these two tables indicate, the picked threshold values are effective to keep most broccoli and soybean areas and remove some background area. An example is the selected curvature threshold value 0.25, which can keep 79.54% of the broccoli area while removing 77.61% of the background. The combination of the threshold of all features is powerful to segment the inner area of the broccoli and soybean leaves, which is introduced in section 2.5.8.

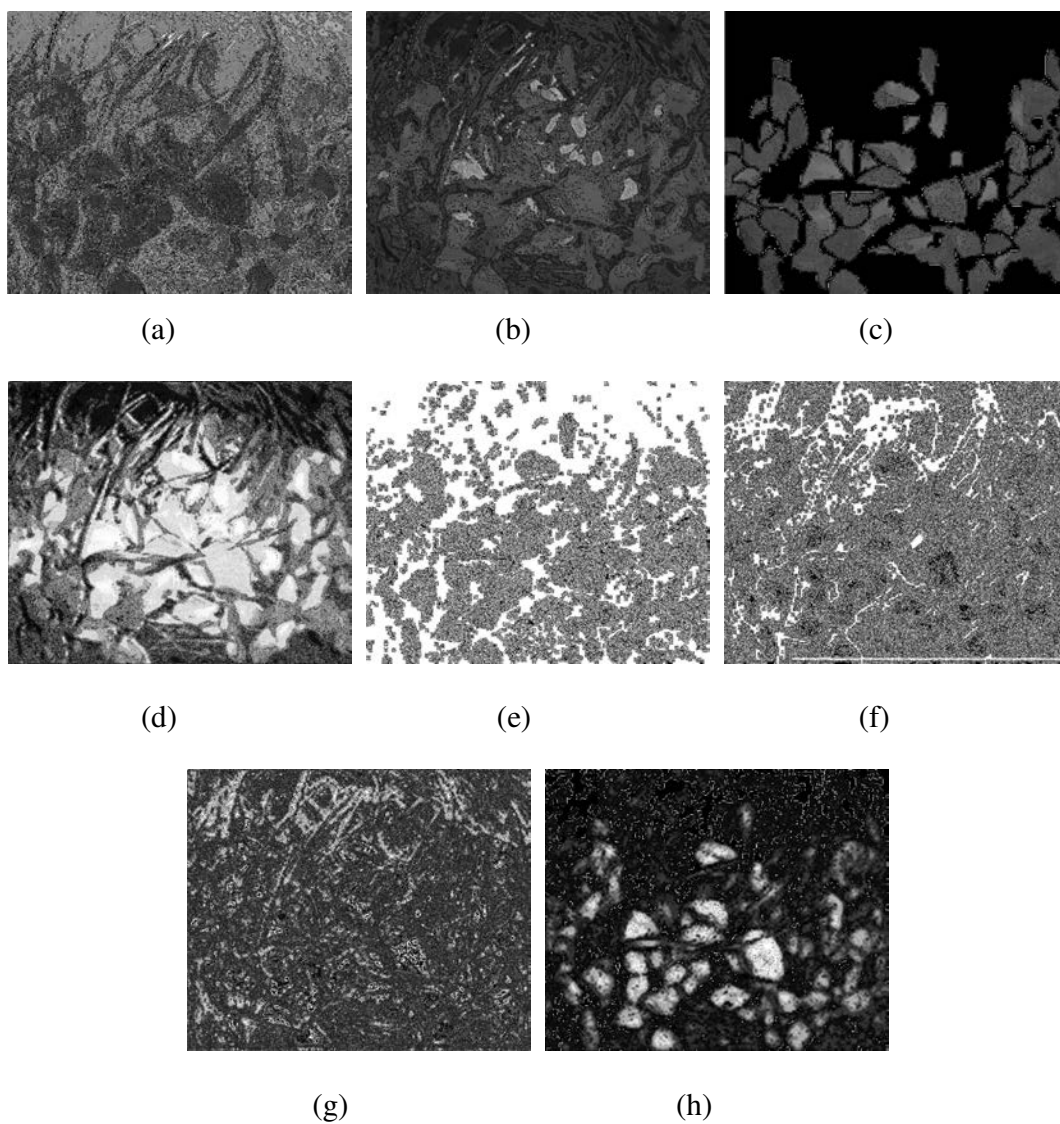


Figure 2.13. Sample image sets of soybean for feature distribution analysis: (a) depth image, (b) amplitude image, (c) manual segmentation image of soybean leaves, (d) API image, (e) curvature image, (f–g) depth and amplitude gradient images, and (h) neighbor count image

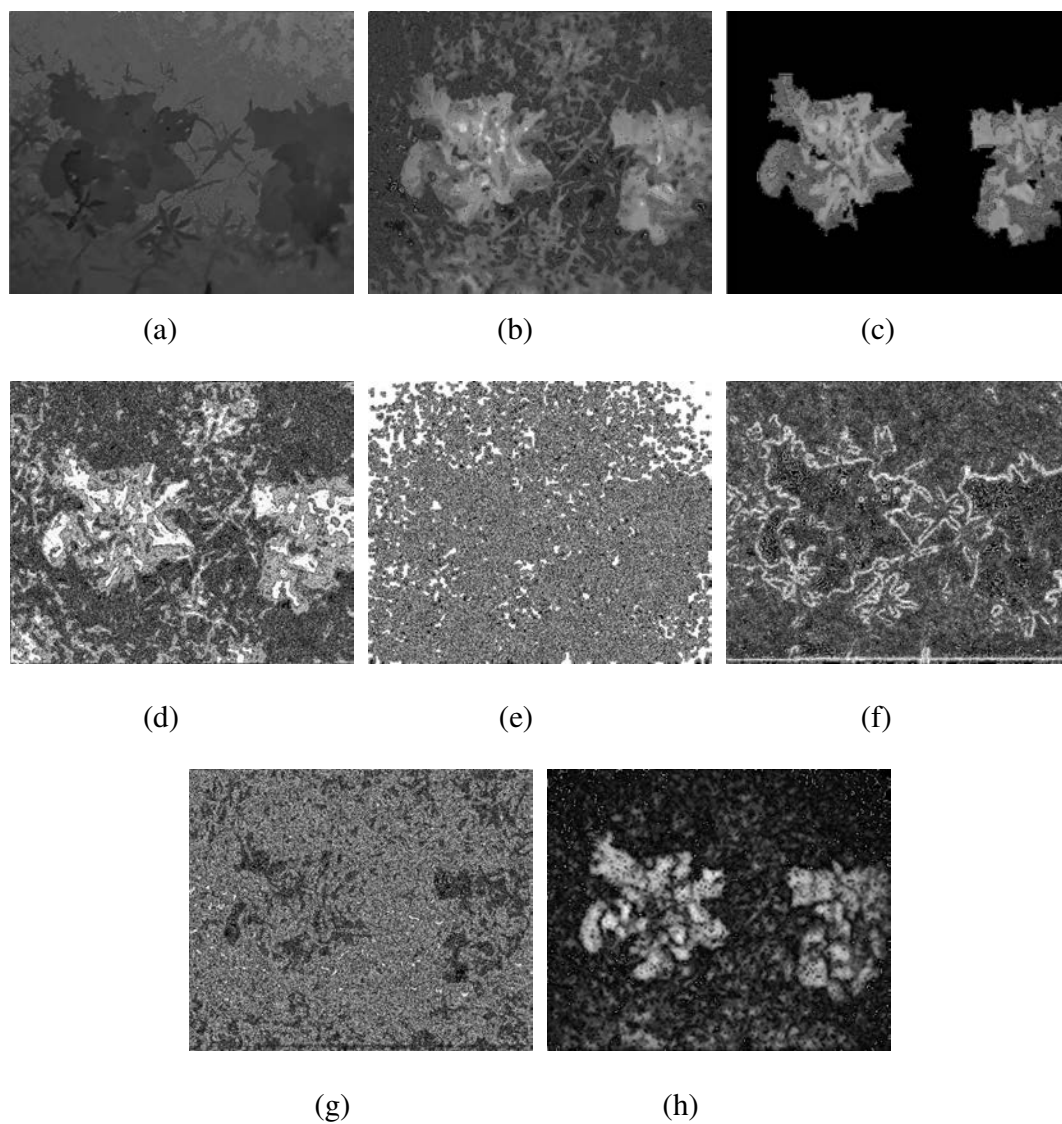


Figure 2.14. Sample image sets of broccoli for feature distribution analysis: (a) depth image, (b) amplitude image, (c) manual segmentation image of soybean leaves, (d) API image, (e) curvature image, (f–g) depth and amplitude gradient images, and (h) neighbor count image



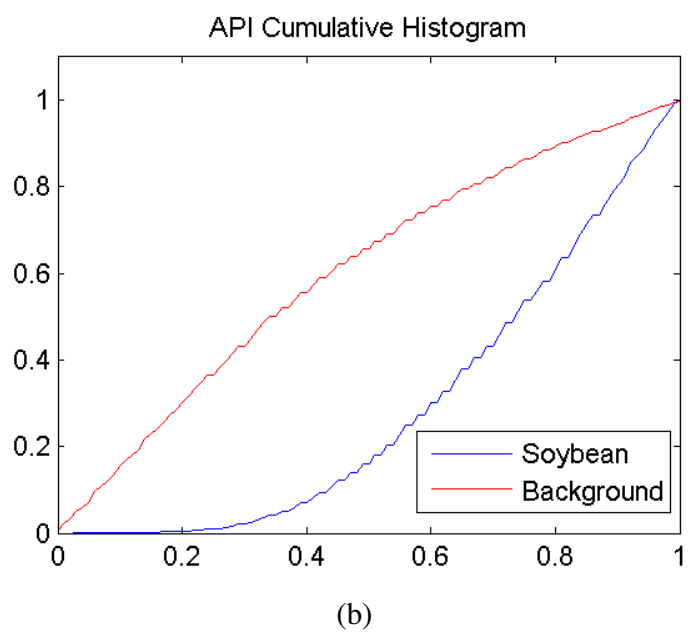
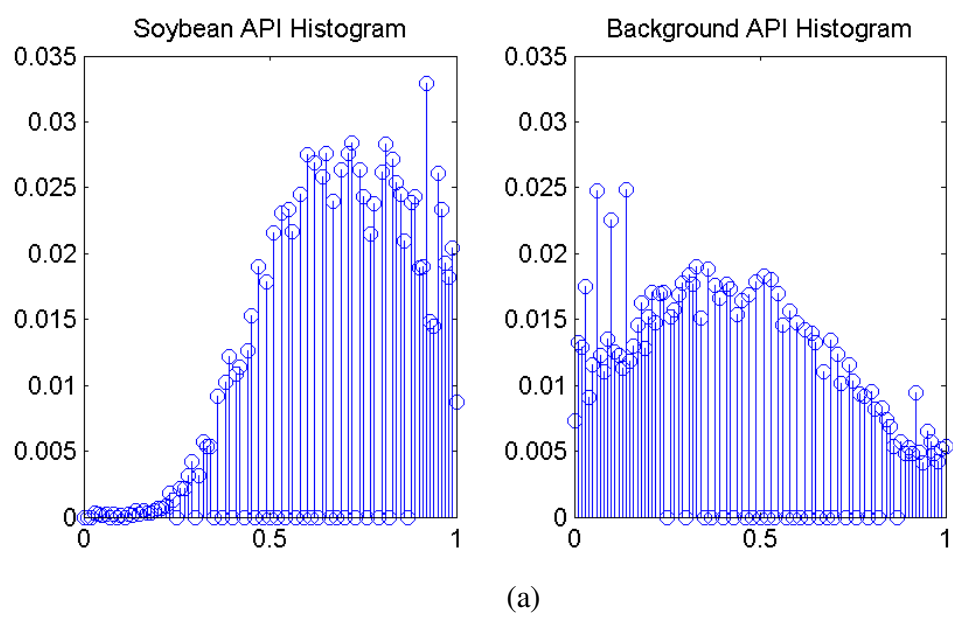


Figure 2.15 continued

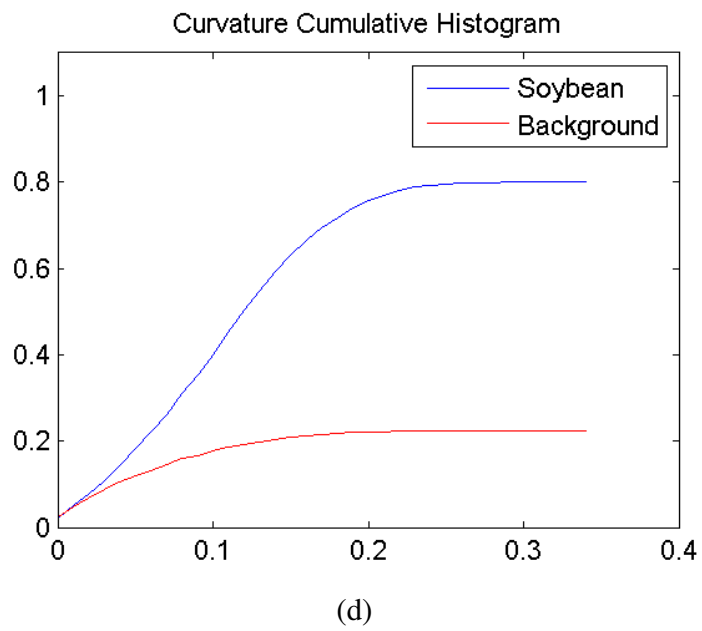
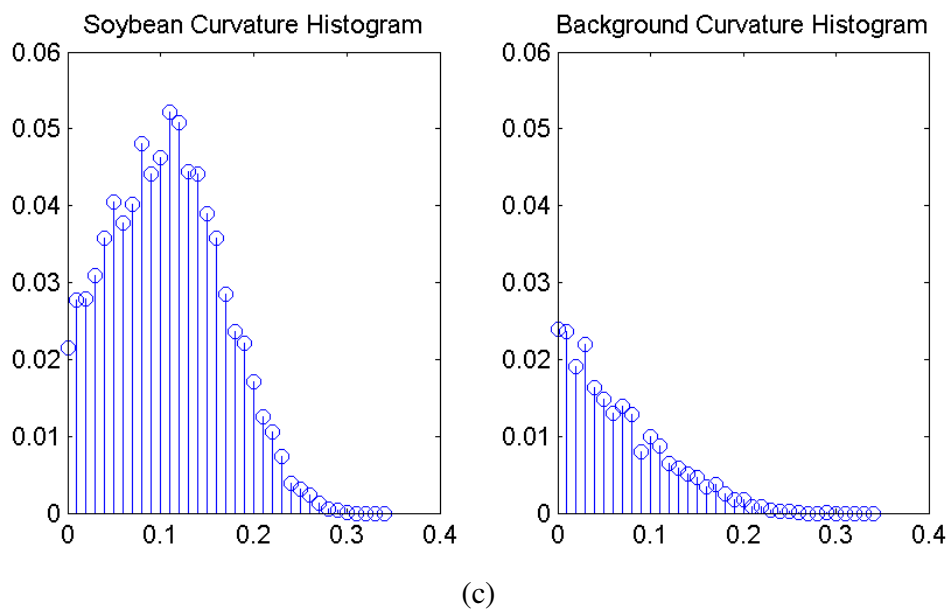


Figure 2.15 continued

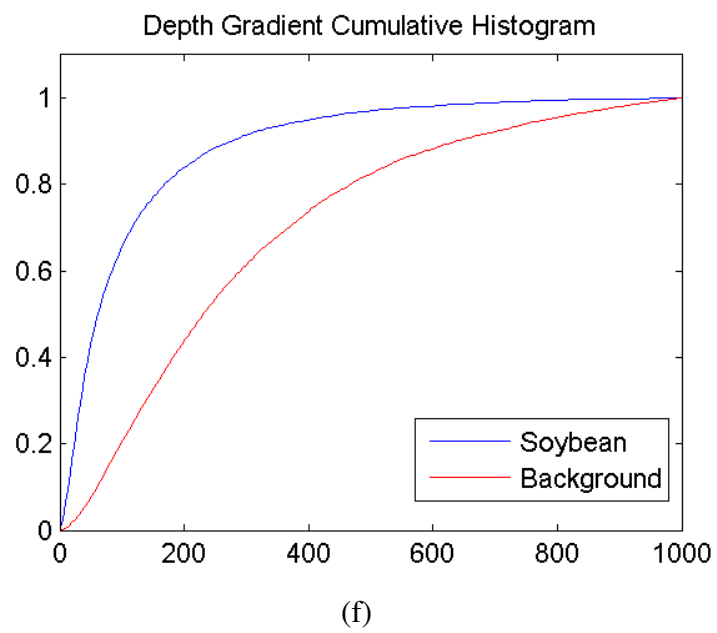
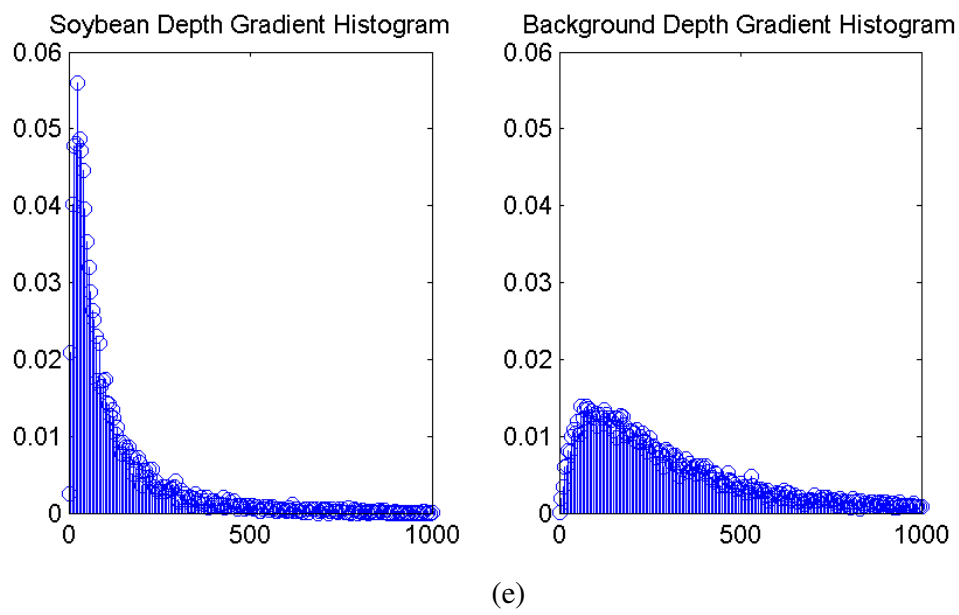


Figure 2.15 continued

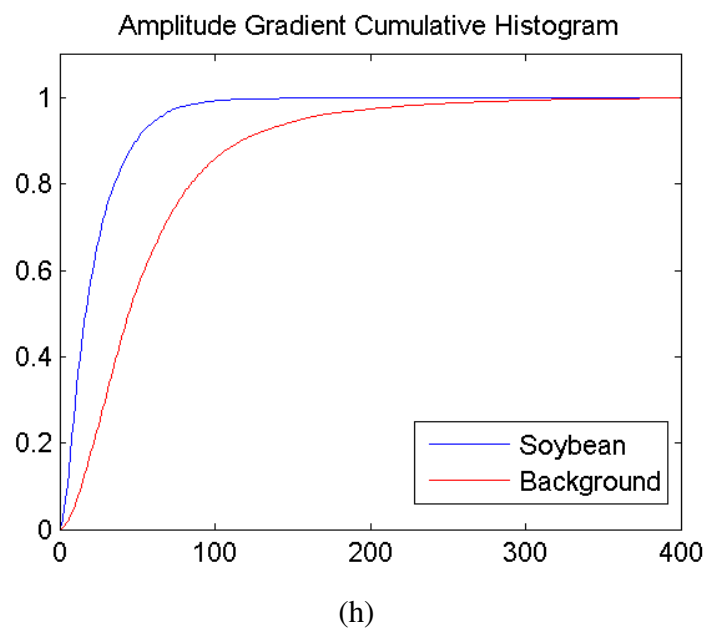
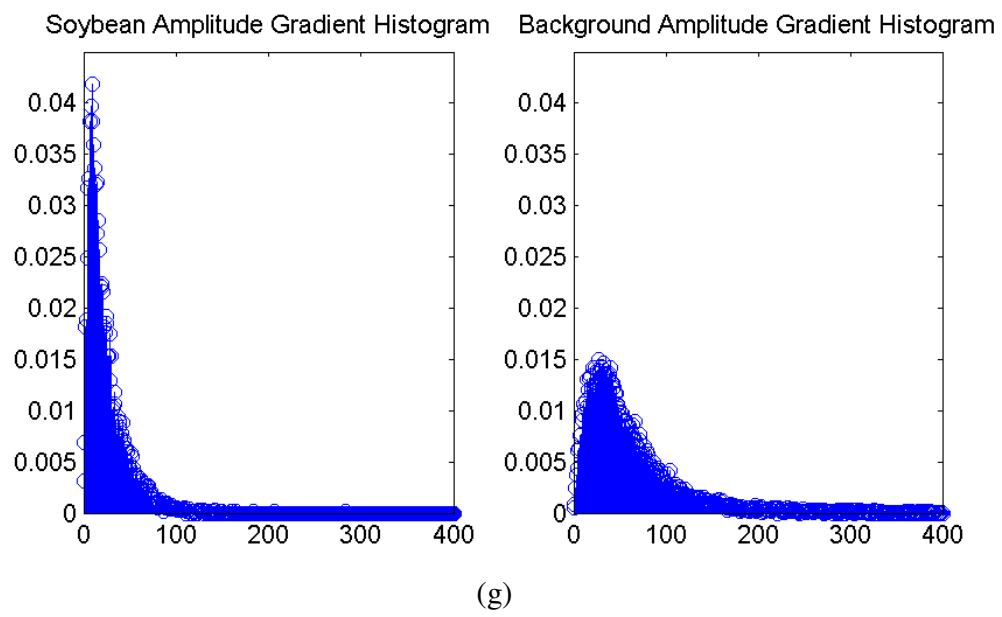


Figure 2.15 continued

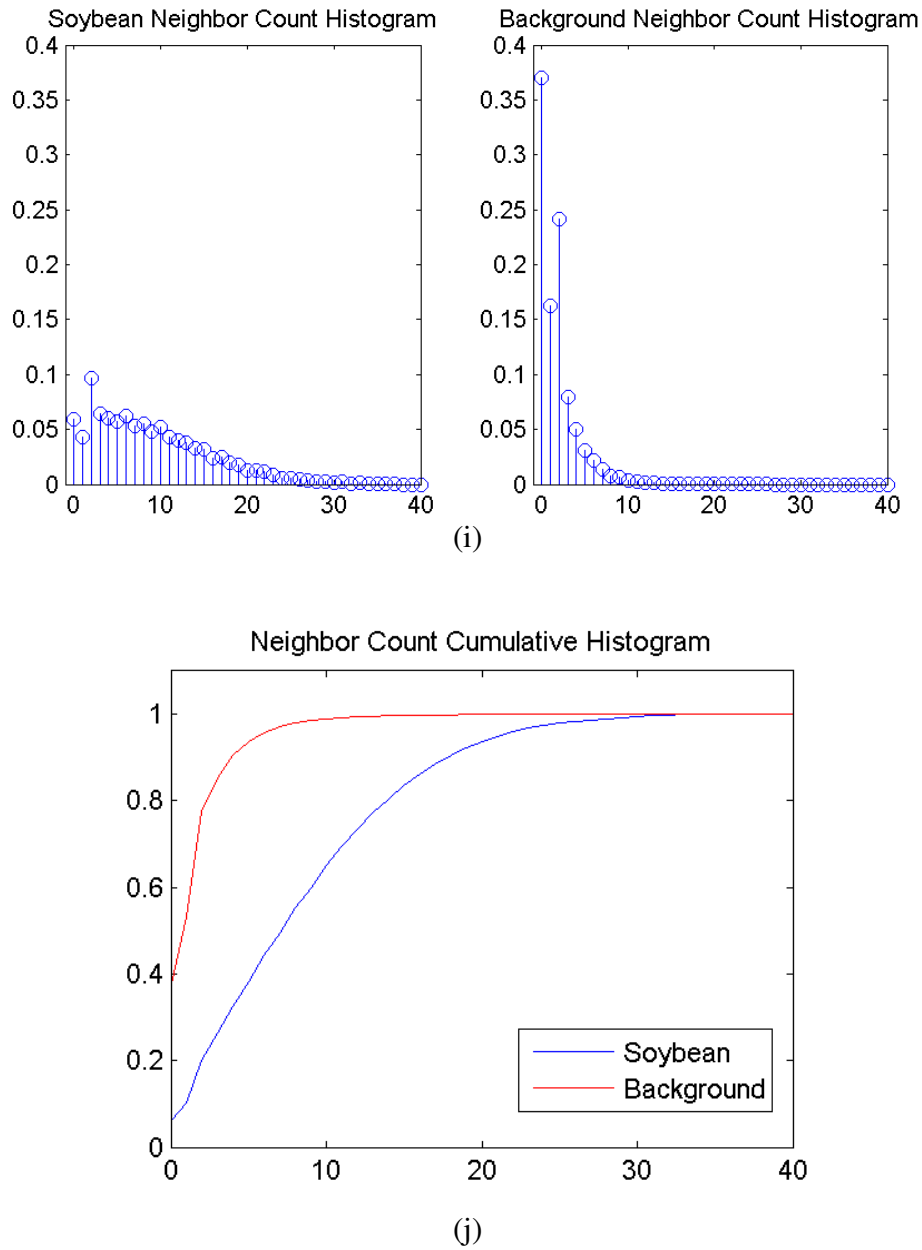
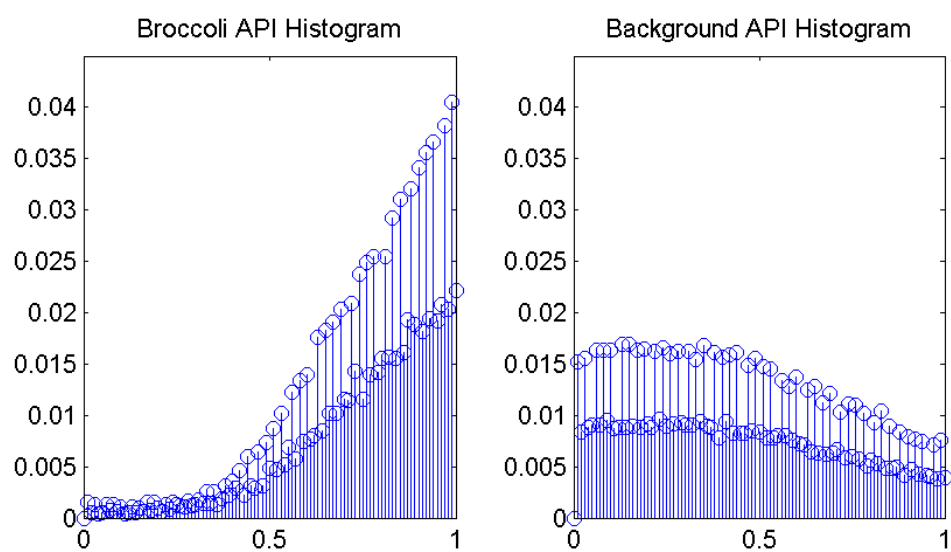
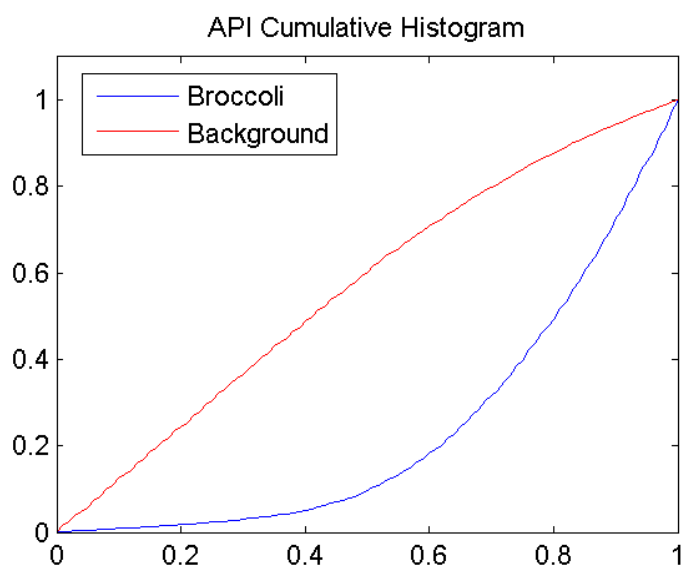


Figure 2.15. Histograms and cumulative histograms of various features of soybean and background: (a–b) histogram and cumulative histogram of API feature of soybean and background; (c–d) histogram and cumulative histogram of curvature feature of soybean and background, (e–f) histogram and cumulative histogram of depth gradient of soybean and background, (g–h) histogram and cumulative histogram of intensity gradient of soybean and background, and (i–j) histogram and cumulative histogram of neighbor count feature of soybean and background

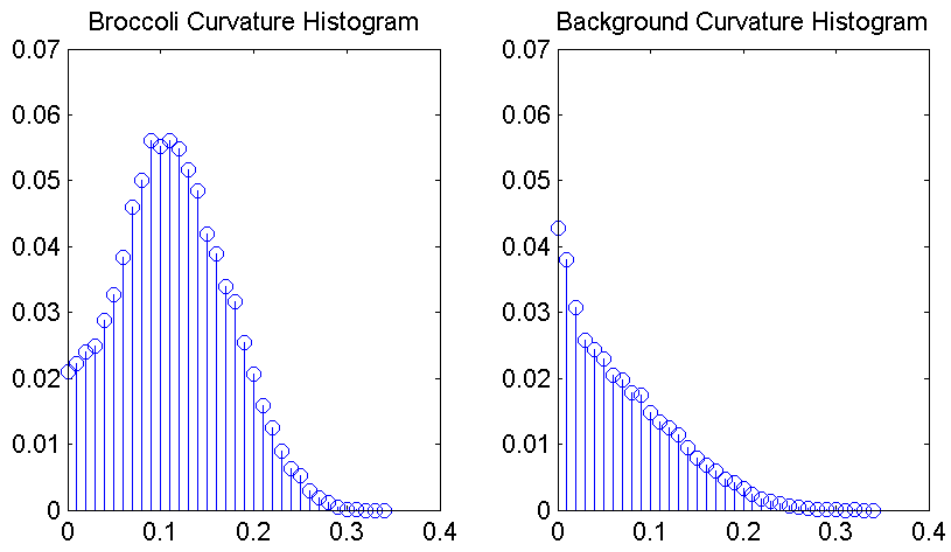


(a)

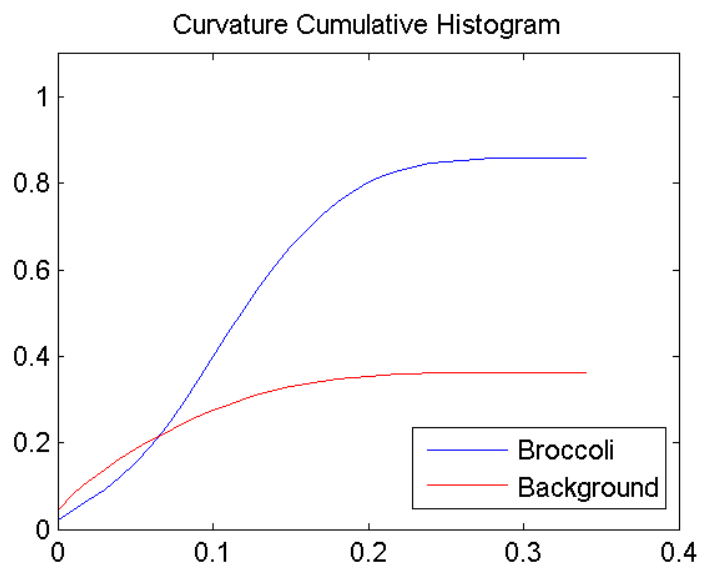


(b)

Figure 2.16 continued



(c)



(d)

Figure 2.16 continued

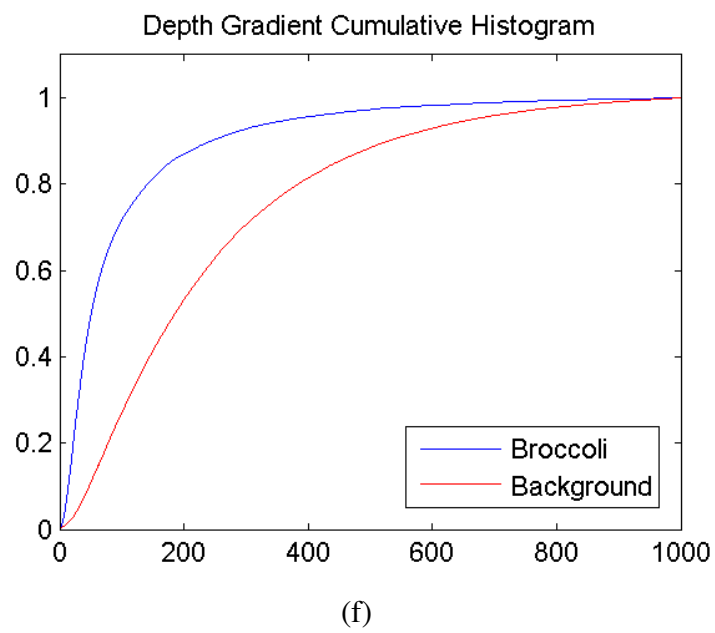
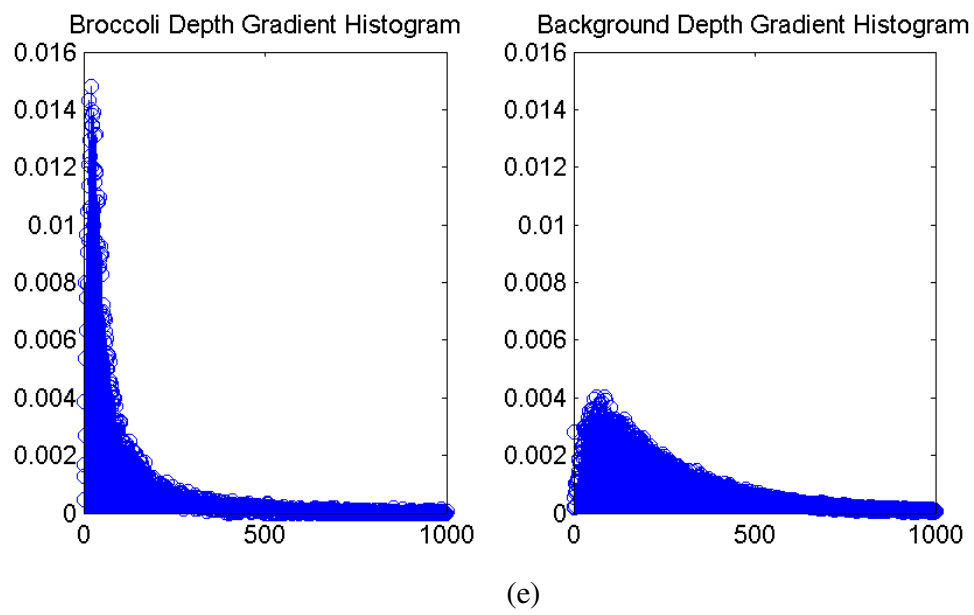




Figure 2.16 continued

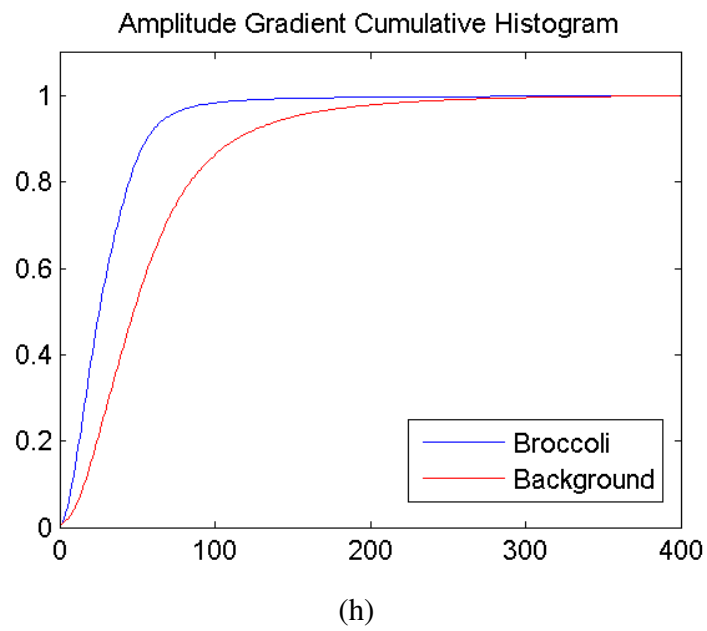
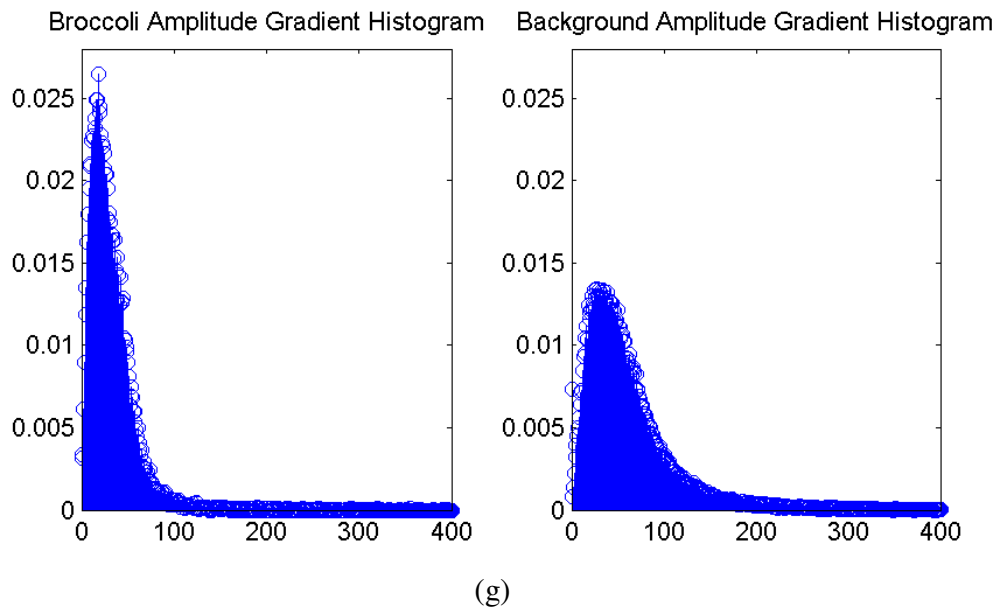


Figure 2.16 continued

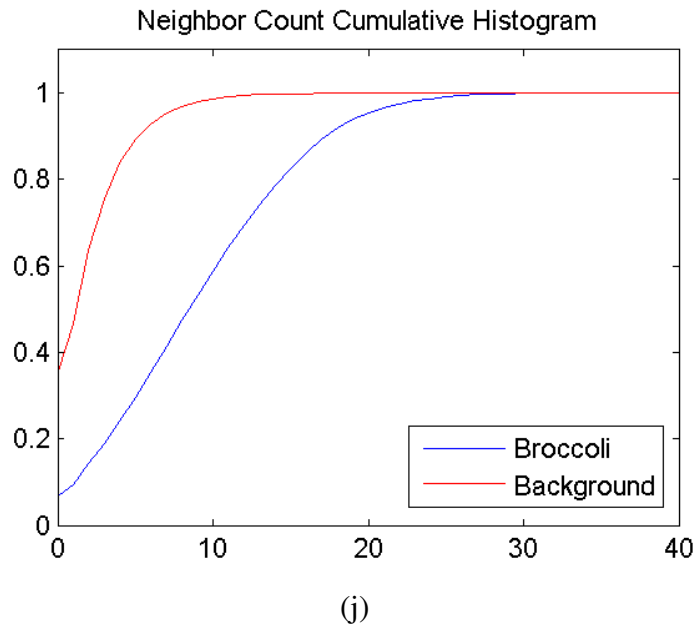
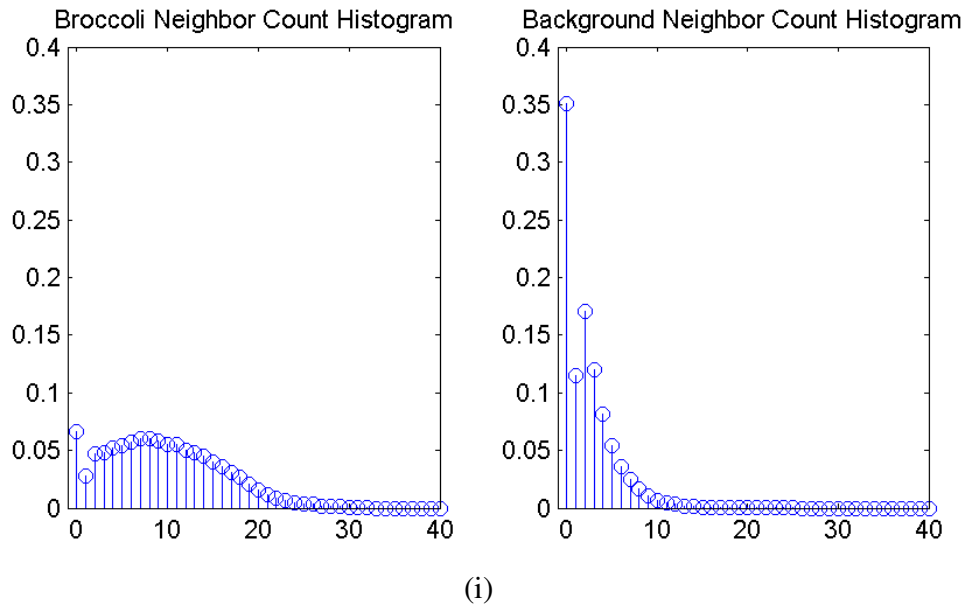


Figure 2.16. Histograms and cumulative histograms of various features of broccoli and background: (a–b) histogram and cumulative histogram of API feature of broccoli and background, (c–d) histogram and cumulative histogram of curvature feature of broccoli and background, (e–f) histogram and cumulative histogram of depth gradient of broccoli and background, (g–h) histogram and cumulative histogram of intensity gradient of broccoli and background, and (i–j) histogram and cumulative histogram of neighbor count feature of broccoli and background

Table 2.2. Feature distribution table for soybean

|                    | Feature Range | Percentage |            | Feature Range | Percentage |            |
|--------------------|---------------|------------|------------|---------------|------------|------------|
|                    |               | Soybean    | Background |               | Soybean    | Background |
| API                | $\geq 0.3$    | 0.979      | 0.5683     |               |            |            |
| Curvature          | $\leq 0.25$   | 0.7954     | 0.2239     | $\leq 0.3$    | 0.8        | 0.2241     |
| Depth gradient     | $\leq 80$     | 0.5767     | 0.1483     | $\leq 100$    | 0.6499     | 0.2002     |
| Amplitude gradient | $\leq 25$     | 0.6718     | 0.2417     | $\leq 36$     | 0.7999     | 0.3807     |
| Neighbor count     | $\geq 5$      | 0.6756     | 0.095      | $\geq 7$      | 0.5564     | 0.0423     |

Table 2.3. Feature distribution table for broccoli

|                    | Feature Range | Percentage |            | Feature Range | Percentage |            |
|--------------------|---------------|------------|------------|---------------|------------|------------|
|                    |               | Broccoli   | Background |               | Broccoli   | Background |
| API                | $\geq 0.65$   | 0.7742     | 0.2565     | 0.6           | 0.83       | 0.3035     |
| Depth gradient     | $\leq 150$    | 0.8134     | 0.4147     | $\leq 220$    | 0.8848     | 0.5743     |
| Amplitude gradient | $\leq 65$     | 0.9396     | 0.6750     |               |            |            |
| Neighbor count     | $\geq 5$      | 0.7581     | 0.1621     |               |            |            |

### 2.5.8 Crop Plant Segmentation Algorithm

After the noise clearance and feature extraction, the crop plant recognition algorithm segments the crop plants. The recognition algorithms of the broccoli and soybean plant share great similarity, but some differences exist because of their differences in 2D and 3D features described below.

#### 2.5.8.1 Soybean Segmentation Algorithm

The raw soybean images used to introduce the soybean segmentation algorithm in this section are given by Figure 2.17 (a-b). This algorithm is based on various feature information calculated by the previously described steps, including curvature  $\delta$ , neighbor count image  $N_C$ ,

amplitude and depth gradient image  $G_I$  and  $G_Z$ , and amplitude percentile image  $API$ ; all of them are also provided in Figure 2.17 (c–g).

The soybean segmentation algorithm consists of multiple steps, which are introduced below.

- 1) By applying the feature images achieved in the previous step, two intermediate images,  $S1$  and  $S2$ , were constructed based on the Equation 2.13 and 2.14, respectively. The threshold values of these two equations were picked based on the feature distribution discussion in section 2.5.7. The main idea of this step was to apply these features to extract the soybean candidates areas based on several observations. First, the 3D geometry of soybean leaf was relatively flat and smooth. Therefore, the depth gradient value  $G_Z$  and curvature  $\delta$  were small. Second, the amplitude value of a soybean leaf was nearly uniform, so its amplitude gradient was small. Third, the  $API$  value of soybean was not too small because its amplitude value is at least higher than soil and some weed background. Fourth, the points of the soybean leaf in the 3D point cloud data were relatively intense because of their flat 3D geometry; therefore the neighbor point count value  $N_c$  was relatively high for the soybean points. Based on these criteria, image  $S1$  and  $S2$  are achieved as Figure 2.17 (h) and (i) indicate.

$$S1 = \begin{cases} 1 & \text{when } G_Z \leq 80, G_I \leq 25, N_c \geq 7, \delta \leq 0.25, API \geq 0.3 \\ 0 & \text{else} \end{cases} \quad (2.13)$$

$$S2 = \begin{cases} 1 & \text{when } G_Z \leq 100, G_I \leq 35, N_c \geq 5, \delta \leq 0.3, API \geq 0.3 \\ 0 & \text{else} \end{cases} \quad (2.14)$$

- 2) The white pixels of the intermediate result image  $S1$  and  $S2$  can be separated into different regions in 2D space. In this study, the regions of size less than 10 and 8 pixels were considered as noise and were removed in image  $S1$  and  $S2$ ,

respectively. Because the amplitude image of a soybean leaf was smooth, the average amplitude gradient was small. Therefore, the regions with average amplitude gradient over 25 were also removed as background. The resulting images of  $S1$  and  $S2$  after this step are represented with  $S11$  and  $S21$ , respectively, which are shown by Figure 2.17 (j) and (k).

- 3) Next, image  $S11$  was dilated and processed in 3D space. Each pixel of  $S11$  was checked to determine whether there was neighbor point in 3D space that was removed in  $S11$  but not in  $S21$  image. If there was a neighbor point, the qualified neighbor point was added into image  $S11$ . In addition, this operation was iterated to process new  $S11$  until there was no longer any new point to be added. After that, all of the points of the new  $S11$  image were separated into different regions in 3D space. For the region separation in 3D space, if the distance between two points was no bigger than 10 mm, they were considered as connected points, belonging to the same region. It also means that between any two points in the same region, there is a link consisting of neighbor points in 3D space. Each region was a soybean leaf candidate. The regions with size smaller than 13 were removed because they were too small to be a soybean leaf. The final result of this step is represented with  $S12$ , as Figure 2.17 (l) shows.
- 4) As Figure 2.17 (l) indicates, the pixels extracted by the intermediate result image  $S12$  are only from the inner part of the soybean leaves instead of the complete results. Because the boundary area of the leaf has relatively high gradient value for both depth and amplitude, they were filtered out by the three steps mentioned above. In order to recover the boundary area of the soybean leaves, the

intermediate image  $S12$  was expanded in 3D space. All of the removed neighbor points in 3D space of any point in image  $S12$  were added to create a new  $S12$  image. In addition, this process was iterated five times to update image  $S12$ , so that the boundary points, which are close to the extracted inner part of soybean leaf, could be recovered, as Figure 2.17 (m) shows. The resulting image of this step is  $S13$ .

- 5) The intermediate result image  $S13$  still cannot fully recover the boundary part of the soybean leaves for some cases, so another algorithm was adopted to solve this problem. If a removed point  $q$  was a neighbor point of a valid point  $p$  in  $S13$  in 3D space, and it satisfied Equation 2.15–2.17, the removed point  $q$  was added to create a new  $S13$  image, and this process was iterated five times.

$$|Z_q - Z_p| \leq 7 \text{ mm}, \quad (2.15)$$

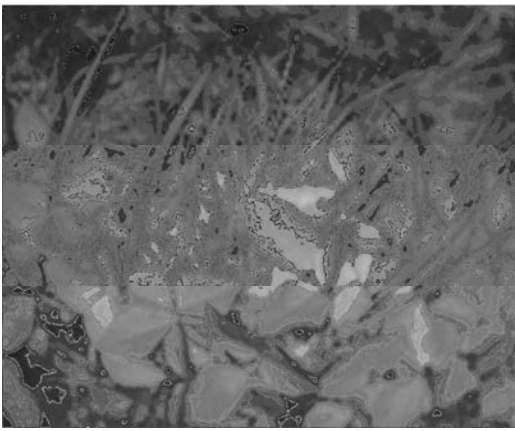
where  $Z_q$  and  $Z_p$  are the  $Z$  value of the 3D coordinates of point  $q$  and  $p$ , respectively. Equation 2.15 specifies that only the point  $p$  whose depth is similar to that of point  $q$  can be added, in order to ensure the newly added points and the existing points are at a single smooth surface.

$$G_{zq} \leq 130 \quad (2.16)$$

$$G_{Iq} \leq 70, \quad (2.17)$$

where  $G_{zq}$  and  $G_{Iq}$  represent the depth gradient value and amplitude gradient value of point  $q$ . Equation 2.16 and 2.17 specify the upper limit of the corresponding gradient value for candidate soybean leaf boundary points to be recovered, and they are designed to avoid adding the points where the 3D geometry and amplitude image are not smooth. The threshold values in these two

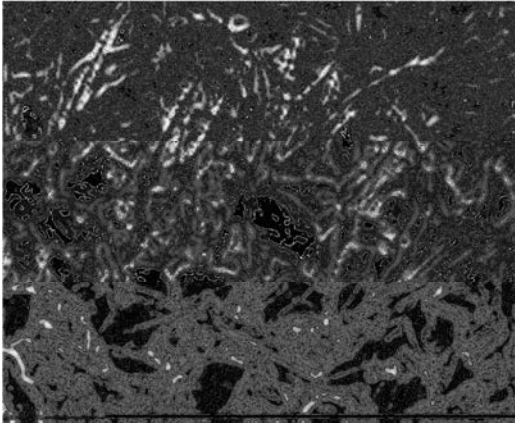
equations are higher than those in Equation 2.13 and 2.14 because of the nature differences between the inner part and the boundary area of the soybean leaves for both the 3D geometry and the 2D amplitude image. The resulting image of this step is the final soybean segmentation result, which is shown in Figure 2.17 (n).



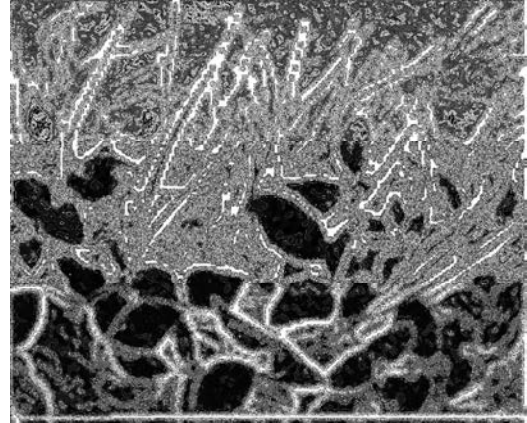
(a)



(b)

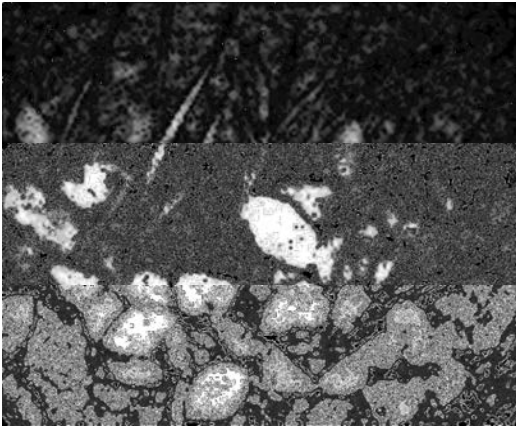


(c)

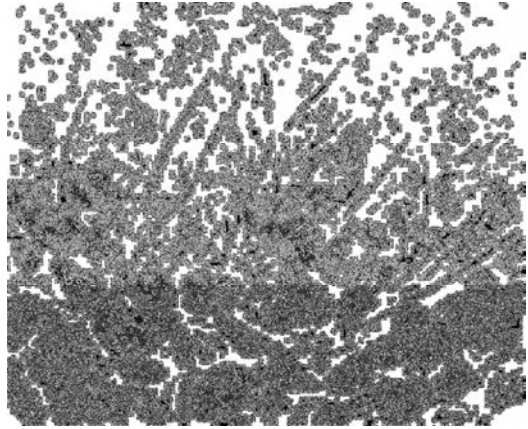


(d)

Figure 2.17 continued



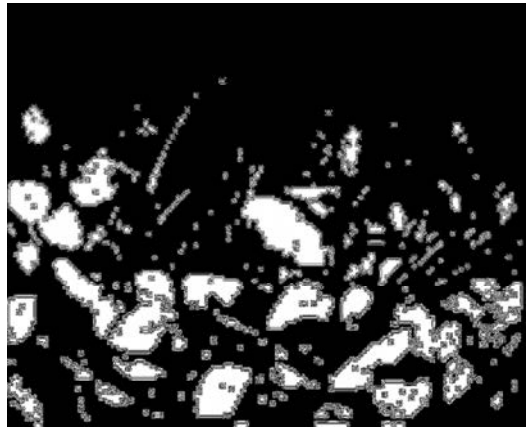
(e)



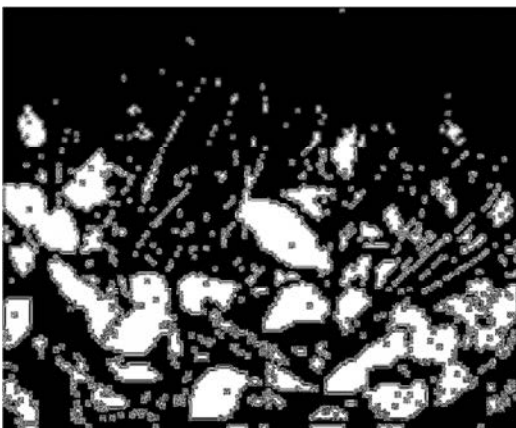
(f)



(g)



(h)



(i)



(j)



Figure 2.17 continued

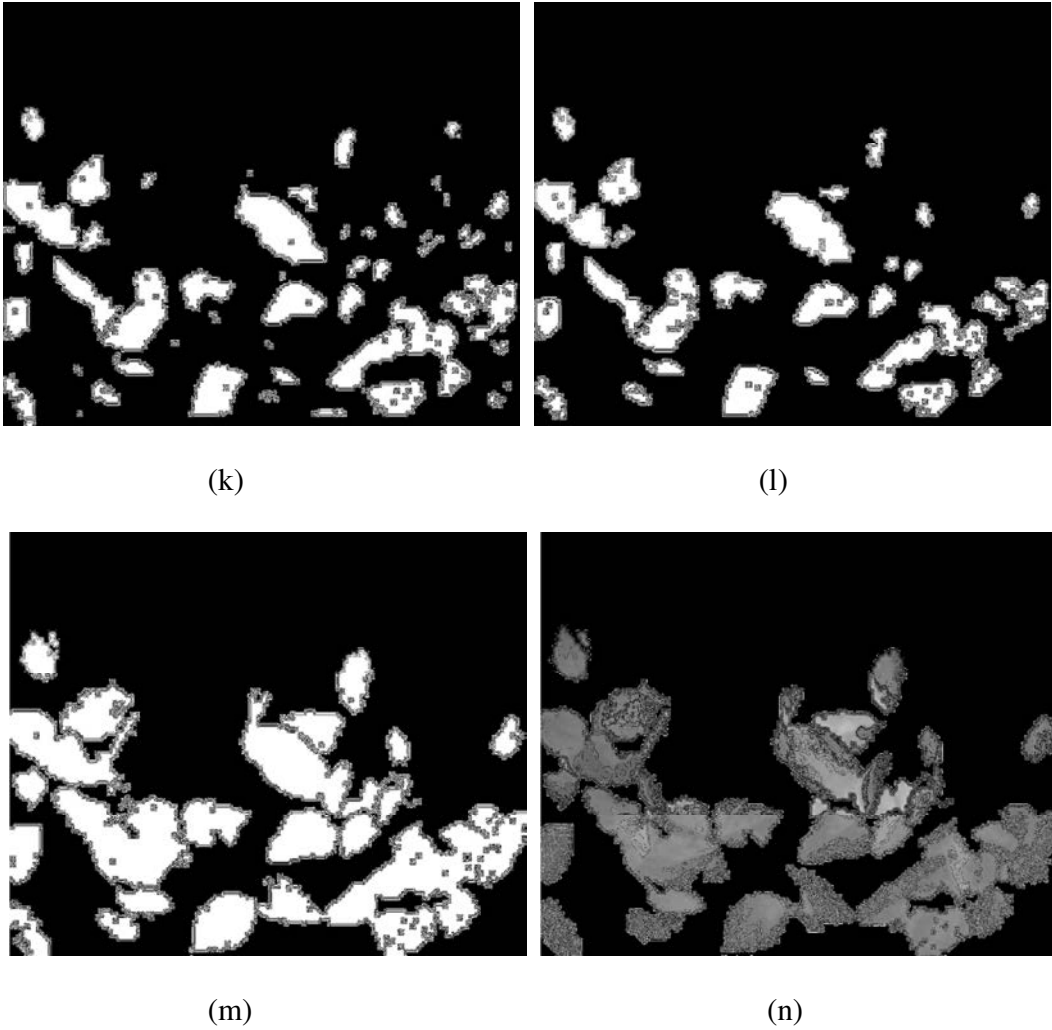


Figure 2.17. Images of the soybean segmentation procedure: (a) amplitude image, (b) depth image, (c) amplitude gradient image  $G_I$ , (d) depth gradient image  $G_z$ , (e) neighbor count image  $N_c$ , (f) curvature image  $\delta$ , (g) amplitude percentile image API, (h–i) intermediate image  $S1$  and  $S2$ , (j–k) image  $S11$  and  $S21$ , (l) image  $S12$ , (m) Image  $S13$ , and (n) final soybean segmentation result image

### 2.5.8.2 Broccoli Segmentation Algorithm

The broccoli segmentation algorithms shares great similarity with that of the soybean plant, but they still have some differences because of their various features. The detailed steps are introduced below:

- 1) Being similar to the first step of soybean segmentation, intermediate images  $S1$  and  $S2$  were constructed based on Equation 2.18 and Equation 2.19. The threshold values of these two equations were picked based on the feature distribution discussion in section 2.5.7. Compared to the first step of soybean segmentation, these two equations set a higher gradient threshold for both depth and amplitude, and they do not have any limitation in terms of the curvature of the 3D surface because of the inherent characteristics of the broccoli plant. As Figure 2.18 (a–g) indicates, the 3D geometry of broccoli leaves is relatively complicated, instead of being simply flat, and this means that some broccoli area has high depth gradient and curvature value. Figure 2.18 (c) also shows that the amplitude gradient of some parts of broccoli is relatively high compared to that of soybean leaves. Additionally, high  $API$  threshold value, which are 0.65 and 0.6 for generating  $S1$  and  $S2$ , respectively, were applied in Equation 2.18 and 2.19 for broccoli segmentation. This is because the relatively high  $API$  value resulted from the intensity of the broccoli plant's amplitude image, which is higher than that of the weeds and soil background. The resulting images  $S1$  and  $S2$  are shown in Figure 2.18 (h) and (i), respectively.

$$S1 = \begin{cases} 1, & \text{when } G_z \leq 150, N_c \geq 5, API \geq 0.65 \text{ or} \\ 0 & \text{else} \end{cases} \quad (2.18)$$

$$S2 = \begin{cases} 1, & \text{when } G_z \leq 220, G_l \leq 65, AN_c \geq 5, PI \geq 0.6 \text{ or} \\ 0 & \text{else} \end{cases} \quad (2.19)$$

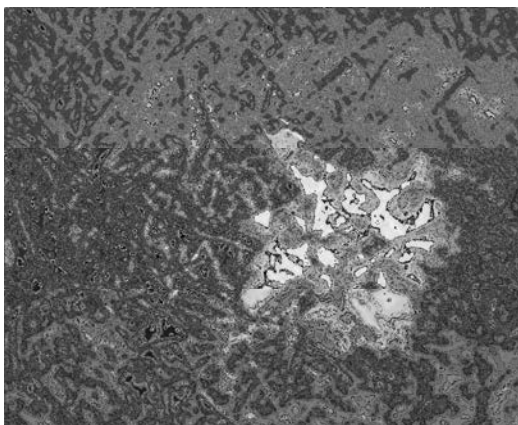
- 2) As Figure 2.18 (h) and (i) show, the intermediate result images  $S1$  and  $S2$  have many small regions, which are noise after threshold. In order to get rid of the noise of  $S1$ , the opening operation, a 2D morphological processing method, was applied. The structuring element of the opening operator was a  $3 \times 3$  square element. The result image of  $S1$  after the opening operation is represented with  $S11$ , which is shown by Figure 2.18 (j). The figure indicates that the small regions are successfully removed. The intermediate image  $S11$  can be separated into different regions in 2D space; regions with sizes smaller than 20 pixels were removed. The result image  $S12$  is shown by Figure 2.18 (k).
- 3) The intermediate result image  $S12$  did not contain a full broccoli plant. Most likely, the broccoli area extracted was only the inner part of a broccoli leaf. To recover the boundary area of the broccoli leaf and to get a more complete broccoli plant image, a method similar to the third step of the soybean segmentation algorithm was adopted. Each pixel of  $S12$  was checked to determine its neighbor point in 3D space, which was removed in  $S12$  but not in the  $S2$  image. If true, these qualified neighbor points were added into image  $S12$ . This operation was iterated to process new  $S12$  until there were no new points to be added. After this process, all of the points of the final  $S12$  image were separated into different regions in 3D space. Regions smaller than 30 points were removed because they were too small to be considered as parts of a broccoli leaves. The final result of this step is represented with  $S13$ , as Figure 2.18 (l) shows.
- 4) A method similar to the fourth step of the soybean segmentation algorithm was applied to further recover the boundary area of broccoli. Any background point

with a distance to any foreground point of image *S13* that was smaller than 10 mm was added to generate a new *S13* image. This operation was iterated twice, and the final result is called image *S14*, which is shown in Figure 2.18 (m).

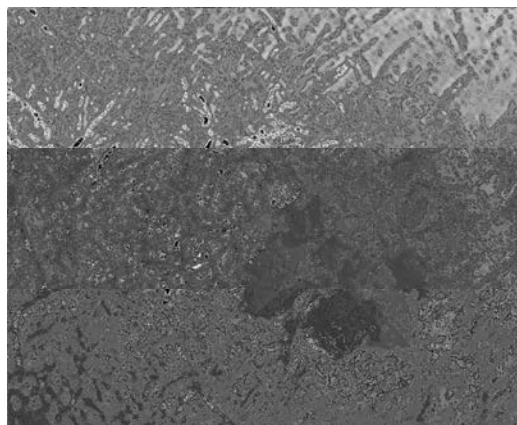
- 5) As Figure 2.18 (m) shows, the broccoli image extracted in intermediate image *S14* is broken into unconnected pieces. This problem made it hard to analyze the broken pieces together as a whole plant. To solve this problem, 2D dilation with a structural kernel of  $3 \times 3$  square element was applied to process image *S14*. Being different from normal 2D dilation, only the pixels with *API* value over than 0.3 were added in this process. This is because the area with *API* value smaller than 0.3 is very likely to be soil and weed background, as discussed previously. This dilation process was iterated five times, and the result image is called *S15*. As Figure 2.18 (n) indicates, the broken pieces of broccoli are merged together as a whole part. This procedure also took undesired soil or weed areas as part of the extracted broccoli image. The problem was solved in next step.
- 6) This study developed a method inspired by the flood fill algorithm to process intermediate image *S15* to remove the undesired background area from the extracted broccoli image. It is based on observation of the 3D geometry structure of the broccoli plant. First, the outer area of the broccoli plant is normally higher than the weeds and soil around it, this makes the boundary area of broccoli like a “dam” that can prevent the “water” from flooding the inner part of broccoli for the flood fill algorithm. Second, the inner leaves are on top of the outer ones, and this means that there are several levels of “dam” from the outside to the inside area to protect the island (broccoli). Therefore, even if the outer leaves are covered by the

“flood,” the inner leaves on top still have multiple chances to block the water.

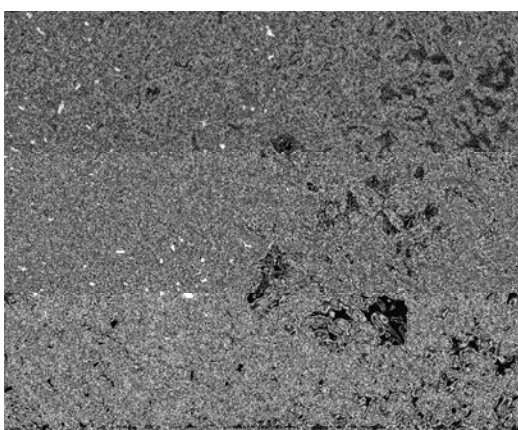
Because of the 3D structure of broccoli, the “flood” can easily cover the soil and weed background, but the broccoli image is more or less safe from “flood.” The difference between the algorithm in this study and the normal flood algorithm is that the “water” in this study flows to the neighbor pixel only when the height of the neighbor pixel is not more than 15 mm higher than that of the pixel where water it is. Moreover, the water continues to flood new areas unless the neighbor area is over 15 mm than where it is. In this study, all of the background pixels at the boarder of image S14 were used as the start point, in which “water” was added at the beginning. The recursive algorithm will make the “water” continuously flood the new background area until it is blocked by the “dam,” which is the boundary of broccoli or other objects which are significantly higher than neighbor area. After this flood fill algorithm, broccoli and some other objects which were significantly higher than neighbor area were kept. By removing the point cloud region in 3D space whose size was smaller than 100 points, the objects other than broccoli can be reliably cleaned as Figure 2.18(o) shows. In Figure 2.18(o), the white pixels represent the area covered by “flood,” and the black part is the extracted broccoli area, which is safe from “water.” By rendering the amplitude value to the black area of Figure 2.18 (o), the final broccoli segmentation result is achieved as Figure 2.18 (p) shows.



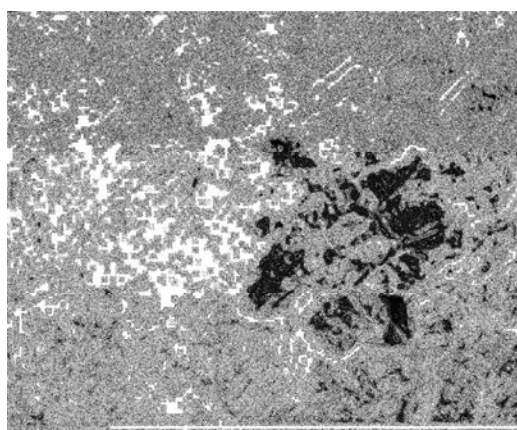
(a)



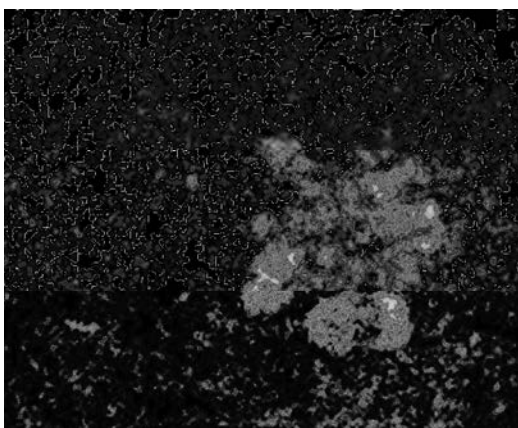
(b)



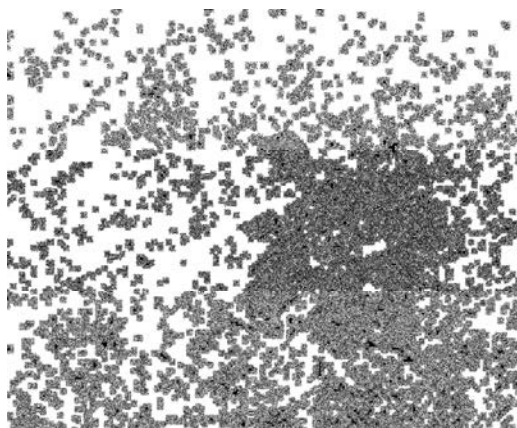
(c)



(d)



(e)

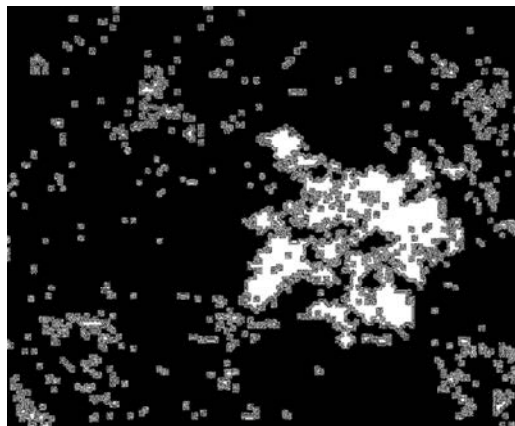


(f)

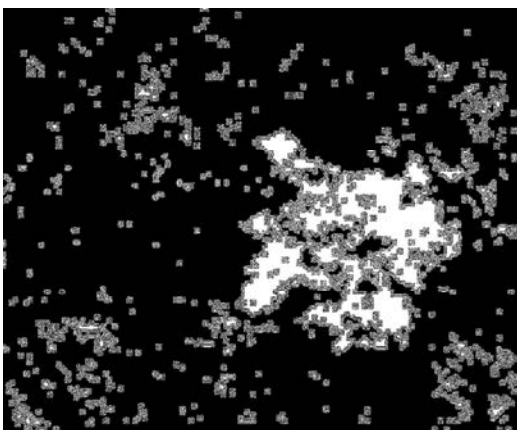
Figure 2.18 continued



(g)



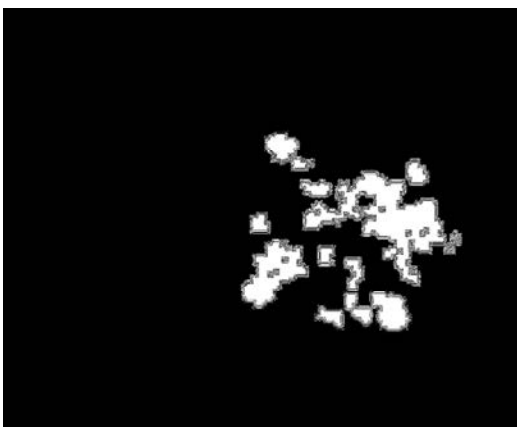
(h)



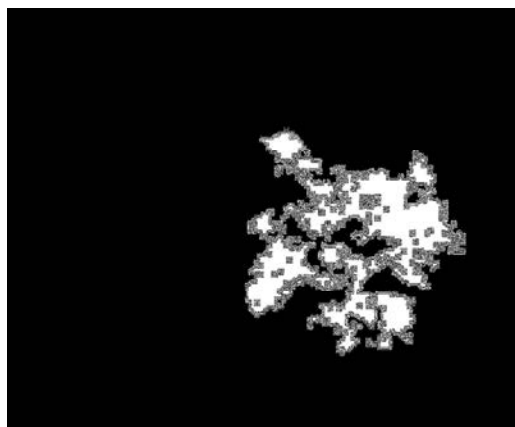
(i)



(j)



(k)



(l)

Figure 2.18 continued

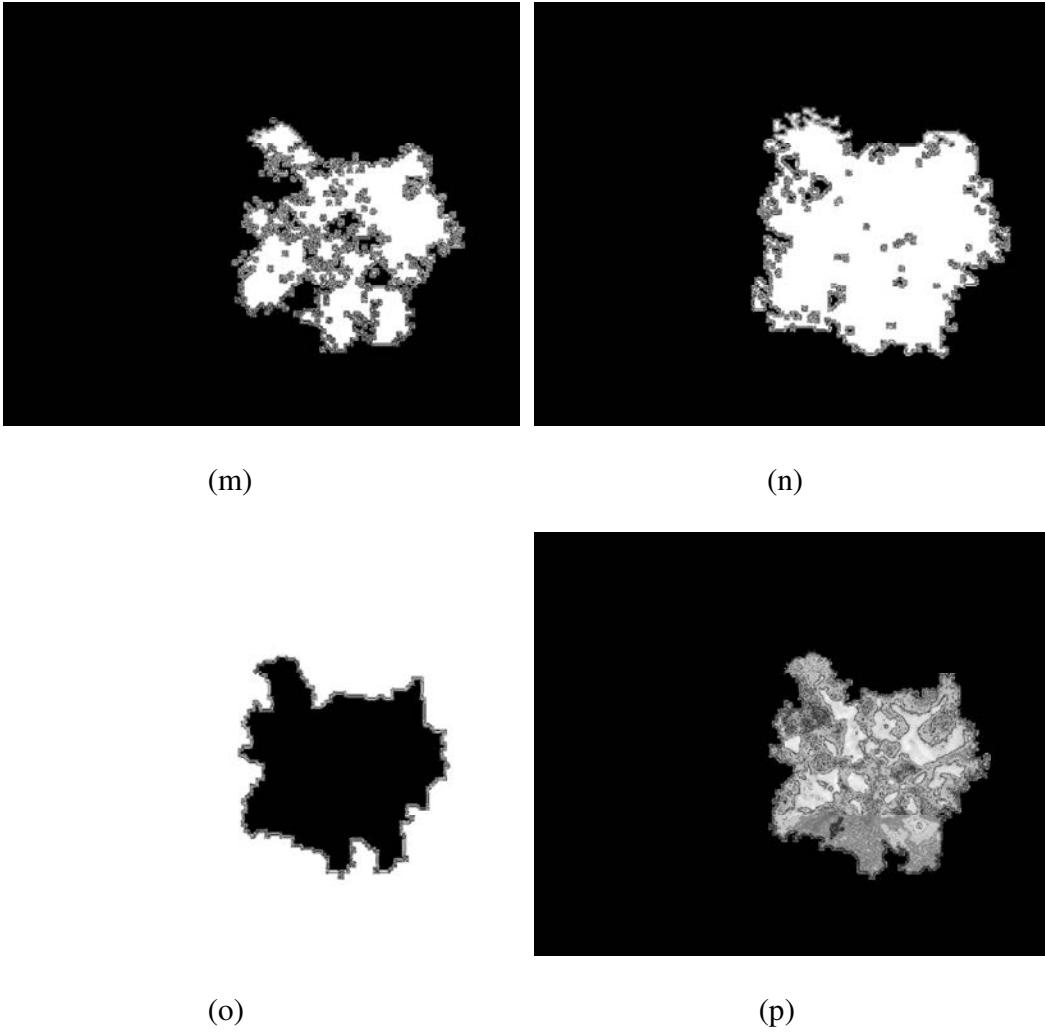


Figure 2.18. Images of broccoli segmentation procedure: (a) amplitude image, (b) depth image, (c) amplitude gradient image  $GI$ , (d) depth gradient image  $Gz$ , (e) neighbor count image  $N_c$ , (f) curvature image  $\delta$ , (g) amplitude percentile image  $API$ , (h–i) intermediate images  $S1$  and  $S2$ , (j) image  $S11$ , (k) image  $S12$ , (l) image  $S13$ , (m) image  $S14$ , (n) image  $S15$ , (o) result image of flood fill, and (p) final broccoli segmentation result image



## 2.6 Results & Discussion

As introduced in the experiment design section, in this research, data from 206 broccoli plants and 3 rows of soybean plants were collected during June and July. The data consisted of a total of 49,179 and 18,793 sequential image sets captured for broccoli and soybean plants, respectively. As it was impractical to manually check the processing result of all of these large size image sets, part of the data set was randomly selected to test and evaluate the system performance.

### 2.6.1 Broccoli Recognition Result

For each of the 206 broccoli plants, of each of three data capturing dates, three image sets were randomly selected to evaluate the recognition performance of this system for the broccoli plant at different growing stages. For each data capturing date, the broccoli plant recognition accuracy of the 718 randomly selected images was collected. Additionally, for practical applications, when the mobile platform moved through the field, the platform had multiple chances to detect a crop plant by capturing and analyzing multiple sequential images of the plant. Therefore, the rate of recognizing a broccoli plant accurately from at least one of its three image sets was also evaluated. The false detection rate, which mistakenly recognized the background areas as a broccoli plant, was also collected.

The statistical results are given in Table 2.4. As it indicates, from one image, the broccoli plant detection rate is over 84.3%, and the rate of the broccoli plant being recognized from at least one of its three images is over 88.3%. The false detection rate is less than 1.5%. As Table 2.4 shows, the recognition accuracy of the later image sets is slightly

higher than those of earlier image sets, which indicates that broccoli plants at a higher growth stage are relatively easy for this system to recognize.

The major reasons for the broccoli plants being missed by this system were the plants' small size and the low resolution of the ToF camera. The diameter of all of the missed plants was less than 12 cm. As the resolution of the ToF camera used in this research was only  $144 \times 176$  pixels, and the camera was over 0.6 m away from the top of the plants, the missed plants appeared too small in the captured image, as Figure 2.20 shows. The occlusion problem caused by weeds also led to miss detection.

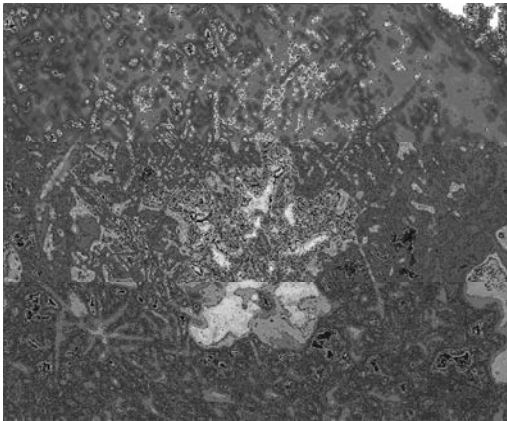
The broccoli plants were extracted with relatively accurate shape. Table 2.5 provides 20 sets of randomly selected original input images and the broccoli segmentation result. Each image set belongs to a different broccoli plant. As the images indicate, the broccoli plant can be segmented out with relatively precise and complete shape. Because of the small resolution of input images and occlusion caused by weeds, some small parts of broccoli plant may be missed by this system. However, the major areas of the broccoli plant were accurately kept in the segmentation result for all samples, as Table 2.5 shows. Additionally, the segmentation result was clean, and in only approximately 1% of the 3090 randomly selected images was the background mistakenly recognized as a broccoli plant.

Table 2.4. Broccoli plant recognition accuracy rate

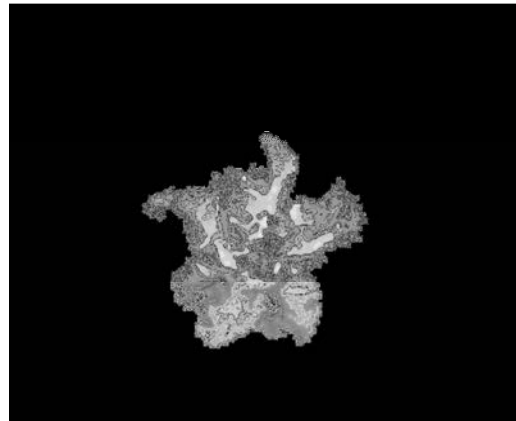
| Data collection date | Height of broccoli plant (cm) | Broccoli recognition rate from one image | Detection rate of one plant from 3 images | False detection rate |
|----------------------|-------------------------------|--|---|----------------------|
| June 26, 2012        | 18–23                         | 84.3%                                    | 88.3%                                     | 1.45%                |
| July 6, 2012         | 18–23                         | 86.1%                                    | 90.3%                                     | 0.65%                |
| July 18, 2012        | 18–30                         | 88.0%                                    | 92.7%                                     | 0.97%                |



(a)

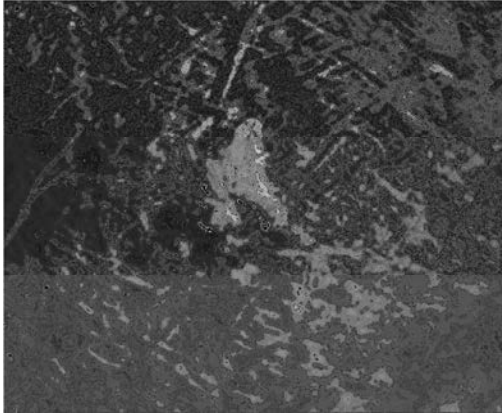


(b)



(c)

Figure 2.19. Broccoli segmentation result: (a) 2D color image, (b) amplitude image from ToF camera, and (c) broccoli segmentation result



(a)



(b)

Figure 2.20. Small plant missed by system: (a) amplitude image captured by ToF camera and (b) 2D color image

Table 2.5. Broccoli segmentation result

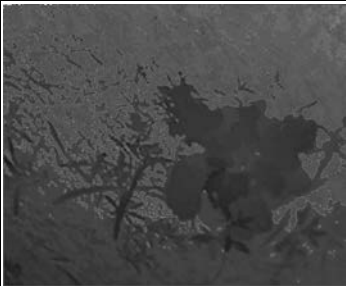
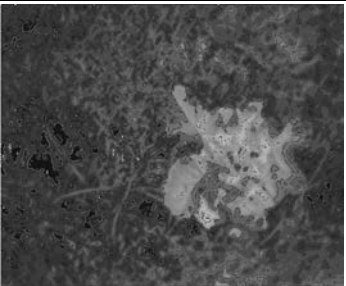
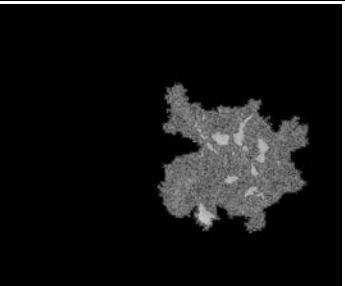
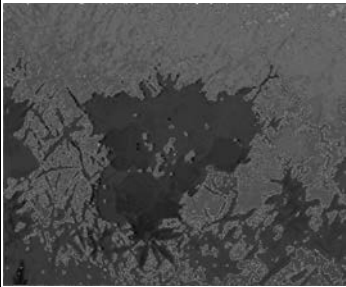
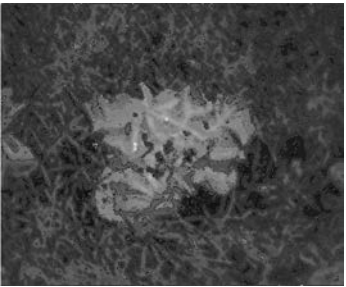
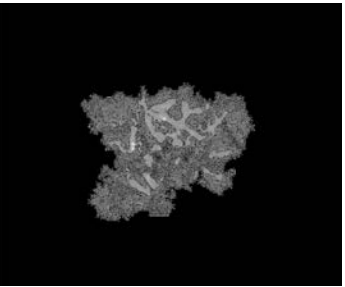
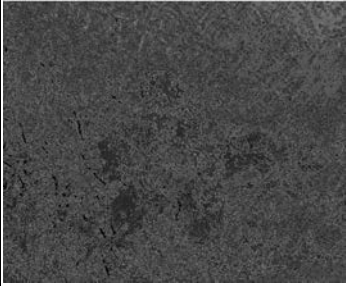
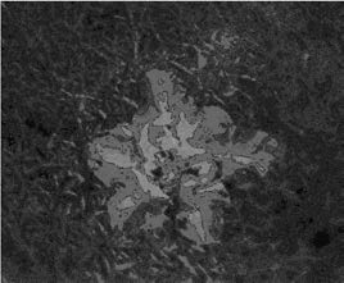
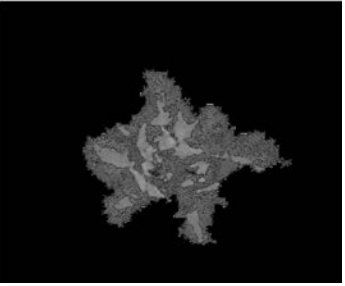
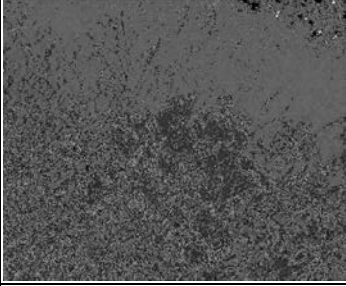
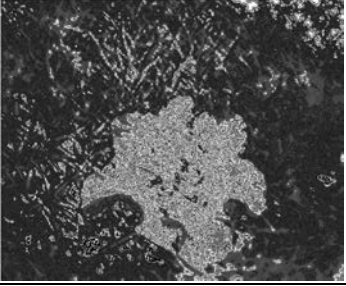
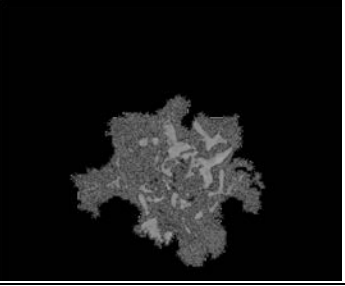
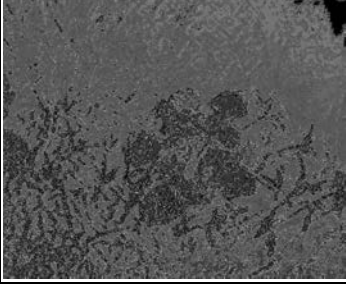
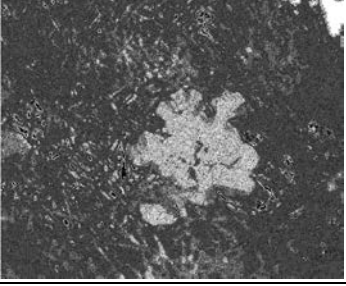
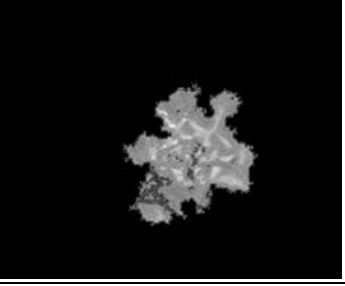
| ID | Depth Image   | Amplitude Image  | Segmentation Result   |
|----|---|--|---|
| 1  |    |    |    |
| 2  |    |    |    |
| 3  |   |   |   |
| 4  |  |  |  |
| 5  |  |  |  |

Table 2.5 continued

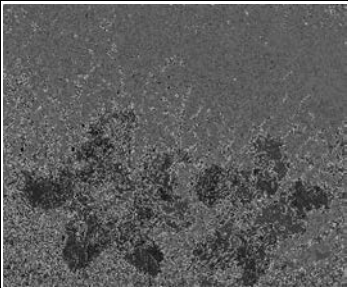
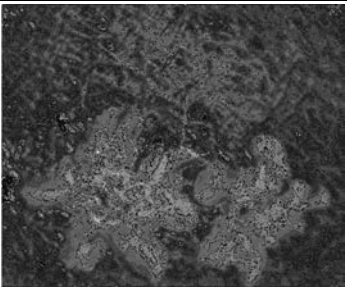
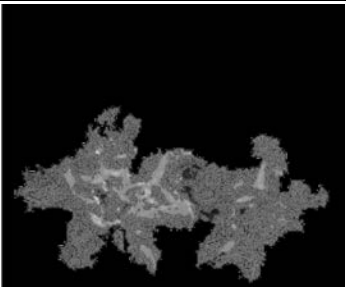
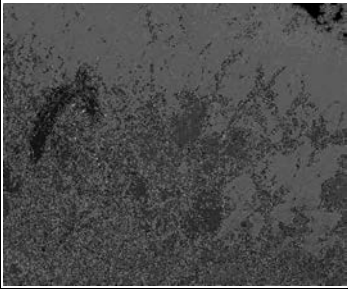
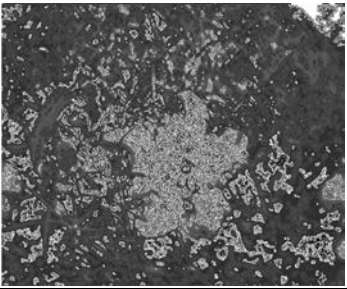
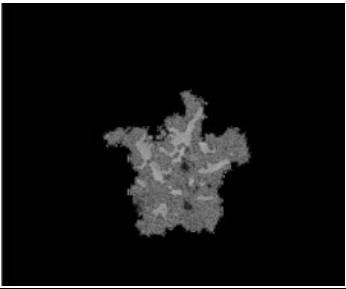
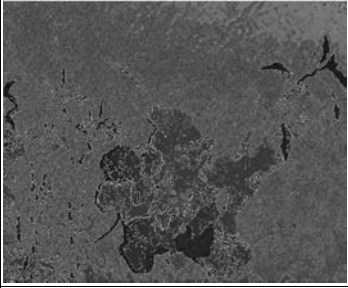
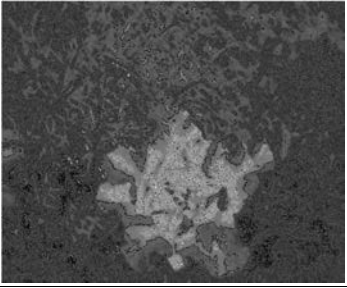
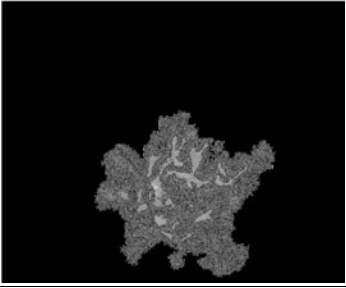
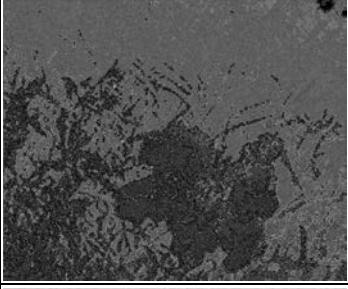
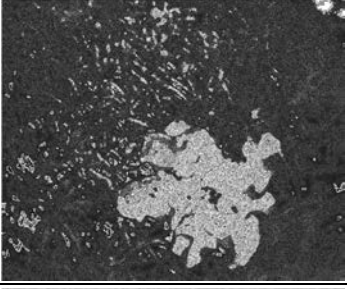
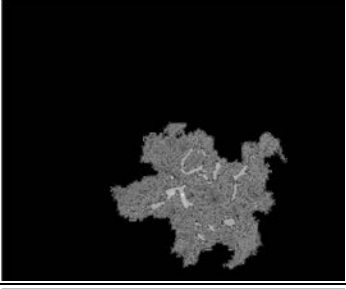
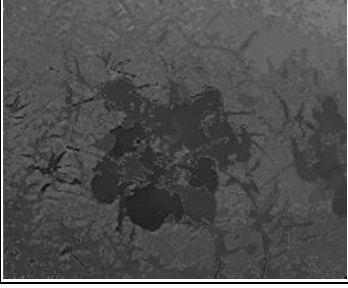

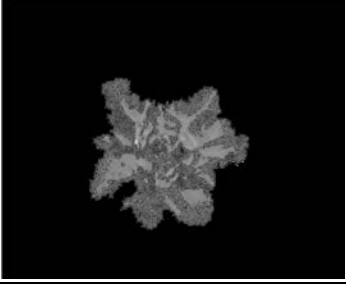
| ID | Depth Image   | Amplitude Image  | Segmentation Result   |
|----|---|--|---|
| 6  |    |    |    |
| 7  |    |    |    |
| 8  |   |   |   |
| 9  |  |  |  |
| 10 |  |  |  |

Table 2.5 continued

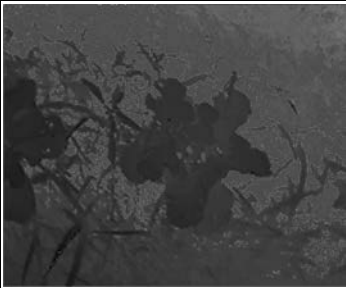
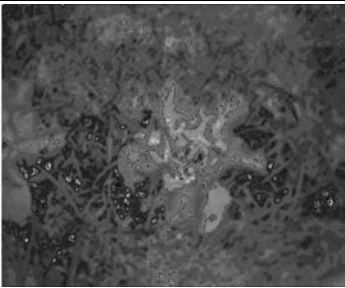
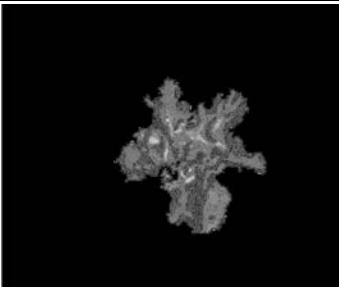
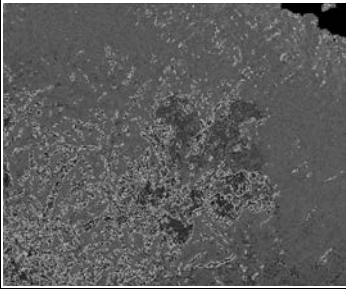
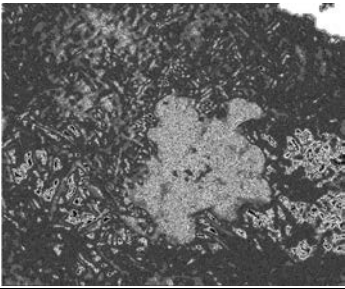
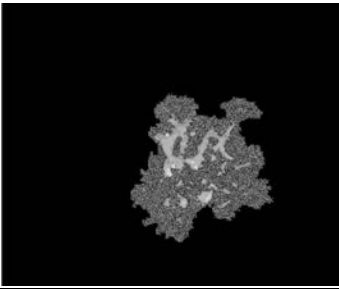

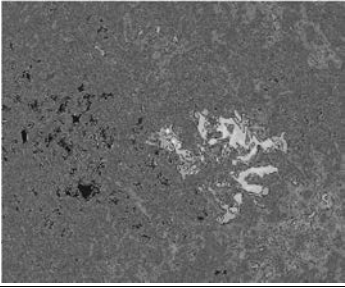
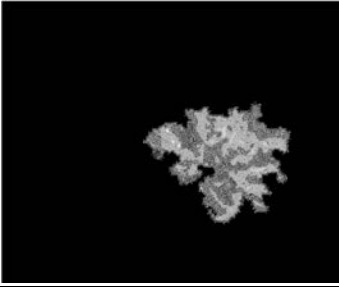
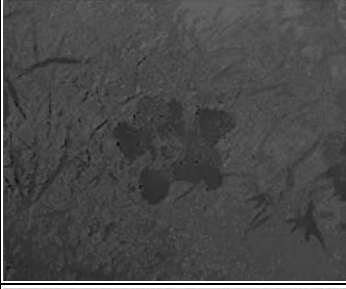
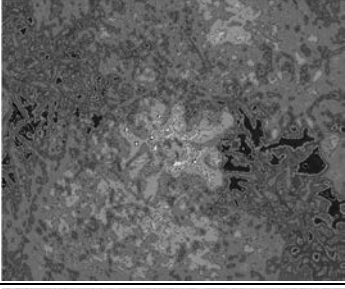
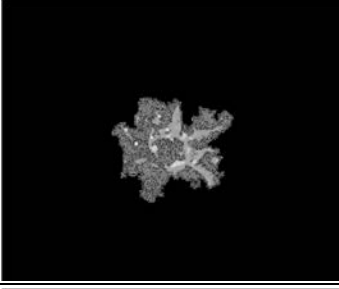
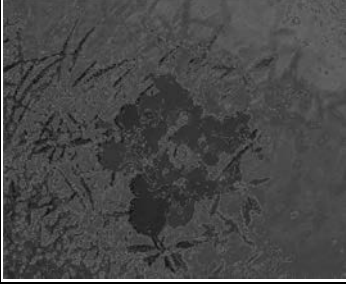
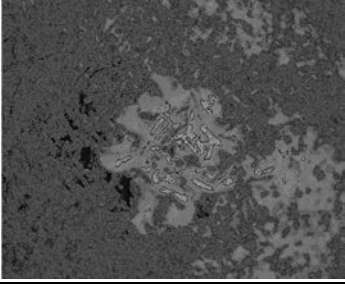
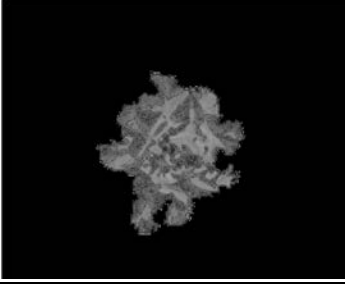
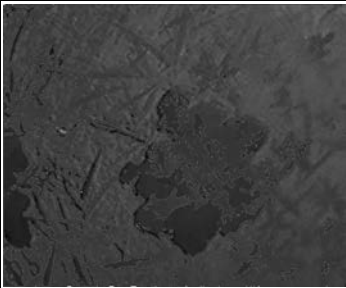
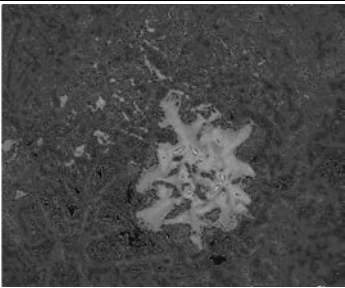
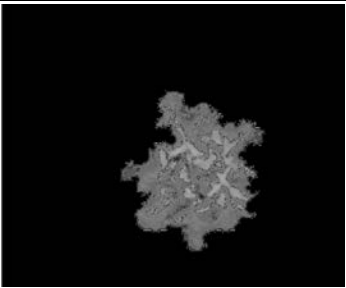
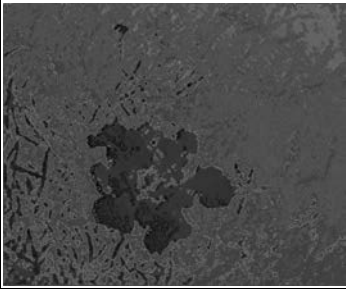
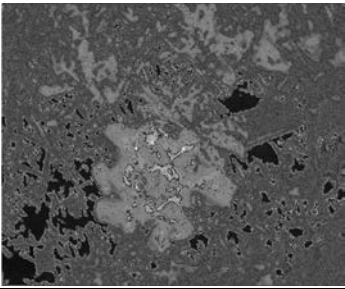
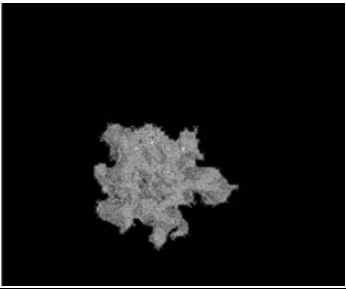
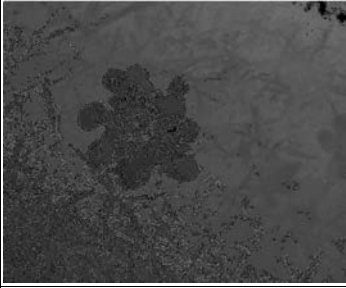
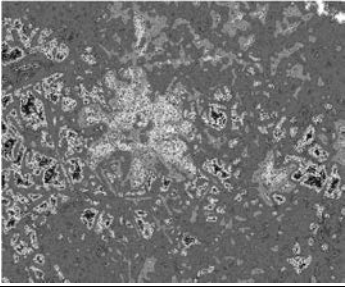
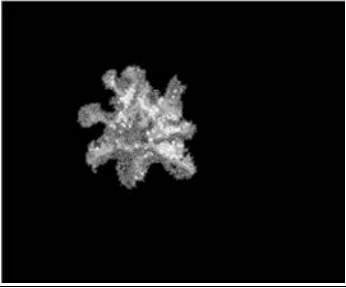
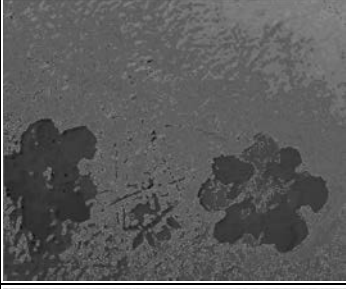
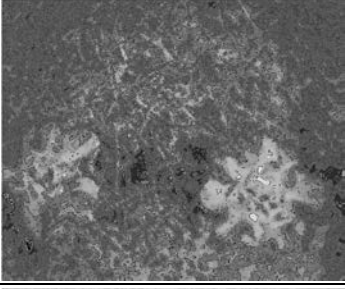
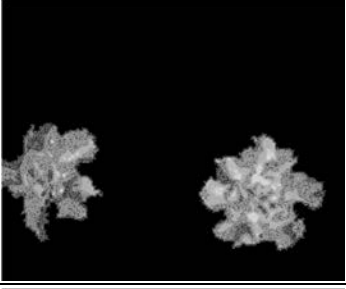
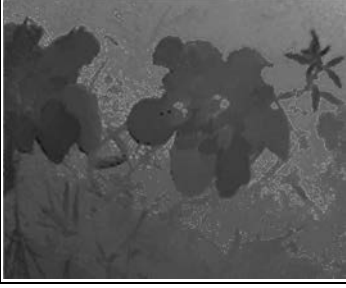
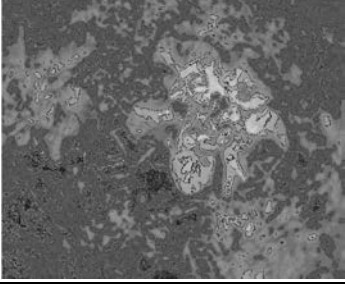
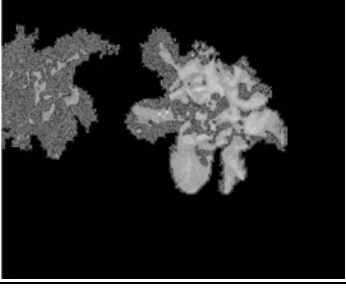
| ID | Depth Image   | Amplitude Image  | Segmentation Result   |
|----|---|--|---|
| 11 |    |    |    |
| 12 |    |    |    |
| 13 |   |   |   |
| 14 |  |  |  |
| 15 |  |  |  |



Table 2.5 continued

| ID | Depth Image   | Amplitude Image  | Segmentation Result   |
|----|---|--|---|
| 16 |    |    |    |
| 17 |    |    |    |
| 18 |   |   |   |
| 19 |  |  |  |
| 20 |  |  |  |



### 2.6.2 Soybean Recognition Result

Soybean images were collected on June 26 and July 5, 2012. Because of the high competition between weeds and grass, there was little maturity difference in the soybean between the two data capture dates. The soybean data collected on July 5, 2012 was representative enough to be used to evaluate the soybean recognition accuracy of the system. Among the total 6967 image sets of three rows of soybean plant, 100 image sets were randomly selected to evaluate the soybean detection and segmentation accuracy. For the 100 images, there were about 3938 soybean leaves according to the manual counting result, and 3593 soybean leaves were accurately detected by the system. Therefore, the detection rate of soybean leaves is 91.2%. Most of the leaves missed by the system were either too small or occluded. The main idea of the soybean detection algorithm is to search for the flat surface with relatively uniform amplitude value. Small or partially occluded leaves did not have enough pixels in the low resolution images captured by the ToF camera and can hardly be recognized as flat surface in this system; therefore, they were missed. Moreover, there was no weed or soil area recognized as soybean leaves for all of these 100 images.

The soybean leaves were extracted with relatively accurate shape by the segmentation algorithm of this research. Table 2.6 provides 20 randomly selected soybean input data and the corresponding segmentation result. It shows that the big soybean leaves that are not occluded by weeds can always be extracted without significant loss of area. However, for the small soybean leaves or the ones occluded, if they are not missed, their loss of fragment is more obvious in the segmentation result.

Table 2.6. Soybean segmentation result


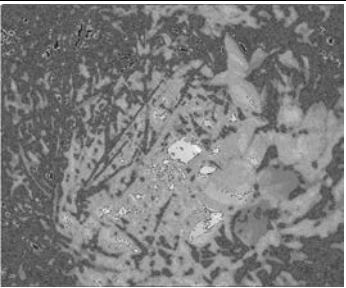
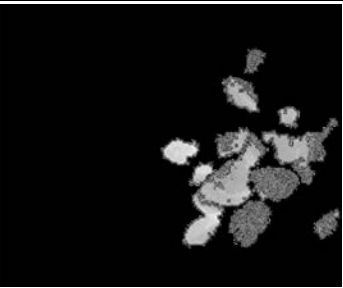
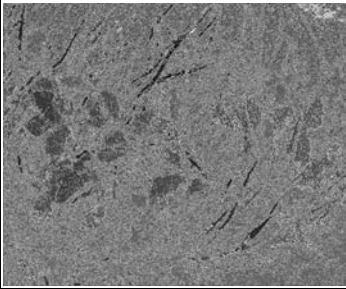

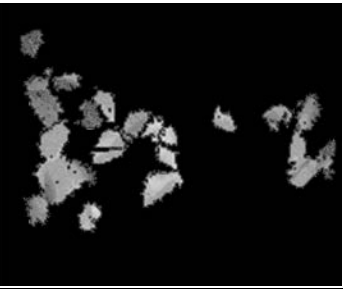
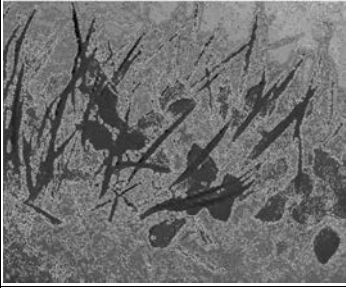
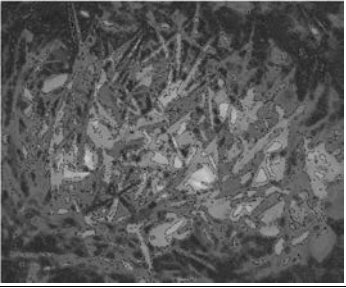
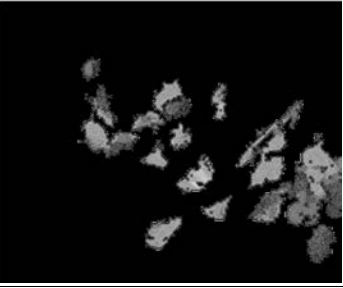
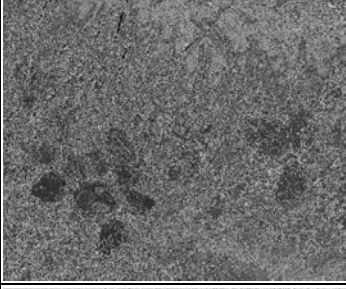
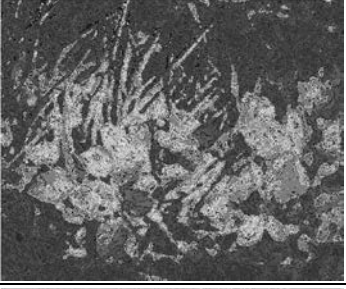
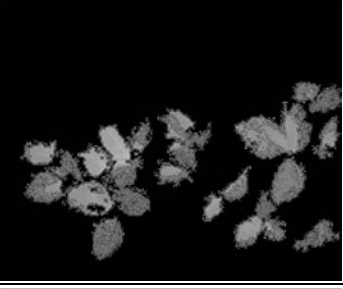
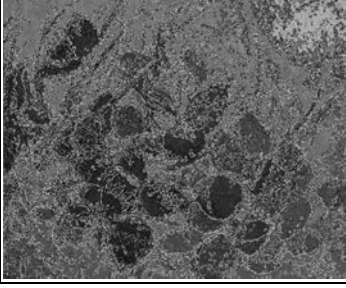
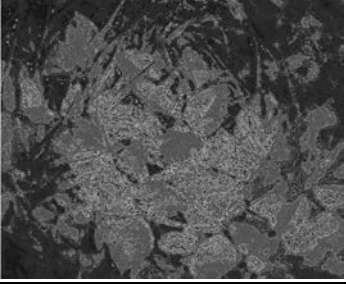
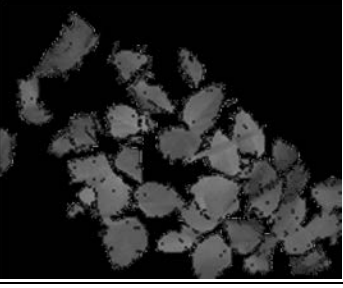
| ID | Depth Image   | Amplitude Image  | Segmentation Result   |
|----|---|--|---|
| 1  |    |    |    |
| 2  |    |    |    |
| 3  |   |   |   |
| 4  |  |  |  |
| 5  |  |  |  |

Table 2.6 continued

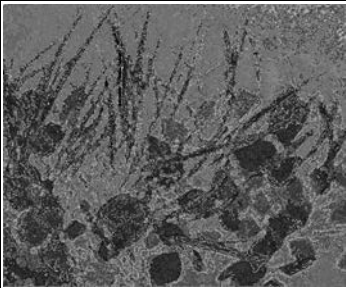
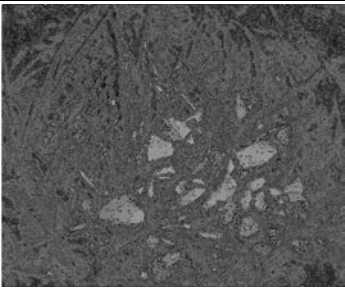
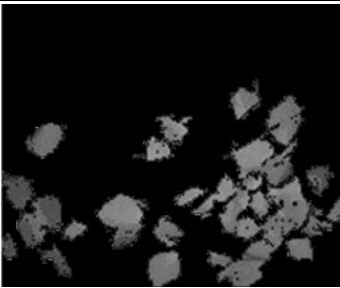

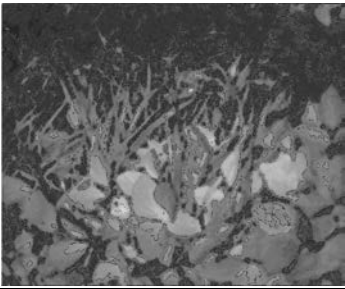
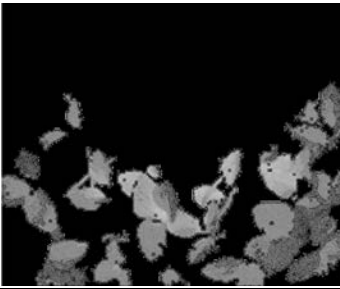
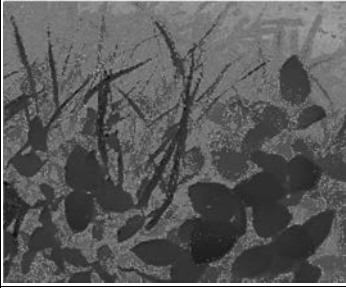
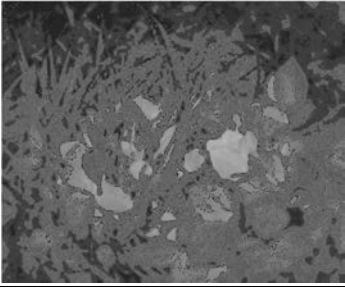
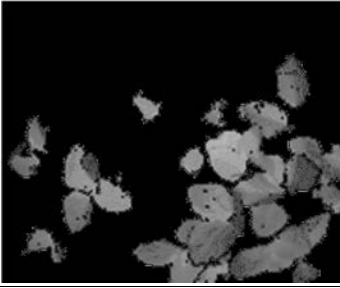
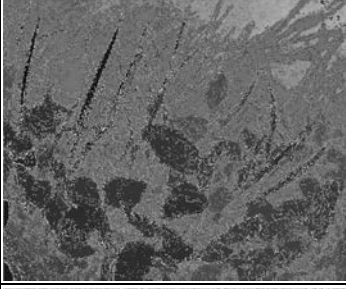
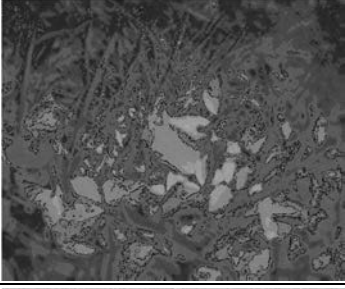
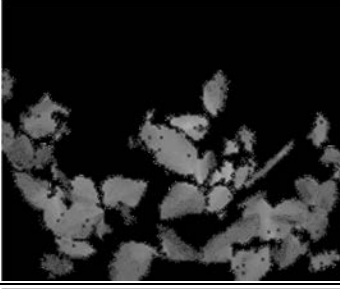
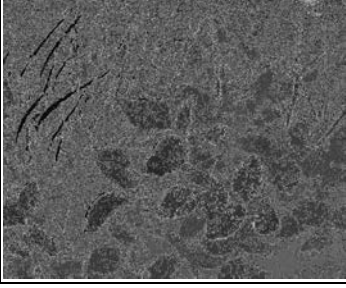
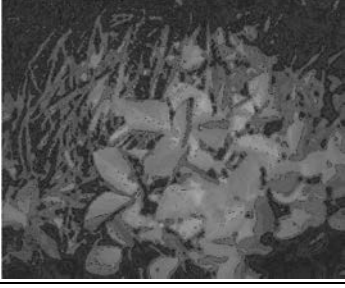
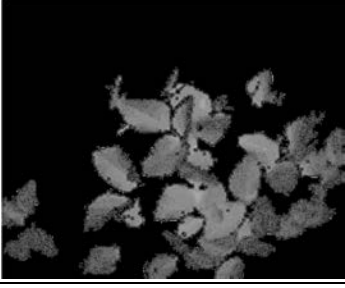
| ID | Depth Image   | Amplitude Image  | Segmentation Result   |
|----|---|--|---|
| 6  |    |    |    |
| 7  |    |    |    |
| 8  |   |   |   |
| 9  |  |  |  |
| 10 |  |  |  |

Table 2.6 continued

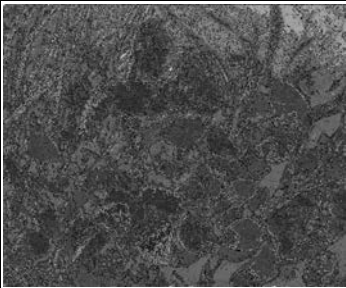
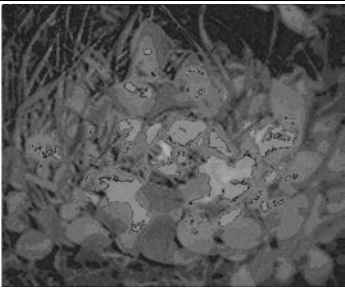
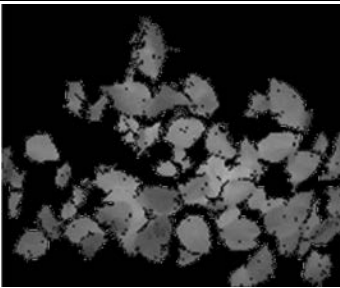
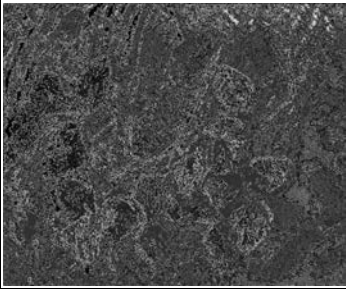
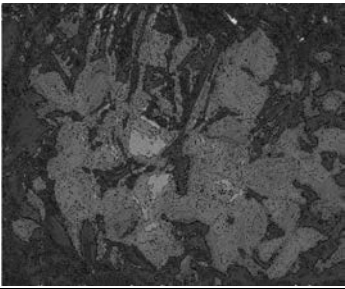
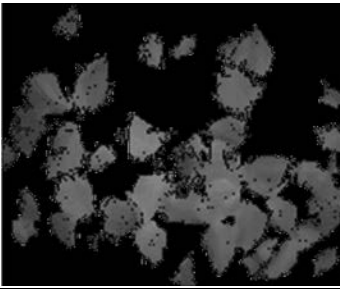
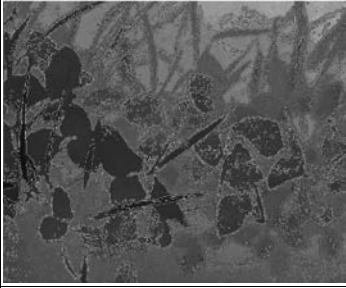

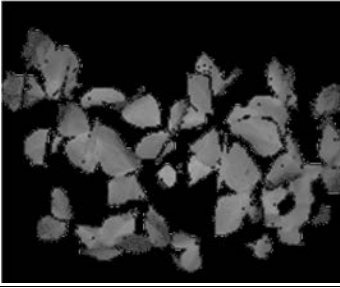
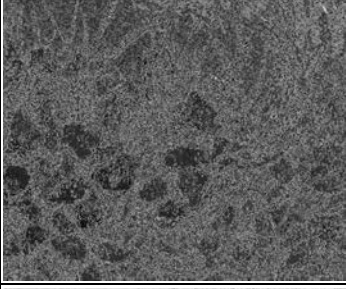
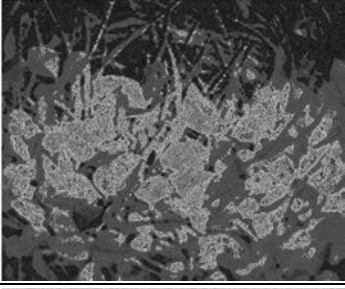
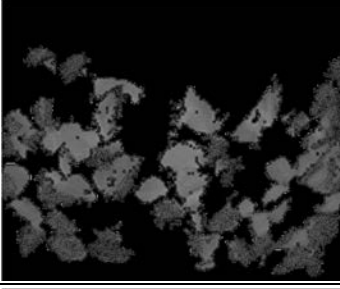
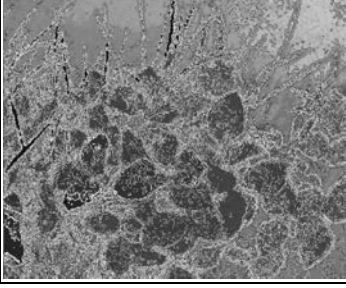
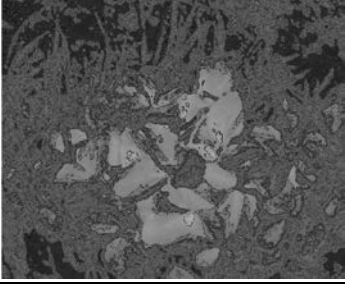
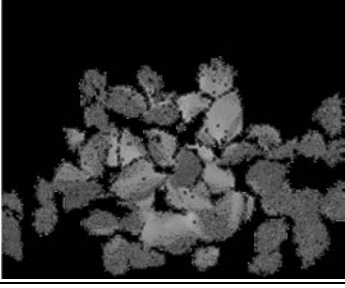
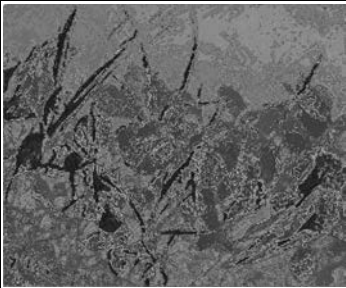
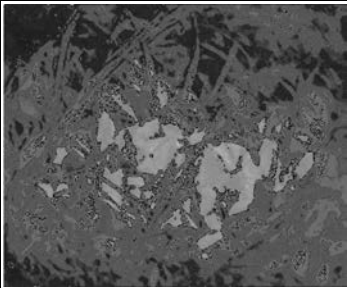
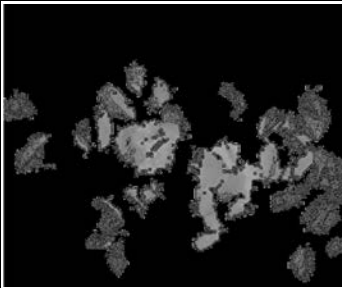
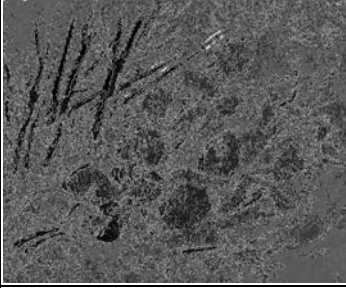
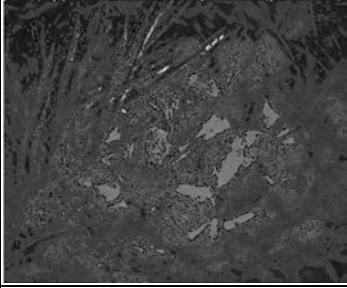
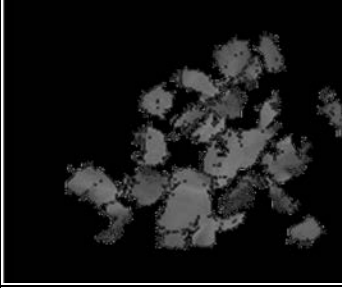
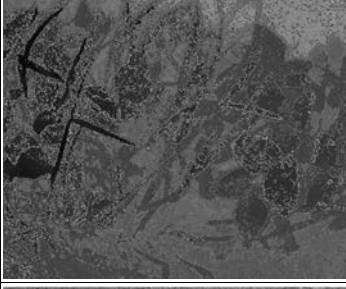
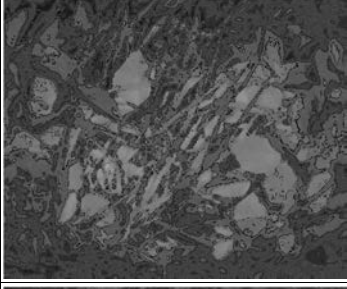

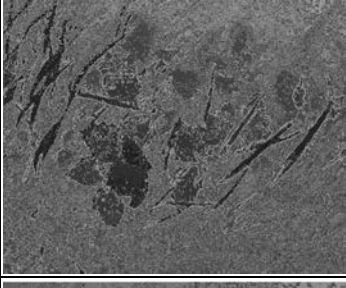
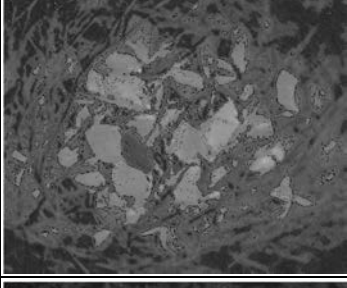
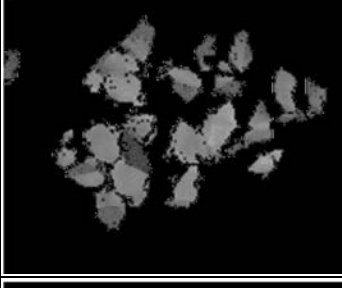
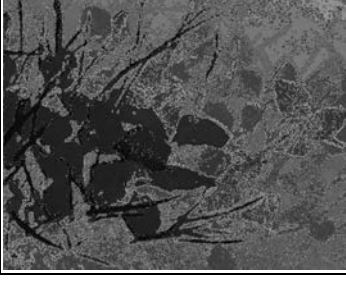
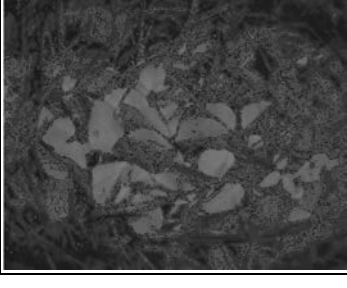
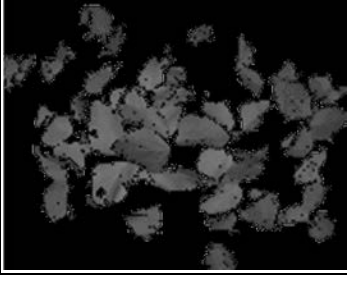
| ID | Depth Image   | Amplitude Image  | Segmentation Result   |
|----|---|--|---|
| 11 |    |    |    |
| 12 |    |    |    |
| 13 |   |   |   |
| 14 |  |  |  |
| 15 |  |  |  |

Table 2.6 continued

| ID | Depth Image   | Amplitude Image  | Segmentation Result   |
|----|---|--|---|
| 16 |    |    |    |
| 17 |    |    |    |
| 18 |   |   |   |
| 19 |  |  |  |
| 20 |  |  |  |

### 2.6.3 Processing Speed

Automated weeding application requires satisfactory processing speed for the crop plant detection task. The algorithm of this research is highly optimized. To test the processing speed of the system, it was run on a 3.4 GHz Intel Xeon CPU to process 1091 broccoli images and 2184 soybean images. According to the experiment result, the processing time range of one broccoli image was between 16.82 ms and 48.53 ms; the mean value was 26.82 ms; and the standard deviation was 6.6 ms. For soybean data, the processing time of one image ranged from 24.11 ms to 59.01 ms; the mean value was 32.68 ms; and the standard deviation was 5.34 ms. The average processing speed for broccoli and soybean were 37.29 and 30.60 frames per second (fps), respectively.

## 2.7 Conclusions

This research developed a soybean and broccoli detection system based on the use of a 3D ToF camera for automated weeding application. From the result of the research, it can be concluded that the 3D imaging based crop plant and weed recognition exhibited promising potential for automated robotic weeding application. First, the sparse noise filter of this research was effective and efficient. Second, the 2D and 3D features, including gradient of amplitude and depth, surface curvature, *API*, normal direction, and neighbor point count in 3D space, were effective to discriminate broccoli and soybean from weeds and soil. Third, according to the 3D geometry and 2D amplitude characteristics of broccoli and soybean plants, one segmentation algorithm was developed for each crop plant. The detection rate of this system reached 88.3% and 91.2% for broccoli and soybean, respectively. Crop plants that were not too small sized in the images were extracted and recovered with relatively

complete shape. Fourth, both the 2D and 3D machine vision algorithms were highly optimized, and the image processing speed of this system was over 30 Hz for both types of plants.

It was found that using a low spatial resolution ToF camera is a limitation to achieve better crop plant detection and segmentation accuracy. This problem makes the small size broccoli plant and soybean leaves do not have enough pixels to accurately analyze their 2D characteristics and 3D geometry, leading them more likely to be missed. The resolution of other 3D ToF cameras is comparable to that of the one used in this study. A better camera with higher resolution could be used in the future to help resolve this problem.

## 2.8 Reference

- Arya, S., D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu. 1998. An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *Journal of the ACM (JACM)* 45(6):891-923.
- Åstrand, B., and A.-J. Baerveldt. 2005. A vision based row-following system for agricultural field machinery. *Mechatronics* 15(2):251-269.
- Du, J.-X., X.-F. Wang, and G.-J. Zhang. 2007. Leaf shape based plant species recognition. *Applied Mathematics and Computation* 185(2):883-893.
- Ehsani, M., S. K. Upadhyaya, and M. L. Mattson. 2004. Seed location mapping using RTK GPS. *Transactions of the ASAE* 47(3):909-914.
- Furbank, R. T., and M. Tester. 2011. Phenomics - technologies to relieve the phenotyping bottleneck. *Trends in Plant Science* 16(12):635-644.
- Griepentrog, H.-W., M. Nørremark, H. Nielsen, and B. Blackmore. 2005. Seed mapping of sugar beet. *Precision Agriculture* 6(2):157-165.

- Jin, J., and L. Tang. 2009. Corn Plant Sensing Using Real-Time Stereo Vision. *Journal of Field Robotics* 26(6-7):591-608.
- Jones, H. G., and R. A. Vaughan. 2010. *Remote sensing of vegetation: principles, techniques, and applications*. Oxford University Press, Oxford, UK.
- Jørgensen, M., H. T. Søgaaard, and P. S. Nielsen. 2002. *Præcision ved automatisk styring af radsensere*. Grøn Viden-Markbrug(268).
- Kapur, J., P. K. Sahoo, and A. Wong. 1985. A new method for gray-level picture thresholding using the entropy of the histogram. *Computer Vision, Graphics, and Image Processing* 29(3):273-285.
- Kise, M., Q. Zhang, and F. Rovira Más. 2005. A stereovision-based crop row detection method for tractor-automated guidance. *Biosystems Engineering* 90(4):357-367.
- Kittler, J., and J. Illingworth. 1986. Minimum error thresholding. *Pattern recognition* 19(1):41-47.
- Klasing, K., D. Althoff, D. Wollherr, and M. Buss. 2009. Comparison of surface normal estimation methods for range sensing applications. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*. IEEE.
- Kumar, L., K. Schmidt, S. Dury, and A. Skidmore. 2002. Imaging spectrometry and vegetation science. In *Imaging spectrometry*, 111-155. Springer.
- Nagasaka, Y., N. Umeda, Y. Kanetai, K. Taniwaki, and Y. Sasaki. 2004. Autonomous guidance for rice transplanting using global positioning and gyroscopes. *Computers and Electronics in Agriculture* 43(3):223-234.
- Nakarmi, A. D., and L. Tang. 2012. Automatic inter-plant spacing sensing at early growth stages using a 3D vision sensor. *Computers and Electronics in Agriculture* 82:23-31.
- Noh, H., Q. Zhang, S. Han, B. Shin, and D. Reum. 2005. Dynamic calibration and image segmentation methods for multispectral imaging crop nitrogen deficiency sensors. *Transactions of the ASAE* 48(1):393-401.



- Otsu, N. 1975. A threshold selection method from gray-level histograms. *Automatica* 11(285-296):23-27.
- Pauly, M., M. Gross, and L. P. Kobbelt. 2002. Efficient simplification of point-sampled surfaces. In *Proceedings of the conference on Visualization'02*. IEEE Computer Society.
- Slaughter, D., D. Giles, and D. Downey. 2008a. Autonomous robotic weed control systems: A review. *Computers and electronics in agriculture* 61(1):63-78.
- Slaughter, D. C., D. K. Giles, and D. Downey. 2008b. Autonomous robotic weed control systems: A review. *Computers and Electronics in Agriculture* 61(1):63-78.
- Søgaard, H. T., and H. J. Olsen. 2003. Determination of crop rows by image analysis without segmentation. *Computers and electronics in agriculture* 38(2):141-158.
- Thompson, J., J. Stafford, and P. Miller. 1991. Potential for automatic weed detection and selective herbicide application. *Crop protection* 10(4):254-259.
- Veeken, M. v. d., L. Tang, and J. W. Hofstee. 2006. Automated corn plant spacing measurement at early growth stages using active computer vision. In *ASABE 2006 Annual International Meeting*. Portland.
- Vrindts, E., and J. d. Baerdemaeker. 1997. Optical discrimination of crop, weed and soil for on-line weed detection. *Precision agriculture '97*. Volume II. Technology, IT and management. Papers presented at the First European Conference on Precision Agriculture, Warwick University, UK, 7-10 September 1997.
- Weiss, U., P. Biber, S. Laible, K. Bohlmann, and A. Zell. 2010. Plant species classification using a 3d lidar sensor and machine learning. In *Ninth International Conference on Machine Learning and Applications (ICMLA)*. IEEE.
- Zwiggelaar, R. 1998. A review of spectral properties of plants and their potential use for crop/weed discrimination in row-crops. *Crop Protection* 17(3):189-206.

## CHAPTER 3. MACHINE VISION BASED REAL-TIME MOTION AND ATTITUDE ESTIMATION

### 3.1 Abstract

For the development of a 3D imaging based plant phenotyping system to extract complex 3D phenotypic features of plants, a methodology for 3D plant reconstruction is the foundation. When using a vision sensor to capture images from multiple views, the main challenge for a complete 3D reconstruction of the entire plant (holographic 3D reconstruction) is to precisely track the posture of the image sensor with a high accuracy so that those images of different views can be registered correctly in a 3D space. In this research, a real-time target (chessboard pattern) detection machine vision algorithm was developed and then utilized for the motion and attitude estimation for a 2D camera and its carrier. Compared to previously reported chessboard pattern detection algorithms, the speed of our new algorithm was greatly improved through effective elimination of the background image. The consistent high-response speed of the target detection algorithm with over 97.49% detection accuracy was achieved, regardless of the existence of the target in the camera view under both indoor and outdoor lighting conditions. Centimeter-level positioning accuracy of the vision sensor was obtained within a  $4.27 \times 7.27$  m testing area in front of one beacon by analyzing the target's appearance in the image. The objects with known position relative to the target can be precisely derived from the target location detected by the algorithm. In addition, the attitude of the 2D camera and its carrier, including pitch, roll, and yaw rotational angles, can also be estimated with an error rate less than  $1^\circ$ . Moreover, the usability of the proposed positioning system was demonstrated by the reliability of the target detection algorithm in identifying different targets, implying that a large number of targets

can be deployed to construct a beacon infrastructure to cover a large area, where the posture of the camera and its carrier can be tracked.

### 3.2 Introduction

The foundation for the development of a sophisticated phenotyping system that can characterize complex plant morphological traits in a complete 3D space (holographic 3D space) is 3D reconstruction. To do this, the main technical challenge is to develop a reliable and accurate 3D registration method that can precisely align different image views of a plant together to form its 3D model.

Tracking the camera's position and attitude and estimating the spatial relationship between different views is critical to 3D registration. Alenya et al. (2011) applied a robotic arm to control and measure imaging sensors' position and attitude for a 3D reconstruction based phenotyping task. Their approach achieved high accuracy and reliability if the robotic arm is of high positioning accuracy and precision, which however will lead to high equipment cost. On the contrary, Rusu et al. (2008) developed a robust multiple-view point cloud alignment approach based on 3D feature histograms. Their algorithm calculated the feature histogram of each point of every view and looked for the key feature point whose feature histogram is significantly different than others. By searching for the correspondence of key feature points between different views, the relationship between different views can be found and used for alignment. The advantage of this approach is that it is purely based on 3D image processing and does not depend on any other devices such as high-precision robotic arms, thus benefiting system cost. However, its calculation is heavy, and thus it requires long processing time.

Additionally, a posture estimation system of an imaging sensor can also be used in other applications; an example is to mount the camera on a robot to provide a localization solution for the robot itself.

Many research efforts were found that developed various positioning systems. Ultrasound was applied for position measurement. For example, Addlesee et al. (2001) developed a system called the active bat positioning system. It installed over 900 receivers inside a building, which sensed the short pulse of ultrasound signals sent by tags to estimate the position of tags based on triangulation calculations. Their system could provide the position measurement inside the building with an approximately 3 cm error. Another advanced acoustic local positioning system called 3D-LOCUS (Prieto et al., 2009) was reported to have a positioning accuracy at the sub-centimeter level. An ultrasound based indoor positioning study (Cho et al., 2012) achieved approximately centimeter-level accuracy. However, these systems are vulnerable to the reflected sound, and they cannot provide accurate attitude information for 3D registration.

Radio frequency (RF) technologies, including radio frequency identification (RFID) (Ni et al., 2004); WLAN (Ladd et al., 2004); and Bluetooth (Munoz-Organero et al., 2012), have been applied for indoor positioning application. However, they suffer from a multipath problem, and their accuracy is not better than meter level. The ultrawideband (UWB) is based on sending ultra-short pulse RF signal for less than 1 ns, which allows researchers to solve the multipath problem by filtering the reflected signal. A complete system level evaluation of a UWB based communication and location system was presented (Choliz et al., 2011). The accuracy of the RF based location system could hardly reach the centimeter level.

The machine vision based position and attitude estimation system is promising. Zhang (2000) presented a camera calibration method by observing a planar pattern from a few different orientations. This research also provided the method to calculate the extrinsic parameters, which are the rotation and translation parameters that relate the world coordinate system defined by the planar pattern to the camera coordinates system. Once the camera intrinsic parameter calibration is done, which needs to be done only once for each camera, the pose of the target, including the position and attitude (pitch, yaw, and roll rotational angles), related to the camera coordinates system can be precisely calculated when the target image from the camera is analyzed. In this way, the motion and attitude of the camera and its carrier can be precisely evaluated by calculating the relationship between the 2D camera mounted on a carrier and the planar target serving as the beacon. Though this approach has great potential for motion and attitude estimation, it will not become practically useful until the challenge of accurate and rapid target detection under unpredictable image background conditions is resolved.

Among the planar patterns mentioned, the chessboard pattern was widely used for camera calibration. A target detection algorithm was proposed based on the characteristics of local intensity and the grid-line architecture of the planar checkerboard pattern image (Wang et al., 2007). However, this algorithm needed around nine seconds to process an image with a resolution of only  $640 \times 480$ . Another inner corner detection based algorithm was reported (Ha, 2009), but no information about its speed was provided. Another interesting target localization approach was to localize the target through detecting the double-triangle feature of inner corners of a chessboard pattern using a template (Yu and Peng, 2006), but it failed to reach a satisfactory speed. The Hough transform was applied for chessboard pattern detection

by extracting the straight lines of the target image (de la Escalera and Maria Armingol, 2010). However, it can fail when the chessboard appears small in the image. OpenCV achieved a robust target detection algorithm. However, its processing speed is still too low for real-time processing requirements. In all, because of the problems of previous chessboard target detection algorithms, they were mostly applied to camera calibration, which has relatively low speed requirements; and none of them were applied or suitable for real-time motion and attitude estimation.

Although satisfactory accuracy, robustness, high update rate, and low cost are important for motion and attitude estimation systems, the literature review revealed that previous local positioning systems have limitations to meeting these requirements both indoors and outdoors. The objective of this research is to greatly improve the chessboard pattern target detection speed so that the detected target can be used as the beacon for a machine vision based sensor motion and attitude estimation system, which is used for a 3D reconstruction system. Additionally, a precise object tracking system based on the reference of beacon target images can be used to help a mobile robot to localize itself and manipulate objects, such as buttons, knobs, handles, and many other devices, enabling the robot to automatically manage facilities, for instance.

### 3.3 Sensor

The camera is critical for the precision of the motion and attitude sensing system based on machine vision technique. To achieve precise measurement results, the camera should satisfy following requirements.

- 1) The lens of camera must be fixed. The system proposed by this research calculates the pose of the target by analyzing its captured image, which is related to the optical system. The information about the camera model, such as focal length and distortion efficiency, must be known before it can be applied as positioning sensor. Therefore, it is better to fix the optical system of the camera so that the camera can be applied to precision measurement after its model is calibrated once.
- 2) As the focal length of the camera must be fixed, the camera must have a wide focus range to capture the target board either close or far away.
- 3) The high pixel resolution of the camera benefits bigger image size of the target, which helps to increase the detection distance of the target as well as the target pose measurement precision.
- 4) Even though the distortion effect of the camera can be measured through calibration, and its negative impact for target pose calculation can be compensated for to some degree, having a well-designed camera optical system with low distortion effect is helpful for system precision.

For these reasons, the Logitech HD Webcam C310 was selected for this research. It features wide focus range, a fixed focal length of 4.4 mm, 60° field of view, and 1280 × 720 pixel resolution. In addition, it has the satisfactory distortion effect defined for this study.

### 3.4 Target Detection Algorithm

#### 3.4.1 Challenges of Target Detection

The inner corners of the chessboard pattern target image should be extracted to calculate the camera extrinsic parameters, which describe the translation and rotation relationship between the target and camera. As OpenCV function *cvFindChessboardCorners*, which detect the chessboard pattern target and locate its inner corners, shows great robustness but it does not have enough speed for real-time processing, and the corresponding algorithm of this research was developed based on it in order to greatly improve the speed while keeping its robustness. The main steps of the target detection algorithm of OpenCV are:

- 1) Image normalization, which is carried out first to equalize the input image histogram, so that the contrast between the black and white areas of the target is visible.
- 2) Adaptive threshold is applied to convert the grayscale image to the binary image, so that the black and white squares of the chessboard target can be distinguished.
- 3) Dilation is applied with some iteration times.
- 4) The algorithm searches for the quadrangles that are square-like and big enough, and their corners are also retrieved. The black squares of the chessboard target should be contained by the achieved quadrangles candidates.
- 5) The neighbor quadrangles are organized and ordered into groups. Further analysis is conducted to search for the quadrangle group, which forms the chessboard target pattern.

For each input image, the OpenCV function iterates steps two to five and tries different window size and other parameters of the adaptive threshold of step two and



different dilation iteration times in step three to search for target patterns until it succeeds or reaches the maximum iteration time. Therefore, while it is relatively faster to localize the target when one is available in the image, much more time is required for this function to determine that no target is available because the function must try all of the different parameters before it finishes. The localization system of the mobile vehicle requires real-time response speed of the beacon detection algorithm, no matter whether there is target available in the image captured or not. According to the performance of the OpenCV, the speed of the chessboard detection algorithm should be very high, especially in cases when no target is available in the image.

This research makes two contributions to significantly increase the target detection speed. First, a pre-analysis algorithm was developed to effectively identify most parts of the background area in the image. The target searching area in the steps described in the steps previously was greatly reduced. Second, as the adaptive threshold of step two of the target detection function is time consuming and it may be called multiple times during the target searching process, a new implementation was proposed to accelerate it.

### **3.4.2 Background Identification Algorithm for Target Detection Acceleration**

A key to increasing the target detection speed is to effectively identify as much background image area as possible in advance, with low calculation cost, in order to save the further target detection time cost by reducing the searching area. This research developed an algorithm that is capable of doing so with low calculation cost before calling the target searching algorithm mentioned previously. Based on that, the edges of all of the squares of the target were connected together as a frame, and the squares were enclosed inside and

separated as neighbors with similar properties. The main idea was to detect the edge first, and then remove the edges that cannot belong to a target based on some criteria. After that, the images were divided into separated areas by the edges. Finally, the properties of all of the areas and the similarity between neighboring ones were analyzed, and the ones that did not satisfy the criteria proposed were regarded as the background. The details are illustrated below.

#### 3.4.2.1 Grayscale Down-Sampled Image

A down-sampled grayscale image A was created from the original input color image captured by the camera using Equation 3.1.

$$A(x, y) = \frac{R(2x, 2y) + G(2x, 2y) + B(2x, 2y)}{3} \quad (3.1)$$

where R, G, and B are the red, green, and blue color component of the input color image, respectively. The width and height of the down-sampled image were 640 and 360, respectively, which were half of those of the input image. There were two benefits of down-sampling actions. First, the decreased image size helped to increase the image processing speed. Second, it made the edge detection robust to the fuzzy edges of target, which were caused by the movement of camera or target. Down sampling also helped to avoid the missing detection of fuzzy edges of the solid black squares of the target.

#### 3.4.2.2 Edge Detection

To do the edge detection, the Sobel operator was used to process gray image A to calculate the approximate gradient magnitude image D, as expressed by Equation 3.2-3.4.

$$D_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} * A \quad (3.2)$$

$$D_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} * A \quad (3.3)$$

$$D = |D_x| + |D_y| \quad (3.4)$$

The two matrixes in Equation 3.2 and 3.3 are Sobel convolution kernels, and they were applied to input gray image A to generate the horizontal and vertical derivative images  $D_x$  and  $D_y$ , respectively. In addition, the approximate gradient magnitude image D is the sum of their absolute value, as Equation 3.4 indicates.

A constant threshold 170 was applied for edge detection. The pixels above 170 were regarded as edges. Figure 3.2 shows one edge detection result.

#### 3.4.2.3 Edge Screening

To judge whether the edges achieved in the previous step were possibly to be those of the chessboard target, each detected edge was analyzed using the criteria represented by Equation 3.5-3.8.

$$w \geq 30, \text{ where } w = X_{max} - X_{min} + 1 \quad (3.5)$$

$$h \geq 30, \text{ where } h = Y_{max} - Y_{min} + 1 \quad (3.6)$$

$$nPixel > 250 \quad (3.7)$$

$$nPixel < pwh, \text{ where } p = \min\left(0.6, \frac{25w+25h}{wh}\right), \quad (3.8)$$

where  $X_{max}$ ,  $X_{min}$ ,  $Y_{max}$ , and  $Y_{min}$  are the maximum and minimum value of the X and Y coordinates of all of the pixels belonging to the corresponding edge, respectively, and  $nPixel$  is the total number of its pixels. The edge of the chessboard target should be a frame, whose width, height, and total number of pixels has their minimum threshold, respectively, as illustrated by Equation 3.5–3.7. Since the edge of the target image should have some blank

area inside rather to being solid as Figure 3.2 shows, Equation 3.8 was used to check this criteria.

Only the edges that satisfied all of Equation 3.5–3.8 were kept for further analysis. The white lines in Figure 3.3 are the edge selection result of Figure 3.2.

#### 3.4.2.4 Area Analysis

After the edges were screened, the image was divided into different disconnected areas. The areas with a number of pixels that was equal or smaller than 150 were kept for the future target detection algorithm, without any further inspection in this step. These areas were so small that their shapes could deviate from the original ones after the edge detection. For the rest of the areas with sizes of more than 150 pixels, Equation 3.9–3.10 were applied to preliminarily screen the target squares candidates based on the fact that the squares of the chessboard target should be enclosed by the target edge detected with the rectangular or square-like shape in the image.

$$5h \geq w \geq 0.2h, \text{ where } h = Y_{max} - Y_{min} + 1, \text{ and } w = X_{max} - X_{min} + 1 \quad (3.9)$$

$$nPixel > 0.4wh, \quad (3.10)$$

where  $X_{max}$ ,  $X_{min}$ ,  $Y_{max}$ , and  $Y_{min}$  are the maximum and minimum value of the X and Y coordinates of all the pixels of the corresponding area, respectively, and  $nPixel$  is the total number of its pixels. Equation 3.9 defines a range between 0.2 and 5 for the ratio between the width and height of a candidate area. Equation 3.10 could screen out areas whose shapes are not rectangular or square.

The relatively big areas that passed Equation 3.9–3.10 were further inspected. Observing the chessboard target, the inner squares have more than four neighboring squares

with similar width, length, and number of pixels. Equation 3.11–3.13 were applied to check whether neighbor areas had similar shapes.

$$0.75(X_{ij\_max} - X_{ij\_min}) < X_{i\_max} - X_{i\_min} < 1.33(X_{ij\_max} - X_{ij\_min}) \quad (3.11)$$

$$0.75(Y_{ij\_max} - Y_{ij\_min}) < Y_{i\_max} - Y_{i\_min} < 1.33(Y_{ij\_max} - Y_{ij\_min}) \quad (3.12)$$

$$0.75 \times nPixel_{ij} < nPixel_i < 1.33 \times nPixel_{ij}, \quad (3.13)$$

where  $X_{i\_max}$ ,  $X_{i\_min}$ ,  $Y_{i\_max}$ , and  $Y_{i\_min}$  are the maximum and minimum value of the X and Y coordinates of  $i_{th}$  area, respectively, and  $nPixel_i$  is its size.  $X_{ij\_max}$ ,  $X_{ij\_min}$ ,  $Y_{ij\_max}$ ,  $Y_{ij\_min}$ , and  $nPixel_{ij}$  are the parameters of the  $j_{th}$  neighbor area. For the areas with at least four neighbors, which satisfied Equation 3.11–3.13, these areas and their similar neighbors were kept as the final candidate area for target detection. The rest were regarded as not belonging to the target.

Figure 3.3 shows the result of area analysis. The gray part is the area identified as background, which was eliminated for later target searching. The white lines are the edges kept by the edge selection step, and the color areas are those kept for later target searching.



Figure 3.1. Raw sample image captured from camera

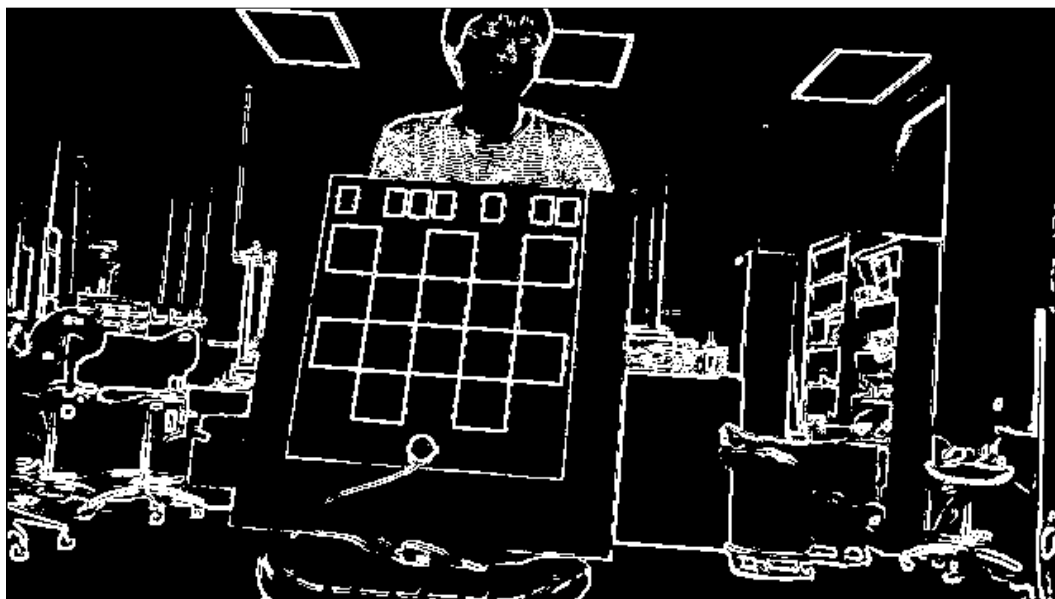


Figure 3.2. Edge extraction result

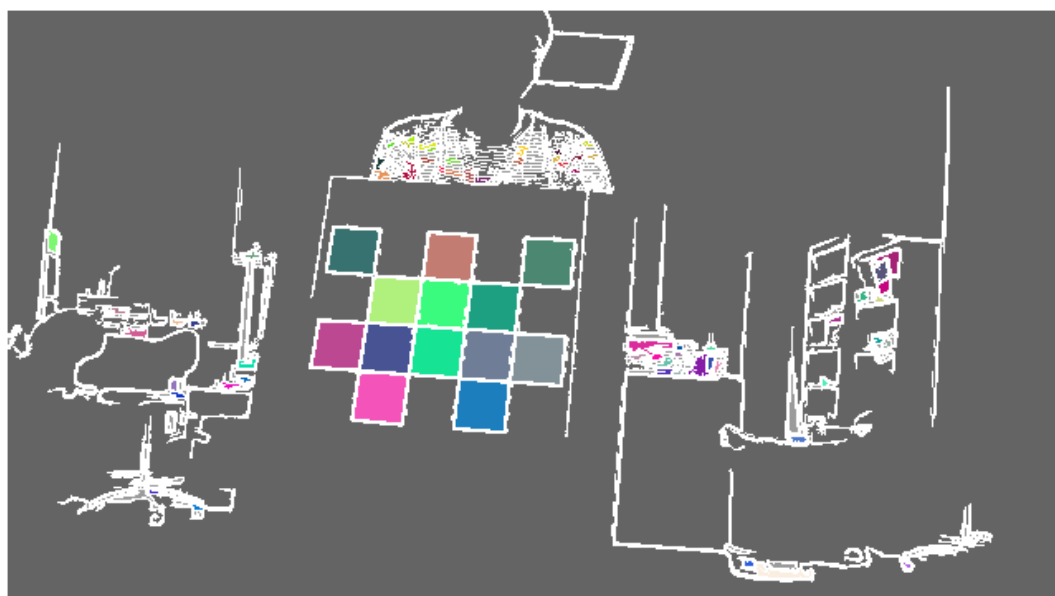


Figure 3.3. Result of area analysis: The gray part is the area eliminated for later target searching; the white lines are the edges kept by the edge selection algorithm; and the color areas are kept for later target searching



Figure 3.4. Target detection result

#### 3.4.2.5 Summary of Background Identification Algorithm

The four steps detailed in the previous section were developed to identify most parts of the background image area. Low CPU time cost was achieved because of the simplicity of this algorithm. Edge and area screening criteria represented by Equation 3.5–3.13 alone were simple, but their combination showed great power to robustly and effectively reduce the target searching candidate area and thus significantly reduced the detection time.

#### **3.4.3 Optimization of Adaptive Threshold**

After the pre-analysis algorithm for identifying the background image area, a similar target detection process with the OpenCV function *cvFindChessboardCorners* illustrated earlier was called. During the iterations of the target detection process, adaptive threshold was called repeatedly to try different window size  $w$ , which was an odd integer, and

parameter  $p$ , to accurately extract the squares of the chessboard pattern. The process compared the intensity value of each pixel of the gray image with the mean value of  $w \times w$  pixel neighborhood minus parameter  $p$  to determine the threshold result as shown by Equation 3.14 and 3.15.

$$dst(x, y) = \begin{cases} 1, & \text{if } src(x, y) > T(x, y) \\ 0, & \text{otherwise} \end{cases}, \quad (3.14)$$

where

$$T(x, y) = \frac{\sum_{i=x-\frac{w-1}{2}}^{x+\frac{w-1}{2}} \sum_{j=y-\frac{w-1}{2}}^{y+\frac{w-1}{2}} src(x, y)}{w^2} - p \quad \text{and} \quad (3.15)$$

where  $src$  is the input grayscale image and  $dst$  is the output binary image of the threshold action. According to Equation 3.15, for each image with  $1280 \times 720$  pixels, the mean value of the corresponding neighborhood needs to be calculated to determine the threshold value. It is time consuming to sum the intensity value of  $w^2$  neighbor pixels. If all of these calculations were processed every time during the multiple iterations of the target detection, it would create a bottleneck to speed.

To speed up the adaptive threshold algorithm, an image  $C$  was proposed to be created in advance before target detection. It had the same dimensions as the grayscale input image  $src$ , which was  $1280 \times 720$  pixels, and it was generated using Equation 3.16.

$$C(x, y) = \sum_{i=1}^x \sum_{j=1}^y src(x, y) \quad (3.16)$$

The  $C(x, y)$  is the sum of the rectangle areas with corners  $(1, 1)$ ,  $(x, 1)$ ,  $(1, y)$ , and  $(x, y)$  of input image  $src$ . Having the image  $C$ , Equation 3.15 could be replaced with Equation 3.17 for iterations of adaptive threshold.

$$T(x, y) = \frac{C\left(x+\frac{w-1}{2}, y+\frac{w-1}{2}\right) - C\left(x+\frac{w-1}{2}, y-\frac{w+1}{2}\right) - C\left(x-\frac{w+1}{2}, y+\frac{w-1}{2}\right) + C\left(x-\frac{w+1}{2}, y-\frac{w+1}{2}\right)}{w^2} - p \quad (3.17)$$



Therefore, the heavy-load calculation of Equation 3.15 was greatly reduced by Equation 3.17, which required few operations and thus saved much CPU time. Figure 3.4 shows the target found by the accelerated detection algorithm, whose inner corners are marked by the connected color circles.

### 3.5 Position and Attitude Estimation

#### 3.5.1 Positioning Theory

Having the chessboard pattern target detected, the translation and rotation relationship between camera and beacon target can be calculated. For example, the size of the target image decreases when their distance increases. In addition, another example is that the image of the square of the target deforms to a trapezoid or other shape when it is not parallel to the imager. Knowing the dimensions of the target and the camera model, which describes the relationship between the object and its projection on the imager, the target's rotation and translation related to the camera can be calculated based on its captured image. Therefore, calibrating the camera model in advance is indispensable for the positioning system based on the machine vision technique proposed.

The simplest model of a camera is a pinhole model. Equation 3.18 summarizes the projection of the points in the physical world into the camera.

$$q = sMQ_c, \text{ where } q = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, M = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, Q_c = \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix}, \quad (3.18)$$

where  $M$  is the *camera intrinsic matrix*,  $(C_x, C_y)$  is the principal point,  $f_x$  and  $f_y$  are the focal lengths related to two directions expressed in pixel related units, and  $Q_c (X_c, Y_c, Z_c)$  is the coordinate value of a physical point in the camera coordinate system.

The simple pinhole camera model is not accurate enough to use for precise measurement. The distortion of the lens cannot be ignored. In practice, no camera is perfect because of its nonideal lens and assembly, which lead to radial and tangential distortions. To correct the error caused by the lens distortion, a  $5 \times 1$  distortion vector to describe these two kinds of distortions needs to be calculated.

Camera calibration was done to calculate the camera intrinsic matrix and distortion vector using the OpenCV function *cvCalibrateCamera2* by inputting the inner corners of the chessboard pattern target image captured at different viewpoints. In practice, there is difference between the lens shape and assembly of different cameras, even for those of same model, so the camera individuals' intrinsic and distortion parameters are not exactly the same. Thus every camera needs to be calibrated before being applied for positioning measurement in advance.

The mapping relationship between the viewed physical point  $Q$  and the point  $q$  on the imager to which  $Q$  is mapped can be expressed in terms of matrix multiplication. If it is defined that:

$$\tilde{Q} = [X \ Y \ Z \ 1]^T \quad (3.19)$$

$$\tilde{q} = [\tilde{x} \ \tilde{y} \ 1]^T, \quad (3.20)$$

where  $(X, Y, Z)$  of  $\tilde{Q}$  represents the world coordinates of one physical point, and  $(\tilde{x}, \tilde{y})$  of  $\tilde{q}$  is the coordinates of the mapped point on the image after correction for lens distortion.

$$\tilde{q} = sMW\tilde{Q}, \text{ where } W = [R \ | \ t] \quad (3.21)$$

Then, the mapping between  $\tilde{Q}$  and  $\tilde{q}$  can be represented by Equation 3.21.  $W$  is a matrix of extrinsic parameters, which is the combination of rotation matrix  $R$  and translation vector  $t$ , for describing the relationship between the world coordinate system and the camera

coordinates system. In this research, if not mentioned, the world coordinate system is the coordinate system defined by the target; the target on the plane follows the equation  $Z = 0$ ; and the origin is the bottom-left inner corner, as Figure 3.5 indicates. Figure 3.5 also shows the camera coordinate system.

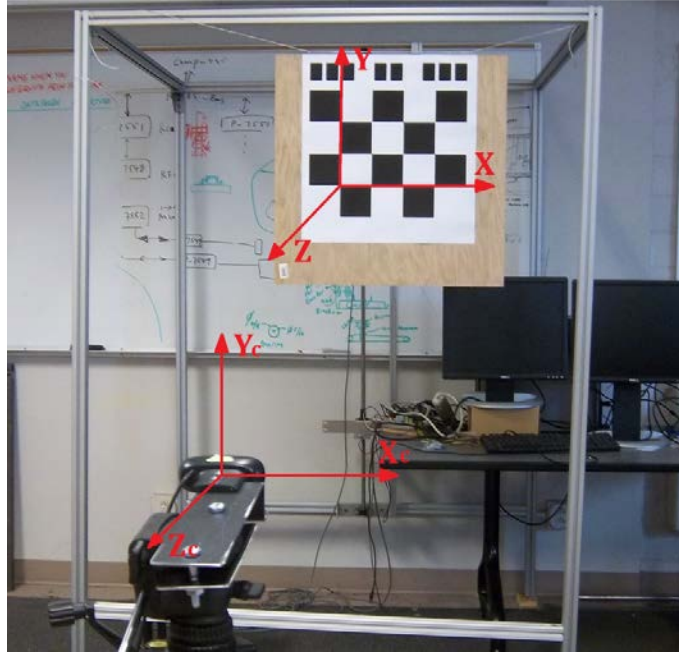


Figure 3.5. The world coordinate system defined by the target and camera coordinate system

The mapping relationship between the world coordinate system and camera coordinate system can be expressed by Equation 3.22.

$$Q_c = W\tilde{Q} = [R \mid t]\tilde{Q} \quad (3.22)$$

Equation 22 equals another form represented by Equation 3.23, which expresses how to calculate the world coordinate value of a physical point based on its position related to the camera coordinate system. It is useful to evaluate the motion of the camera's carrier, as the world coordinate value of a physical point that has a fixed position related to the mounted camera can be applied as the position of carrier. In this research,  $Q_c$  of this reference point was  $[0, 0, 0]^T$ .

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = R^{-1}(Q_c - t) \quad (3.23)$$

After the chessboard target was detected, its inner corners were extracted and inputted to the function *cvFindExtrinsicCameraParams2* of the OpenCV to calculate the rotation matrix  $R$  and the translation vector  $t$  related to the camera coordinate system and the world coordinate system in Equation 3.22 and 3.23. By substituting  $Q_c$  in Equation 3.23 with a reference point  $[0, 0, 0]^T$ , the camera/carrier position can be calculated as Equation 3.24 expresses:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = -R^{-1}t \quad (3.24)$$

The image point  $q$  of the objects should be located and tracked. The distortion corrected image point  $\tilde{q}$  of the point  $q$  in the original image related to a physical point with a known world coordinate value  $Q$  can be calculated by solving Equation 3.21. For this equation, the matrix  $W$  can be achieved after the target has been detected and the camera's intrinsic matrix has been obtained using the calibration process in advance; it still has three unknown variables, which are  $s$ ,  $\tilde{x}$ , and  $\tilde{y}$ . These can be solved by Equation 3.25-3.27.

$$s = \frac{1}{Z_c} \quad (3.25)$$

$$\tilde{x} = s(f_x X_c + c_x Z_c) \quad (3.26)$$

$$\tilde{y} = s(f_y Y_c + c_y Z_c) \quad (3.27)$$

where  $Q_c (X_c, Y_c, Z_c)$  can be calculated using Equation 3.23. After the distortion corrected image point  $(\tilde{x}, \tilde{y})$  is achieved, the OpenCV function can be applied to calculate the corresponding point before distortion correction in the original image.

### 3.5.2 Attitude Evaluation

The attitude of the camera can be evaluated by measuring the rotation relationship between the coordinate system of camera mounted and that of the target, which is expressed by Equation 3.22, whose rotation matrix  $R$  and translation vector  $t$  can be achieved using the OpenCV after the target has been detected. Using the *cvRodrigues2* function of the OpenCV, the rotation matrix  $R$  can be translated to pitch, yaw, and roll rotation angles from the target coordinate system to the camera coordinate system.

### 3.6 Target Identification

To identify different chessboard pattern targets, up to ten black rectangles are proposed to encode a 10 bit binary ID, as the top part of the two targets of Figure 6 indicates. The existence and absence of a black rectangle represents one and zero at the corresponding bit, respectively. Therefore, the ID of the left and right target of Figure 3.6 is 1023 and 813, respectively.

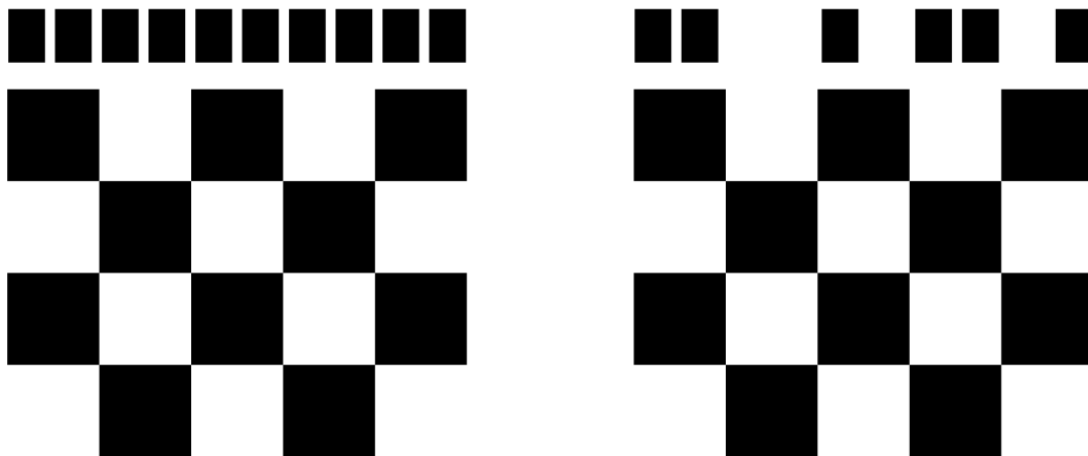


Figure 3.6. Two target samples

As the ID's physical position related to the chessboard pattern is known when making the target, the corners of each square can be precisely located on the original image captured by the camera using Equation 3.25–3.27 and relative to the earlier discussion when the target is detected, as the red circles on the top of the target of Figure 3.4 shows. Therefore, it is straightforward to decode the ID through identifying the existence of the rectangle for every bit.

### 3.7 Experiments and Results

A set of experiments were done to evaluate the performance of the target detection algorithm, position and rotation measurement, ID recognition, and object tracking on the image proposed. The chessboard pattern of the target used in this research had four rows and five columns squares, and the side of every square was 82.9 mm. Moreover, in order to prevent the target from appearing glossy under sunlight, matte finish paint was applied to the target to ensure that it could be reliably detected outdoors.

#### 3.7.1 Target Detection and ID Recognition Performance Experiment and Result

To apply the chessboard pattern target as the beacon of the positioning solution, the speed and reliability of its detection algorithm are crucial. There are four criteria to evaluate the target detection speed. First, the algorithm should respond quickly when the target available in the view. Second, the positioning algorithm needs to be acknowledged as soon as possible if no beacon is available in the captured image. Third, the detection speed is better when it is consistently fast enough to ensure that it does not drop significantly at any time for some condition to impact the operation of the application. Fourth, the detection algorithm needs to work reliably in both indoor and outdoor conditions.

For this system, it could not detect the target when it was too far or moving too quickly. When the target distance was so far that the side of the square of the chessboard target was less than 8 pixels in the image, the target could not be detected, and its ID could not be recognized by this system, as it was filtered by the design of this algorithm. For the target size and Logitech HD Webcam C310 used in this research, this condition typically happened when the target was over approximately 10 m from the camera. Further, if the target moving speed is too fast, the edge of the beacon becomes fuzzy in the image, thus failing this target detection system by breaking its edge detection algorithm. The maximum target moving speed allowed by this system was relative to the camera exposure time, and thus was related to the camera's light sensitivity and environment illumination condition. For all of the following target detection and target ID recognition experiments, the distance between the target and the camera was within 7 m and the target moving speed was less than 0.7 m/s if the target was available in the view.

The target detection performance was examined by comparing the algorithms of this research and OpenCV. Four videos were applied to test the speed and reliability performance of the target detection and ID recognition algorithm. To test the performance when the target was in the view, three videos were captured indoors and outdoors. One of them was captured indoors, as Figure 3.7 shows, and it has 1952 frames. The other two were captured outdoors. The first outdoor video has 2987 frames, and the camera direction is about  $135^\circ$  to the sun, which means the sunlight was on the rear right of camera, as Figure 3.8 (a) shows. The second outdoor video has 2709 frames, and the camera direction is about  $45^\circ$  to sun, which means the sun light was on the front right of the camera, as Figure 3.8 (b) shows.

Moreover, in order to test the performance of the target detection algorithm when there is no target in view, two videos were captured, indoors and outdoors, respectively for no target condition, as Figure 3.9 and 3.10 show. For both videos, the camera was rotated  $360^\circ$  around the vertical axis to test the system performance for different images. The video captured indoors has 1596 frames, and the other one has 1954 frames.

For all five videos, each frame was processed 20 times by the algorithms of OpenCV and this research, respectively, and the mean time cost of every frame of each method was recorded for the performance analysis.

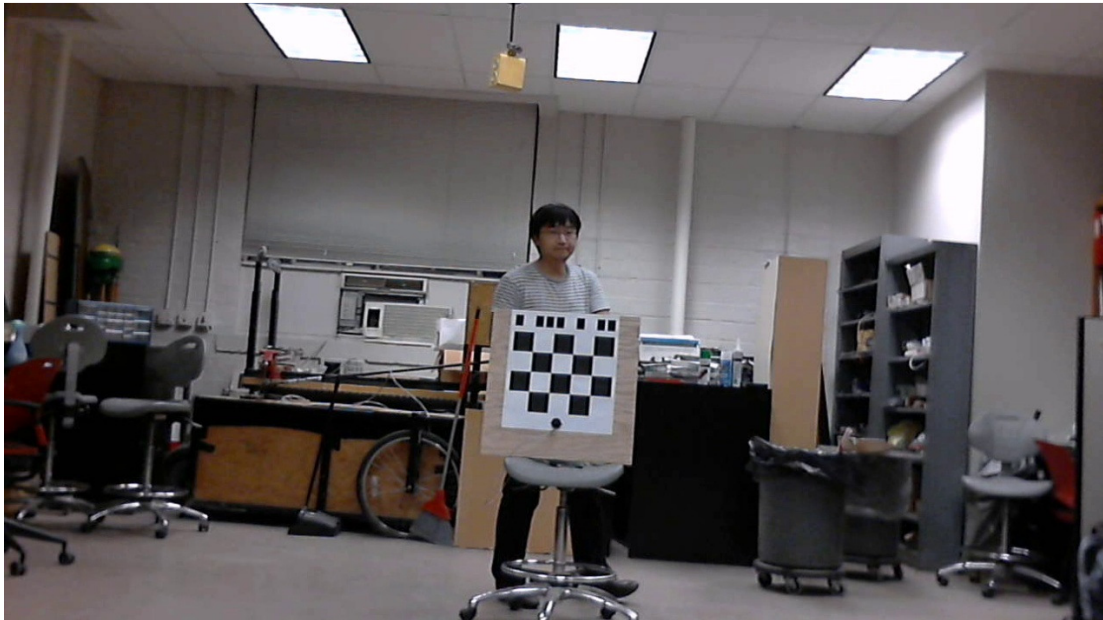


Figure 3.7. Video frame of moving target captured indoors





(a)



(b)

Figure 3.8. Frames of the videos of a moving target captured outdoors: (a) a frame of the video whose camera direction is around  $135^\circ$  to the sun and (b) a frame of the video whose camera direction is around  $45^\circ$  to the sun



Figure 3.9. A frame of the video without a target in view, captured indoors



Figure 3.10. A frame of the video without a target in view captured outdoors



For the experiment with the target in view, both OpenCV and this research reported 100% accuracy in locating the target with the indoor video and the outdoor video that were made with the camera direction at about  $135^\circ$  to the sun. For the outdoor video made with the camera direction at about  $45^\circ$  to the sun, the target detection rate of the OpenCV and this research is around 97.2% and 97.49%, respectively. The target detection time cost of both algorithms to process the indoor and outdoor videos is given by the boxplot shown by Figure 3.11, and Table 3.1 also provides the relative statistics. Table 3.1 indicates that the mean time cost to locate the target available by this research is 24.84 ms for the indoor video and 29.75 ms for the two outdoor videos, and those of OpenCV were 36.29 ms and 99.23 ms. The mean speed of this research is 1.46 and 3.34 times of that of OpenCV to process the videos captured indoors and outdoors, respectively. The interquartile range, which is the difference between the 25% and 75% percentile, of the detection time cost of this research is 13.49% and 14.59% of that of OpenCV for the indoor and outdoor video, respectively, supporting that it has more consistent high target detection speed with less variation. Further, the maximum target detection time cost of the videos, captured indoors and outdoors by this research, was 39.92 ms and 71.37 ms, respectively, and they were much smaller than those of the OpenCV, which were 72.88 ms and 412.7 ms, respectively. For all of the videos captured, both indoors and outdoors with a moving target in the view, the target ID was 100% accurately recognized by the proposed method when the target was detected in these target detection experiments.

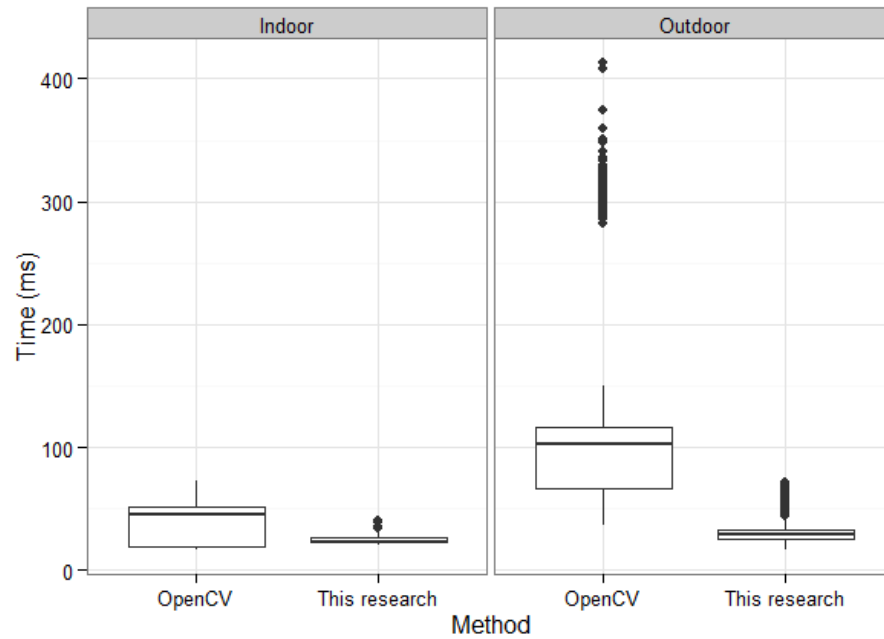


Figure 3.11. Target detection time cost when target is available

For the two videos without a target in the view that captured images indoors and outdoors, no false positive detection result were reported by this research. Comparatively, OpenCV algorithm made one false positive detection when processing the 1954 frames of one video captured outdoors. The target detection time cost of these two videos by two algorithms is provided in the Figure 3.12 and Table 3.1. The mean response time for the condition that no target is available in the view of this research is 39.33 ms and 62.61 ms for the indoor and outdoor video, respectively, and those of OpenCV were 664.4 ms and 1104.5 ms. The mean speed of this research is 16.89 and 17.64 times of that of the OpenCV to process the indoor and outdoor videos respectively. The interquartile range of the response time cost of this research is 11.43% and 21.05% of that of the OpenCV for the indoor and outdoor video respectively, showing its consistent higher processing speed with less variation. Further, the maximum response time cost of the videos, captured indoors and outdoors in this

research, was 74.06 ms and 169.74 ms, respectively, and they were much smaller than those of OpenCV, which were 846.3 ms and 1791.5 ms, respectively. The result shows that the target detection algorithm of OpenCV is slow when no target is available in the image, and it is a bottleneck to applying this machine vision based technique as a real-time positioning solution for 3D reconstruction or autonomous robot. The greatly improved speed of this research solved the limitation.

The variation of processing time was mainly the result of the difference of background image. The complexity image led to longer response time, especially for OpenCV algorithm. The experiment result indicated that this research reduced the influence of background complexity and helped to achieve consistent high performance.

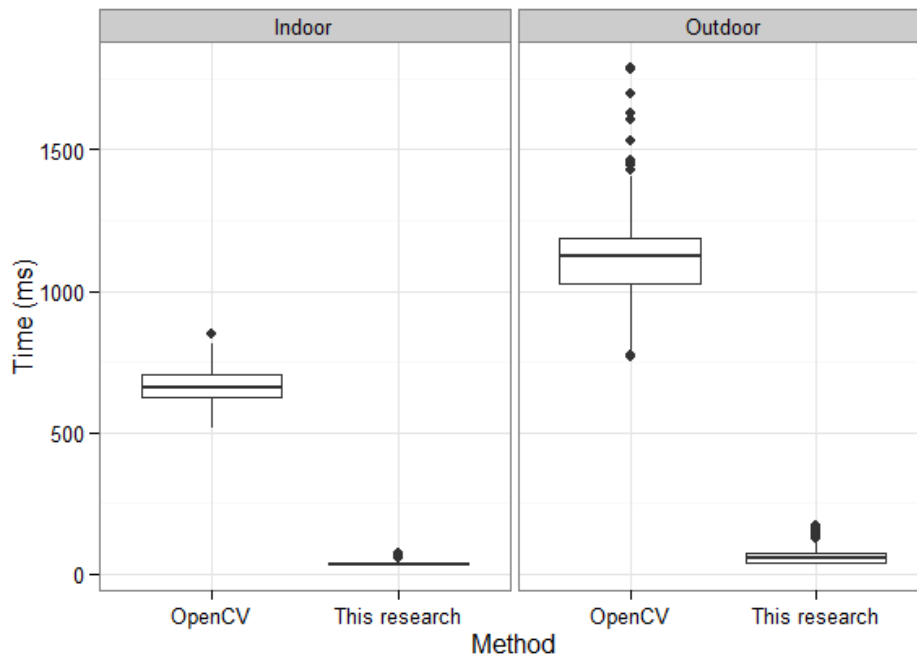


Figure 3.12. Response time of target detection algorithm when no target is available

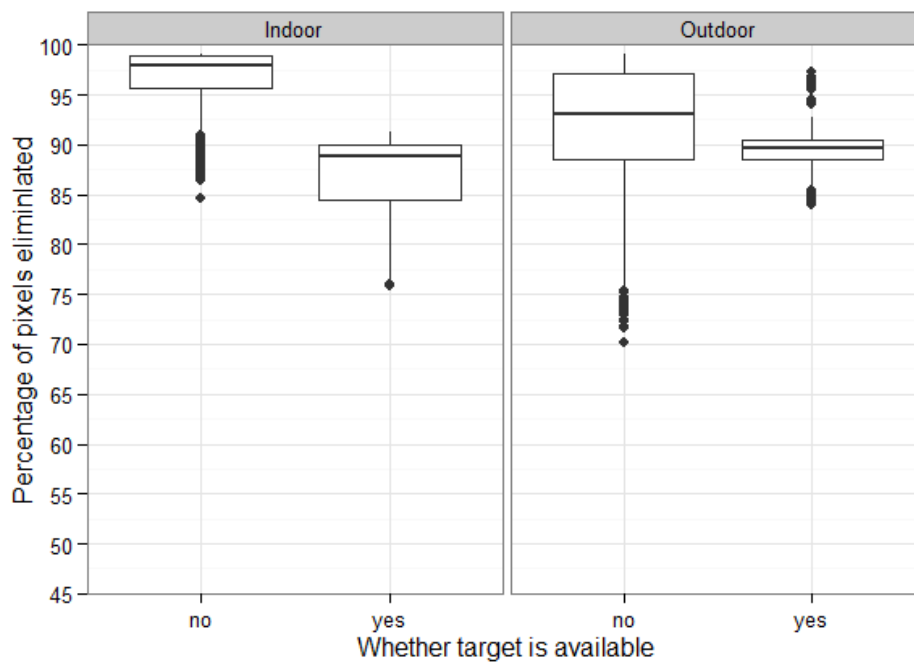


Figure 3.13. Percentage of image areas eliminated for the later target searching

Table 3.1. Response time cost of different target detection algorithm

| Algorithm     | Condition | Target | Time cost percentile (ms) |        |        |        |        |        |        |
|---------------|-----------|--------|---------------------------|--------|--------|--------|--------|--------|--------|
|               |           |        | Min                       | 25%    | median | Mean   | 75%    | 95%    | Max    |
| This research | Outdoor   | Yes    | 17.11                     | 25.55  | 29.04  | 29.75  | 32.76  | 37.26  | 71.37  |
| OpenCV        | Outdoor   |        | 36.66                     | 66.68  | 102.30 | 99.23  | 116.10 | 134.71 | 412.7  |
| This Research | Indoor    |        | 19.84                     | 22.69  | 23.46  | 24.84  | 27.01  | 30.72  | 39.92  |
| OpenCV        | Indoor    |        | 16.67                     | 18.94  | 45.7   | 36.29  | 50.97  | 57     | 72.88  |
| This Research | Outdoor   | No     | 33.52                     | 41.6   | 56.87  | 62.61  | 75.3   | 117.09 | 169.74 |
| OpenCV        | Outdoor   |        | 765.9                     | 1025.7 | 1124.7 | 1104.5 | 1185.8 | 1266   | 1791.5 |
| This Research | Indoor    |        | 31.4                      | 32.71  | 35.42  | 39.33  | 42.13  | 60.38  | 74.06  |
| OpenCV        | Indoor    |        | 515.1                     | 623.9  | 659.8  | 664.4  | 706.3  | 759.5  | 846.3  |

In summary, the target detection speed was greatly improved by this research for all cases. Especially when no target was available in view, the response speed was increased to over 16 times that of OpenCV algorithm. According to the best and worst processing speed cost record of 5 videos by this research, the processing speed between 14 and 58.4 fps (frame per second) was achieved when there was a target in view, and speed between 5.9 and 31.8 fps was achieved when there was no target available in the image. Considering the slowest speed record was only a few outliers, based on the 95<sup>th</sup> percentile time cost record of each experiment, it can be found that the processing speed of 95% of the frames of the videos with target captured indoors and outdoors was above or equal 32.6 and 26.8 fps, respectively, and the 95% speed records of the videos without target captured indoors and outdoors were equal or above 16.6 and 8.5 fps, respectively. The statistics about the reliability performance and the improved speed supports that the target detection algorithm of this research satisfies the requirement of a real-time positioning system. The improved speed was benefited from the pre-analysis algorithm proposed, which showed strong power to effectively eliminate great percentage of image areas as background for later target searching task. As indicated by the boxplot of Figure 3.13, over 84% image area was removed as background for most conditions.

### **3.7.2 Localization Experiment and Result**

An experiment was set up to evaluate the position measurement accuracy of proposed method as Figure 3.14 indicates. There were two targets. The left one in the picture served as the beacon for position calculation, and it was hang on a frame to make sure it was vertical to the ground. The right target served as the object to be located on the image based on its known world coordinates using Equation 3.25-3.27 in order to evaluate the accuracy of



proposed object tracking method. The inner corners' world coordinates related to the beacon target were calculated based on a single picture which contains both of their images. If the two targets are captured by a camera simultaneously, the relationship between two coordinate values of a physical point corresponding to two targets is expressed by Equation 3.28.

$$\begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \end{bmatrix} = R_1^{-1} R_2 \left( \begin{bmatrix} X_2 \\ Y_2 \\ Z_2 \end{bmatrix} + t_2 \right) - t_1 \quad (3.28)$$

Where the subscript represents which target's coordinate system it relates to. X, Y, Z are the coordinate value. R and t are the rotation matrix and translation vector related to the coordinate system of camera and corresponding target. As the position of the inner corners related to their own target is known, its coordinates related to the other target appears on a same image simultaneously can be calculated by Equation 3.28.

Having the right target's inner corners' position related to the beacon, the program of this research calculated their pixels in the image indirectly through analyzing the pose of the beacon in the view by Equation 3.25-3.27, and the result was compared with the those achieved by the inner corners directly extracted by the target pattern detection algorithm to determine the accuracy.



Figure 3.14. The setup of localization experiment

The bottom of the beacon target was around 1.5 m above the ground, and the camera was mounted on a tripod with the height of around 0.5 m. Square grids with side length of two feet (60.96 cm) were defined on the ground. And 88 testing points, which were the corners of the grid, were picked in the  $4.27 \times 7.27$  m area in front of the beacon target to place the camera for location measurement experiment. For each testing point, 20 images were captured to provide 20 position measurements of camera related to the beacon. The localization error was achieved by calculating the difference between the average positioning solution and the true value at each testing point, and the result is shown in Table 3.2, in which the X and Z values are the camera's true position related to the world coordinate

system defined by the beacon target. For better visual illustration, Figure 3.15 was created to show the localization error changes versus the camera position related to beacon.

The error increases when the camera is farther from the beacon because the lower resolution of target resulted from its smaller image size. When the Z coordinate of camera was smaller than three meters, the error was less than 17 mm. For the Z value over six meters, there is an obvious tendency of increased error. And the maximum positioning error was 65 mm, which was gotten where the camera was 727 cm away from the target at the Z direction.

Table 3.3 and Figure 3.16 provide the standard deviation of the position measurement result at each testing point. The tendency that the standard deviation increased along with the Z coordinate of camera can be observed. For the Z value less than or equal to 544 cm, the standard deviation was smaller than 20mm. The maximum standard deviation, 61.8 mm, was gotten at the last row, whose Z coordinate was 727 cm.

Table 3.2. Localization error (mm)

| Z (mm) \ X (mm) | -1839 | -1230 | -620 | -10 | 599 | 1209 | 1819 | 2428 |
|-----------------|-------|-------|------|-----|-----|------|------|------|
| 1169            | 4     | 2     | 3    | 8   | 8   | 9    | 3    | 5    |
| 1779            | 8     | 1     | 4    | 10  | 11  | 9    | 10   | 8    |
| 2388            | 8     | 2     | 4    | 13  | 11  | 16   | 13   | 10   |
| 2998            | 15    | 16    | 5    | 16  | 16  | 8    | 17   | 14   |
| 3608            | 7     | 16    | 11   | 9   | 5   | 24   | 16   | 24   |
| 4217            | 23    | 22    | 20   | 14  | 30  | 29   | 20   | 13   |
| 4827            | 25    | 34    | 11   | 5   | 40  | 25   | 7    | 13   |
| 5436            | 4     | 31    | 52   | 27  | 62  | 20   | 34   | 10   |
| 6046            | 37    | 44    | 17   | 9   | 46  | 30   | 38   | 33   |
| 6656            | 32    | 22    | 44   | 3   | 26  | 41   | 37   | 60   |
| 7265            | 23    | 16    | 32   | 65  | 65  | 64   | 61   | 30   |

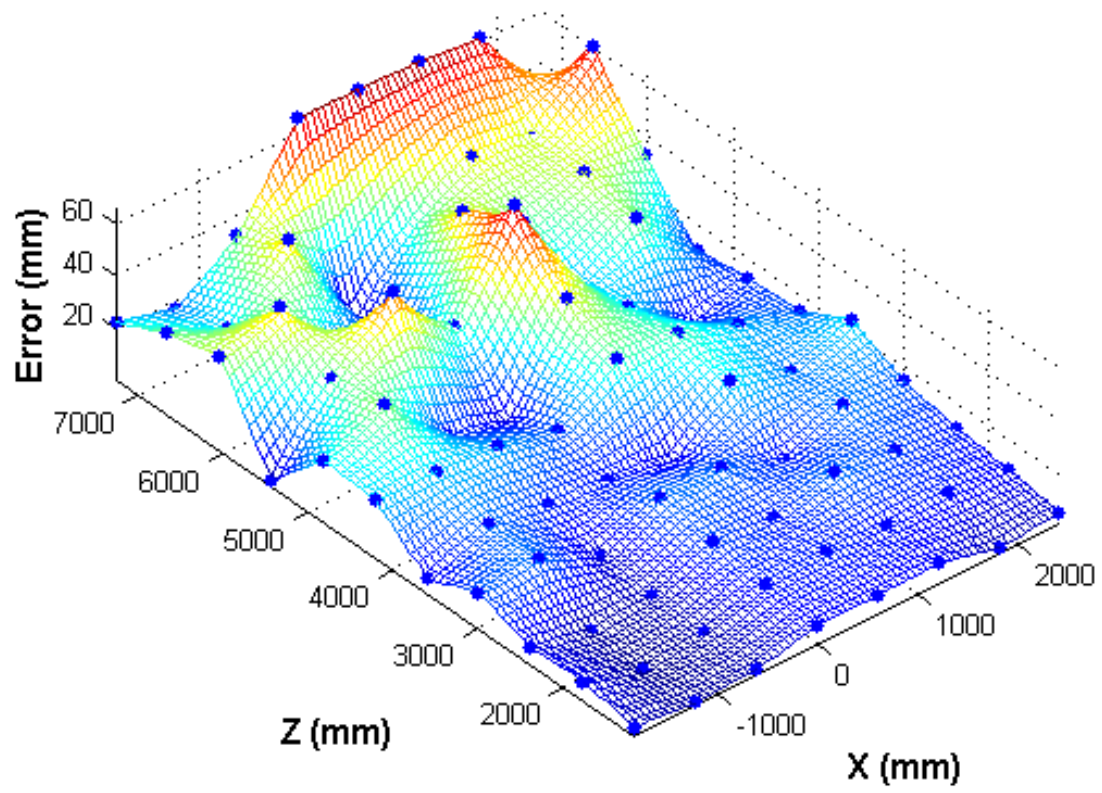


Figure 3.15. Location error map

Table 3.3. Standard deviation of localization result (mm)

| Z (mm) \ X (mm) | -1839 | -1230 | -620 | -10  | 599  | 1209 | 1819 | 2428 |
|-----------------|-------|-------|------|------|------|------|------|------|
| 1169            | 0.9   | 0.5   | 0.5  | 0.5  | 0.6  | 0.8  | 1.4  | 1.3  |
| 1779            | 1     | 1.1   | 1.3  | 0.9  | 1    | 1.1  | 1.2  | 1.2  |
| 2388            | 1.5   | 1.5   | 1.6  | 1.7  | 2.2  | 2.4  | 1.7  | 2    |
| 2998            | 2.7   | 2.9   | 2.9  | 2.5  | 2.2  | 3.4  | 3.2  | 2.8  |
| 3608            | 3.5   | 3.8   | 4.3  | 4.4  | 6.3  | 3.9  | 3.6  | 3.3  |
| 4217            | 4.5   | 4.9   | 6.6  | 7.3  | 7    | 4.9  | 4.1  | 3.8  |
| 4827            | 6.6   | 11.2  | 8.5  | 10.3 | 12.5 | 9.3  | 6.8  | 5.5  |
| 5436            | 7.1   | 10.7  | 15.1 | 13.4 | 19.1 | 14   | 6.4  | 10.7 |
| 6046            | 11.8  | 12.9  | 16.5 | 17.9 | 23.1 | 17.7 | 14.4 | 12.4 |
| 6656            | 9.9   | 15.1  | 25.3 | 29.7 | 22.1 | 21.3 | 12.7 | 18.8 |
| 7265            | 15.6  | 14.7  | 35   | 52.7 | 61.8 | 41.4 | 36.1 | 29.2 |

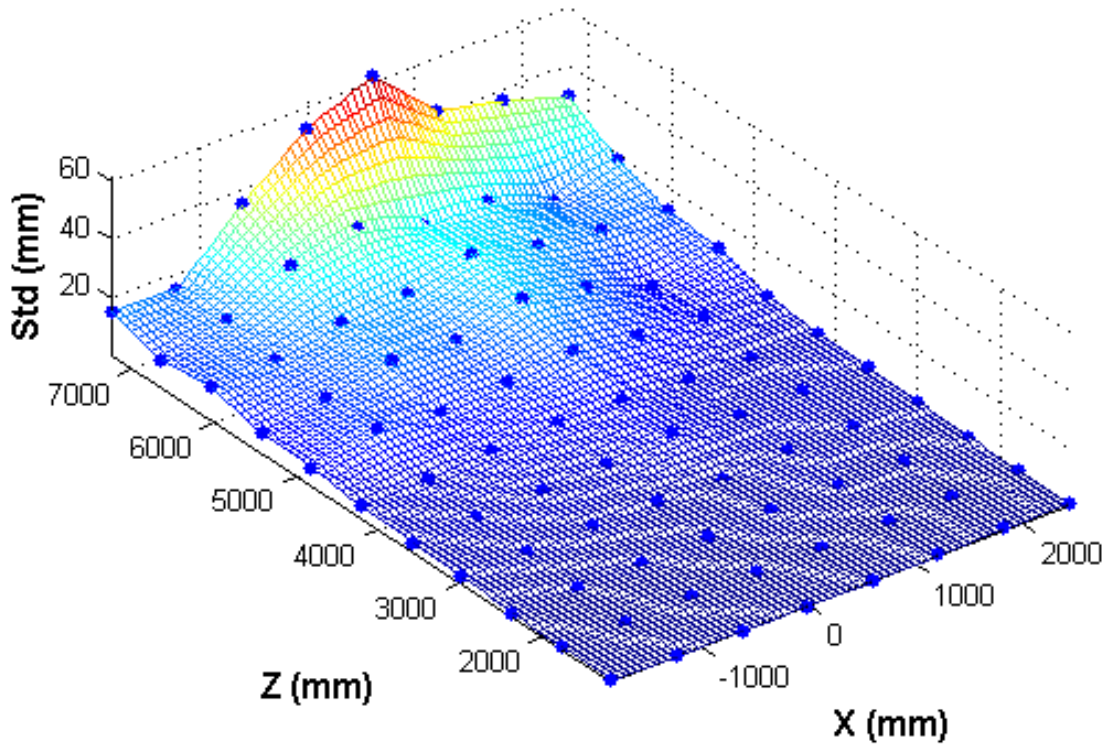


Figure 3.16. Standard deviation of positioning result

Table 3.4 gives the error result of locating the corners of the target other than beacon with known world coordinates in the image based on Equation 3.25-3.27. It is the mean difference between the corners' pixels calculated based on the world coordinates and those directly extracted by target detection algorithm at each testing point. The error was smaller than one pixel at most testing points, and maximum error was 3.8 pixels. The maximum standard deviation of the error at all testing point was 0.938 pixels. This experiment result shows that the objects with known position related to the beacon target can be reliably and precisely found in the image based on the proposed method. It is promising to store the button, handle, and other operable objects in the database and guide the robot to locate them in the image using this technique.

Table 3.4. Corner positioning error (pixel)

| Z (mm) \ X (mm) | -1839 | -1230 | -620  | -10   | 599   | 1209  | 1819  | 2428  |
|-----------------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1169            | 0.777 | 0.895 | N/A   | N/A   | 3.751 | 2.109 | 1.555 | 0.578 |
| 1779            | 0.527 | 0.713 | 0.821 | 0.776 | 1.852 | 1.235 | 0.95  | 0.829 |
| 2388            | 0.624 | 0.436 | 0.41  | 0.23  | 0.987 | 0.733 | 0.611 | 0.709 |
| 2998            | 0.488 | 0.458 | 0.343 | 0.323 | 0.857 | 0.744 | 0.528 | 0.415 |
| 3608            | 0.393 | 0.61  | 0.574 | 0.373 | 0.541 | 0.707 | 0.528 | 0.436 |
| 4217            | 0.5   | 0.288 | 0.399 | 0.401 | 0.6   | 0.297 | 0.264 | 0.334 |
| 4827            | 0.415 | 0.374 | 0.309 | 0.337 | 0.78  | 0.659 | 0.345 | 0.399 |
| 5436            | 0.553 | 0.393 | 0.241 | 0.992 | 0.585 | 0.431 | 0.328 | 0.3   |
| 6046            | 0.61  | 0.294 | 0.599 | 0.318 | 0.622 | 0.264 | 0.38  | 0.468 |
| 6656            | 0.719 | 0.222 | 0.348 | 0.254 | 0.564 | 0.293 | 0.29  | 0.597 |
| 7265            | 0.226 | 0.207 | 0.301 | 0.664 | 1.15  | 0.408 | 0.265 | 0.385 |



### 3.7.3 Attitude Estimation Experiment and Result

To examine the rotation measurement accuracy of this research, the X3M multi-axis absolute MEMS (micro electro-mechanical system) inclinometer (US Digital, Vancouver, Washington, USA) was applied as the reference rotation measurement sensor. Figure 17 shows the sensor and its three axes. With Axis 2 or Axis 0 parallel to gravity, the other two axes will give the rotation angles about them, and the one parallel to gravity will be invalid. The angular error is  $\pm 1.2^\circ$ .

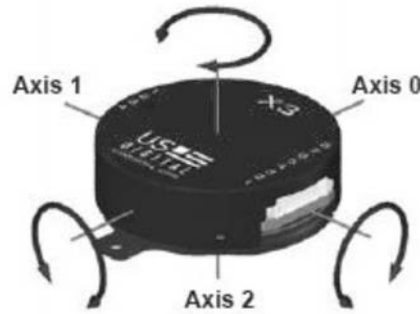


Figure 3.17. X3M multi-axis absolute MEMS inclinometer and its three axis

The inclinometer was mounted on the target with axis 0 and axis 1 parallel to axis X and Y of the target, respectively as shown by Figure 3.18. This research defined the rotation angle around X, Y, and Z axis of the target as  $\alpha$ ,  $\beta$ , and  $\gamma$  respectively. With the Z axis parallel to gravity, which means the target is placed horizontally, as Figure 3.19 shows, the X3M inclinometer can provide rotation related to its axis 0 and axis 1 as reference to evaluate the measurement of angle  $\alpha$  and  $\beta$  calculated by this research.

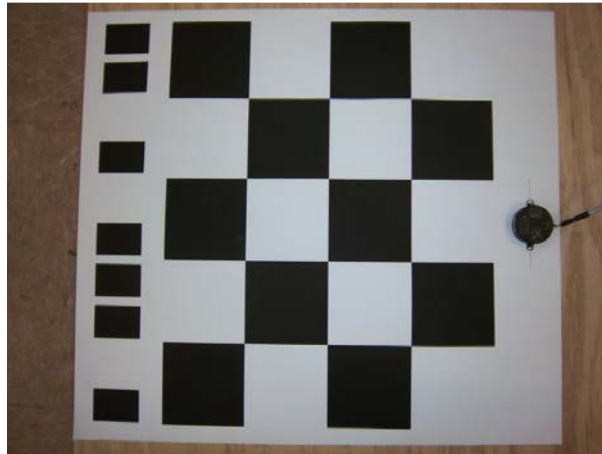


Figure 3.18. Method for mounting the inclinometer on the target



Figure 3.19. The setup of the angle measurement experiment with the Z axis of the target parallel to gravity

When placing the target on the horizontal ground, the reading of angle  $\alpha$  and  $\beta$  from the inclinometer was zero. The rotation matrix related the target and camera at this moment, which was called time zero and was stored as  $R_0$ . Then the target was moved and rotated, and the rotation matrix related target and camera at time  $t$  was named  $R_t$ . As the camera was stationary for this experiment, the rotation matrix from the target coordinate system at time zero to the one at time  $t$  was  $R_t^{-1} R_0$ , from which the three rotation angles  $\alpha$ ,  $\beta$ , and  $\gamma$  can be extracted. The rotation angles  $\alpha$  and  $\beta$  were compared with the corresponding valid results from the X3M inclinometer. Figure 3.20 shows the rotation trajectories achieved by the two methods. The two trajectories match well. The maximum difference between two trajectories calculated by Equation 3.29 was  $2.744^\circ$ .

$$d = |\alpha_c - \alpha_i| + |\beta_c - \beta_i|, \quad (3.29)$$

where the subscript  $c$  stands for the angle measured by the camera, and subscript  $i$  means the angle reading from the inclinometer.

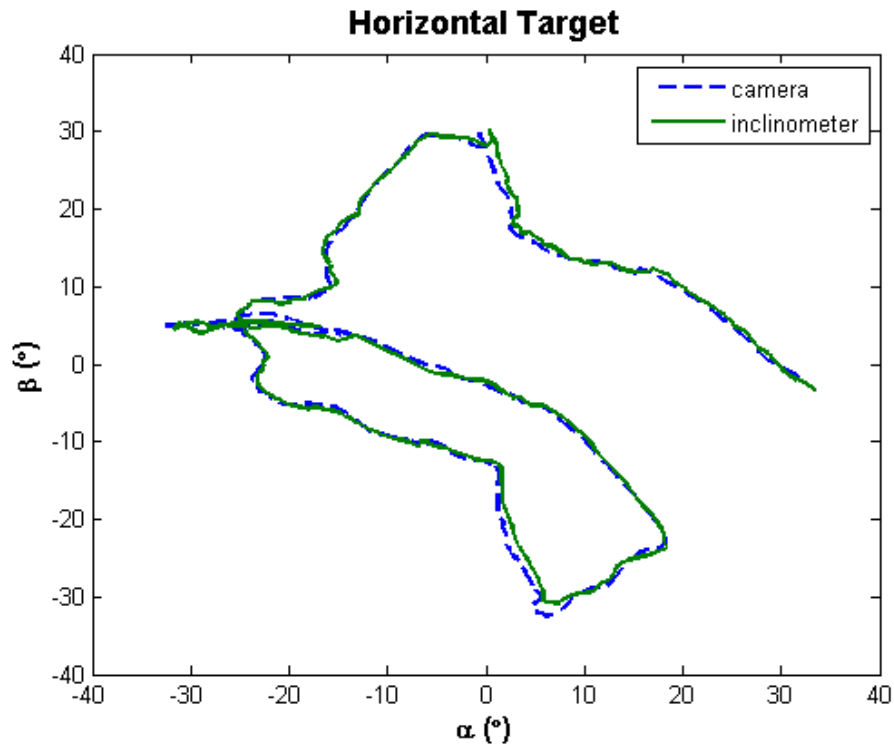


Figure 3.20. The angle measurement trajectory of two methods when the Z axis of the target is parallel to gravity

Moreover, with the X axis of the target coordinate system parallel to gravity as Figure 3.21 shows, the X3M inclinometer can provide a valid reading about angle  $\beta$  and  $\gamma$  to compare the result based on the camera. Similar to the above mentioned experiment, when placing the target vertically with its X axis parallel to gravity, the reading of angle  $\beta$  and  $\gamma$  from the inclinometer was zero, and the rotation matrix of the target at that moment, which was called time zero, was stored as  $R_0$ . Then the target was moved and rotated, and the angle  $\beta$  and  $\gamma$  were calculated using the same method of previous experiment and compared with the readings from the X3M inclinometer. Figure 3.21 shows their rotation trajectories, which

match well. The maximum difference between two trajectories calculated by Equation 3.30 was  $3.941^\circ$ .

$$d = |\beta_c - \beta_i| + |\gamma_c - \gamma_i| \quad (3.30)$$



Figure 3.21. The setup of the angle measurement experiment with the X axis of the target parallel to gravity

Considering the angular error of the inclinometer is  $\pm 1.2^\circ$ , the angle measurement difference between the two methods in these two experiments is reasonable.

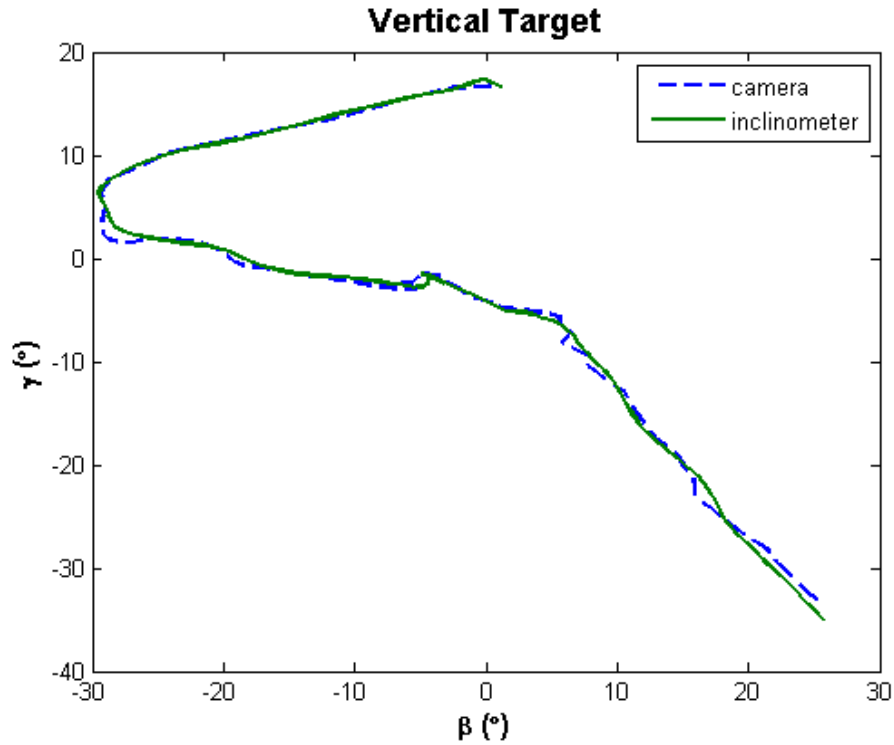


Figure 3.22. The angle measurement trajectory of two methods when the X axis of the target is parallel to gravity

### 3.8 Conclusions

This research developed a real-time detection algorithm of a chessboard pattern target, and proposed a motion and attitude estimation system based on analyzing the target image. From the result of this research, it can be concluded that this system has promising potential for indoor and outdoor position and attitude estimation application. First, the target detection algorithm of this research was effective in identifying most of the background image areas, and therefore greatly reduced the target search area. This significantly helped to decrease the target detection time regardless of whether there was a target in the image view for both indoor and outdoor conditions. Second, the target detection algorithm achieved over 97.49% accuracy for both indoor and outdoor experiments, showing the great reliability and robustness to different illumination conditions. Third, the system demonstrated centimeter-

level accuracy for position estimation within a  $4.27 \times 7.27$  m testing area in front of the beacon, and it achieved subdegree level accuracy for attitude estimation. Fourth, the proposed positioning system shows great scalability. The proposed ID was effective to identify different targets, and the manufacturing cost of the target is low. Therefore, it is feasible to deploy a large number of beacons to provide position and attitude estimation service in large areas.

The high performance of this position and attitude estimation solution enable tracking the 3D camera with high accuracy for the 3D reconstruction based phenotyping system introduced in the next chapter, and it helps to align different 3D image views precisely for the 3D reconstruction of a plant.

### 3.9 Reference

- Addlesee, M., R. Curwen, S. Hodges, J. Newman, P. Steggles, A. Ward, and A. Hopper. 2001. Implementing a sentient computing system. *Computer* 34(8):50-56.
- Alenya, G., B. Dellen, and C. Torras. 2011. 3D modelling of leaves from color and tof data for robotized plant measuring. In 2011 IEEE International Conference on Robotics and Automation. Shanghai, China: IEEE.
- Cho, S., J. Park, and J. Lee. 2012. A dynamic localization algorithm for a high-speed mobile robot using indoor GPS. *Robotica* 30:681-690.
- Choliz, J., A. Hernandez, and A. Valdovinos. 2011. A Framework for UWB-Based Communication and Location Tracking Systems for Wireless Sensor Networks. *Sensors* 11(9):9045-9068.
- de la Escalera, A., and J. Maria Armingol. 2010. Automatic Chessboard Detection for Intrinsic and Extrinsic Camera Parameter Calibration. *Sensors* 10(3):2027-2044.
- Ha, J.-E. 2009. Automatic detection of chessboard and its applications. *Optical Engineering* 48(6).

- Ladd, A. M., K. E. Bekris, A. P. Rudys, D. S. Wallach, and L. E. Kavraki. 2004. On the feasibility of using wireless ethernet for indoor localization. *IEEE Transactions on Robotics and Automation* 20(3):555-559.
- Munoz-Organero, M., P. J. Munoz-Merino, and C. Delgado Kloos. 2012. Using bluetooth to implement a pervasive indoor positioning system with minimal requirements at the application level. *Mobile Information Systems* 8(1):73-82.
- Ni, L. M., Y. H. Liu, Y. C. Lau, and A. P. Patil. 2004. LANDMARC: Indoor location sensing using active RFID. *Wireless Networks* 10(6):701-710.
- Prieto, J. C., A. R. Jiménez, J. Guevara, J. L. Ealo, F. Seco, J. O. Roa, and F. Ramos. 2009. Performance evaluation of 3D-LOCUS advanced acoustic LPS. *IEEE Transactions on Instrumentation and Measurement* 58(8):2385-2395.
- Rusu, R. B., N. Blodow, Z. C. Marton, and M. Beetz. 2008. Aligning point cloud views using persistent feature histograms. In *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE.
- Wang, Z., W. Wu, X. Xu, and D. Xue. 2007. Recognition and location of the internal corners of planar checkerboard calibration pattern image. *Applied Mathematics and Computation* 185(2):894-906.
- Yu, C., and Q. Peng. 2006. Robust recognition of checkerboard pattern for camera calibration. *Optical Engineering* 45(9):093201.
- Zhang, Z. Y. 2000. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(11):1330-1334.



## CHAPTER 4. DEVELOPMENT OF A 3D RECONSTRUCTION BASED PLANT PHENOTYPING SYSTEM

### 4.1 Abstract

A holographic three-dimensional (3D) plant phenotyping system was developed for plant 3D reconstruction and morphological traits characterization. Corn plant seedlings were used as research objects for algorithm development and validation. In this application, precise alignment of multiple 3D views generated by a 3D time-of-flight (ToF) sensor is critical to the holographic 3D reconstruction of a plant. Previous research indicated that there is a strong need for high-throughput, high-accuracy, and low-cost 3D plant reconstruction and trait characterization phenotyping systems. This research contributed a 3D reconstruction system of a plant by innovatively integrating a low-cost 2D camera, a low-cost 3D ToF camera, and a chessboard pattern beacon array to track the position and attitude of the 3D ToF sensor and thus accomplished precise 3D point cloud registration over multiple views. Specifically, beacon target detection, camera pose tracking, and spatial relationship calibration between 2D and 3D cameras algorithms were developed as a low-cost but high-performance 3D reconstruction solution. A plant analysis algorithm in a holographic 3D space was developed to extract the morphological trait parameters of the plants by analyzing its 3D reconstruction. The phenotypical data obtained by this novel holographic 3D reconstruction based phenotyping system were validated by the experimental data generated by instrument and manual measurement; and the results demonstrated that the developed phenotyping system has achieved satisfactory measurement accuracy with a processing time of less than 5 seconds per plant on average.

## 4.2 Introduction

A plant's phenotype is the result of the dynamic interaction of the plants' genotype and environment. Phenotypic parameters, such as leaf size, crop height, cereal yield, photosynthesis rate, nutrient intake rate, resistance to disease and drought, etc. are important for breeders (Foundation and Mcb, 2011). Understanding the linkage between a particular genotype and a specific phenotypic parameter is a core goal of modern biology; however, it is generally difficult due to the large number of genes and the interaction with complex and changeable environmental influences (Foundation and Mcb, 2011).

The fast development technology enables rapid genomes sequencing at steadily declining costs and rapidly increasing speed. Scientists have collected abundant information of plant genotype due to the recent revolution of genomic technologies (Foundation and Mcb, 2011). The genomic information could not be fully capitalized without correct linkage between genotype and phenotype (Cobb et al., 2013; Foundation and Mcb, 2011; Furbank and Tester, 2011).

Phenomics is the science of large-scale phenotypic data collection and analysis to reveal the relationship between phenotypic feature and genotype (Allen et al., 2010; Foundation and Mcb, 2011; Heffner et al., 2011; Lu et al., 2011; Nichols et al., 2011; Speliotes et al., 2010; Winzeler et al., 1999). To extract and quantify sophisticated phenotypic features on a large scale is challenging. Traditional phenotyping is labor intensive, expensive, and destructive (Furbank and Tester, 2011). Imaging based systems provide a remote and noninvasive method to capture not only the morphological phenotype data but also the physiological status for the plant (Foundation and Mcb, 2011).

A 2D imaging based phenotyping system has achieved a degree of development and commercial success. Researchers have reported extracting the projected leaf area from 2D color images to estimate the growth rate and drought tolerance for rosette plants such as *Arabidopsis* (Granier et al., 2006; Walter et al., 2007). LemnaTec Scanalyser (LemnaTec GmbH, Germany) is a commercial product that can estimate the morphological features of a plant, such as plant height and canopy's diameter, roundness, circumference, etc., based on top view or side view 2D images. Chlorophyll fluorescence is another 2D imaging technique. It is effective to estimate photosynthetic responses in various conditions such as drought stress, cold, heat, and ultraviolet light (Jansen et al., 2009), and it also successfully detected pathogen infections that affect photosynthesis of a plant (Chaerle et al., 2009; Scholes and Rolfe, 2009; Swarbrick et al., 2006). Moreover, thermography could estimate the transpiration of a plant by distinguishing by measure the temperature difference between leaf and backgrounds, and it has been used to phenotype plant traits and disease related transpiration such as root fungal pathogen infection (Sirault et al., 2009).

Because of the inherent limitation of the above mentioned 2D imaging technique, their applications are mostly limited to plants with a simple canopy, and their performance drops greatly for plants with a complex canopy when occlusion becomes problematic (Furbank and Tester, 2011). Moreover, this problem calls for the use of a 3D reconstruction of the plant (Bellasio et al., 2012; Fiorani et al., 2012; Furbank and Tester, 2011).

Although the development of advanced 3D sensors, including active stereo vision, Lidar, and 3D ToF cameras, makes high-quality 3D spatial data available, 3D reconstruction algorithms for crop plants phenotyping are still primitive because 3D reconstruction requires aligning different 3D data views of the plant precisely into a complete 3D model of the plant.

The alignment of 3D data views is called 3D registration. Alenya et al. (2011) reported a system that applied an expensive robotic arm to control and track the position and attitude of a 3D camera for 3D registration. This system had high accuracy and reliability for 3D registration by a high-quality robotic arm, which however lead to high equipment cost. On the contrary, Rusu et al. (2008) developed a software based 3D registration method. Their algorithm estimated the 3D feature histogram of each point of every 3D view and looked for the key points whose feature histogram is unique. By searching for the correspondence of key points between different views, the relationship between different views can be found for 3D registration. This software based approach has a low system cost. However, it is calculation intensive and thus requires a long processing time, making it less favorable for large-scale phenotyping applications.

This research aims to apply a 3D ToF camera to capture multiview images of the maize plant, and to derive algorithms for a 3D model reconstruction, leaf and stem segmentation, as well as leaf phenotypic parameters quantification. Ultimately, this research aims to provide new imaging and an image processing system as tools to build a low-cost, accurate, robust maize phenotyping system.

### 4.3 Experimental Design

This study developed a 3D reconstruction based phenotyping system. It is able to reconstruct the 3D model of the plant and measure some of its physical characteristics.

Three corn plants were used to test the 3D reconstruction based phenotyping system in this research. They were at growth stage V5 and were about 0.5 m high (Figure 4.1). Their 3D models were reconstructed by this system. The corn plants' physical parameters,

including height, width, length, and area of each leaf, are automatically estimated by analyzing their 3D models. The leaf height in this study is the height of its collar related to the soil surface of its pot.

In order to evaluate the accuracy of this 3D reconstruction based phenotyping system, the collar height, length, and width of each leaf of the corn plant were measured manually as the ground true value. The area of each leaf was measured using Leaf Area Meter (LI-3100C Area Meter, LI-COR Corporate, Lincoln, Nebraska, USA).



Figure 4.1. Corn plant used in this phenotyping research

#### 4.4 Sensors

The key goal of this 3D reconstruction based phenotyping research is 3D plant reconstruction. In order to get the complete 3D model, 3D point cloud data views of the plant should be captured and aligned precisely, for three reasons. First, because of the occlusion of the canopy, the 3D camera could hardly observe every part of the plant at a single viewpoint. Second, in order to get a high-quality 3D data view of the leaf, the imaging sensor should

observe the leaf surface along its normal direction. As different parts of different leaves have different surface direction, the viewpoint of the 3D camera should be adjusted to get the good quality 3D data correspondingly. Third, the resolution of the ToF camera is only  $120 \times 165$  pixels. It is too low to capture the details of the whole plant in a single view. Therefore, it is necessary to capture the 3D dense point cloud of different parts of the plant at a close viewpoint and to align them together to reconstruct a complete 3D model.

This research applied both 2D and 3D imaging sensors to develop a 3D reconstruction based phenotyping system. As Figure 4.2 (a) indicates, two 2D web cameras were mounted beside a 3D ToF camera as the data collection system. However, one of the web cameras (web camera 2 in Figure 4.2) was not used for this research yet; the 2D camera mentioned in this research only means web camera 1. As Figure 4.5 shows, the 3D imaging sensor was used to capture the point cloud data of different parts of the corn plant, and the 2D web camera was used to capture the image of chessboard pattern targets array simultaneously. By analyzing the 2D images of the chessboard pattern beacon, the 2D camera's pose related to the beacon can be estimated as discussed in chapter 3. The position and attitude of the 3D sensor can be derived from the calibrated pose relationship between the 2D camera and the 3D camera. Based on the 3D sensor's pose corresponding to each point cloud data view, different views can be aligned precisely to reconstruct the complete 3D model. Therefore, the relationship between the 2D and 3D camera is critical for this research. Figure 4.2 (b) provides the manual measurement result of the translation relationship between the imaging sensors. Because it was difficult to find the exact origin of each camera's coordinate system, the manual measurement could only get the approximate range of each parameter.

The 3D camera used in this study is the PMD Camboard nano (Pmdtechnologies, Germany), which is a ToF camera, as Figure 4.2 (a) shows. Its resolution is  $120 \times 165$  pixels, and its depth measurement receptivity is 5 mm. Its standard depth measurement range is 0 to 2 m, according to the datasheet; however, our experiment indicated the best working distance between the object and sensor was between 0.2 and 0.5 m in this research. Because the 3D camera can be placed as close as around 0.20 m to the observed object, it can provide dense point cloud data with detailed information for parts of plant at close viewpoint despite its low resolution. Dense point cloud data views were available to reconstruct the complete 3D model with detailed information.

The 2D camera used to track the 3D camera's pose is the Logitech HD Webcam C310. The working principal to estimate the pose of the camera is similar to that described in chapter 3. The model of the 2D camera used here is the same as that used chapter 3. It features a wide focus range, fixed focal length of 4.4 mm,  $60^\circ$  field of view, and  $1280 \times 720$  pixel resolution. The distortion effect is satisfactory according to the previous study. All of these features indicate it is a good sensor for the 2D imaging based position and attitude system. Before the 2D camera was used in the experiment, it was calibrated to get its intrinsic matrix and distortion vector using the method introduced in chapter 3.



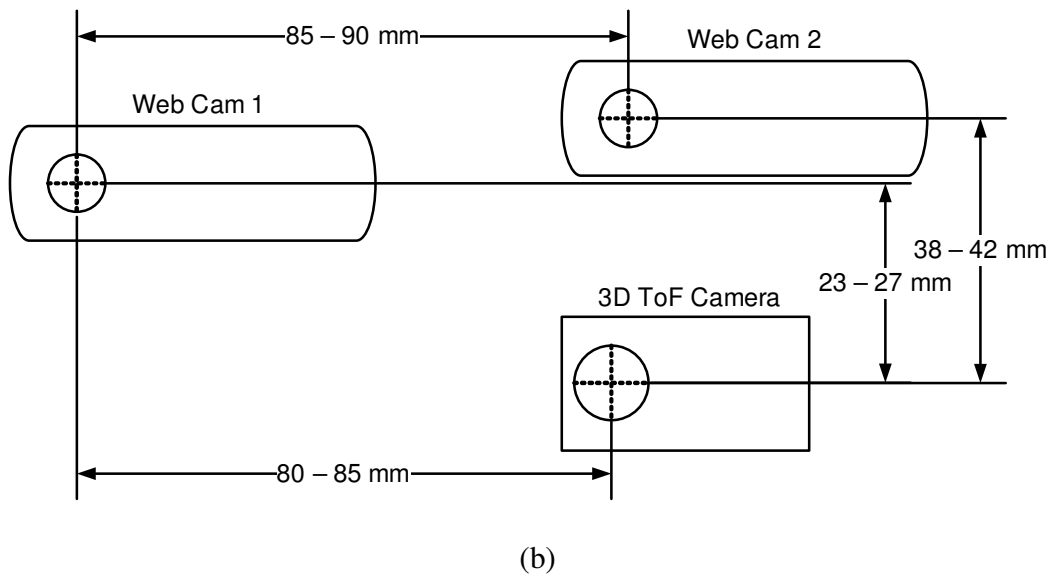
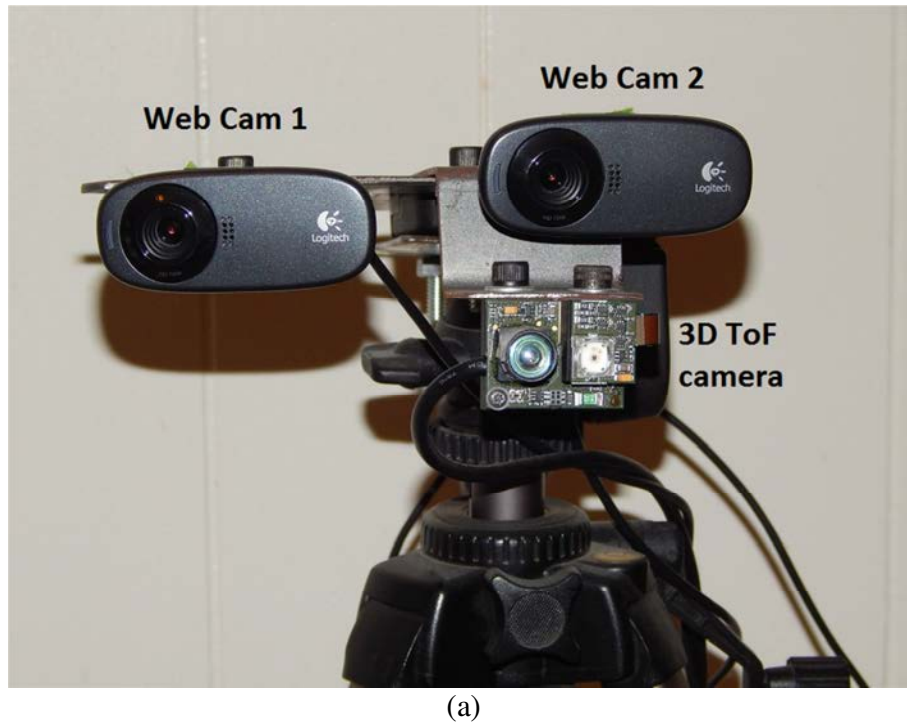


Figure 4.2. Data collection system built for this 3D reconstruction based phenotyping research: (a) picture of system and (b) translation relationship between sensors

#### 4.5 Camera Posture Tracking Infrastructure

A chessboard pattern array was designed to be used as the infrastructure to track the position and attitude of the 2D and 3D cameras. As chapter 3 discussed, the 2D camera could estimate its pose as long as it captured a chessboard pattern beacon. However, a single chessboard beacon was not enough for this research. To capture the 3D point cloud of the plant at different viewpoints, the pose of the imaging sensors changes. To ensure at least one beacon was visible to the 2D camera at different viewpoints, a beacon array, which contained seven rows by five columns of chessboard pattern beacons, was designed as the infrastructure of the pose estimation system, as Figure 4.3 shows. Figure 3.1 shows a sample beacon unit. By placing the corn plant pot about 0.7 m above the center of the beacon array and holding the data collection system as Figure 4.4 indicates. The 2D camera can capture at least one chessboard pattern beacon of the target array when the 3D imaging sensors are observing the corn plant at different viewpoints.

The chessboard beacon target defined the world coordinate system of the research platform. As Figure 4.3 shows, the origin of the world coordinate system is the bottom-left inner corner of the bottom-left beacon in the target array, and its X and Y axis are parallel to the up and right direction of the target array, respectively. In addition, the direction of the Z axis is vertically going inside the target array. In this study, the position and attitude estimation result of the 2D and 3D cameras was related to the world coordinate system.

All of the beacons in the target array have identical chessboard pattern design, and everyone consists of  $5 \times 4$  squares, as Figure 4.3 shows. The side of each square was 52.36 mm. The translation relationship between two neighbor beacons at the same row was

$[292.72, 0, 0]^T$  mm, and that between the two neighbor beacons at the same column was  $[0, 303.43, 0]^T$  mm.

Each chessboard pattern beacon had a column of small rectangles on the right, as Figure 4.3 shows. The rectangles were the ID of the beacon; IDs were used to inform the system which beacon was captured by the 2D camera. By applying IDs, the system can assign the right world coordinate value to each inner corner of the detected beacon based on its position in the beacon array according to the ID, and then the system could calculate the position and attitude of the 2D camera related to the world coordinate system using the techniques introduced in chapter 3.

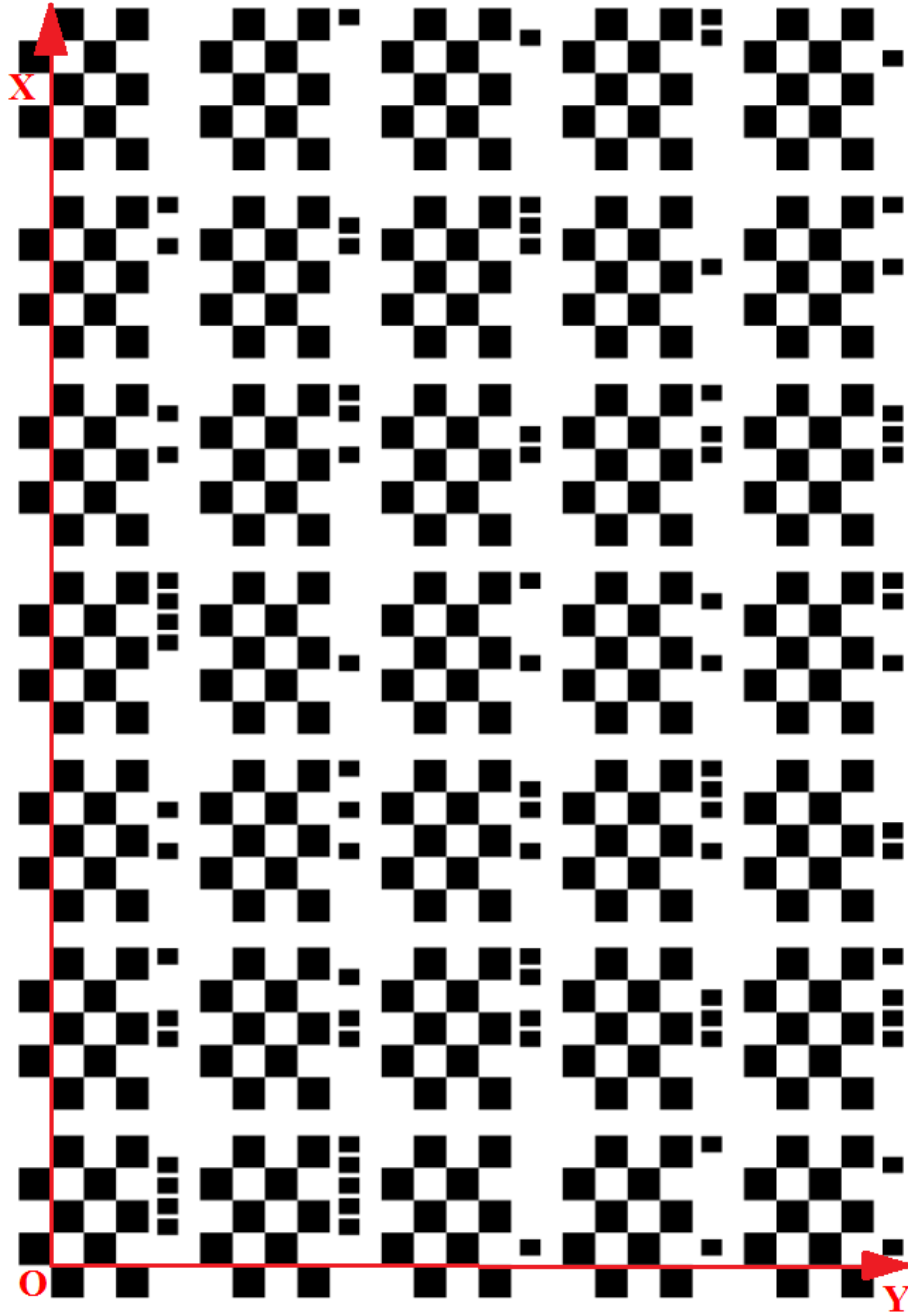


Figure 4.3. World coordinate system and chessboard pattern array used as the camera's pose estimation infrastructure

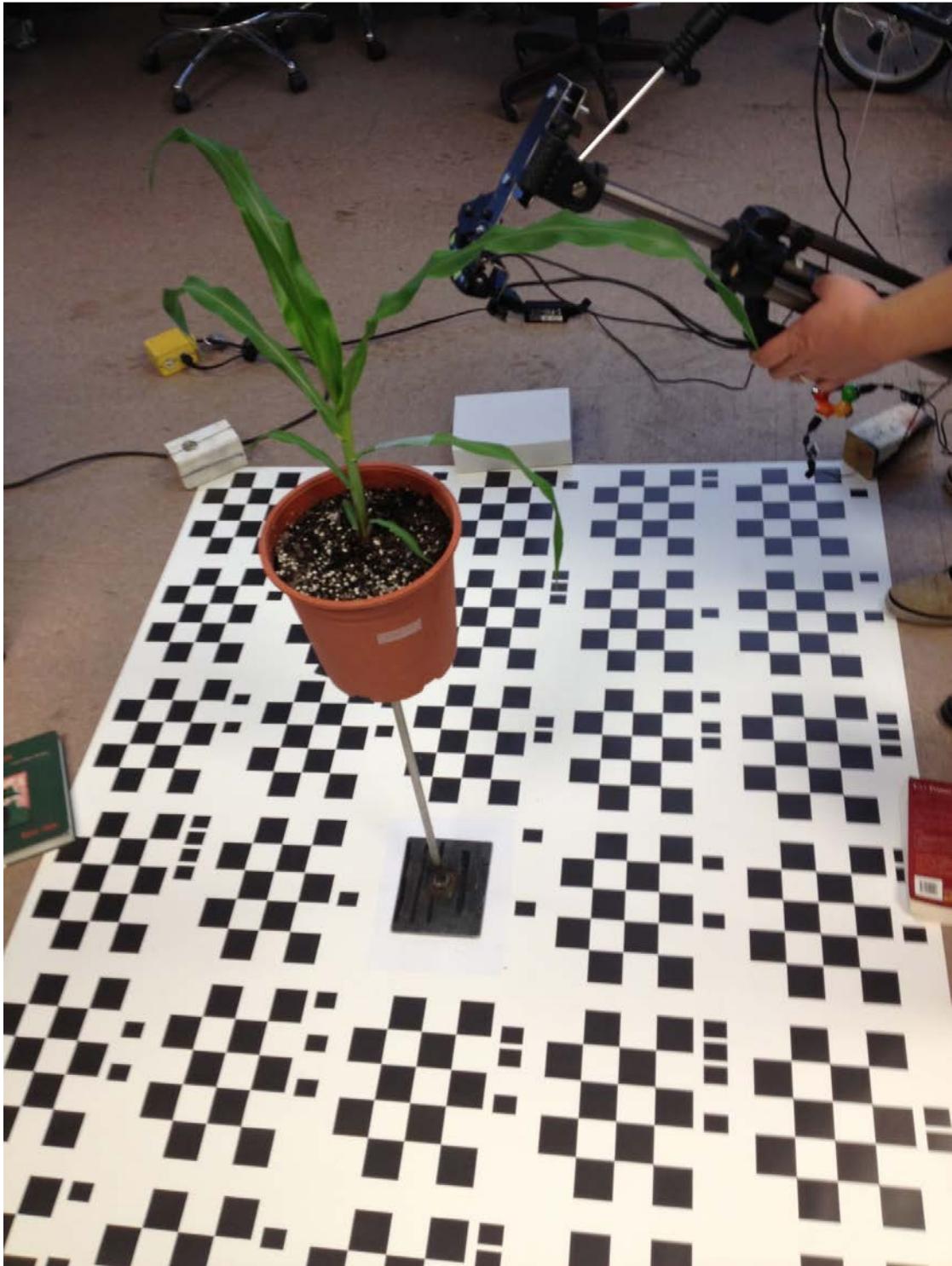


Figure 4.4. Infrastructure setup

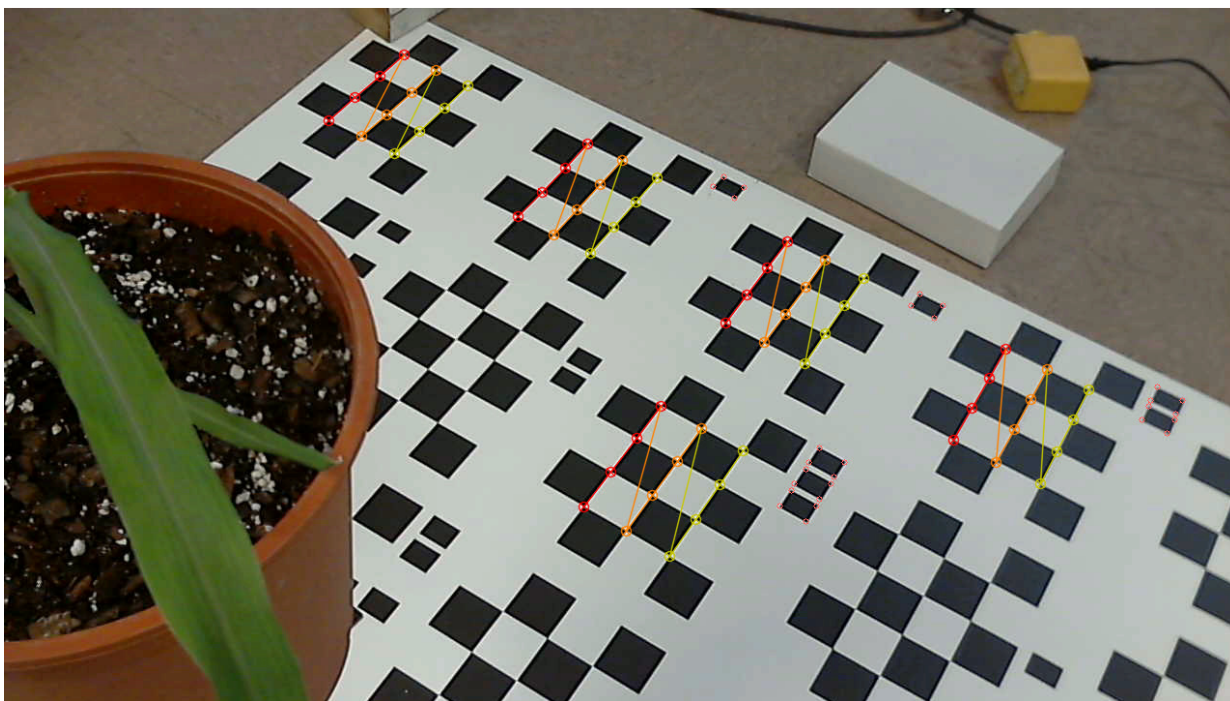
#### 4.6 Position and Attitude Estimation of 3D Camera

To reconstruct the complete 3D model of the corn plant, the position and attitude of the 3D camera corresponding to every point cloud data view is required. The pose of the 3D camera was derived from the pose of the 2D camera plus the position and attitude relationship between the two cameras. Therefore, the pose estimation of the 2D camera and the calibration of the relationship between the 2D and 3D cameras are critical.

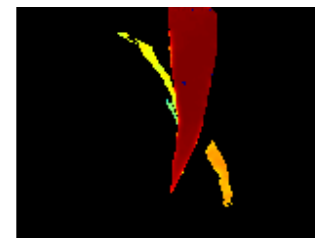
##### 4.6.1 Position and Attitude Estimation of 2D Camera

When the 3D ToF camera was collecting the point cloud data of the corn plant, the 2D imaging sensor beside it captured the image of the beacon targets simultaneously. Using the techniques introduced in chapter 3, the beacons were detected from the 2D image, and their IDs were recognized, as Fig 4.5 (a) shows. Based on the ID information and pre-knowledge of the chessboard pattern array, the system achieved the world coordinates of the inner corners of the detected beacons, and applied the coordinates to calculate the position and attitude of the 2D camera related to the world coordinate system by using the method introduced in chapter 3.

The 2D camera can capture one or more beacons during the data collection process depending on the viewpoint. Although one beacon is enough to estimate the 2D camera's position and attitude, more beacons can improve the measurement accuracy.



(a)



(b)



(c)

Figure 4.5. Images collected by 2D and 3D camera: (a) chessboard pattern beacon detected by 2D camera, (b) depth image collected by 3D ToF camera, and (c) intensity image captured by 3D ToF camera

#### 4.6.2 Calibration between 2D and 3D Cameras

To calibrate the relationship between the 2D and 3D cameras, two chessboard pattern targets were used. The design of the data collection system made it impossible for the 2D camera and the 3D camera to observe the same chessboard target simultaneously. To solve this problem, two chessboard targets were used and assembled side by side, as Figure 4.6 (a) indicates. The calibration procedure is discussed next.

First, the relationship between the two chessboard targets was calibrated. The data collection system was moved to a relatively farther viewpoint A, and the 2D camera captured an image that contains two targets, as Figure 4.6 (a) shows. The system detected two targets and estimated the relationship between the coordinate system of the 2D camera and each target separately, based on the method introduced in chapter 3. The rotation matrix and translation vector of the 2D camera's coordinate system related to the left target are represented with  $R_{LA}$  and  $t_{LA}$ , respectively, and those related to the right target are represented with  $R_{RA}$  and  $t_{RA}$ .  $Q_{2D}$  is the coordinate vector  $[x, y, z]^T$  of a point related to the 2D camera's coordinate system, and  $\widetilde{Q}_L$  and  $\widetilde{Q}_R$  are the coordinate vector of a point related to the coordinate system defined by the left target and right target, respectively. The relationship between  $Q_{2D}$ ,  $\widetilde{Q}_L$  and  $\widetilde{Q}_R$  of a same point can be expressed by Equation 4.1 and 4.2. In addition, the relationship between the coordinates systems defined by two targets can be expressed by Equation 4.3.

$$Q_{2D} = R_{LA}\widetilde{Q}_L + t_{LA} \quad (4.1)$$

$$Q_{2D} = R_{RA}\widetilde{Q}_R + t_{RA} \quad (4.2)$$

$$\widetilde{Q}_R = R_{RA}^{-1}(R_{LA}\widetilde{Q}_L + t_{LA} - t_{RA}) \quad (4.3)$$



Based on Equation 4.3, the rotation matrix  $R_{L2R}$  and the translation vector  $t_{L2R}$  of left target related to right target is expressed by Equation 4.4 and 4.5.

$$R_{L2R} = R_{RA}^{-1} R_{LA} \quad (4.4)$$

$$t_{L2R} = R_{RA}^{-1} (t_{LA} - t_{RA}) \quad (4.5)$$

Then Equation 4.3 can be written with other format as Equation 4.6 indicates.

$$\widetilde{Q}_R = R_{L2R} \widetilde{Q}_L + t_{L2R} \quad (4.6)$$

Second, the data collection system was moved closer to two targets, and the position of the data collection system at this time was viewpoint B. At viewpoint B, the 2D camera can only capture the image of the right target as Figure 4.6 (b) shows, while the 3D ToF camera gets the intensity image and 3D data of the left target as Figure 4.6 (c) and (d) show. By extracting the right chessboard target from the color image captured by the 2D camera as Figure 4.6 (b) indicates, the rotation matrix  $R_{RB}$  and the translation vector  $t_{RB}$  of the 2D camera at viewpoint B related to right target can be achieved. Moreover, the rotation matrix  $R_{L22D}$  and the translation vector  $t_{L22D}$  of the left target related to the 2D camera at viewpoint B can be derived based on Equation 4.3–4.6,  $R_{RB}$ , and  $t_{RB}$ . The derivation results are expressed by Equation 4.7 and 4.8.

$$R_{L22D} = R_{RB} R_{L2R} = R_{RB} R_{RA}^{-1} R_{LA} \quad (4.7)$$

$$t_{L22D} = R_{RB} R_{L2R} t_{L2R} + t_{RB} = R_{RB} R_{RA}^{-1} (t_{LA} - t_{RA}) + t_{RB} \quad (4.8)$$

Third, the relationship between the left target and the 3D camera at viewpoint B was estimated. The point cloud data of the chessboard target from the 3D camera was processed by linear regression algorithm to estimate the plane of the target board; and then the original

point cloud data were replaced with their projection points on the plane to reduce the measurement error of the 3D data. Additionally, the inner corners of the left chessboard target were extracted from the intensity image captured by the 3D camera, as Figure 4.6 (c) shows, to get the inner corners' coordinate value related to the 3D camera. The relationship between the coordinate systems of the left target and the 3D camera can be estimated based on two steps, as follows:

- 1) The target plane in the 3D camera's coordinate system was achieved in the plane regression calculation, and the target plane in its own coordinate system is  $Z = 0$ . The rotation matrix  $R_1$  between the target plane in the 3D camera's coordinate system and the plane  $Z = 0$  was estimated first. The normal direction of the target plane related to the coordinate system of the target is  $[0, 0, 1]^T$ , and the normal direction of the target plane related to the 3D camera's coordinate system is unit vector  $[a, b, c]^T$ . The value of  $[a, b, c]^T$  was achieved in the target plane regression step. The rotation matrix for the rotation from  $[0, 0, 1]^T$  to  $[a, b, c]^T$  is  $R_1$ . The corresponding rotation axis  $L$  is perpendicular to both  $[0, 0, 1]^T$  and  $[a, b, c]^T$ ; therefore,  $L$  is the cross product of two vectors, as Equation 4.9 indicates, and the rotation angle is expressed by Equation 4.10.

$$L = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \times \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} -b \\ a \\ 0 \end{bmatrix} \quad (4.9)$$

$$\theta = \arccos \left( \frac{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix}}{\sqrt{a^2 + b^2 + c^2}} \right) = \arccos(c) \quad (4.10)$$

The rotation matrix  $R_1$  can be derived by applying the Rodrigues' rotation formula with rotation axis  $L$ , and the rotation angle  $\theta$ ; the result is described by Equation 4.11.

$$R_1 = \begin{bmatrix} b^2 + a^2 \cos(\theta) & ba(\cos(\theta) - 1) & \sin(\theta)a \\ ba(\cos(\theta) - 1) & a^2 + b^2 \cos(\theta) & \sin(\theta)b \\ -\sin(\theta)a & -\sin(\theta)b & (b^2 + a^2) \cos(\theta) \end{bmatrix} \quad (4.11)$$

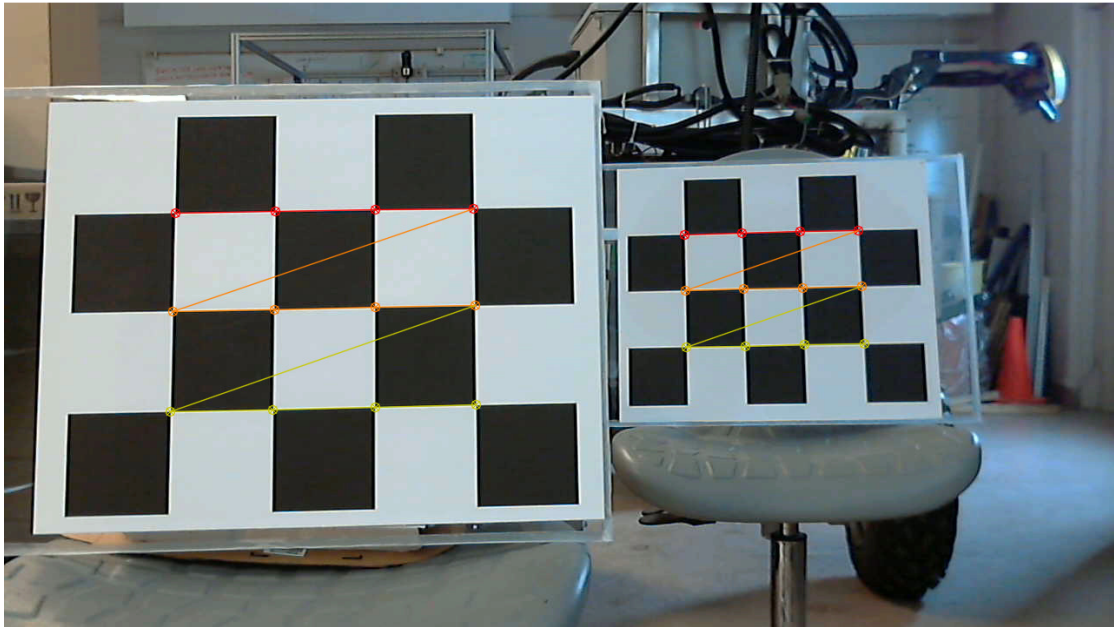
- 2) By applying the rotation matrix  $R_1$  to the 3D camera's coordinate system  $C_{3d}$ , a new coordinate system  $C'_{3d}$  was achieved. The Z axis of  $C'_{3d}$  and the Z axis of the coordinate system of the left target  $C_L$  are parallel. Therefore, by rotating the coordinate system  $C'_{3d}$  around its Z axis with an angle  $\beta$ , the X, Y, Z axis of the new coordinate system  $C''_{3d}$  is parallel to the X, Y, Z axis of coordinate system  $C_L$ . Figure 4.7 shows the coordinate system  $C_L$  defined by the left target. Point  $P_1$  and  $P_2$  in Figure 4.8 are two inner corners of the target.  $P_1$  is the origin point of the coordinate system  $C_L$ , and  $P_2$  is on the X axis of  $C_L$ . The original XYZ coordinate of  $P_1$  and  $P_2$  provided by the 3D ToF camera are represented by vector  $V_1$  and  $V_2$ , respectively, and the  $\cos(\beta)$  and  $\sin(\beta)$  can be calculated by Equation 4.12–4.14.

$$V = R_1(V_2 - V_1) \quad (4.12)$$

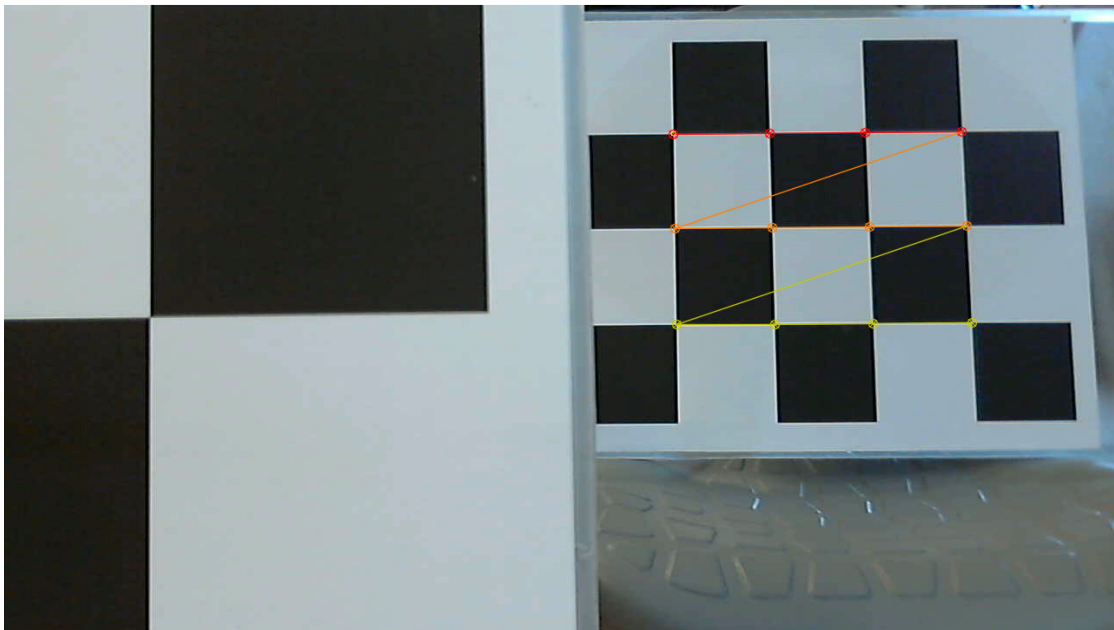
where vector  $V$  can be presented with  $[u, w, 0]^T$ .

$$\cos(\beta) = \frac{V \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}}{|V|} = \frac{u}{\sqrt{u^2 + w^2}} \quad (4.13)$$

$$\sin(\beta) = \frac{V \times \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}}{|V|} = \frac{-w}{\sqrt{u^2 + w^2}} \quad (4.14)$$



(a)

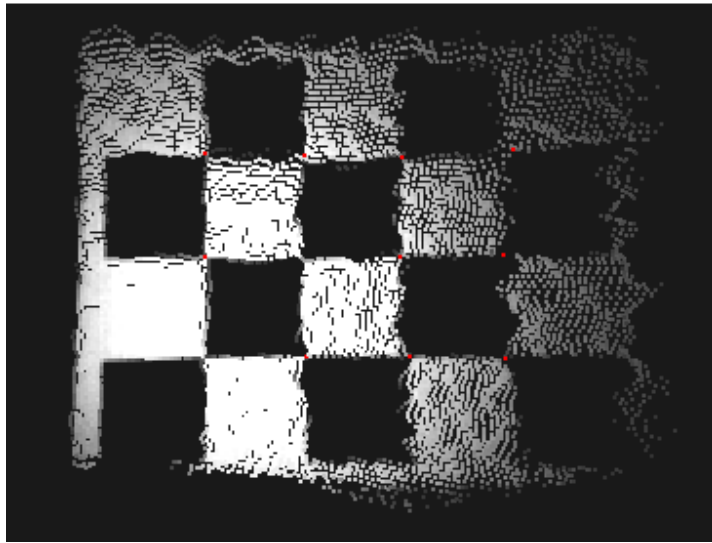


(b)



(c)

(d)



(e)

Figure 4.6. 2D and 3D camera calibration images: (a) images of two chessboard targets captured by 2D camera at a farther viewpoint A, (b) image of right chessboard target captured by 2D camera at close viewpoint B, (c–e) intensity and depth image and point cloud data of left chessboard target captured by 3D camera at close viewpoint B

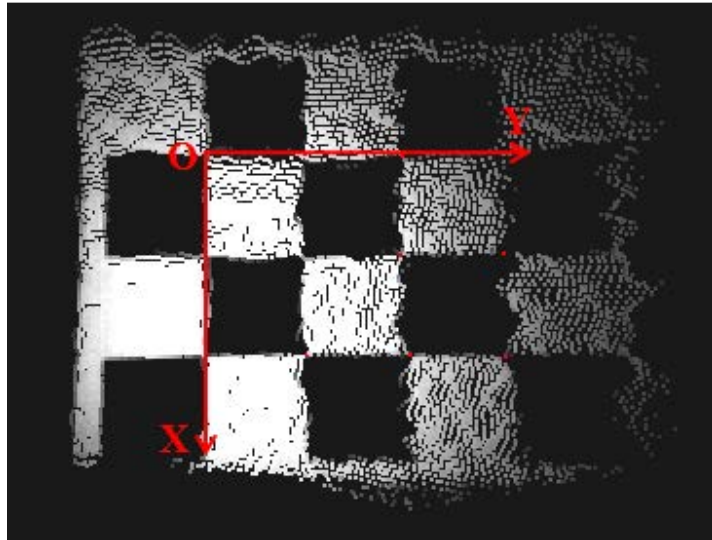


Figure 4.7. Coordinate system  $C_L$  defined by left target board

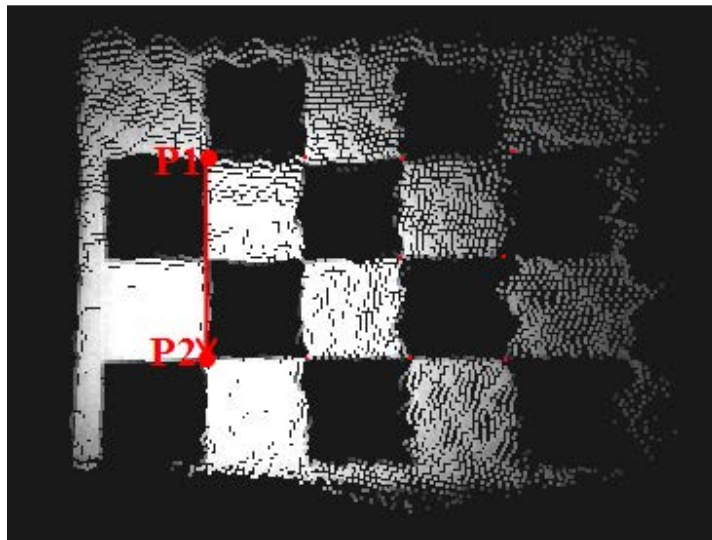


Figure 4.8. Inner corner points  $P_1$  and  $P_2$  on the X axis of the coordinate system defined by the left target board

The rotation matrix  $R_2$  from coordinate system  $C_{3d}''$  to  $C_L$  is expressed with Equation 4.15.

$$R_2 = \begin{bmatrix} \cos(\beta) & -\sin(\beta) & 0 \\ \sin(\beta) & \cos(\beta) & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{u}{\sqrt{u^2+w^2}} & \frac{w}{\sqrt{u^2+w^2}} & 0 \\ \frac{-w}{\sqrt{u^2+w^2}} & \frac{u}{\sqrt{u^2+w^2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (4.15)$$

Having rotation matrix  $R_1$  and  $R_2$ , the rotation matrix  $R_{3D2L}$  of coordinate system  $C_{3d}$  related to the coordinate system of left target  $C_L$  can be achieved as Equation 4.16 expresses.

$$R_{3D2L} = R_1 R_2 \quad (4.16)$$

Finally, the translation vector  $t_{3D2L}$  of coordinate system  $C_{3d}$  related to the coordinate system of left target  $C_L$  can be achieved as Equation 4.17 expresses.

$$t_{3D2L} = -R_{3D2L}P_1 = -R_1 R_2 P_1 \quad (4.17)$$

Fourth, the rotation matrix  $R_{3D22D}$  and the translation vector  $t_{3D22D}$  from the coordinate system of 3D camera  $C_{3d}$  to the coordinate system of 2D camera  $C_{2d}$  can be expressed by Equation 4.18 and 4.19.

$$R_{3D22D} = R_{L22D} R_{3D2L} \quad (4.18)$$

$$t_{3D22D} = R_{L22D} t_{3D2L} + t_{L22D} \quad (4.19)$$

For a point, if its coordinates in the coordinate system of the 3D camera and the 2D camera are represented using  $Q_{3D}$  and  $Q_{2D}$ , respectively,  $Q_{3D}$  and  $Q_{2D}$  satisfy Equation 4. 20.

$$Q_{2D} = R_{3D22D} Q_{3D} + t_{3D22D} \quad (4.20)$$

For the data collection system of this research, the value of  $R_{3D22D}$  and  $t_{3D22D}$  is given by Equation 4.21 and 4.22. The distance between the 2D and 3D cameras is  $|t_{3D22D}|$ , which is 86.75 mm according to the value of  $t_{3D22D}$ . Their actual horizontal and vertical distances are 82.769 and 25.394 mm. Figure 4.9 provides the translation relationship between

the cameras of the data collection system using the calibration result  $t_{3D22D}$ . It matches the manual measurement result given by Figure 4.2 (b).

$$R_{3D22D} = \begin{bmatrix} 0.961945772 & 0.02478376 & -0.272113949 \\ -0.0630362481 & 0.989143193 & -0.132748678 \\ 0.265869647 & 0.14485009 & 0.953064382 \end{bmatrix} \quad (4.21)$$

$$t_{3D22D} = \begin{bmatrix} -82.7686462 \\ 25.3944626 \\ 5.52676392 \end{bmatrix}, \quad \text{where unit is mm.} \quad (4.22)$$

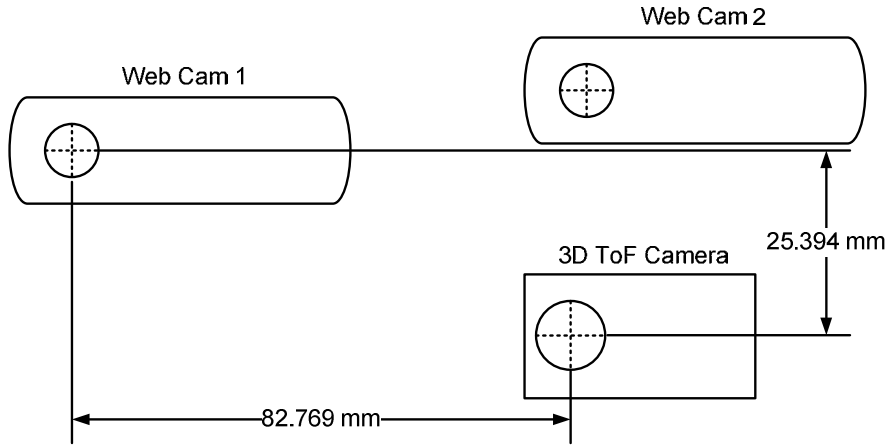


Figure 4.9. Translation relationship between cameras of the data collection system with calibration result

#### 4.7 3D Reconstruction

The 3D camera collects the 3D point cloud data view of different parts of the corn plant, and the 2D camera beside the 3D camera captures the images of beacons simultaneously for pose estimation of the 3D camera corresponding to each point cloud view, as Figure 4.5 indicates. The 3D reconstruction recovered the complete 3D model of the corn plant through 3D registration, which aligned different point cloud data views together according to the position and attitude of corresponding viewpoints estimated by the 2D



camera. In this research, a corn plant has around 20 point cloud data views to be aligned for 3D reconstruction.

#### **4.7.1 Preprocessing**

Before 3D registration, the preprocessing to clean the noise point of each point cloud data view should be accomplished. Points that qualify any of the following criteria were recognized as noise and were removed.

- 1) The point which is identified as noise by the corresponding flag provided by the PMD camera.
- 2) The point with depth over 0.7 m, because the interested object is always within 0.7 m in front of the 3D camera.
- 3) In the 3D space, the sparse point in the region whose size is smaller than 12 points. The radius used to search for neighbor points in 3D space is 10 mm in this study. For the ToF camera, at the edge of the object, sparse point noise is common. If sparse noise clearance processing was not accomplished, the final 3D model would be noisy, as Figure 4.10 (a) indicates. By applying this processing, the reconstructed 3D model is as clean as Figure 4.10 (b) shows.

After the noise clearance, each point cloud data view was ready for 3D registration.



(a)



(b)

Figure 4.10. (a) The 3D model's achievement without sparse point noise clearance and (b) the 3D model's achievement with sparse noise clearance

#### 4.7.2 3D Registration

As Figure 4.5 shows, the 2D camera of the data collection system captures the 2D images of chessboard targets simultaneously when the 3D camera collects the point cloud view. As discussed previously, with the world coordinate value of the inner corners of

detected targets, the system estimates the rotation matrix  $R_{2D}$  and the translation vector  $t_{2D}$  of the 2D camera related to the world coordinate system defined by target array based on the approach introduced in chapter 3. For a point in 3D space, its 3D coordinate vector  $Q_{2D}$  related to the 2D camera and its coordinate related to target array  $Q_w$  should satisfy Equation 4.23.

$$Q_{2D} = R_{2D}Q_w + t_{2D} \quad (4.23)$$

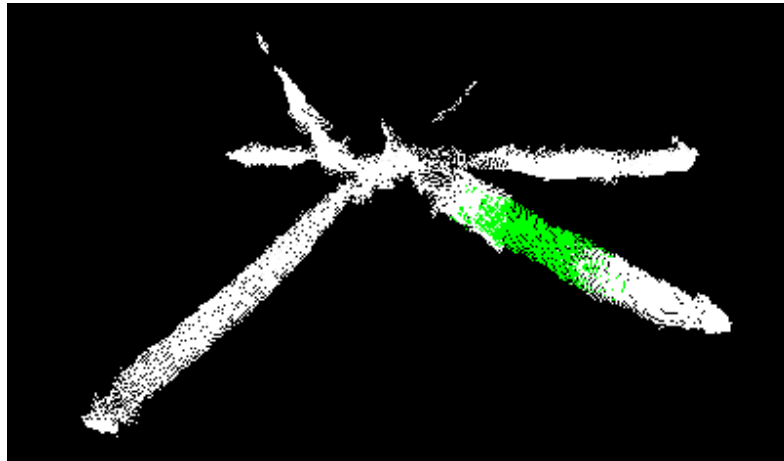
Based on the relationship between the 2D and 3D cameras, the 3D coordinate vector  $Q_{3D}$  of the point cloud data view from the 3D camera can be converted to the coordinate values related to the world coordinate system as Equation 4.24 describes based on Equation 4.20 and 4.23.

$$Q_w = R_{2D}^{-1}(R_{3D22D}Q_{3D} + t_{3D22D} - t_{2D}) \quad (4.24)$$

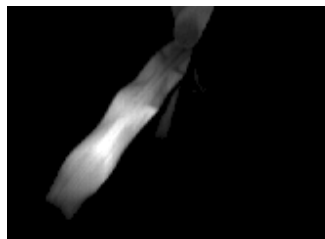
For each 3D point cloud view, the position and attitude of the 3D camera vary. However, the world coordinate system defined by the target array is consistent because the plant and target array keep static during the data collection process. By applying Equation 4.24 to convert the original 3D information of different point cloud data views from the ToF camera to those related to the consistent world coordinate system, the 3D registration is done. Figure 4.11 (a) and (b) shows the side view and top view of a 3D model reconstruction of a corn plant. It was obtained by aligning 23 point cloud data views together. Among all of the 23 data views, one is shown with green color points in Figure 4.11 (a) and (b), and its intensity and depth image are provided by Figure 4.11 (c) and (d).



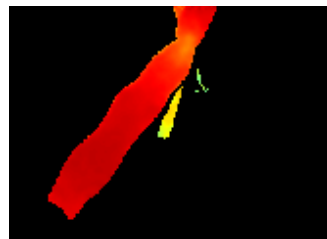
(a)



(b)



(c)



(d)

Figure 4.11. Complete 3D model of corn plant achieved through 3D registration: (a) side view and (b) top view of 3D model, (c) the intensity and (d) depth image of the point cloud data view corresponding to the green points in (a) and (b)

#### 4.8 Leaf and Stem Segmentation

After obtaining the 3D model of the corn plant, this research separated leaves and stem in order to measure their physical parameters.

##### 4.8.1 Stem Segmentation

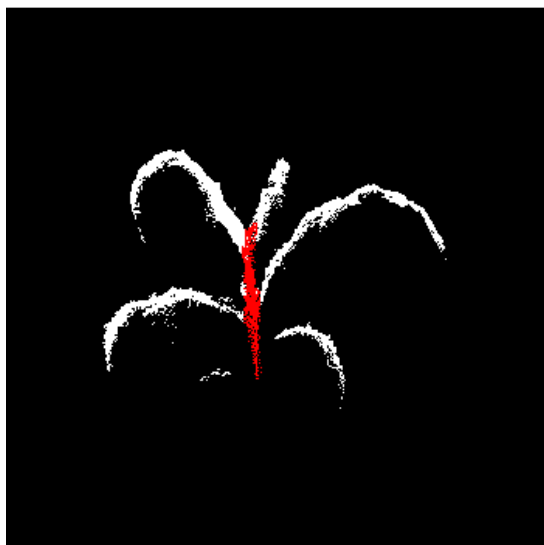
Stem segmentation was carried out first by generating six 2D side view images of a corn plant based on the 3D model as Figure 4.12 shows. Suppose the range of  $x$ ,  $y$ , and  $z$  coordinate value of the 3D model are  $[x_0, x_1]$ ,  $[y_0, y_1]$ , and  $[z_0, z_1]$ , the width and height of the 2D side view image are  $W$  and  $H$ . The  $i_{th}$  side view image is achieved by setting its point  $(u, v)$  and corresponding every point of the 3D model to white, and the relationship between  $(u, v)$  of the side view image and the point  $(x, y, z)$  of the 3D model is expressed by Equation 4.25–4.27.

$$\varphi = i * 60, \text{ where } i \in \{0, 1, 2, 3, 4, 5\} \quad (4.25)$$

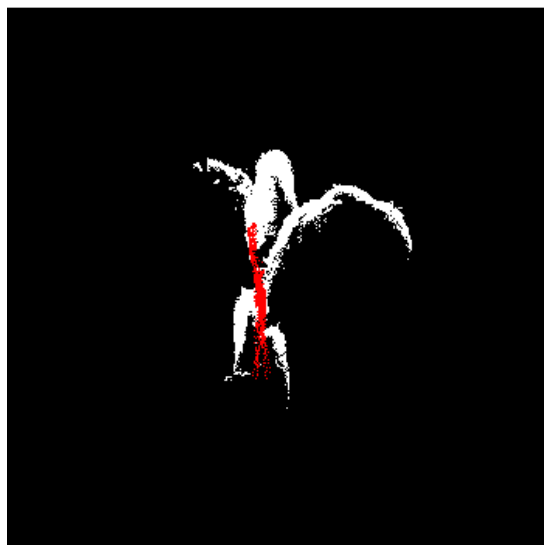
$$u = 0.4 \left[ \left( x - \frac{x_0+x_1}{2} \right) \sin(\varphi) + \left( y - \frac{y_0+y_1}{2} \right) \cos(\varphi) \right] + \frac{W}{2} \quad (4.26)$$

$$v = 0.4 \left( z - \frac{z_0+z_1}{2} \right) + \frac{H}{2} \quad (4.26)$$

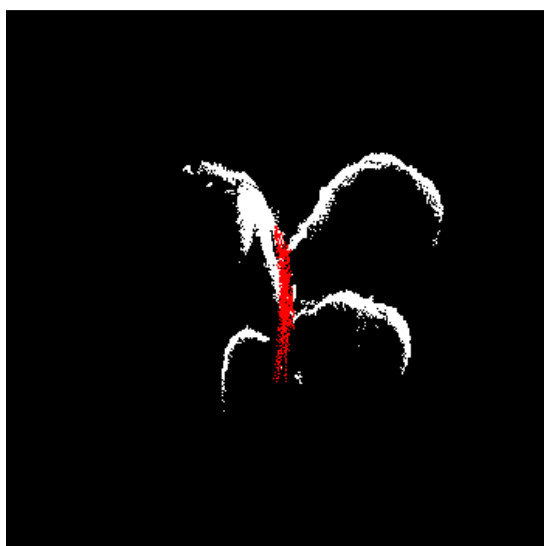
For each side view image, this system searched for the straight lines with the length over 50 pixels and the angle range from  $-5^\circ$  to  $+5^\circ$  around the vertical direction. Pixels on these detected straight lines are displayed with red color in the side view images provided in Figure 4.12. The points of the 3D model whose corresponding points in all 2D side view image that were on the detected straight lines were recognized as the stem of the corn plant.



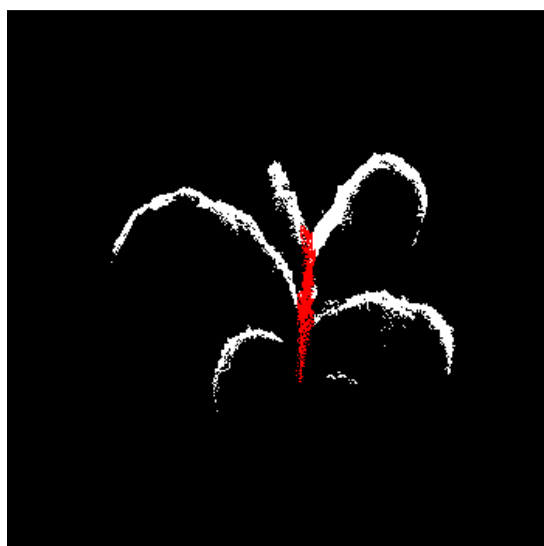
(a)



(b)



(c)



(d)

Figure 4.12 continued

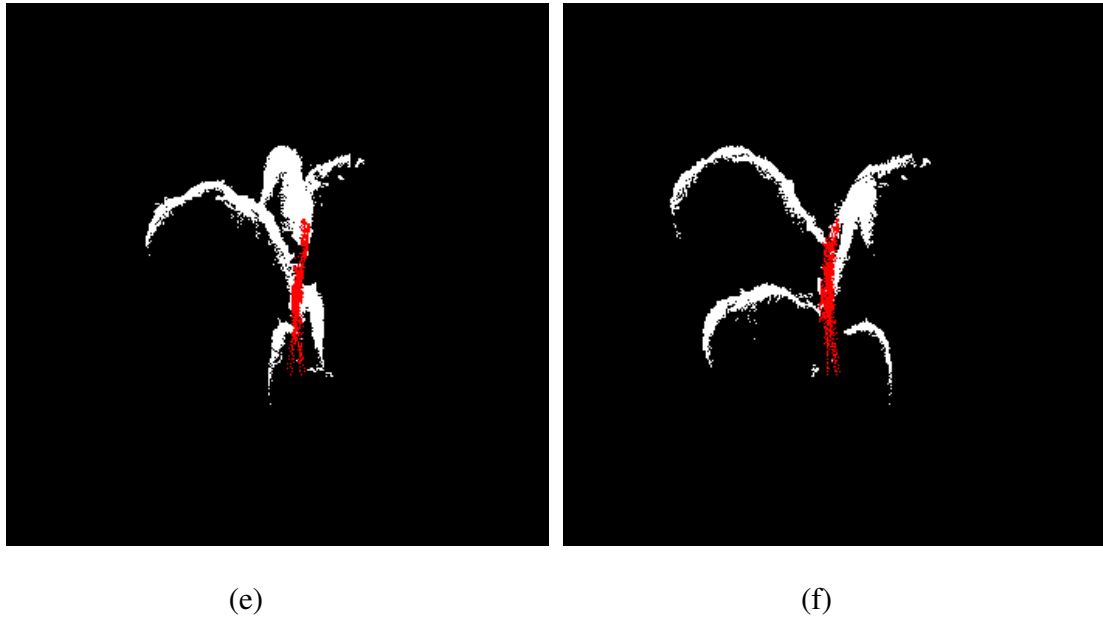


Figure 4.12. The stem detected in different side views at different viewpoints

#### 4.8.2 Leaf Segmentation

The stem segmentation achieved in the previous section separated the 3D model into several regions in 3D space, and each big region is one leaf. As the point cloud data of the reconstructed 3D model was not organized, and separating the data into regions in 3D space is time consuming. To solve this problem, a 2D image processing based algorithm was developed to process multiple side views and top view images for region separation in 3D space.

Figure 4.12 shows an example of side view images, and Figure 4.13 is the top view image. The detected stem is displayed with red color in these images. It is clear that the leaves were successfully separated into different parts by the stem in these 2D images. Although some leaves may fuse together for some image view, this occlusion problem can be solved by other image views.

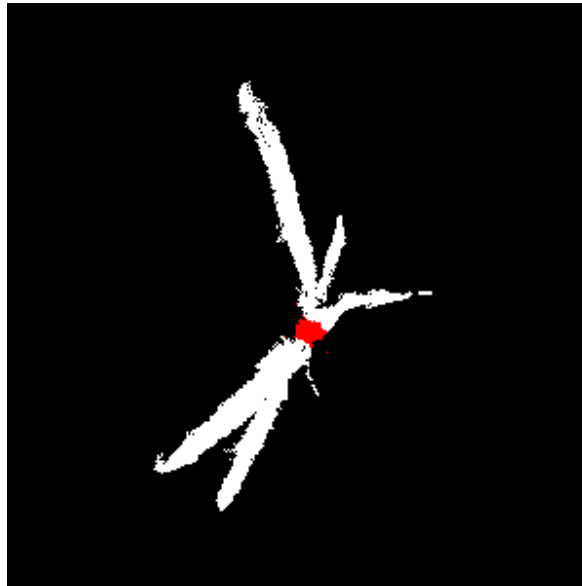


Figure 4.13. The top view image of the point cloud of the 3D model; the stem is marked with red color

The region separation algorithm started from separating the white points of the first image view into regions in 2D space, and points of the 3D reconstruction model were organized into different groups accordingly. Then the system repeated this method to process the next image to check whether the points grouped previously should be separated into different regions. This procedure was iterated until all the image views were processed. For the final regions, those whose sizes were smaller than 40 points were removed as noise data, and each of the remaining regions was recognized as a leaf. The stem and leaves of the 3D reconstruction model of the corn plant are well segmented by this approach, as Figure 4.14 indicates.



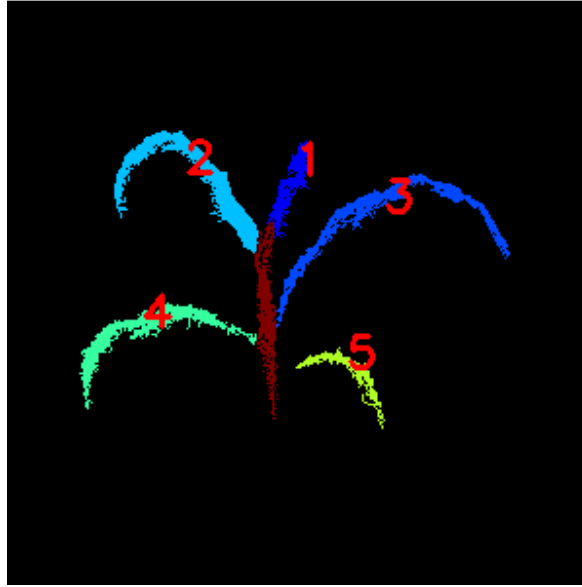


Figure 4.14. Leaves and stem segmentation result

#### 4.9 Leaf Parameter Estimation

After the points of the stem and every leaf of 3D reconstruction model were determined, this system was ready to estimate the physical parameters of the corn plant. This research developed the algorithm to estimate the collar height as well as the width, length, and area of leaves.

##### 4.9.1 Leaf Points Regression

To begin the physical parameter estimation of leaves, this research employed Equation 4.27 and 4.28 to describe the curve of the skeleton of each leaf. In these two equations,  $x$ ,  $y$ , and  $z$  are the known world coordinate value of the skeleton point of the corresponding leaf. Variables  $\varphi$ ,  $a$ ,  $b$ ,  $c$ ,  $d$ , and  $e$  of these two equations are unknown, and they are solved by applying singular value decomposition (SVD) regression method to process the 3D information of all points of the corresponding leaf.

$$\tilde{y} = x \sin(\varphi) + y \cos(\varphi) \quad (4.27)$$

$$z = a\tilde{y}^4 + b\tilde{y}^3 + c\tilde{y}^2 + d\tilde{y} + e \quad (4.28)$$

#### 4.9.2 Leaf Parameter Estimation

Having  $\varphi$  solved, Equation 4.27 was applied to calculate  $\tilde{y}$  value of every point of the corresponding leaf. The white pixels of Figure 4.15 show the transformation result of leaf 3 in Figure 4.14. For Figure 4.15, the horizontal direction linearly related to  $\tilde{y}$  value of the point, and vertical direction related to its  $z$  value. The red curved line shows the linear regression result represented by Equation 4.28. Moreover, Equation 4.29 was used to calculate  $\tilde{x}$  value of all of the point of the corresponding leaf.

$$\tilde{x} = x \cos(\varphi) - y \sin(\varphi) \quad (4.29)$$

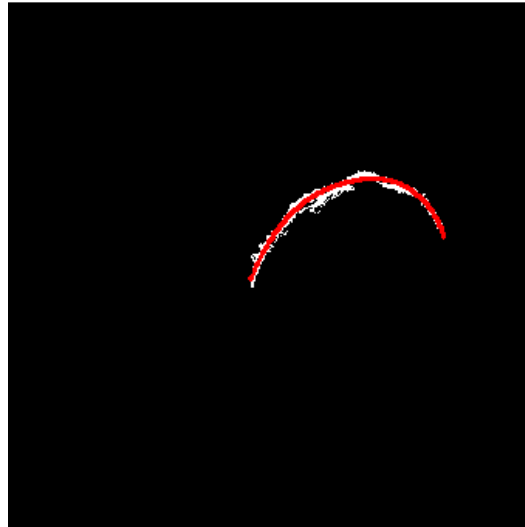


Figure 4.15. Regression result of leaf 3 of plant 1

Based on the leaf skeleton curve described by the solved Equation 4.28, the leaf length and area was estimated. The range of  $\tilde{y}$  value of a leaf is represented with  $[\tilde{y}_{min}, \tilde{y}_{max}]$ . To estimate the length and area of the leaf, the leaf was divided into 50

fractions along the direction parallel to the  $\tilde{x}$  axis. In addition, the  $i_{th}$  fraction, which is represented with  $F_i$  in this research, contains all of the points whose  $\tilde{y}$  is within the range between  $\tilde{Y}_{i-1}$  and  $\tilde{Y}_i$ , where  $\tilde{Y}_i$  is represented by Equation 4.30.

$$\tilde{Y}_i = \frac{i \times (\tilde{y}_{max} - \tilde{y}_{min})}{50} \quad i \in \{0, 1, 2, \dots, 50\} \quad (4.30)$$

Then by applying  $\tilde{y} = \tilde{Y}_i$  to Equation 4.28, the corresponding result value  $z$  is represented with  $\tilde{Z}_i$ .

Then the leaf length is achieved using Equation 4.31.

$$Length = \sum_{i=1}^{50} \sqrt{(\tilde{Y}_i - \tilde{Y}_{i-1})^2 + (\tilde{Z}_i - \tilde{Z}_{i-1})^2} \quad (4.31)$$

Additionally, if the minimum and maximum  $\tilde{x}$  value of the points of  $F_i$  are represented with  $\tilde{x}_{imin}$  and  $\tilde{x}_{imax}$ , their difference is the width of fraction  $F_i$ . In addition, the area and width of the leaf can be achieved as shown in Equation 4.32 and 4.33:

$$area = \sum_{i=1}^{50} \left[ \sqrt{(\tilde{Y}_i - \tilde{Y}_{i-1})^2 + (\tilde{Z}_i - \tilde{Z}_{i-1})^2} \times (\tilde{x}_{imax} - \tilde{x}_{imin}) \right] \quad (4.32)$$

$$width = \max(\tilde{x}_{imax} - \tilde{x}_{imin}) \quad i \in \{0, 1, 2, \dots, 50\}. \quad (4.33)$$

The leaf collar height was also estimated. As the coordinate of the point cloud of the 3D reconstruction model is related to the world coordinate system defined by the target array on the ground, the  $Z$  coordinate of a point actually is its height from the ground. The leaf collar was achieved by finding the conjunction point between the stem and the leaf skeleton, which is expressed by the solved Equation 4.27 and 4.28. In addition, the leaf collar height is the difference between the  $Z$  coordinate of the conjunction point and the height measurement result of the soil surface of the pot related to the ground.

#### 4.10 Results and Discussion

Three corn plants at vegetative stage V5 were used to test this system. The plants were around half a meter and they had seven leaves.

Figure 4.16–4.18 are the 2D color pictures and the corresponding 3D reconstruction results of the three corn plants. As they indicate, this system achieved a relatively clean and complete 3D model of the corn plant. Visually, the 2D color images and 3D reconstruction result match well together. However, the 3D reconstruction images show that the bottom one or two leaves of corn plants were either missed or incomplete in the reconstruction. Noisy 3D information of the bottom leaves caused by their small sizes led to the incomplete reconstruction. Additionally, the reflectance of the soil surface of the pot to the light source of the 3D camera was very low, which resulted in noisy 3D data. The noisy 3D data of the soil surface fused with the point cloud of the bottom leaves, making it difficult to extract bottom leaves. Therefore, the bottom leaves were removed as noise.

Figure 4.19 provides the leaf and stem separation result of the 3D reconstruction of three corn plants. The leaves and stem are accurately separated, and they are displayed with different color in the result images. The IDs of different leaves are also given by Figure 4.19.

To quantitatively evaluate the accuracy of the 3D model, this system estimated the parameters of each leaf of every plant, including width, length, area, and collar height, which were compared with the reference measurement results. The leaves' area measurement result provided by LI-3100C Leaf Area Meter was used as the ground true value, and the other three parameters were measured manually. The parameters measured by this system and by the reference methods are listed in Table 4.1–4.3. The corresponding error rate is also listed.

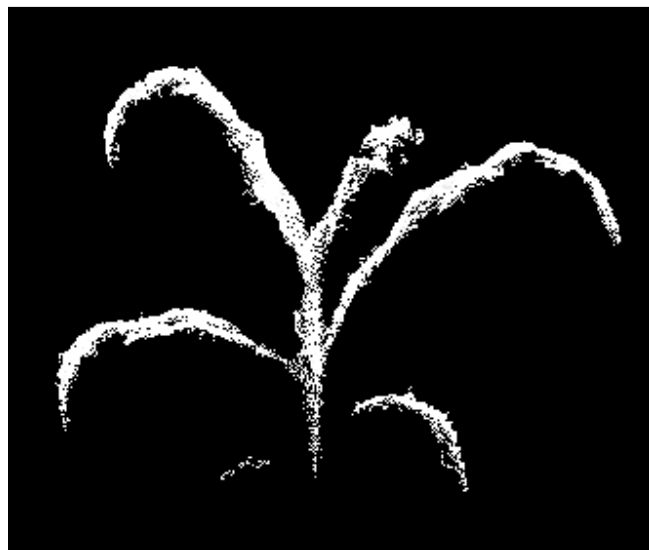
For the bottom one or two leaves of corn plants, the measurement results of this system are not available because they were missed in their 3D reconstruction model.

Statistics were done to analyze the measurement error rate, and the result was given by Table 4.4. As it indicates, the median value of each parameter's error rate is smaller than 7.18%. The average measurement error rate of a leaf's area, length, width, and collar height are 10.46%, 10.42%, 11.10%, and 8.18%, respectively. The third quartile value of these measurement errors are 11.14%, 11.75%, 13.48%, and 6.88%. It means that the error rate of 75% of the corresponding measurements is smaller above the third quartile values. Therefore, a big part of the measurement has relatively good accuracy. However, there are some big outliers for each parameter's measurement. For example, the error rate of width and area measurement of leaf 1 of plant 2 is 41.2% and 31.13%, respectively. Another example is that the area and length measurements of leaf 5 of plant 2 are 61.87% and 27.55% smaller than the reference value, which is because this leaf is too small for the ToF camera, thus causing incomplete reconstruction.

When running on a 3.4 GHz Intel Xeon CPU, the system's average processing time cost for 3D reconstruction and leaf parameter estimation of a corn plant was 4.73 second. The processing time cost for a corn plant was less or at least comparable to the time needed to move the imaging sensor to around 20 viewpoints to collect different point cloud data views. Therefore, the image data processing and collection can be processed simultaneously.



(a)

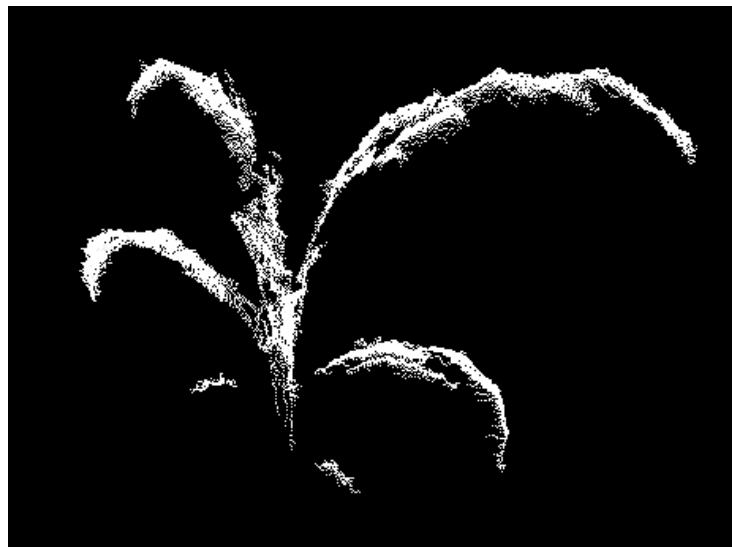


(b)

Figure 4.16. Corn plant 1 and its 3D reconstruction result: (a) 2D color picture and (b) 3D reconstruction result



(a)

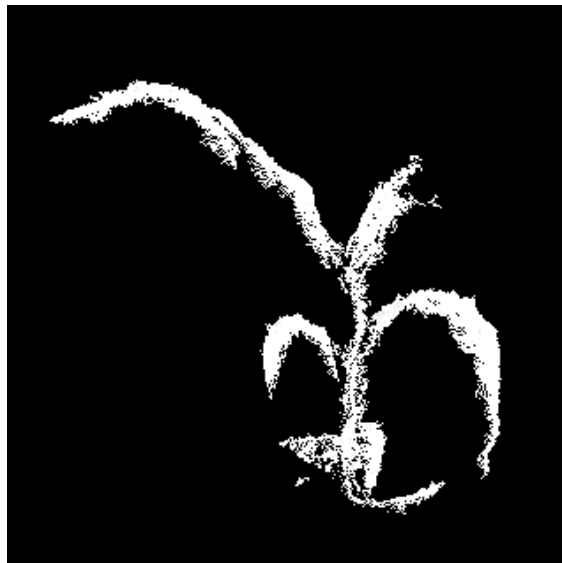


(b)

Figure 4.17. Corn plant 2 and its 3D reconstruction result: (a) 2D color picture and (b) 3D reconstruction result



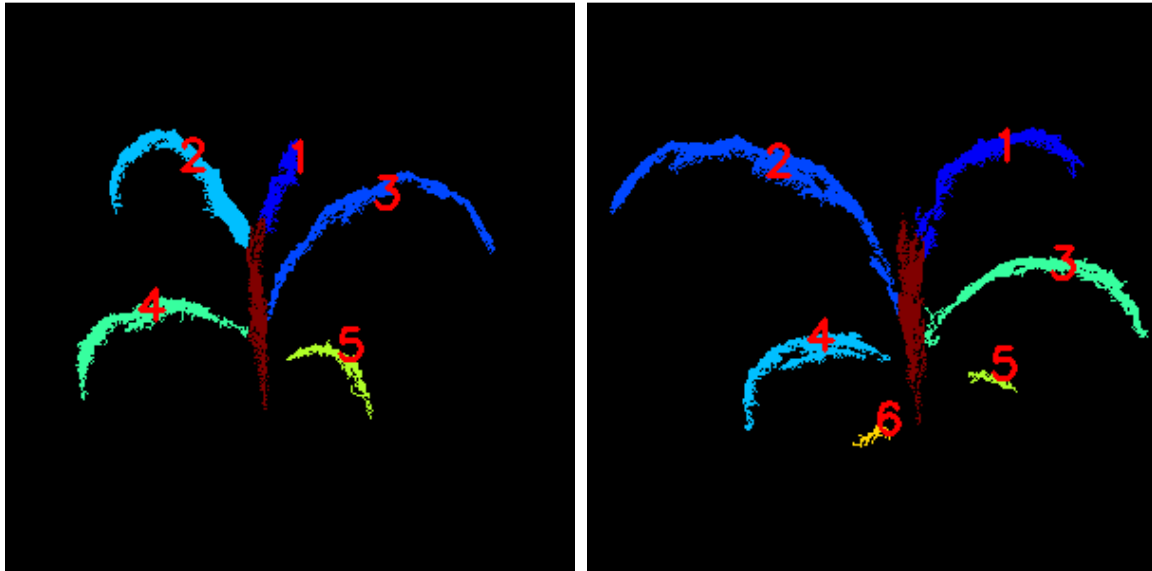
(a)



(b)

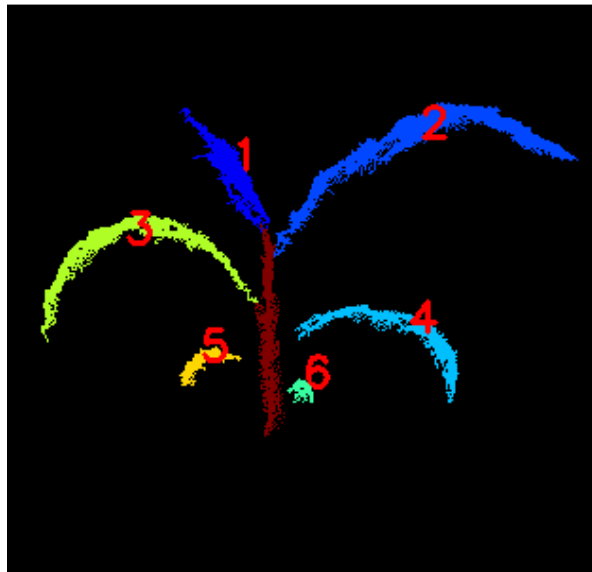
Figure 4.18. Corn plant 3 and its 3D reconstruction result: (a) 2D color picture and (b) 3D reconstruction result





(a)

(b)



(c)

Figure 4.19. Leaf and stem segmentation result: (a) plant 1, (b) plant 2, and (c) plant 3

Table 4.1. Measurement result of corn plant 1

| Leaf | Reference Measurement Result |             |            |                    | Estimation Result of This System |             |            |                    | Error Rate % |        |       |               |
|------|------------------------------|-------------|------------|--------------------|----------------------------------|-------------|------------|--------------------|--------------|--------|-------|---------------|
|      | Area (mm <sup>2</sup> )      | Length (mm) | Width (mm) | Collar Height (mm) | Area (mm <sup>2</sup> )          | Length (mm) | Width (mm) | Collar Height (mm) | Area         | Length | Width | Collar Height |
| 1    | 4364                         | 210         | 40         | 254.0              | 4114.3                           | 222.9       | 38.3       | 269.3              | 5.72         | 6.14   | 4.25  | 6.03          |
| 2    | 13436                        | 435         | 40         | 254.0              | 12080.4                          | 417.1       | 37.7       | 240.3              | 10.09        | 4.11   | 5.75  | 5.39          |
| 3    | 11659                        | 490         | 32         | 163.9              | 11393.0                          | 469.3       | 30.1       | 170.3              | 2.28         | 4.22   | 5.94  | 3.92          |
| 4    | 5716                         | 370         | 24         | 140.76             | 6083.8                           | 363.3       | 23.8       | 132.1              | 6.43         | 1.81   | 0.83  | 6.14          |
| 5    | 2386                         | 227         | 15         | 87.13              | 2089.4                           | 236.6       | 16.7       | 87.7               | 12.43        | 4.23   | 11.33 | 0.67          |
| 6    | 858                          | 110         | 10         | 44.76              | N/A                              | N/A         | N/A        | N/A                | N/A          | N/A    | N/A   | N/A           |
| 7    | 423                          | 50          | 12         | 21.48              | N/A                              | N/A         | N/A        | N/A                | N/A          | N/A    | N/A   | N/A           |

Table 4.2. Measurement result of corn plant 2

| Leaf | Reference Measurement Result |             |            |                    | Estimation Result of This System |             |            |                    | Error Rate % |        |       |               |
|------|------------------------------|-------------|------------|--------------------|----------------------------------|-------------|------------|--------------------|--------------|--------|-------|---------------|
|      | Area (mm <sup>2</sup> )      | Length (mm) | Width (mm) | Collar Height (mm) | Area (mm <sup>2</sup> )          | Length (mm) | Width (mm) | Collar Height (mm) | Area         | Length | Width | Collar Height |
| 1    | 13186                        | 430         | 50         | 265.88             | 9081.2                           | 367.3       | 29.4       | 259.51             | 31.13        | 14.58  | 41.20 | 2.40          |
| 2    | 17653                        | 590         | 40         | 210.06             | 16360.4                          | 520.7       | 37.5       | 224.51             | 7.32         | 11.75  | 6.25  | 6.88          |
| 3    | 11364                        | 480         | 32         | 159.22             | 11481.2                          | 430.2       | 35.4       | 153.21             | 1.03         | 10.38  | 10.63 | 3.77          |
| 4    | 5771                         | 350         | 23         | 128.92             | 5635.3                           | 323.1       | 26.1       | 121.71             | 2.35         | 7.69   | 13.48 | 5.59          |
| 5    | 2401                         | 220         | 15         | 78.99              | 915.5                            | 159.4       | 12.7       | 99.21              | 61.87        | 27.55  | 15.33 | 25.60         |
| 6    | 891                          | 110         | 11         | 39.8               | 916.4                            | 102.1       | 10.9       | 24.71              | 2.85         | 7.18   | 0.91  | 37.91         |
| 7    | 371                          | 45          | 12         | 9.99               | N/A                              | N/A         | N/A        | N/A                | N/A          | N/A    | N/A   | N/A           |

Table 4.3. Measurement result of corn plant 3

| Leaf | Reference Measurement Result |             |            |                    | Estimation Result of This System |             |            |                    | Error Rate % |        |       |               |
|------|------------------------------|-------------|------------|--------------------|----------------------------------|-------------|------------|--------------------|--------------|--------|-------|---------------|
|      | Area (mm <sup>2</sup> )      | Length (mm) | Width (mm) | Collar Height (mm) | Area (mm <sup>2</sup> )          | Length (mm) | Width (mm) | Collar Height (mm) | Area         | Length | Width | Collar Height |
| 1    | 6012                         | 250         | 43         | 273.06             | 6325.6                           | 234.5       | 38.5       | 264.22             | 5.22         | 6.20   | 10.47 | 3.24          |
| 2    | 15161                        | 680         | 44         | 273.06             | 14956.4                          | 528.1       | 35.4       | 259.42             | 1.35         | 22.34  | 19.55 | 4.99          |
| 3    | 12598                        | 450         | 37         | 155.54             | 13169.9                          | 469.4       | 37.5       | 162.02             | 4.54         | 4.31   | 1.35  | 4.17          |
| 4    | 6797                         | 390         | 26         | 122.98             | 6893.3                           | 362.8       | 24.4       | 127.52             | 1.42         | 6.97   | 6.15  | 3.69          |
| 5    | 2789                         | 248         | 19         | 82                 | 2351.5                           | 177.3       | 20         | 88.52              | 15.69        | 28.51  | 5.26  | 7.95          |
| 6    | 1036                         | 119         | 12         | 46.19              | 920.6                            | 108.1       | 15.6       | 51.12              | 11.14        | 9.16   | 30.00 | 10.68         |
| 7    | 412                          | 48          | 13         | 8.95               | N/A                              | N/A         | N/A        | N/A                | N/A          | N/A    | N/A   | N/A           |

Table 4.4. Statistics of leaf parameter's measurement error rate

| Measurement   | Min   | 1 <sup>st</sup><br>Quartile | Mean   | Median | 3 <sup>rd</sup><br>Quartile | Max    |
|---------------|-------|-----------------------------|--------|--------|-----------------------------|--------|
| Area          | 1.03% | 2.35%                       | 10.76% | 5.72%  | 11.14%                      | 61.87% |
| Length        | 1.81% | 4.31%                       | 10.42% | 7.18%  | 11.75%                      | 28.51% |
| Width         | 0.83% | 5.26%                       | 11.10% | 6.25%  | 13.48%                      | 41.42% |
| Collar Height | 0.67% | 3.77%                       | 8.18%  | 5.39%  | 6.88%                       | 37.91% |

#### 4.11 Conclusions

In this research, a 3D reconstruction based phenotyping system of plant was developed. From the results of this research, we can conclude that this system exhibited promising potential for developing a maize phenotyping system.

The 3D reconstruction approach is effective. The chessboard pattern target array can provide precise position and attitude estimation of the 2D camera. Moreover, the proposed calibration method was proven effective in getting the spatial relationship between the 2D and 3D cameras, which were installed side by side as the data capturing system of this study, and therefore enabled deriving the pose of the 3D camera based on that of the 2D camera. According to the position and attitude of the 3D camera related to each 3D image view, different views were aligned precisely into a complete 3D reconstruction of a corn plant.

It also can be concluded that this processing algorithm of the reconstructed 3D model of the corn plant is promising. The segmentation algorithm was effective to extract the stem and each leaf from the 3D model of the corn plant. Leaf points regression and leaf parameter estimation algorithms automatically quantified the leaf phenotypic parameters, such as leaf width, leaf length, leaf area, and collar height. The measurement of the leaf parameter is relatively accurate, while big outliers still exist.

Especially, the fast processing speed, high accuracy, low cost, and nondestructive nature of this phenotyping system may benefit the high-throughput, large-scale phenotyping system that can collect and process data throughout the life cycle of the maize plant.

#### 4.12 Reference

- Alenya, G., B. Dellen, and C. Torras. 2011. 3D modelling of leaves from color and tof data for robotized plant measuring. In 2011 IEEE International Conference on Robotics and Automation. Shanghai, China: IEEE.
- Allen, H. L., K. Estrada, G. Lettre, S. I. Berndt, M. N. Weedon, F. Rivadeneira, C. J. Willer, A. U. Jackson, S. Vedantam, and S. Raychaudhuri. 2010. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 467(7317):832-838.
- Bellasio, C., J. Olejníčková, R. Tesař, D. Šebela, and L. Nedbal. 2012. Computer reconstruction of plant growth and chlorophyll fluorescence emission in three spatial dimensions. *Sensors* 12(1):1052-1071.
- Chaerle, L., S. Lenk, I. Leinonen, H. G. Jones, D. Van Der Straeten, and C. Buschmann. 2009. Multi-sensor plant imaging: Towards the development of a stress-catalogue. *Biotechnology Journal* 4(8):1152-1167.
- Cobb, J. N., G. DeClerck, A. Greenberg, R. Clark, and S. McCouch. 2013. Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. *Theoretical and Applied Genetics* 126(4):867-887.
- Fiorani, F., U. Rascher, S. Jahnke, and U. Schurr. 2012. Imaging plants dynamics in heterogenic environments. *Current opinion in biotechnology* 23(2):227-235.
- Foundation, N. S., and G. Mcb. 2011. Phenomics : Genotype to Phenotype A report of the Phenomics workshop sponsored by the USDA and NSF.
- Furbank, R. T., and M. Tester. 2011. Phenomics - technologies to relieve the phenotyping bottleneck. *Trends in Plant Science* 16(12):635-644.

- Granier, C., L. Aguirrezabal, K. Chenu, S. J. Cookson, M. Dauzat, P. Hamard, J. J. Thioux, G. Rolland, S. Bouchier-Combaud, A. Lebaudy, B. Muller, T. Simonneau, and F. Tardieu. 2006. PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in *Arabidopsis thaliana* permitted the identification of an accession with low sensitivity to soil water deficit. *New Phytologist* 169(3):623-635.
- Heffner, E. L., J.-L. Jannink, and M. E. Sorrells. 2011. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *The Plant Genome* 4(1):65-75.
- Jansen, M., F. Gilmer, B. Biskup, K. A. Nagel, U. Rascher, A. Fischbach, S. Briem, G. Dreissen, S. Tittmann, S. Braun, I. De Jaeger, M. Metzlauff, U. Schurr, H. Scharr, and A. Walter. 2009. Simultaneous phenotyping of leaf growth and chlorophyll fluorescence via GROWSCREEN FLUORO allows detection of stress tolerance in *Arabidopsis thaliana* and other rosette plants. *Functional Plant Biology* 36(10-11):902-914.
- Lu, Y., L. J. Savage, M. D. Larson, C. G. Wilkerson, and R. L. Last. 2011. Chloroplast 2010: a database for large-scale phenotypic screening of *Arabidopsis* mutants. *Plant physiology* 155(4):1589-1600.
- Nichols, R. J., S. Sen, Y. J. Choo, P. Beltrao, M. Zietek, R. Chaba, S. Lee, K. M. Kazmierczak, K. J. Lee, and A. Wong. 2011. Phenotypic landscape of a bacterial cell. *Cell* 144(1):143-156.
- Rusu, R. B., N. Blodow, Z. C. Marton, and M. Beetz. 2008. Aligning point cloud views using persistent feature histograms. In *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*. IEEE.
- Scholes, J. D., and S. A. Rolfe. 2009. Chlorophyll fluorescence imaging as tool for understanding the impact of fungal diseases on plant performance: a phenomics perspective. *Functional Plant Biology* 36(10-11):880-892.
- Sirault, X. R. R., R. A. James, and R. T. Furbank. 2009. A new screening method for osmotic component of salinity tolerance in cereals using infrared thermography. *Functional Plant Biology* 36(10-11):970-977.
- Speliotes, E. K., C. J. Willer, S. I. Berndt, K. L. Monda, G. Thorleifsson, A. U. Jackson, H. L. Allen, C. M. Lindgren, J. a. Luan, and R. Mägi. 2010. Association analyses of

- 249,796 individuals reveal 18 new loci associated with body mass index. *Nature genetics* 42(11):937-948.
- Swarbrick, P. J., P. Schulze-Lefert, and J. D. Scholes. 2006. Metabolic consequences of susceptibility and resistance (race-specific and broad-spectrum) in barley leaves challenged with powdery mildew. *Plant Cell and Environment* 29(6):1061-1076.
- Walter, A., H. Scharr, F. Gilmer, R. Zierer, K. A. Nagel, M. Ernst, A. Wiese, O. Virnich, M. M. Christ, B. Uhlig, S. Juenger, and U. Schurr. 2007. Dynamics of seedling growth acclimation towards altered light conditions can be quantified via GROWSCREEN: a setup and procedure designed for rapid optical phenotyping of different plant species. *New Phytologist* 174(2):447-455.
- Winzler, E. A., D. D. Shoemaker, A. Astromoff, H. Liang, K. Anderson, B. Andre, R. Bangham, R. Benito, J. D. Boeke, and H. Bussey. 1999. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* 285(5429):901-906.



## CHAPTER 5. GENERAL CONCLUSIONS

### 5.1 Conclusions

#### 5.1.1 3D Imaging Based Crop Plant and Weed Recognition System

Crop and weed differentiation is key for an automated weeding system, and a practical solution has not yet been achieved. This research developed a soybean and broccoli detection system based on processing the point cloud data generated by a 3D ToF camera, which exhibited promising potential for automated robotic weeding application. First, the 3D ToF camera in this study was proven to work well when direct sunlight is blocked from shining onto plants. It is less susceptible to the change of ambient light due to seasonal or weather variation compared to color camera and spectral imaging. Second, the sparse point noise filter was developed to process the 3D image, and it was proven effective and efficient. Third, several 3D and 2D features, including gradient of amplitude and depth, surface curvature, API, normal direction, and neighbor point count in 3D space, were found to discriminate effectively broccoli and soybean plants from weeds and soil, and their optimized calculation algorithms were contributed. Fourth, according to the 3D geometry and 2D amplitude image characteristics of broccoli and soybean plants, one segmentation algorithm was developed for each. Moreover, this system showed robustness in outdoor field conditions. When tested in the field where weed plants were dense and had comparable height with crop plants, this system reached the detection rate of 88.3% and 91.2% for broccoli and soybean plants, respectively. The crop plants that were not too small in the images were extracted and recovered with relatively complete shape. Fifth, the 2D and 3D machine vision algorithm was highly optimized, and the image processing speed of this system was over 30 Hz for both types of plants.

### **5.1.2 3D Reconstruction Based Plant Phenotyping System**

A 3D reconstruction based plant phenotyping system was developed, and it exhibited promising potential when it was tested on maize plants. The main contribution of this system was the 3D reconstruction method. The chessboard target array provided real-time and precise position and attitude estimation of a 2D camera. In addition, the proposed calibration method proved effective in getting the spatial relationship between the 2D and 3D cameras, which were installed side by side as the data capturing system of this study, and therefore enabled deriving the pose of the 3D camera based on that of the 2D camera. According to the position and attitude of the 3D camera related to each 3D image view, different views were aligned precisely into a complete 3D reconstruction of a corn plant.

Moreover, this phenotyping system contributed the algorithm to further process the 3D reconstruction of the corn plant. The segmentation algorithm was effective to extract the stem and each leaf from the 3D model of the corn plant. Leaf points regression and leaf parameter estimation algorithms automatically quantified the leaf phenotypic parameters, including leaf width, leaf length, leaf area, and collar height. The measurement of the leaf parameter is relatively accurate, although big outliers still exist. The average time cost for 3D reconstruction and phenotypic data estimation was less than 5 seconds per plant.

In particular, the fast processing speed, high accuracy, low cost, and nondestructive nature of this phenotyping system may benefit the high-throughput, large-scale phenotyping system that can collect and process data through the life cycle of the maize plant.

## 5.2 Recommendations

There are multiple ways to further improve the crop plant recognition system in future research:

1. The low resolution of the 3D ToF camera is the main limitation to recognizing small broccoli plants and soybean leaves. Some alternative 3D imaging sensors such as the active stereo camera features significantly higher spatial resolution and measurement accuracy with lower noise level, making them more favorable to the algorithm proposed in this research. Although they cannot work outdoors during day time, they may be applied to develop the fully automated weeding system that works in the evening only.
2. Further study is needed to systematically test the proposed algorithm with various crop plant species at different growth stages.
3. Current work mainly focuses on the local geometry features instead of that of whole plant. For future research, the crop plant recognition algorithm could be improved by analyzing the 3D geometry of the whole plant.
4. The system could be improved by combining it with a GPS mapping system that records the precise position of crop plants during the planting or transplanting process. The preknowledge of the crop plant map from the GPS mapping system can help to reduce the searching area of the crop plant recognition system, therefore increasing the processing speed and recognition accuracy.

The 2D imaging based position and attitude estimation system could be improved by the following approaches:

1. A 2D camera with higher resolution, better optical properties, and short exposure time would improve the system performance. Higher resolution can introduce better measurement accuracy and longer pose tracking range. In addition, short exposure time will enable the system to work well with higher camera moving speed and will benefit better measurement accuracy.
2. Hardware acceleration techniques such as GPU and FPGA (field programmable field array) can be used to increase the target detection rate for bigger resolution images.

There are multiple ways to further improve the 3D reconstruction based phenotyping system in future research:

1. A better 3D camera can help to capture the small leaves for 3D reconstruction, can reduce the noise level of the final 3D reconstruction of the plant, and get better physical parameter measurement accuracy. Alternative 3D imaging sensors, such as the active stereo camera, features significant higher spatial resolution and accuracy with less noise, making it more favorable to the indoor phenotyping application.
2. The 3D reconstruction of this research is purely based on the 2D camera's pose estimation system. In order to improve the 3D reconstruction accuracy, the algorithm to evaluate the misalignment of the 3D image views would need to be developed to adjust the 3D registration accordingly.
3. A meshing algorithm to construct plant surfaces from the point cloud is needed.

4. In order to meet the need of phenomics to collect diverse phenotypic features, it is necessary to extend the phenotypic feature estimation algorithms to collect other data, such as plant volume, stem diameter, etc.
5. To expand the power of the 3D reconstruction based phenotyping system, in future research, the 3D reconstruction of the plant could be combined with other texture, such as color, chlorophyll fluorescence image, and thermography, to study pathogen infection and tolerance, transpiration, and photosynthetic response corresponding to a specific part of the plant.