Automatic Plant Annotation Using 3D Computer Vision

Nielsen, Michael

Publication date:
2011

Document Version
Early version, also known as pre-print

Media Technology

Ph.D. Dissertation

Automatic Plant Annotation Using 3D Computer Vision

Michael Nielsen

Faculty of Engineering and Science

Aalborg University 2011

About the Author

Michael Nielsen started his master's studies in Electronic Engineering at Aalborg University in 1997. He graduated in 2002 at Aalborg University with his major in outdoor Computer Vision for traffic surveillance. He worked as a research assistant on HCI gesture interfaces before starting his Ph.D studies. Key interests are interdisciplinary computer vision applications, combining human-centered ideas with the technical, investigating domains that are often overlooked.

His goal is to create innovative media products that will merge the real world with the electronic world for enhanced productibility and immersion. The systems should be a natural part of modern life and thus be aware of the dynamic environment that surrounds them.

Below is a list of papers that are related to the ph.d project documenting the results achieved during the work. Furthermore, there are a series of other papers from the work as research assistant and post-doc listed as "unrelated publications".

Related Publications

Michael Nielsen and Hans Jørgen Andersen. Plant and leaf analysis based on 3d reconstruction. In *Agricultural and Biosystems Engineering for a Sustainable World, International Conference on Agricultural Engineering and Industry Exhibition, Hersonissos, Crete, 23-25 June*, 2008.

Michael Nielsen, Hans Jørgen Andersen, David Slaughter, and Erik Granum. Ground truth evaluation of computer vision based 3d reconstruction of synthesized and real plant images. *Precision Agriculture*, 8(1-2):49–62, 2007.

Michael Nielsen, Hans Jørgen Andersen, and Erik Granum. Comparative study of disparity estimations with multi-camera configurations in relation to descriptive parameters of complex biological objects. In O. Hellwich, I. Niini, C. Ressl, V. Rodehorst, D. Scharstein, and P. Sturm, editors, *BenCOS - Towards Benchmarking Automated Calibration, Orientation and Surface Reconstruction from Images. ISPRS Workshop in conjunction with ICCV 2005*, pages 63–68. ISPRS Working Groups - WG III/1 Automatic Calibration and Orientation of Optical Cameras - WG III/2 Surface Reconstruction, October 2005a.

Michael Nielsen, Hans Jørgen Andersen, David C. Slaughter, and Erik Granum. Ground truth evaluation of 3d computer vision on non-rigid biological structures. In J.V. Stafford, editor, *Precision Agriculture 05*, pages 549–556. Wageningen Academic Publishers, The Netherlands, June 2005b.

Michael Nielsen, Lene Krøl Christensen, and Hans Jørgen Andersen. Sub-leaf scale remote sensor for npk discrimination using stereo vision. In *Engineering the Future, International Conference on Agricultural Engineering, Leuven, Belgium, 12-14 September, Session 10, no. 327*, 2004a.

Michael Nielsen, Hans Jørgen Andersen, David C. Slaughter, and D. Ken Giles. Detecting leaf features for automatic weed control using trinocular stereo vision. In *International Conference on Precision Agriculture, Minneapolis, MN, USA*, July 2004b.

Unrelated Publications

M. Nielsen, D.C. Slaughter, and C. Gliever. Stereo vision blossom mapping for automated thinning in peach. In *Industrial Electronics (ISIE), 2010 IEEE International Symposium on*, pages 499 –504, 2010. doi: 10.1109/ISIE.2010.5637846.

Michael Nielsen, Moritz Störring, Thomas B. Moeslund, and Erik Granum. *Gesture User Interfaces*, page In Press. Morgan Kauffman, Elsevier, 2008. ISBN 978-0-12-374017-5.

Claus B. Madsen and Michael Nielsen. Towards probe-less augmented reality : a position paper. In Joo Madeiras Pereira Jos Braz, Nuno Jardim Nunes, editor, *Proceedings: GRAPP 2008*, pages 255–261. Institute for Systems and Technologies of Information, Control and Communication, 2008.

Michael Nielsen and Claus B. Madsen. Segmentation of soft shadows based on a daylight- and penumbra model. In A. Gagalowicz and W. Philips, editors, *Proceedings of the MIRAGE 2007, Computer Vision / Computer Graphics Collaboration Techniques and Applications*, pages 341–352. Springer, March 2007a.

Michael Nielsen and Claus B. Madsen. Graph cut based segmentation of soft shadows for seamless removal and augmentation. In *Proceedings on the Image Analysis, 15th Scandinavian Conference, SCIA 2007, Aalborg, Denmark*, pages 918–927. Springer, June 2007b.

Michael Nielsen and Claus B. Madsen. Shadow segmentation and augmentation using α-overlay models that account for penumbra. In Søren I. Olsen, editor, *Proceedings fra den 15. Danske Konference i Mønstergenkendelse og Billedanalyse*, pages 60–69. DIKU Technical Report No. 06/08, August 2006.

Michael Nielsen, Moritz Störring, Thomas B. Moeslund, and Erik Granum. A procedure for developing intuitive and ergonomic gesture interfaces for hci. In Antonio Camurri and Gualtiero Volpe, editors, *Gesture Workshop*, volume 2915 of *Lecture Notes in Computer Science*, pages 409–420. Springer, 2003. ISBN 3-540-21072-5.

Automatic Plant Annotation Using 3D Computer Vision

A Ph.D. dissertation

by

Michael Nielsen

Section for Media Technology

Department of Architecture, Design & Media Technology

Faculty of Engineering and Science

Aalborg University, Denmark

E-mail: michael@sequoiagrove.dk

URL: http://www.cvmt.dk/∼mnielsen

March 2011

This report was typeset by the author using LaTeX 2_ε.

ISBN 978-87-992732-3-2

This dissertation was submitted in July 2008 to the Faculty of Engineering and Science, Aalborg University, Denmark, in partial fulfillment of the requirements for the Doctor of Philosophy degree.

The defense took place at Aalborg University, Niels Jernes Vej 14, DK-9220 Aalborg on 7th November 2008. The session was moderated by Associate Professor Thomas B. Moeslund, Section for Media Technology, Department of Architecture, Design & Media Technology, Aalborg University.

While the first edition was approved, this second edition includes revisions in accordance with comments from the adjudication committee.

The following adjudication committee was appointed to evaluate the thesis. Note that the supervisor was a non-voting member of the committee.

Professor Horst Bischof
Institute for Computer Graphics and Vision
Graz University of Technology
Austria

Associate Professor Henning Tangen Sgaard
Engineering College of Aarhus
Denmark

Associate Professor Claus B. Madsen
Department of Media Technology
Aalborg University
Denmark

Professor Erik Granum, Ph.D. (supervisor)
Computer Vision and Media Technology Laboratory
Department of Media Technology
Aalborg University
Aalborg, Denmark

Abstract

In this thesis 3D reconstruction was investigated for application in precision agriculture where previous work focused on low resolution index maps where each pixel represents an area in the field and the index represents an overall crop status in that area. 3D reconstructions of plants would allow for more detailed descriptions of the state of the crops analogous to the way humans evaluate crop health, i.e. by looking at the canopy structure and check for discolorations at specific locations on the plants.

Previous research in 3D reconstruction methods based on cameras has focused on rigid frontoplanar scenes such as buildings and rooms with billboard-like figures. Rigidness allows for advanced methods using structured light, laser scanning, object models for reference, etc.

Plants can be described as non rigid biological objects with complex structures, i.e. not frontoplanar but rather sloped surfaces. Plants wave in the wind, have complex structures with overlapping semitransparent leaves and have little texture variation and specular highlights. Little work had been done in this area previously.

When analyzing 3D results from complex structures it is very difficult to obtain good quality dense ground truth of disparities. Therefore, a test framework was developed based on ray tracing. The goal was to analyze existing methods for disparity map generation. The major problem for existing methods was the steepness of the leaves relative to the closeness of overlapping leaves. Both sum-of-squared difference methods and energy-minimizing methods had this problem.

Following the test a series of disparity estimation techniques were developed and tested in the test framework using a set of ray traced images and a hand-annotated set of real plants with similar plant shapes.

Novel similarity measures were developed that could lead to better disparity estimations on sloped surfaces for a multi-baseline 5-camera setup and trinocular setups. In the multi-baseline setup the developed method showed improvements mainly in areas with specular highlights. The methods using a trinocular setup showed better reconstruction in occluded areas.

The trinocular setup was used for both window correlation based and energy minimization based algorithms. A novel adaption of symmetric multiple windows algorithm with trinocular vision was developed. The results were promising and allowed for better disparity estimations on steep sloped surfaces.

Also, a novel adaption of a well known graph cut based disparity estimation algorithm with trinocular vision was developed and tested. The results were successful and allowed for better disparity estimations on steep sloped surfaces.

After finding the disparity maps each individual leaf had to be seperated using a simple labeling algorithm based on connected components analysis (based on Rosenfeld and Pfaltz union-find algorithm) where height information was also included, described as a NURBS surface and annotated with Number of leaves on plant, Area of plant and individual leaves, Leaf steepness, Height of leaf relative to ground, and Size (bounding box).

The new disparity map estimation methods were able to handle piecewise smooth sloped surfaces better to such a degree that individual separation of the leaves was possible and the extracted information was more accurate than using existing methods.

In order to allow for spectral reflection sampling at designated spots on the plants it was necessary to find tips and bases of each leaf. The results were promising but could be refined using knowledge about surface normals.

2D computer vision research has been done in active shape modeling of weeds for weed detection. Occlusion and overlapping leaves were main problems for this kind of work. Using 3D computer vision it was possible to separate overlapping crop leaves from weed leaves using the 3D information from the disparity maps.

The results of the 3D reconstruction and extracted features can be used to locate sample locations for multi-spectral reflection analysis. It can be employed during early growth stages in low cost crops or in high cost crops where the grown plants are placed separately. The image acquisition can be done from manually operated machinery or a field robot or a self guided tractor following a sample strategy based on overview maps of the field.

Preface

The work in this dissertation is part of the project ACROSS which is part of the National Research program "Sustainable Technology in Agriculture" ("Bæredygtig teknologi i Jordbruget"), supported by the Danish Technical Research Council, the Danish Agricultural and Veterinary Research Council and the Danish Ministry of Food, Agriculture and Fisheries.

ACROSS is an acronym for "Autonomous spatial-temporal crop and soil surveying". The general vision is effective precision farming, which in harmony with the environment utilises resources optimally. This requires continuous selective and adaptive control of growth, weeds, diseases and pest. In turn such control is conditioned on corresponding continuous monitoring in the field using appropriate methods of measuring the current conditions of and for the plant growth.

The technical field of study is computer vision applied in precision agriculture. Hence, there are two main disciplines that are relevant in this project:

- Scientific discipline: Computer Vision
- Application domain: Precision Agriculture

During my work I made some external connections. On practical experiments I worked together with Lene K. Christensen from the Royal Veterinary and Agricultural University and with David Slaughter and Ken Giles from University of California in Davis.

The work was done in 2003-2005. In the beginning of this thesis there is a list of definitions.

Thanks to all those I worked with and made this possible.

Michael Nielsen

Aalborg, Denmark, July 2008

Contents

Preface i

Table of Contents iii

List of Definitions vii

1 Introduction 1

1.1 The Challenge for Computer Vision 2
1.2 Aim of the study 3
1.3 3D Reconstruction 3
1.4 Objectives 4
1.5 Focus 5
1.6 Collaboration 5
1.7 Outline 6

I Analysis 9

2 3D Computer Vision in Precision Agriculture 11

2.1 Precision Farming 12
2.1.1 Taxonomies and Concepts 12
2.1.2 State of Precision Farming 12
2.1.3 Current Development 14
2.1.4 The Standing Problems 15
2.2 3D based computer vision sensor 16
2.2.1 Sampling Strategies 17
2.2.2 Usage of the 3D based computer vision sensor 18
2.2.3 Establishing ACROSS in the Community 19

2.3 Summary 20

3 Preliminaries **21**
3.1 Descriptive Object Parameters 21
3.2 Structural Descriptions and Other Annotations 22
3.3 Camera setup 23
3.4 Leaf Color Segmentation 24
3.5 Window Correlation based Disparity Estimation 25
3.6 Energy Minimization 26
3.7 Dense versus Sparse Ground truth 27
3.8 High Quality Disparity Maps 28

4 Ray-Traced Plant Images for Dense Ground Truth **31**
4.1 Test Framework 31
4.1.1 Additional Image processing 33
4.1.2 Occlusion Mask 33
4.1.3 Highlight Mask 33
4.1.4 Shadows Masks 34
4.1.5 Quality Metrics 34
4.1.6 Statistical and Graphical Testing 35
4.1.7 Image examples 35
4.2 Comparative Study of Disparity Estimations 35
4.2.1 Results 36
4.2.2 Discussion 40
4.2.3 Future Work 40
4.2.4 Conclusions 41

II Improving Disparity Estimations **43**

5 Multi-baseline Camera Setup **45**
5.1 Methods and Material 46
5.1.1 Introducing SISSD 46
5.1.2 Comparable Methods 50
5.1.3 Experimental Setup 50
5.2 Results and Discussion 51
5.3 Conclusions 54

5.3.1 Perspectives on Future Work 55

6 Combined SMW **59**
6.1 Method 60
6.2 Results 62
6.3 Conclusion 62

7 Trinocular Graph Cuts for Piecewise Smooth and Sloped Surfaces **65**
7.1 Theory 65
7.1.1 Energy Formulation 66
7.1.2 Graph Construction 68
7.2 Sloped Extension 72
7.3 Results 74
7.4 Conclusion 81

III Structural Representations **85**

8 Surface Reconstruction **87**
8.1 Segmenting Disparity Maps 87
8.2 Non-Uniform Rational B-Splines 88
8.3 Fitting NURBS 89
8.3.1 Fitting Multiple Views 90
8.3.2 Matching the labels in multiple views 91
8.3.3 Choosing the Weights 92
8.4 Generation of the Mesh 92
8.4.1 Looking behind the disparity map 93
8.4.2 Reconstruction Performance 95
8.5 Extraction of Information 95
8.6 Conclusion 99

9 Automatic Annotation of Structures **103**
9.1 Automatic Annotation of Barley for Classification of NPK Deficiencies 104
9.1.1 Materials and Methods 105
9.1.2 Results 108
9.1.3 Discussion 110
9.1.4 Summary 113
9.2 Classification of Weeds in Tomato fields 114

9.2.1 Materials and Limitations . . . 115
9.2.2 Methods . . . 115
9.2.3 Test . . . 123
9.2.4 Results . . . 123
9.2.5 Discussion . . . 125

IV Results **127**

10 Contributions **129**
10.1 Results at Disparity Level . . . 129
10.2 Results at Plant Level . . . 130
10.3 Discussion . . . 135
10.3.1 Test Framework . . . 135
10.3.2 Disparity Estimation and Surface Reconstruction . . . 135
10.3.3 Structural Representations . . . 137

11 Summary **139**
11.1 Perspectives . . . 141
11.2 Acknowledgments . . . 141

Bibliography **143**

List of Figures **147**

A Trinocular Rectification **155**

B Reconstructed Plant Models **159**
B.1 Virtual Plant Set . . . 159
B.2 Real Plant Set . . . 166

C Another Construction of a Graph **173**
C.1 Practical example . . . 174

List of Definitions

ACROSS	Acronym for the research project Autonomous spatial-temporal CROp and Soil Surveying.
Combined SMW	One of the contributions in this thesis based on the SMW algorithm but it works better for steep sloped surfaces.
EM38	A soil surveying sensor based on inductance measuring soil salinity.
GrC	The energy minimizing graph-cut based algorithms designed by Kolmogorov and Zabih in 2001. In literature called KZ1.
GrC Sloped Extension	One of the contributions in this thesis based on the GrC algorithm but it works better for steep sloped surfaces.
Multi-baseline	Camera setup using a number of cameras sharing the same baseline.
NDVI	Normalized Difference Vegetation Index. An index roughly correlated to crop health in an area based on spectral reflectance measurements.
NPK	Nitrogen, Phosphorous, Potassium (K).
PBMP	Percentage of Bad Matching Pixels. A disparity quality measure that tells how well the structure is preserved.
Precision Agriculture	The art of using information systems for variable rate application and management of large farms. Typical uses of technology includes satellite-, airplane, or miniature model plane photography, GPS systems, auto guided vehicles, and soil-, crop-, and yield- sensors.
Precision Farming	See Precision Agriculture.
Remote Sensing	Non-destructive no-touch sensors for application in agriculture.
RMS error	Root-means-Squared error. A disparity quality measure not well suited for telling how well the structure is preserved.
SISSD	Sum of Independent Sums of Squared Difference. One of the contributions in this thesis based on the SSSD algorithm but it works better for specular highlights and theoretically for steeped sloped surfaces where projection distortions occur more excessively.
SMW	A SSD based disparity algorithm using Symmetric Multiple Windows. When searching for optimal window match the SSD window is tested with a centered window as well as off-centred windows. Each pixel pair is thus tested 9 different ways.
SMW1	Short for using the SMW algorithm with only a centered window.
SMW5	Short for using the SMW algorithm with only a centered window, and corner-placed windows.
SMW9	Short for using the SMW algorithm with centered window, corner-placed windows, and side-placed windows.

SSD ..... Sum of Squared Difference. Here used as a window based similarity measure for photo-consistency.

SSSD ..... Sum of Sums of Squared Difference. Multi-baseline disparity map algorithm where all the sums of SSD across all cameras along the baseline is used as a similarity measure.

Trinocular ..... Camera setup using 3 cameras with 2 orthogonal baselines. Also called L-setup.

Chapter 1

Introduction

Precision agriculture is an interdisciplinary field that utilizes Information technology, sensor technology, robotics, and agronomic sciences, etc. Precision agriculture is about handling spatial and temporal variability in fields and automation of the agricultural tasks.

Remote sensing is a valuable tool for managing the variability of crop health in precision agriculture. Remote Sensing techniques are based on satellite images, digital photography or spectral sensing from planes, tractors, or miniature model planes/helicopters. They calculate simple indexes such as Normalized Differential Vegetation Index (NDVI) from two spectral wavebands. For the sake of simplicity NDVI is correlated with nitrogen or any other nutrient deficiency or disease in focus of a given study. The spatial resolution is very low, i.e. at best 1 m^2 per pixel, because each pixel in the image will be a spectral averaging of the area it represents.

An approach such as NDVI maps cannot distinguish the countless causes for the measured symptoms. In [Christensen and Jørgensen, 2003] a fixed set of locations are isolated on barley where N, P, and K stress can be detected using hyper spectral reflectance analysis. A human being can differentiate between many stresses, and can do so because of detailed features on the plants, and the possibility to check the roots and soil. Many factors affect the measured symptom. For example, water stress and soil compaction affects the ability to utilize the nutrients. The features that can be important for diagnosis include:

- General color of leaves
- Locations of discolored areas on individual leaves
- Curliness of leaves and their edges
- Posture of leaves/canopy

Consequently, it is necessary to create a high resolution sensor that can base its decision on more features than the existing NDVI based sensors, which base their decisions on the general reflectance in an area. The hypothesis is that computer vision can be used to create such a sensor through 3D stereo vision. With this technology a sensor with a resolution below 1 mm^2 can be developed.

In this work the term *remote sensing* is extended to include a non-destructive no-touch sensor based on a camera which is under 1 m away from the crops.

1.1 The Challenge for Computer Vision

There are many challenges in computer vision when working with precision farming, many of which need research. One aspect is the fact that it takes place outdoors with arbitrary sunlight and weather conditions. Many computer vision solutions are solved by controlling the light source. Another aspect is that computer vision techniques rely on prior knowledge of the observed object, such as optimizing the detection of a known surface or the corners. This is especially easy to do when the objects are rigid man-made structures such as buildings. In agriculture the sensors are looking at biological, free-shaped, semitransparent structures.

Spectral analysis of reflections a 3D reconstruction together with analysis of geometrical structural features of the crops will enable the system to distinguish the causes for the variability in the field or to detect alien obstacles in the field.

This will not be the first computer vision based sensor in precision agriculture. Some projects research computer vision based weed detection for automatic weeding, which can be chemical weeding or mechanical weeding [Lee et al., 1999] [Assémat and Chapron, 2003] [Woebbecke et al., 1995] [Zhang and Chaisattapagon, 1995]. Computer vision methods are also used for row tracking, stone detection [Chapron and Huet, 2003], and fertilizer density measurement. Relevant and interesting research at CVMT and the Silsoe Institute has been devoted to spectral analysis with changing illumination, specifically the case with sunlight[Andersen, 2001].

It is notable that most work has been in the 2D domain on image processing and weed detection, whereas humans often look at the (3D) canopy and color spots on the crops for evaluating crop health combined with other knowledge about growth conditions.

In stereo vision based 3D reconstruction algorithms rely on the ability to correlate points in images captured from different viewpoints. More viewpoints usually means better reconstruction. Temporal information can also be used. When looking at outdoor plants that are waving in the wind it is necessary to capture all views simultaneously. If the images are to be captured from a moving vehicle, the lighting also have to be good enough to use a fast shutter speed.

The actual correlation of points between the views is difficult because the leaf material is mostly just green with little variation. Using lighting with a bit of shading and a color space which is not independent of this shading might be an advantage. Furthermore, the correlation is more difficult because the surfaces are often glossy which makes specular highlights.

The shapes of the leaves may be upright and closely overlapping which makes occlusion a dominant issue. The also have a high degree of natural variability and may be broken by a e.g. rabbit, tractor, or the wind, which makes it difficult to use knowledge about the shape. However, they can be described is sloped piecewise smooth surfaces in general terms, which is a new research area within the field of 3D reconstruction.

1.2 Aim of the study

The aim in this thesis is to investigate the possibility to use 3D computer vision to describe and extract features from crop plants. These representations would allow for automatic localization of spectral sampling points. The spatial information together with the spectral reflectance measurements would allow for more accurate diagnostics of crop health. This makes the approach analogous to the farmer manually inspecting the individual plants, except that the cameras can see the near infrared colors as well.

In the following a set of considerations are presented regarding the problem at hand.

1.3 3D Reconstruction

In order to achieve 3D reconstruction of the scene, it is necessary to see the scene from different viewpoints, either by moving the camera or to have multiple cameras. Then the correspondence between each pixel (or feature) in the different views need to be found. If the scene is expected to change during the image capture, it is best to use multiple cameras so that the different views are captured at the same time. This is the case with outdoor plants because the wind causes the crops to move.

Existing algorithms can be divided into sparse (few correspondences) and dense (correspondence for most or all pixels) correspondence algorithms. The ones giving sparse correspondence maps (disparity maps) use more complex similarity measures (such as SIFT features [Lowe, 2004]) and find correspondence between the features. The ones giving dense correspondence maps use simpler similarity measures, such as sum of squared difference correlation windows. The best dense algorithm is called Symmetric Multiple Windows (SMW) [Fusiello et al., 2000] which features a test for occlusion and is relatively fast to compute and does not assume rigid surfaces. Furthermore, the algorithms can be based on energy minimization, but is very computationally comprehensive. The best of these treats the disparity estimation as a segmentation problem and minimizes the energy with graph cuts [Kolmogorov and Zabih, 2002], but it assumes fronto-planarity in the scenes. Lin and Tomasi [2004] proposes an extensive adaption that allows sloped surfaces by fitting b-spline surfaces to the segmentation and minimize the energy function for the surfaces. Theses are all tested on a common test set[1] with ground truth where the scenes are more simple than a crop scene and they are only binocular.

3D analysis of free-shaped biological objects such as plants is not trivial because of the following problems:

Lack of geometrical A-priori knowledge and rigidness

Biological objects can have arbitrary orientations, occlusions, and disparity discontinuities, so there is little knowledge about surfaces, corners, straight lines, etc. to take advantage of. Furthermore, they may wave in the wind or by the airflow of a computer fan. This makes it impossible to use any multi-shot approach such as structured light and rotating camera approaches.

[1] The typical image sets used to test binocular stereo are called Tsukuba, Sawtooth, Venus, and Map. From http://cat.middlebury.edu/stereo/data.html

Occlusions and point of view

Free-shaped objects have potentially many occlusions. These may be reduced by using multiple views. They are different problems depending if they are in the reference view or the other views. In the reference view they become hidden from the system, i.e. those parts are unknown. If a structure in the reference view is partially or fully occluded in one of the other views it will lead to bad matches unless the occluded regions are detected as such.

Depth information versus correspondence accuracy

The baseline is the distance between pairs of camera views. A large baseline produces detailed depth information, but it also makes the correspondence between the cameras worse and the search space large. An overlooked problem in the literature is that the 2D projection of an object (e.g. a grass-like leaf) that is pointing toward the cameras has different orientations in the different views. The larger the baseline is, the more different the projections are. This means that the algorithm cannot trust that the similarity measure is best at the correct match.

Orientation of the objects

With free-shaped structures it is impossible to predict their alignment with the camera axes. Leaves may be elongated and possibly aligned with the epipolar line between the camera pair in question. Such leaves are difficult to reconstruct, because the search space in the 2nd camera becomes very large. In other cases the leaves may point toward the camera and look very different to the different views as explained above. Steep slopes may also be a problem for existing algorithms especially energy minimizing algorithms where disparity constancy is assumed. Especially when other overlapping leaves may almost touch each other within few millimeters.

Getting Ground Truth

Quantitative evaluation of computer vision methods is difficult with non-rigid biological structures, because geometric ground truth is very difficult to establish.

Various experimental studies for which ground truth is essential will be carried out in order to investigate the above challenges.

Ground truth is important in different layers of abstraction:

- *low level* Disparity maps, occlusion maps, etc. Very difficult to obtain.
- *high level* Object classification, object area, no. of parts, etc. Easier to obtain, but of course it depends on size of the data set.

1.4 Objectives

The following provides an overview of how the problem has been turned into action. The overall problem statement that encapsulates the research was:

How is it possible to reconstruct and describe objects that consist of overlapping, sloped, piecewise smooth surfaces such as plants and extract appropriate features from such objects?

A number of tasks was addressed to solve the problem in this new domain for 3D reconstruction where the scene consists of freely shaped objects for which ground truth is unavailable.

- Development test framework providing dense ground truth in order to analyze problems with existing correlation based and energy minimizing methods. The aim was a set of virtual plant images with detailed information about disparities, occlusions and highlights.
- Development of new methods for improving the disparity estimations for usage on plant structures. The aim was better structure preserving disparity maps, especially facilitating subsequent 3D representation and annotation of individual leaves.
- Development of methods for representation and annotation of the plant structures. The aim was to extract parameters relevant to precision agriculture and spot location which allows for e.g. hyper-spectral reflectance sampling in future work.
- Case studies of 3D computer vision for specific agricultural tasks in practice. The aim was to demonstrate the usage of 3D computer vision for typical tasks in agriculture in the field, demonstrating that the approach is not limited to laboratory imagery.

1.5 Focus

This project has a very broad definition and there are many questions to investigate. It is therefore not possible to analyze all aspects and possibilities in all areas. The solutions will be application driven and should be seen as examples of how it can be done. The experiences will be discussed in the end and their applicability in general will be evaluated for future researchers to benefit from.

An overview of the process of Plant 3D reconstruction is shown in figure 1.1. The core in the 3D reconstruction and the research is Disparity Estimation and the extraction of features for use in precision agriculture. It follows from this focus that basic algorithms will used for isolation of individual leaves and surface fitting. Entire research programs could be devoted to the surface fitting approaches as well. The focus is chosen with the principle in mind that corrupt input data leads to corrupt output data.

Within this work the extracted features are limited to spatial and structural features, i.e. spectral data is left for future work.

1.6 Collaboration

Test cases from precision agriculture was chosen in collaboration with partners from agricultural engineering for the project. The cases gave insight in the extent of the problems and provided information about the requirements for the computer vision methods regarding which data needed to be extracted. Two cases have been worked with: nutrient deficiency in barley (low cost crop) and weed detection in transplanted tomato fields (high cost crop).

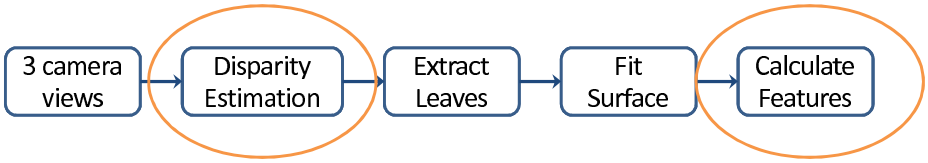


Figure 1.1: Overview of the relevant parts and the focus of this thesis. Emphasis is on generating high quality disparity maps for extraction of individual leaves and the automatic annotation of them.

1.7 Outline

The rest of the dissertation is divided into three parts:

I Analysis

II Improving Disparity Estimations

III Structural Representations

IV Results

Part I contains the problem analysis regarding contextual placement within precision agriculture of the perspectives of a 3D reconstruction based sensor, the algorithms and test approaches that were pursued. Then the development of a ray tracing based test framework that had to be made for the work is presented. It was used to uncover the exact performance problems of existing algorithms. The results of these tests pinpointed that the major problem was the steep slopes in the plant canopies. An annotation tool for real plant images was developed and a set of real plants were manually annotated. The results obtained from the virtual plants could anticipate the performance on the real plants only when all error sources in the real world image acquisition were simulated.

Part II contains the work on improving the performance of the early vision algorithms that would solve the problems found in the benchmark test. Both correlation based and energy minimizing graph cut based algorithms were made for comparison. It also contains an interesting discovery in using a multi baseline camera setup for handling specular highlights.

Part III contains the work on finding automated annotations with structural descriptions of the plants. The disparity maps were labeled using an adaption of Rosenfeld and Pfaltz union-find algorithm that separated individual leaves. Each leaf was then represented as a NURBS surface and was converted to a triangle mesh. Various features of the plant were extracted from this. The tips and bases of grass-leaves were also found for application on NPK stress in barley. There is also an experiment with using leaf height to detect weeds on transplanted tomato fields and to make spray maps.

Part IV contains the test similar to the one in part I but this time with the methods developed in part II. The new methods were found to be an improvement and made it possible to extract individual leaves and features from the plants. Followed by discussion and conclusions on the dissertation.

Each chapter starts with summaries of the motivation for the chapter and ends with concluding remarks.

Part I

Analysis

Chapter 2

3D Computer Vision in Precision Agriculture

Summary:

The 3D computer vision sensor approach was developed as part of a project called ACROSS. The field of application is precision agriculture remote sensing. This chapter places the work within this context and investigates this research field and how it should be applied. Such a sensor was found to be relevant and easy to combine with other research activities in the field. It can be attached to existing field vehicles as well as autonomous vehicles and robots. The data in combination with other sensor data will help make an intelligent decision guiding variable rate application of fertilizers and pesticides.

The development of the 3D based computer vision sensor was part of a project called ACROSS. ACROSS is the acronym for the project "Autonomous spatial-temporal CROp and Soil Surveying". The project is part of the National Research program "Sustainable Technology in Agriculture", supported by the Danish Technical Research Council, the Danish Agricultural and Veterinary Research Council and the Danish Ministry of Food, Agriculture and Fisheries.

This thesis addresses the part of ACROSS which aims to create a sensor using 3D based computer vision which implements the "farmer's eye" for detection of crop status and needs in agriculture, such as nutrition deficiency and crop disease.

Sustainable agriculture integrates three main goals, environmental stewardship, farm profitability, and prosperous farming communities. Precision farming or precision agriculture is an agricultural concept relying on the existence of in-field variability. Precision farming may be used to improve a field or a farm's management from several perspectives:

- Agronomical perspective: Adjustment of cultural practices to take into account the real needs of the crop
- Technical perspective: Better time management at the farm level
- Environmental perspective: Reduction of agricultural impacts

- Economical perspective: Increase of the output and/or reduction of the input, increase of efficiency

The relationship between these two concepts is that precision farming is a tool which assists obtaining the goal, sustainability, in sustainable agriculture. This is done through management and technology, and most importantly through the use of information. Consequently, precision farming needs technology for information gathering, and information analysis and management.

This section will investigate how ACROSS fits into the current state and development in precision farming, and the requirements that needs to be fulfilled for the work to assist sustainable agriculture.

First, this section will contain the analysis of:

- General concepts
- State of precision farming
- Current development
- The standing problems

Then the possible practical scenarios for its usage will be discussed.

2.1 Precision Farming

This section presents the current state and development in precision farming, and explains some of the different principles and concepts used herein.

2.1.1 Taxonomies and Concepts

Precision farming deals with variability [Blackmore et al., 2003]. There are two main types of variability to deal with:

1. Spatial Variability - in-field variation, because uneven yield, caused by e.g. soil variation, static obstacles.
2. Temporal Variability - Variation within a year and from year to year, caused by e.g. weather, pests, nutrition stress.

An important area of precision farming is to automate processes with autonomous vehicles. There are different approaches that is described in table 2.1.

2.1.2 State of Precision Farming

Presently the field variability is dealt with by mainly manual or semi-automatic methods.

Table 2.1: Concepts in precision agriculture

Term	Description
Vehicle	Can be (1) an existing vehicle, e.g. an auto steering tractor, which is augmented for precision farming, or (2) a new designated vehicle (robot).
Inter-row	Process that works between rows in the field.
Intra-row	Process that works between crops within each row in the field. Implementation is difficult as to not destroy crops.
Sub-canopy [Müller et al., 2003]	Vehicle that moves under the canopy of the crops. Can get close to the crops without destruction.
Super-canopy	Vehicle that moves over the canopy of the crops. Risk destruction.
Automated Guided Vehicles (AGV)	Vehicle without a driver which follows a predefined route and actions. It is not aware nor can react to the surroundings.
Self Guided Vehicles (SGV)	Vehicle without a driver which is aware of its surroundings and can react to it.
Remote Sensing	Non-destructive non-contact sensor that measures the state of the crop from a distance.
Map-based Variable Rate Application (VRA)	Information is gathered in the field, and the data is centrally analysed for planning of the actions to be applied in the field.
Sensor-based VRA	The actions to be applied in the field and the information analysis are performed on-the-fly.

Manual soil samples are done to measure soil pH, soil compaction with a penetrometer, and samples for laboratory analysis measure organic material in the soil.

Manual sampling is time costly and expensive, and near impossible in large fields.

Automatic soil measurement with EM-38 gives an inaccurate map of the soil quality, defined by clay content. It measures electrical conductivity, which is highly influenced by water. It is rather expensive.

This data can be presented in soil maps, which can be used mainly to predict spatial variability. Yield maps are collected in the combine with GPS at harvest [Blackmore et al., 2002b]. The map shows what happened when it is too late to do anything about it. Data from a series of years can be used to predict spatial variability, and show the temporal variability in the field. They are also converted into money maps, showing the distribution of profit in the field.

Converting soil maps and yield maps into maps for map-based VRA is still complicated, and involved a lot of error correction. Errors are introduced by the sensors used and interpolation techniques (e.g. Krieging) [Gangloff and Westfall, 2003].

Detecting in-field variation in time to help the situation can be done manually with the handheld N-Tester or perform destructive measurements. This is not a feasible solution for medium-large fields, because it is time consuming. It only measures nitrogen requirement.

Remote Sensing mapping techniques includes satellite and aerial photography. They calculate a Normalized Differential Vegetation Index (NDVI, see Figure 2.1) from two spectral wavebands. NDVI is correlated with nitrogen. Problems arise with varying weather conditions and atmospheric disturbances. Especially the satellite images are expensive and have little control when to get the images. The spatial resolution is bad, because each pixel in the image will be a spectral averaging of the area it represents.

A sensor-based VRA, Hydro N-sensor, is commercially available. It calculates the N application based on a measure similar to NDVI from the area near the vehicle.

In an area where there is little water, or very hard soil, a system like this will waste unnecessary nitrogen. This is in direct conflict with the aim for sustainable agriculture, as the residue nitrogen on the soil will be worse than using conventional application.

2.1.3 Current Development

There is still a lot of research devoted to improving NDVI based decisions, for example, by improving picture quality by taking them from small remote controlled airplanes or choppers [Jensen and Young, 2003]. This way the problems on the images caused by atmospheric disturbances and clouds are removed.

Laser scanning technology is researched as a better estimate of biomass, gap fraction, and leaf angles. These measures will then be correlated with nutrient stress.

For autonomous data collection and application of fertilizer, vehicles are being developed. Super-canopy vehicles are used for data collection or map-based VRA. A sub-canopy vehicle for automatic weeding of Christmas trees is being developed [Blackmore et al., 2002a].

Some projects research computer vision based weed detection for automatic weeding, which can be chemical weeding [Lee et al., 1999] or mechanical weeding.

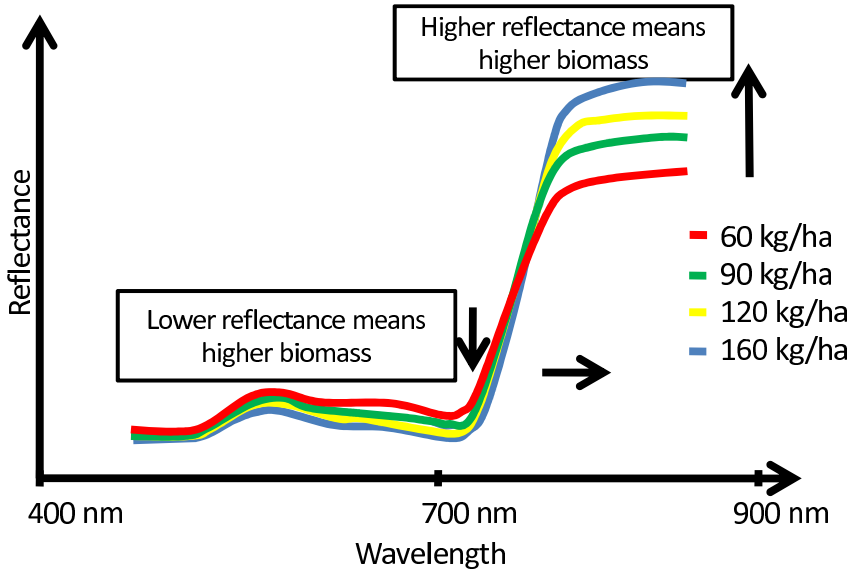


Figure 2.1: Leaf Reflectance at different N-application levels. NDVI is the ratio between 2 wavelengths, e.g. at 800 nm compared to 660 nm.

Computer vision methods are also used for row tracking, stone detection [Chapron and Huet, 2003], and fertilizer density measurement.

2.1.4 The Standing Problems

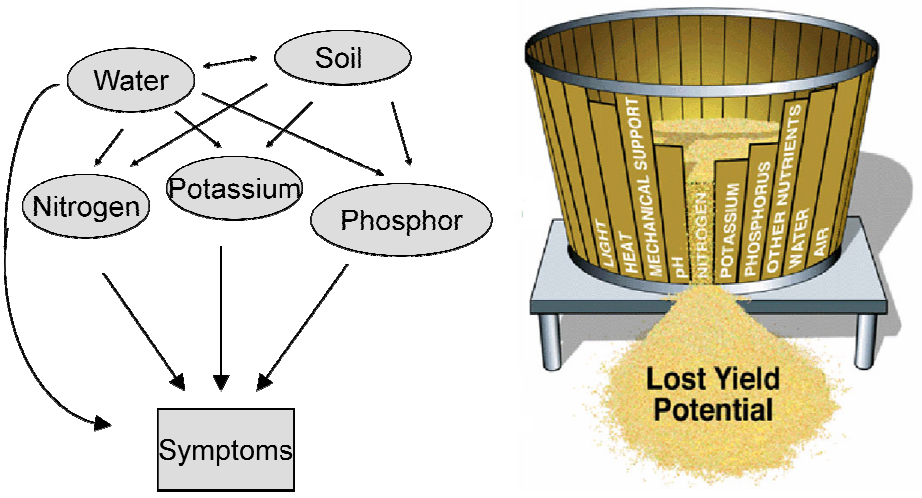
Soil analysis or vegetation reflectance analysis alone does not suffice to conclude nutrient stress. The existing non-destructive tools use too little information with bad spatial resolution. Hydro N-sensor, satellite and aerial photos use only spectral averaging. They all measure a symptom which can result from many causes (Figure 2.2 Left). They cannot separate different nutrient stresses or other factors such as water, soil compaction, and pest.

The limiting factor is the weakest link in crop health (Figure 2.2 Right). In VRA the goal is not necessarily less nutrients but zero residual nutrients. Detecting nitrogen deficiency does not equal that there is no nitrogen available in the soil. If the soil is too compact or dry, the pH is too low, or there are pest on the roots of the crops, the ability to consume the nitrogen is affected. Water and soil erosion can transport nutrients away from the crops, which means that application timing is important, too.

In [Christensen and Jørgensen, 2003] a fixed set of locations are isolated on barley, where N, P, and K stress can be detected using reflectance analysis.

A human being can differentiate between many stresses, and can do so because of detailed features on the plants, and the possibility to check the roots and soil.

Figure 2.2a Many factors affect the measured symptom. For example, water stress and soil compaction affects the ability to utilize the nutrients. Figure 2.2b The limiting factor is the weakest link in crop health. In this example it is Nitrogen which is limiting the yield.



(a) (b)

Figure 2.2: (a) A given symptom can come from many factors such as water, soil compaction, and pest. (b) The limiting factor is the weakest link in crop health.

These features that are important for diagnosis include:

- General color of leaves
- Locations of discolored areas
- Curliness of leaves and edges
- Posture of leaves

Consequently, it is necessary to create a sensor that can base its decision on more features than the existing NDVI or Hydro sensors, which base their decisions on the general reflectance in an area.

2.2 3D based computer vision sensor

This section relates the work to the topics presented in the previous section.

The goal is to make a sensor which can differentiate disease- and stress symptoms on crops.

Some of the problems in the previous section will be addressed by using 3D reconstruction of the plants, using two or three cameras from approx. 50 cm distance. These cameras are also sensitive in the near infrared wavelengths. This way the diagnosis is based on a high level of detail, including pattern recognition of the features mentioned that a human uses to detect the symptoms, plus the sub-canopy locations of hyper spectral information.

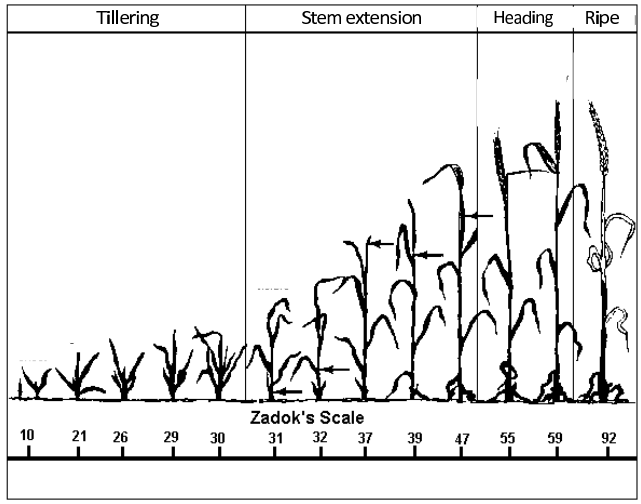


Figure 2.3: Growth Stages. Initial aim of the sensor is that it will be functional in the Tillering period. This is when fertilizer adjustment has most impact. Adapted from Agronomy Guide For Field Crops, Ontario Ministry of Agriculture, Food and Rural Affairs, 2003.

The work can easily be applied to pest, virus, and weed detection, but will focus on nutrition deficiencies. It is expected to perform better than simple reflectance based systems (e.g. NDVI, Hydro N-Sensor, Laser Scanning), because of the extra information used in the detection.

It is not decided whether the system will be used as sensor-based VRA or map-based VRA, where the information can be supplemented by soil map information. Aerial photography can aid the planning of sampling strategy.

The limitation of the sensor is: "if you cannot see it, you cannot detect it". This means that the disease must show visible signs, or visible in the near-infrared spectral response. It also means that it will work best at early growth stages where the canopies do not occlude each other (The tillering period). See Figure 2.3.

Choosing sampling scheme for ACROSS depends on the usage of the sensor.

2.2.1 Sampling Strategies

In precision agriculture soil- or crop sampling there are different sampling schemes [Morgan and Ess, 1997][Mulla et al., 2000]:

- Grid centre sampling
- Grid cell sampling
- Targeted sampling

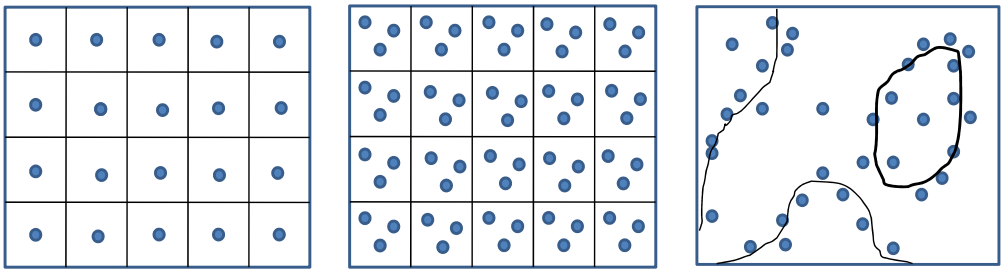


Figure 2.4: Grid sampling: The field is divided into a dense grid of cells, which can be sampled. Multiple samples from each cell can be averaged for better reliability. Targeted sampling: Sparse sample locations are selected intelligently.

Grid sampling divides the field into uniform cells. In grid centre sampling, the cells are samples in their centers. In grid cell sampling, the cells are sampled in clusters, which can be chosen randomly (as in figure 2.4) or systematically.

Grid sampling is expensive and requires many samples, using no a priori information of the spatial variability.

Targeted sampling is used to reduce the number of samples by using known information about the spatial variability to divide the field into zones. An example of targeted sampling is to use knowledge of the spatial variability to focus sampling around borders of the different areas.

In [Mulla et al., 2000] grid sampling schemes of 344-1375 samples were reduced to 17 targeted samples based on information from near-infrared aerial photography. The grid samples identified 3-4 zones, while the targeted samples identified only two, which covered almost the same pattern as the grid sample zones. Likewise, [Taylor and Whitney, 2001] reduced 52 grid samples to 21 targeted samples using yield maps.

Satellite- and aerial remote sensing sample the fields in a pixel grid (raster) . The pixel grid translates to a square sampling grid. A sample is an average measure of a cell.

2.2.2 Usage of the 3D based computer vision sensor

The 3D based computer vision sensor is basically a ground level sensor which can detect sub-canopy features in crops.

It is unlikely that it will be implemented as a sensor-based VRA system from the start. Potential for high processing requirement makes it likely that it will require targeted sampling for map-based VRA.

Best performance is expected if supplemented by soil maps and other information that cannot be extracted visually. When the system has showed good performance, and the processing power is adequate, a sensor-based VRA system can be introduced. In this case the sampling scheme can be grid cell sampling. The grid can be dense if processing power allows it.

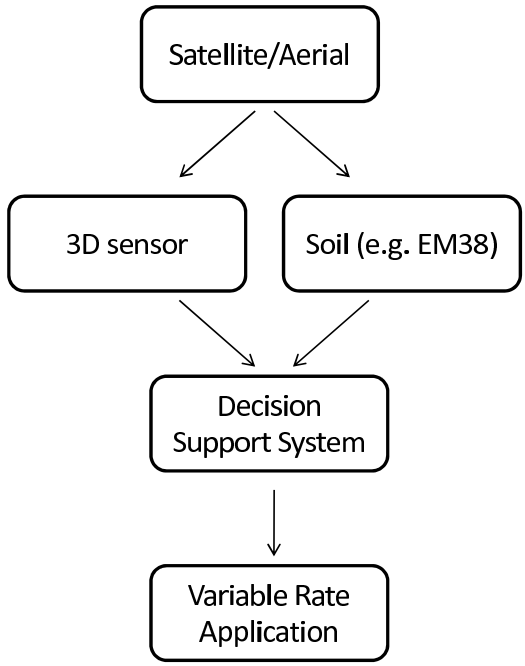


Figure 2.5: Combining resources. The 3D based computer vision sensor as part of a decision support system.

For example, if processing time is 2 seconds per image set, and the vehicle speed is 4 m/s. There will be an 8 meters decision delay, which can be countered by placing the sensor 8 meters ahead of the application sprayer. However, a more robust decision can be achieved with e.g. 3 samples per decision, resulting in a grid cell size of 32 meters, using a 3 sample systematic cluster. However, there will also be a spatial application delay of 32 meters from the first sample of that grid cell.

If the application is performed from an autonomous vehicle, the vehicle speed can easily be controlled. This way the sample grid can be improved. Information about spatial variability can be used to control the grid size near zone borders.

Figure 2.5 shows a potential system in which the 3D based computer vision sensor is part of. Large scale (distant) sensors, satellite/aerial images, can be used to target sampling strategies from the small scale (ground level) sensors. The sensor data comes together in a decision support system, which can be used directly in VRA, or aid the farmer's management.

2.2.3 Establishing ACROSS in the Community

It will be necessary to establish trust in the system. This can be done by having the farmers confirm all decisions manually before application is performed. ACROSS will thus be a decision support system and used with map-based VRA. The data collection can be performed from existing vehicles or autonomous vehicles.

Table 2.2: Establishing the sensor in the community

1.	Manual Operation, Map based VRA	3D based computer vision sensor attached to existing vehicle for data collection. The farmer reviews the data and creates the application map.
2.	(Autonomous Operation, Map based VRA)	An autonomous robot/vehicle collects data and decisions are made as in step 1.
3.	Manual Operation, Sensor based VRA	The piloted application vehicle collects data on-the-go, and the DSS guides the driver to control application rate from the control board in the cabin.
4.	Autonomous Operation, Sensor based VRA	Application rate is chosen by the DSS, and this can be done from an autonomous robot/vehicle.

In order to establish the sensor in agriculture, the implementation should take place in steps (stages), gradually granting the machines more responsibility (step 2 can be omitted), see table 2.2.

2.3 Summary

Current remote sensing is too remote, because their measures do not contain enough information to distinguish stress and disease. The sensor will deal with this by developing a ground level sensor using visible features (symptoms) at approx. 50 cm distance, and sub-canopy locations of NIR spectral responses. Furthermore, decisions can be refined by combining with other information sources that are being researched in parallel to the 3D based computer vision sensor, such as soil information, and to take advantage of scaling between ground-level and aerial/satellite sensing. The current research in precision agriculture will complement the 3D based computer vision sensor, especially the additional sensors and the vehicles from which the 3D based computer vision sensor platform can function. In order to establish the 3D based computer vision sensor in agriculture, the implementation should be in steps, gradually giving the machines more responsibility.

Chapter 3

Preliminaries

Summary

Chapter 1 defined the context and general purpose of the project, and and chapter 2 looked further into the application field and placed the work within that context. This chapter will define the taxonomies for the parameters and syntactical descriptions, and the requirements for the computer vision methods.

3.1 Descriptive Object Parameters

Scenes with plants belong to a general class of scenes that consist of *non-rigid, piecewise smooth, sloped surfaces*. Each of these adjectives is a problem for all 3D reconstruction algorithms.

- *Non-rigid*. means that the object sways in the wind so that all methods involving multiple shots of the same plant (cite structure from motion, shading, and moving light sources) cannot be used.

- *piecewise smooth*. means that there are disparity discontinuities. The positive key word is *Smooth* because that qualifies as an a-priori knowledge about its structure. It means that the disparity within each piece (region) is continuous. However, the smoothness does not mean that each piece is a plane. Global variation can occur.

- *Sloped*. is a problem for close-up imagery because of projection distortion, because it is a problem for window based correlation. It can also corrupt the usage of smoothness constraints because of quantization of the depth plane.

A set of parameters that describe objects will be used in the test of the algorithms. The parameters are chosen by the expectation that its particular feature affects how the algorithms perform in the 3D reconstruction.

Descriptive parameters of the objects are:

Table 3.1: Symptoms of N K P and S deficiency in cereal crops

Nutrient	Typical Symptoms (all of them result in lower growth rate)
N	Lighter green or pale leaves, Steeper leaves, Red bases, symptoms starting at the bottom
P	Red bases. Starting from bottom leaves, darker green at first and then red all over the leaves. If it is very serious the bottom leaves are yellow. Inhibits number of leaves.
K	Starting from the bottom, white-yellow tips, many brown-yellow spots all over the leaves and in the veins.
S	Like N, but when N is not stressed, the symptoms appear on top leaves first.

- *surface shape.* Broad objects versus oblong objects. When discussing leaves, there are grass-leaf and broad leaf shapes. It is to be expected that these must be dealt with separately, because the amount of edges, occlusions and the size of smooth areas differ.
- *surface orientation.* Steepness of the objects is a problem as well as direct alignment with the epipolar axis.
- *presence of texture.* A textured object is always easier to reconstruct than one that is texturally smooth.
- *proportion of changing specular highlight.* The highlighted areas are located differently in the different views. It is the areas where the highlight is different in the different views that are interesting.
- *proportion of occlusion.* A scene with more occlusion if more difficult to reconstruct fully and is a threat to the extracted information about the plant such as no. of leaves.

3.2 Structural Descriptions and Other Annotations

I will establish which information can be extracted from the crops and their value. Research in remote sensing for surveillance of crops are still confined to simple correlations between NVDI indexes and the disease or nutrient deficiency that the given researcher is trying to predict. The problem is that these methods only estimate whether something is wrong and if the controlled crops were provoked into having a certain stress it is obvious that the results are fine.

Consider the case of NPK deficiency in cereal crops. The visible symptoms of NPK deficiency are listed in table 3.1.

Lene K. Christensen [Christensen, 2004] found that the symptoms are distinguishable using manual spectroscopy sampling at the leaf tips and -bases, if each leaf is ranked by its age. Her work also showed that the method can be extended to other crops such as high cost crops

like Maize. This would be a tangible method if the sampling could be done autonomously using 3D computer vision to extract the data.

Considering the case of weed analysis, the actual height of the plants is very useful [Assémat and Chapron, 2003] [Nielsen et al., 2004a].

In conclusion there is a number of valuable descriptive parameters to extract:

- Leaf area
- Leaf height
- Leaf type (grass leaf, broad leaf)
- Number of leaves
- Ranked order of the leaves
- Spectral samples at tips and bases
- Steepness

3.3 Camera setup

This section discusses the camera setup. The biological material used in the practical tests will be discussed in their dedicated chapters.

There are a number of experiments in this thesis using the right-angled trinocular camera L-setup [Mulligan and Daniilidis, 2002] which shown in figure 3.1.

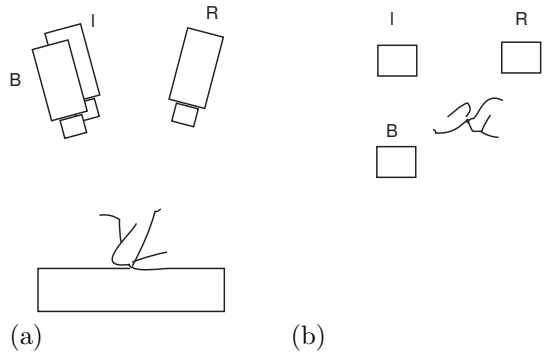


Figure 3.1: Camera Setup looking at winter barley. (a) As seen from the side: Three JAI M4+CL cameras at fixed positions, denoted Left, Right and Bottom Camera, respectively. (b) Seen from above.

The figure shows relative rotation between the cameras which is not a usable constellation. Such camera system requires extrinsic calibration which distorts the images. However, it optimizes the field of common view in all three cameras. This setup is used only in one experiment with barley.

There are commercial trinocular systems in which three cameras are mounted in one case.

These are calibrated in hardware and only intrinsic calibration is required (extrinsically it satisfies epipolar geometry, which was confirmed using matlab calibration toolbox). One such system from Point Grey Research (Digiclops) is used in most experiments.

The basic preliminaries in the presented methods in this thesis is that the camera setup is a pre-calibrated right-angled trinocular setup. The performances will be compared to its equivalent binocular setup.

This means that:

1. The images do not contain intrinsic distortion.
2. There is no rotation between the cameras (optical axis towards infinity).
3. The translations between the two camera pairs are equal and perpendicular.
4. The translation between camera 1 and 2 are aligned with the x axis
5. The translation between camera 1 and 3 are aligned with the y axis

Furthermore, the transformation from disparity to distance from the principal point is given by equation 3.1.

$$z = \frac{fb\lambda}{d} \tag{3.1}$$

where z is the distance from the principal point, f is the focal length, b is the baseline, and d is the disparity. λ is a scale factor given by the image size.

An adaption of Bouquet's Camera Calibration Toolbox for Matlab which is based on [Heikkila and Silven, 1997] was used to achieve this. The intrinsic parameters was calibrated for each camera individually and extrinsic parameters was calibrated each camera pair (using the same image set of the chess board). The principle in this implementation is shown in appendix A.

3.4 Leaf Color Segmentation

Segmentation is not a main issue in this thesis. The typical measures for leaf segmentation are thresholding of excessive green ($\check{g}$) or green chomacity ($\acute{g}$).

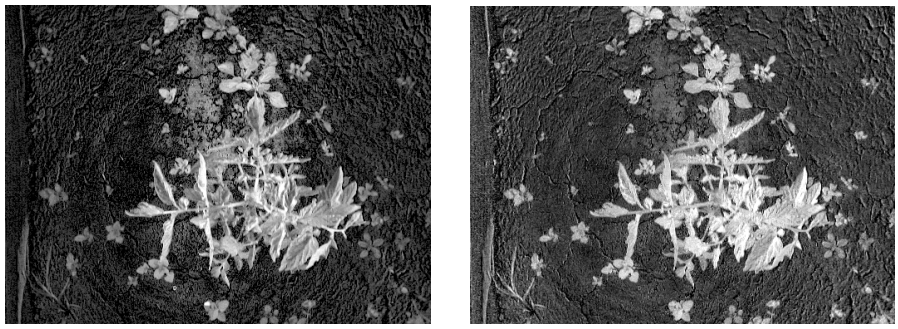
$$\check{g} = 2g - b - r \tag{3.2}$$

$$\acute{g} = \frac{g}{r + g + b} \tag{3.3}$$

where r: red, b: blue, g: green color channels

Figure 3.2 shows an example image.

Very small soil particles can appear in the thresholded image, but these are easilly removed using a median filter.



(a) (b)

Figure 3.2: Example tomato image (a) Excessive green and (b) Green chromacity. The latter is les prone to shading.

Textured leaves need a morphological closing operation to remove the spotted holes in the leaves.

3.5 Window Correlation based Disparity Estimation

In 3D reconstruction the main problem is to find the disparity map, ie. correspondence between image content [Dhond and Aggarwal, 1989]. Literature claims that single pixel correspondence gives a good result and is fast, but it is not a good approach for plants, where most pixels have the same color. For this experiment an area based approach is chosen, because the result is a dense disparity map using neighborhood information. Using an area window can make use of the subtle changes in color which shows the small streaky patterns in the leaves, especially at the skeleton of the leaves.

Binocular Sum of Squared Differences measure (SSD) with Multiple Windows (SMW) [Fusiello et al., 2000]:

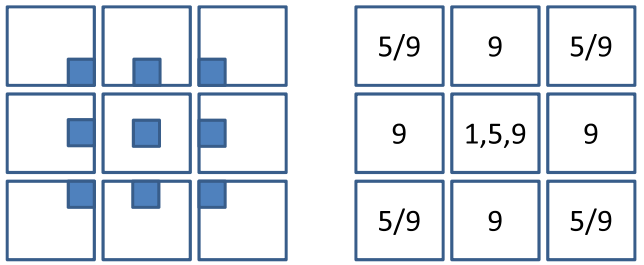
$$E_{i,j}(x,y,d) = \sum_{(u,v)\in W(x,y)} (I_i(u,v) - I_j(u+d,v))^2 \tag{3.4}$$

d is the tested disparity, W is the window around (x,y), I_i is the ith image. The windows can be placed in various ways around the pixel and question.Adding multiple windows can supposedly improve the correspondence near disparity borders.

The SSD is calculated using 9 windows for each pixel pair (referred to as $smw9$), each with their own centers as shown in figure 3.3. It is often only necessary to include the diagonal windows. Later in the thesis this will be referred to as $smw5$.

In [Fusiello et al., 2000] the algorithm was binocular, but it is easy to add a third camera. In principle, there will be two image pairs, where the second pair switches the baseline to the y-axis. I will test two different methods for including the third camera, trinocular sum (T_s eq. 3.6), which is the normal method, and trinocular minimum (T_m eq. 3.5).

$$T_m(x,y,d) = \arg\min_d \min(E_{1,N_x}(x,y,d), E_{1,N_y}(y,x,d)) \tag{3.5}$$



(a) (b)

Figure 3.3: (a) The SMW algorithm uses 9 windows, and the best similarity score is chosen. (b) In this thesis SMW 1 refers to using the centered window alone. SMW 5 refers to using those windows marked with a 5. SMW 9 uses all windows.

$$T_s(x, y, d) = \arg\min_d (E_{1,N_x}(x, y, d) + E_{1,N_y}(y, x, d)) \tag{3.6}$$

In theory the T_m should comparably be more robust to occlusions by choosing the best match in a single image pair. T_s should comparably be more certain of a match if the point is visible in all cameras by choosing the best match where both image pairs are good matches.

The normalization is not necessary to include in the correlation measure and in fact it is not a good idea at all when reconstructing plants, because the reconstruction relies on the shading on the plant. Normalization is done globally by fitting a line $ax+b$ to the correlation plot of the sorted pixel values between the reference camera and the alternative view. Then the normalized alternative view $\hat{I}$ is found using equation 3.7.

$$\hat{I} = \frac{1}{a}(I - b) \tag{3.7}$$

This reduces the brightness/contrast difference between the images.

3.6 Energy Minimization

Energy minimization is usable when there is the assumption that the surfaces are piecewise smooth. [Andersen et al., 2005] used a relaxation scheme lets the correspondence search pick 5 candidates. The best candidate was then found through Simulated Annealing (SA), optimizing the smoothness in the 3D model. Simulated annealing takes a long time to converge towards the global minimum, which is not guaranteed to be found. Another method to minimize an energy function is called Graph Cuts [Kolmogorov and Zabih, 2004]. Graph cuts are especially interesting because they converge to a strong local minimum very fast, given some constraints to the energy functions. The graph cut method will be pursued later in chapter 7. Note that the existing method assumes piecewise constant surfaces which plants are not.

3.7 Dense versus Sparse Ground truth

Outdoor agricultural scenes are unstructured and contain a range of challenging 3D reconstruction situations including semi-transparency, occlusion and specular reflection. There is a need for effective 3D assessment techniques to aid the development of improved 3D algorithms and allow reliable methods of identifying sources of error for use in determining the direction of future efforts for algorithm improvement.

This calls for better quantitative evaluation and understanding of the performance of computer vision methods. It is difficult to evaluate the performance of an algorithm quantitatively, when dense ground truth such as a dense disparity map is not readily available.

A disparity map shows the translation of each pixel from one frame to the next and is related to the inverse of the distance (depth) of all pixels in the image. Obtaining ground truth may require annotation of a thousand pixels by hand. This is in contrast to sparse ground truth data at high levels of abstraction. For example, if the test were about weed/crop classification, it would be easy to annotate a sequence having few plants on each image. However, in this case it can also be difficult if there are hundreds of small weeds in each image. The problem is especially elaborate in agriculture which works with non-rigid biological structures that can be very complex, non-planar and change from time to time (e.g. the height and orientation of a leaf).

It will be necessary develop a test framework based on the complexity found in biological plant structures, which allows statistical analysis and graphical representation for benchmarking 3D reconstruction algorithms and their many combinations of parameter configurations. Previously, the algorithms have been evaluated quantitatively by reconstructing (also called predicting) a known camera view and then comparing the predicted image to the recorded image (Szeliski and Zabih, 1999). However, false correspondences happen when the wrong match looks similar to the correct point. Hence, a large disparity error may not produce a visible error in the projected image.

For simplicity, a manual method has been used to create ground truth disparity maps by pointing out areas of disparities by hand. This method tends to be used with simple scenes and yields disparity maps with a sparse resolution in the depth (Szeliski and Zabih, 1999; Mulligan et al. 2004). Sun (1999) used a frequently used method of generating simple synthetic images with primary geometrical objects such as cubes and spheres, and textures such as gradients or random noise. Another more accurate solution has been to produce ground truth disparity images with a dense depth resolution using a time consuming but highly accurate depth measuring method with structured light (Scharstein and Szeliski, 2003) or laser equipment (Mulligan et al, 2002), which requires rigid scenes.

Most of the methods are mainly appropriate when the context is not close-up sampling of non-rigid scenes. If the camera setup looks at a rather distant scene with fronto-planar disparities, the disparity image and its evaluation is simpler. Scenes like this include indoor office scenes where the disparity planes can be furniture or people at different distances from the camera, the widely used Tsukuba head statue image set, or outdoor scenes of buildings, street signs and vehicles at different distances from the camera (Szeliski and Zabih, 1999).

The context is different for sub-leaf scale 3D reconstruction of plants, because the individual leaves and their poses are important. Consider a grass plant such as barley or wheat at a young growth stage. Their 10-20 leaves are long and intertwining, most leaves having partial

occlusions and large projection distortions from one view to the other. The disparity map must be dense in all three dimensions in order to detect features about their shapes and poses and distinguish closely overlapping leaves. This thesis proposes a test framework based on synthesized plant scenes. This enables objective comparison of dense 3D reconstruction algorithms and camera setups. It also enables testing the feasibility of post processing techniques (such as the following mesh reconstruction) and sample techniques on the ground truth disparity map.

The scenes and cameras can be simulated in ray tracing software. The software simulated typical problems faced in outdoor agricultural scenes:

- Specular surface reflection (highlights) on the leaves
- Occlusions
- Shadows
- Transmission of green light onto the soil
- Light source colours
- Focal blur
- Poisson image noise
- Small offset between baselines
- Small offset in light intensities (brightness/contrast)

The test framework should render ground truth disparity maps, occlusion maps, and mutually exclusive highlight maps.

3.8 High Quality Disparity Maps

The quality of the disparity map depends on its application. Quantitatively the Root Mean Squared Error and Percentage of Bad Matching Pixels is often used [Scharstein and Szeliski, 2002]:

$$rms = \sqrt{\frac{1}{N}\sum_{(x,y)} |d_E(x,y) - d_{GT}(x,y)|^2} \tag{3.8}$$

$$pbmp = \frac{1}{N}\sum_{(x,y)} T(|d_E(x,y) - d_{GT}(x,y)| > \delta) \tag{3.9}$$

where $T(f) = 1$, if f is true and 0, otherwise.

rms measures the average error and is sensitive to large outliers, but is less sensitive to structural deviations. For example it is easily improved by smoothing the disparity image, thus erasing disparity discontinuities..

pbmp counts the proportion of bad matching pixels and does not distinguish between a small error and large error. It is more sensitive to structural deviations than the *rms*. It will penalize over-smoothing and a good *pbmp* performance has better disparity discontinuity borders.

Scharstein also used reprojection error as a metrics. It assumes a third (or fourth) ground truth view, which is reconstructed from the disparity map. The result can be evaluated using *rms* and *pbmp*, but the drawback is that the error does not directly correspond to a disparity error. Especially, when comparing a lot of green pixels to other green pixels. It is likely that the projection hits a similar green pixel.

Qualitatively, the ability to segment the individual pieces of the scene must be evaluated, because that is necessary to rank and count the leaves and to find sampling the correct sampling spots.

Basically, a disparity map with a low *pbmp* is better to work with in the post-processing phase. This is why the tests in this thesis will focus on optimizing *pbmp* scores as a separate problem. However, It is important to keep in mind that a disparity map with a lower score does not always correspond to the one that gives the best result after post-processing.

Another thing to keep in mind when evaluating the methods are the number of parameters and the robustness of the methods when the settings vary from their optimum.

Chapter 4

Ray-Traced Plant Images for Dense Ground Truth

[1]

Summary:

A test framework was developed to facilitate diagnosis of problems with using existing algorithms with plant structures and to compare old algorithms with the ones developed in this research. This chapter describes the plant modeling and Ray-Tracing test framework and features a comparative study of disparity estimations. In the test the following will be compared: binocular smw (1,5,9), trinocular sum, binocular graph cuts. The outcome was a guide to the future research:

- *Improve reconstruction of bad orientation such as steepness*
- *Improve occlusion and mutually exclusive highlights*
- *Improve smoothness inside connected surfaces while retaining discontinuity*
- *Improve Graph Cuts with trinocular setup and steep structures (perhaps also with separate segmentation of disparity map)*

4.1 Test Framework

The software used for creating the synthesized scenes was Plant Studio 2, 3D Studio Max 6, Vrml2Pov , and Pov-Ray. Plant Studio was used to generate four grassy plant models (such as barley) and four broad leaf plant models (such as tomato and beet plants). These models had a very low polygon count. It was possible to obtain a smooth triangle mesh by performing the Tessellate function followed by the NURMS MeshSmooth function in 3D Studio Max. The plant models were inserted into a Pov-Ray scene which contained the three cameras, light, and soil. The light was simulated with sky blue colored ambient light

[1] This chapter was based on paper Nielsen et al. [2007]

and two light sources: Sky blue colored light placed directly above the plant and a powerful white-yellow sun placed besides the scene. The camera focus center was in the middle of the plant, which means the bottom and top of the plant was blurred.

The camera setup was an aligned trinocular L-setup, where the cameras converged at infinity and satisfy epipolar constraints as the Digiclops camera. Figure 4.1 shows the camera setup and geometry in the Ray traced scenes. The following notation is used in this section:

h: Height from the ground plane b: Baseline between the cameras f: Focal Length v: Field of View D: Distance of object from camera d: Disparity I_h: Image vertical resolution λ: Scaling factor

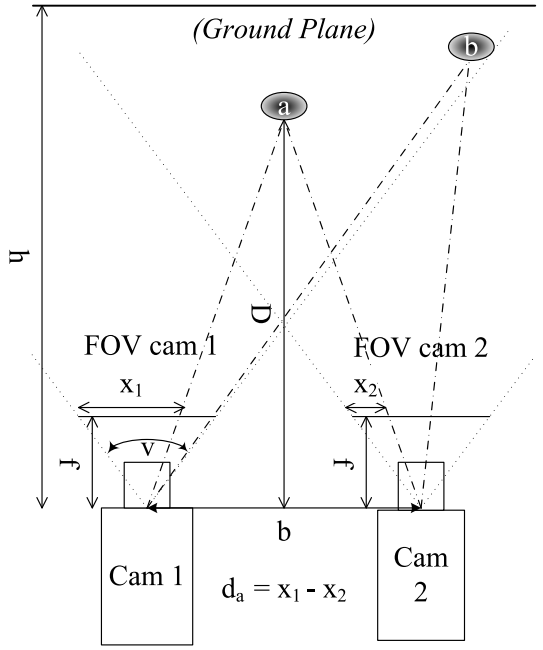


Figure 4.1: An aligned binocular setup where the second camera is transposed [b 0 0] and rotated [0 0 0] from the first. The third camera in a trinocular setup is transposed [0 b 0].

Equations 4.1 through 4.3 show the calculations needed to calculate the ground truth disparity map.

The following equations show how to calculate the ground truth disparity map (d_{GT}) from the depth (or height) map (h_{GT}) generated by the ray tracer. The depth map was a 16 bit map with values between zero and one created by deleting light sources and replacing the textures with a gradient texture, ensuring that the ground plane was at $h = 0$ and the cameras were at $h = 1$. The camera viewpoint in the ray tracer was 4/3 units by 1 unit, so the scaling factor was given by the vertical resolution of the image. Combining the focal length, baseline, height map, camera height, and scaling factor, the ground truth disparity map can be computed as equation 4.3.

$$h_{GT} = \frac{h-z}{h}, 0 < h_{GT} < 1 \quad (4.1)$$

$$f = \frac{1}{2}\frac{r}{tan(\frac{\nu}{2})} \quad (4.2)$$

$$d_{GT} = \frac{fbI_h}{h_{GT} * h} \quad (4.3)$$

4.1.1 Additional Image processing

The synthesized images of plants were manipulated to include the following error sources that are present in real images:

- Poisson [Snyder et al., 1995] image noise (generated in Matlab by the function "imnoise").
- A random small offset in the x and y axis, by removing randomly 1-3 rows and columns from the image and zero padding at the end to preserve the image size.
- A random small change in brightness and contrast by adding a small random number to the image and multiplying by a random number close to 1.0.

4.1.2 Occlusion Mask

Occluded pixels ($O(x, y)$) can be extracted from d_{GT}. Basically the equation find occlusions as *pixels which after the x-translation hits the same pixel as another pixel that were to the left of it.*

$$O(x,y) = \sum_{l=0}^{x-T} T(((l + d_{GT}(l,y)) = (x + d_{GT}(x,y)))) \quad (4.4)$$

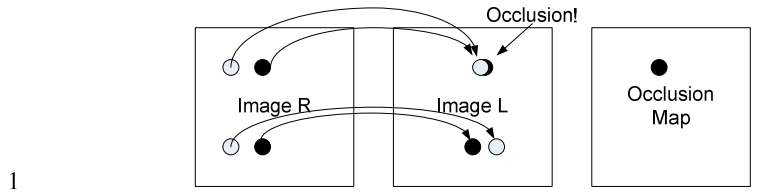
where $T(f) = 1$, if f is true and 0, otherwise.

where T is a tolerance which reflects the expected acceptable gradient in the disparity map. This avoids gradient contours to be denoted as occlusion. Figure 4.2 shows the principle in this detection.

This method can predict *first order* occlusion, but not *second order* occlusion. First order occlusion is defined as *occlusion caused by visible object pixels.* Second order occlusion is *occlusion caused by object pixels that were occluded themselves in the reference.*

4.1.3 Highlight Mask

Prediction of highlights requires rendering the scenes without the *specular* parameter. Then the images with highlights I_h can be compared to the images with no highlights I_n with a tolerance T_h. Highlights in the reference images are H_r, and H_a contains the mask for pixels in the reference image that have highlights in alternative views. It uses d_{GT}, which in



1

Figure 4.2: Illustration of when occlusion occurs. The dots represent objects or pixels. Their positions are shifted to the right (how far is determined by their disparities) on the X axes from image R to image L. Equation 4.4 perceives an occlusion at point (x,y) if pixels to the left of (x,y) would be shifted onto the same position as itself.

Table 4.1: Laplacian of gaussian

-1	-1	-1
-1	8	-1
-1	-1	-1

equation 4.6 is shortened to d. H_d is the union of areas that changes highlight-status from the reference view to the other (those are the interesting regions to explore).

$$H_r(x,y) = T((I_h(x,y) - I_n(x,y)) > T_h) \tag{4.5}$$

$$H_a(x,y) = T((I_h(x + d(x,y)),y) - I_n(x + d(x,y),y)) > T_h) \tag{4.6}$$

$$H_d = (H_r \cup \neg H_a) \cap (H_a \cup \neg H_r) \tag{4.7}$$

where $T(f) = 1$, if f is true and 0, otherwise.

4.1.4 Shadows Masks

It is possible to use a similar method for shadows using the *shadowless* keyword in the light sources. Shadow maps would enable the analysis of stereo algorithm performance in the shadow areas which could cheat an energy minimization algorithm into adding a disparity change. Photo consistency in shadow regions could also be different in the shadow regions because of under saturation and the lack of highlights. Furthermore, shadow regions in plant images in particular undergo complex lighting phenomena. Analysis around edges using edge maps are interesting because this is where disparity changes may occur and where window based methods may wrongly compare photo consistency on two different surfaces. Analysis of edges and disparity discontinuities are easily accessible by applying an edge detector (e.g. laplacian of gaussian (tab. 4.1) or canny) to d_{GT}.

4.1.5 Quality Metrics

The estimated disparity maps d_E were compared to d_{GT} using the same metrics as Scharstein and Szeliski [2002]. These were the Root Mean Squared Error and Percentage of Bad Matching Pixels (section 3.8).

rms measures the average error and is sensitive to large outliers, but is less sensitive to structural deviations. For example it is easily improved by smoothing the disparity image, thus erasing disparity discontinuities.. *pbmp* counts the proportion of bad matching pixels and does not distinguish between a small error and large error. It is more sensitive to structural deviations than the RMS. It will penalize smoothing and good *pbmp* performance has better disparity discontinuity borders. The quality of each measure depends on the application the estimated disparity maps are intended for if the results vary significantly.

4.1.6 Statistical and Graphical Testing

The output included not only arrays of *pbmp* and *rms* values, but also binary indices for all parameters and plant- and texture types corresponding to the tests. For example (these are matlab specific tricks), the index array "winsize12" is true for all tests that are generated with a 12x12 window and false for all others. This makes it easy to compute the mean of all results made from a certain configuration, e.g. window size = 12-16, Leaf type = Spotted Broadleaf: *result = mean(pbmp((winsize12—winsize16) & broadleaf & spotted))*. It is also easy to do a Wilcoxon signrank test: *[h p] = signrank((pbmp((trinocularsum — trinocularmin) & broadleaf), pbmp(binocular & broadleaf))*, where h is 1 when the populations are significantly different at p probability of being equal. The data sets are paired[2] and non-normally distributed.

4.1.7 Image examples

Appendix B shows the image sets used in the majority of the tests in this thesis. There is a test set of virtual plants with dense ground truth and a set of real images of the same types of plants as the virtual set. These are hand annotated in a custom made software named image view (plus more)[3] There are also occlusion-, highlight-, and discontinuity masks for Plant 10. Figure 4.3 shows image examples of the different vurtual and real plant types used.

4.2 Comparative Study of Disparity Estimations

The test will compare different configurations of algorithms and their settings. The symmetric multiple windows will be compared in binocular and trinocular setups. I also include one of the best 3D reconstruction algorithms available that uses a graph cut energy minimization, which yields similar results to the slower simulated annealing. The difference is that graph cuts preserves depth discontinuity Kolmogorov and Zabih [2002]. It does not rely on window sizes which tend to dilate the depth regions and are sensitive to perspective distortion. The main adjustable parameter is the impact of the smoothness constraint, λ. Since it assumes regions of equal depth, it excels at fronto-planar scenes, but may have trouble when it comes to steep leaves on plant structures. It was interesting to see how it performed in this new context. I used Kolmogorov's implementation of the graph cut algorithm Kolmogorov and Zabih [2002] that is referred to as *kz1*. This is only a binocular algorithm which used the 1st

[2]Matching test scores in the pbmp vector come from the same combination of settings other than those used in the selection.

[3]http://www.cvmt.dk/ mnielsen/tools.html

and the Nth camera. λ was given a small value (half of the automatic setting) as to make it better at sloped surfaces.

About 2002 different disparity images were estimated for the ground truth evaluation. They were based on 11 plants (3 smooth textured grassy, 5 smooth textured broadleaf and 3 spotted broad leaf). They were processed with all combinations of the parameters: Window size [4x4, 8x8, 12x12, 16x16, and 20x20], colour space [RBG/Monochrome], Camera setup [Trinocularmin, trinocularsum, and binocular], and Single or Multiple Windows [1-center, 5-corners, and 9-sides+corners]. Plus the graph cut estimations.

Figure 4.3: [Top] Examples of real spotted broadleaf, smooth/glossy broadleaf- and grassy plants [Bottom] Examples of the rendered counterparts.

4.2.1 Results

Table 4.2 shows the overall performances of different setups and parameters. For example, the mean PBMP for the binocular setup is the mean of all other configurations under the binocular setup. The null hypothesis for a given result is that the previous result was not significantly different, i.e. that trinocularmin is the same as binocular, and SMW 5 is the same as SMW 1, etc. If this is rejected using two-sample T-Tests, the means are significantly different. A weak significant difference is denoted + and a very strong significant difference is denoted ++. The first thing to notice is that the means are in the same ranges for the synthesized results and the real results.

The gray cells show discrepancies between the synthesized and real results (Problem P1, P2, and P3). This is regarding smooth grassy performance, trinocular minimum versus sum, and using multiple windows.

Table 4.2: Overall performances (times 100). Mean pbmp (Standard Dev.) Significant improvement over the previous result above. +: Reject H_0 $p < 0.5$, ++ Reject $p < 0.001$ -: Accept H_0.

Parameter	Synthesized	H_0	Real	H_0	
Binocular(GrC kz1)	18.00 (10.04)		20.38 (15.97)		(P1)
Binocular	15.84 (12.67)	++	17.36 (11.98)	++	
Trinocular(min)	10.18 (9.23)	++	8.05 (7.64)	++	(P2)
Trinocular(sum)	7.17 (6.17)	++	8.26 (9.19)	-	
Monochrome	13.30 (12.21)		11.63 (10.86)		
RGB	12.13 (11.31)	+	11.28 (10.69)	+	
SMW 1	11.01 (9.22)		9.66 (9.65)		(P3)
SMW 5	11 (10.64)	-	11.95 (10.9)	+	
SMW 9	11.19 (11.07)	-	12.15 (11.28)	-	
Grassy Smooth	9.84 (8.6)		17.57 (10.51)		(P4)
Broadleaf Smooth	14.81 (9)	++	13.7 (10.34)	++	
Broadleaf Spotted	8.55 (12)	++	2.23 (3.39)	++	

(P1) Graph cuts did not perform convincingly for my test data. It is well known that is works best with fronto-planar scenes. The reason behind the bad performance is staircase approximation of the slopes.

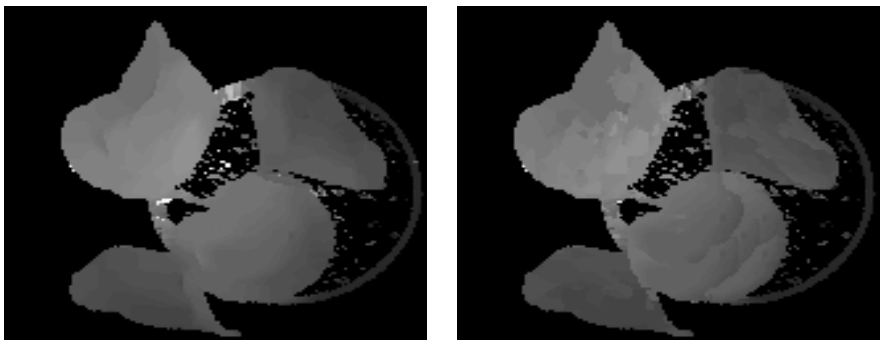
(P2) The discrepancy between synthesized and real results regarding trinocularmin/sum can be explained by a difference in the proportion of occlusion between the models. Further investigation into why trinocular minimum is better for real images shows that it is better for a specific leaf type, smooth broadleaf. Trinocular sum is better in the other circumstances. Taking a look at the structure of the smooth broadleaves, the situation resembles the situation discussed by Nielsen et al. [2005b] where trinocular minimum was found to be an advantage when steep leaves were self-occluding by being parallel to the focal axis and aligned with the baseline.

(P3) It seems to be a waste of computation power to use symmetric multiple windows. Refer to figure 4.4 that shows the disparity maps for the cotton plant with and without multiple windows. Their PBMP scores are the same (2.3% error).

The edges between the overlapping leaves are well preserved, which will make it easier to split the surfaces for the reconstruction. However, the smooth sloped surface of the leaves (especially the lower right leaf) is stepwise sloped. This disparity at a given pixel can be chosen from the best match of the window above the pixel while the next pixel might be chosen from the window below the pixel. If the object under the top window exists at a different disparity than the bottom window, which is the case for a very steep slope, then the result can be like this stepwise sloped disparity map. See figure 4.4.

A combination of these results that preserves the disparity edges and the smooth surfaces would be an advantage. The algorithm that produces a SMW 5/9 disparity map would be able to output the SMW 1 map for free (more later in chapter 6).

(P4) Further investigation into why the real smooth grassy plants were much worse then the synthesized plants showed that it was especially the binocular results that increased the mean error drastically. The mean of the trinocular sum results for smooth grassy plants were

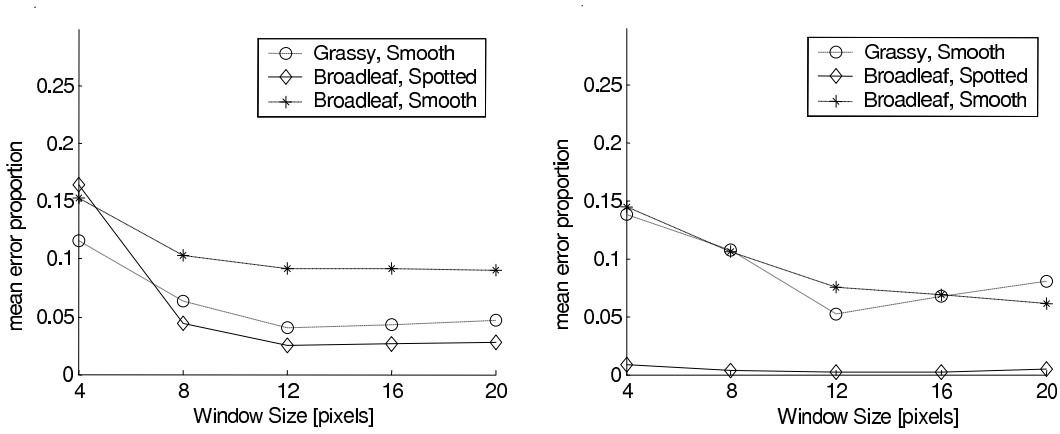


SMW 1 SMW 5

Figure 4.4: Disparity maps of a real cotton plant. Symmetric multiple windows cannot handle the steep slope on the leaves.

10.81. The texture looks more flat on the real images than the synthesized (figure 4.3), so it is easier to match a point on one leaf with a point on another leaf.

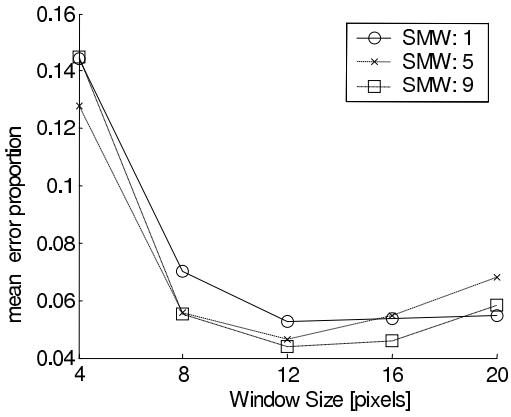
The large standard deviations show that there is a lot of variability within each category, so it is necessary to analyse the relationships further. The test framework easily allows testing on combined configurations. As [Nielsen et al., 2005b] showed, there are complex relationships between object parameters and the technical configurations. Only few of these are shown here in figures 4.5 - 4.7. Finding the optimal window sizes is related to leaf types, symmetric multiple windows and camera setup.



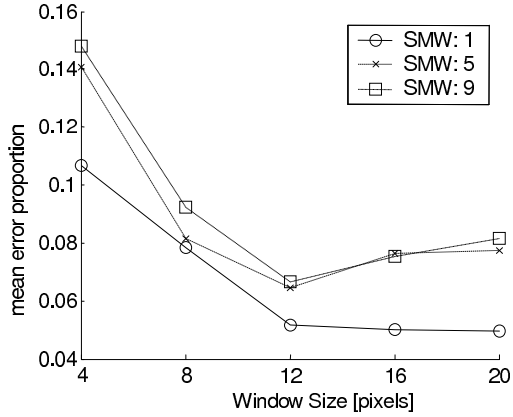
Virtual plants Real plants

Figure 4.5: Comparison of window sizes in relation to leaf types and texture types.

The results look very similar for synthesized plants and real plants. Presence of texture (spots) and grassy shapes call for smaller windows that the smooth broadleaves. Using multiple windows also call for smaller windows. That way the discrepancy between the disparity planes under the top windows and bottom windows around a pixel are minimised.

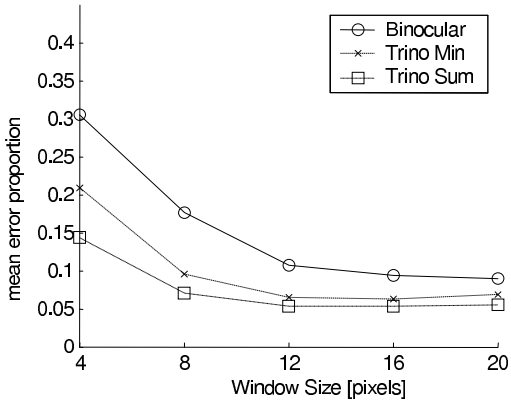


Virtual plants

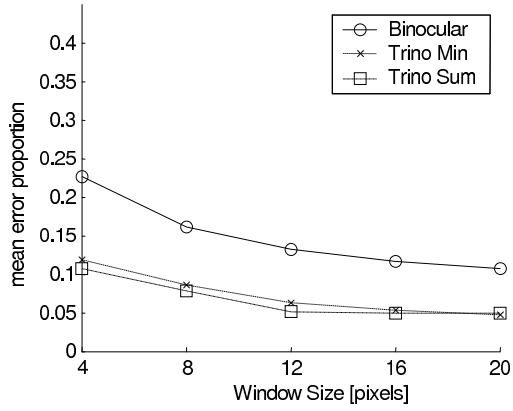


Real plants

Figure 4.6: Comparison of Window Sizes in relation SMW.



Virtual plants



Real plants

Figure 4.7: Comparison of Window Sizes in relation to camera setup.

The window size is less dependent on stereo setup. The binocular setup needs the windows to be as large as possible, though, while the trinocular setups can be more certain of a good match at smaller window sizes.

4.2.2 Discussion

The test framework was satisfying to work with. It was usable for testing vast amounts of parameter configurations and making it easy to do statistics and graphic representation and interpretation of the results. The images rendered by the ray tracer looked lively and near realistic by simulating the well-known computer vision problems such as many difficult structures, Poisson noise, focal blur, imperfect epipolar alignment, imperfect brightness/contrast between cameras, shadows, and colour distortions. It turned out that all the error sources had to be simulated before the results matched. Even with Gaussian noise instead of Poisson noise the results were not as good.

This work shows how to evaluate disparity estimations, but it is not limited to this purpose. Other setup variations such as light source placement can be tested as well as other processing tasks such as structure segmentations and shadow or highlight detection.

It is interesting to note that even if a configuration scores better than another, the final reconstruction is not necessarily improved. Examples of VRML reconstructions are shown in figure 8. They are the same plants that were shown in figure 2, but here they are rotated in a slightly different angle. In order to separate the individual leaves, the disparity maps from SMW 9 results were segmented based on pixel connectivity and discontinuity in the Z-axis. SMW 1 results did not make it possible to segment each leaf. These models are not perfect as a closer look at them shows a wavy surface and they cannot overlap in the Z axis. This has a great impact on the leaf area analysis.

So why did I not use the adaptable windows algorithm?

There are works where the choice of window sizes has been automated using texture presence analysis Kanade and Okutomi [1994]Demoulin and Van Droogenbroeck [2005]. So why did I not use the adaptable windows algorithm? First and foremost I wanted to maintain control of the window sizes as to test it as a parameter relating to crop type as well as aim for as simple an algorithm as possible for future real-time implementation. This is relevant because the crop type in a given field is known. Secondly, the results showed that even the no texture leaves had a tight lower and upper boundary for best results and the automatic texture analysis might in fact overestimate the window sizes which would lead the leaves over-smoothing and making it impossible to separate them.

On the other hand it would be interesting the test the assumption that an adaptive approach would be cheated by smooth leaves. The adaptive algorithms need an upper and lower boundary which could be selected using the results in my experiment.

4.2.3 Future Work

Future work will improve surface reconstruction and extraction of structural information on the plants such as leaf angle, leaf area, leaf rank/age, detection of leaf type and texture,

spectral samples at key positions, etc. Only then can the system be used for diagnostics and classification.

4.2.4 Conclusions

It can be expensive to perform experiments in this application field. It is also very difficult to obtain dense ground truth for complex structures such as plants. Hence, the test framework is a valuable tool that makes it possible to test performance before investing in expensive equipment and field experiments.

Generating simulated images with Ray-Tracing is a versatile approach, which can simulate different camera setups, indoor- and outdoor lighting conditions, focal blur, motion, glossiness, etc. It is not a new idea to use ray traced images for testing computer vision algorithms, but it has usually been simple geometric models with texture.

The test framework allowed statistical- and graphical analysis for benchmarking 3D reconstruction algorithms on synthesized crop scenes, and to find the relationships between their many combinations of parameter configurations and the crop features. The performance of the synthesized images is comparable with real images, when the structure is the same error sources are simulated.

The algorithm showed in particular problems with leaf orientation (steep slopes). The Graph cut algorithm is known to have this limitation, but it has not been well documented that the SMW have this problem. The next part of this thesis will look further into this problem and the deeper analysis tools, such as occlusion and highlight testing, in the test framework will be applied.

Part II

Improving Disparity Estimations

Chapter 5

Multi-baseline Camera Setup

[1]

Summary:

Based on the problems found in the test a new multi-baseline approach was developed trying to minimize the projection distortion within windows when the surfaces are steep slopes. Furthermore, the occlusion and highlight analysis tools were used to get detailed performance information on the algorithms. The multi-baseline method turned out to be better at handling highlights, while the trinocular min term was best at handling occlusions, and trinocular sum was overall best. The Graph Cut algorithm was the worst.

Computer vision based 3D reconstruction of close-up complex biological structures is a difficult discipline. There are various multi-camera configurations to choose from. It would be useful to learn about the performance related to descriptive parameters of the objects at hand, in order to choose the best configuration. The Descriptive parameters of the objects are *surface shape*, *surface orientation*, *presence of texture*, *proportion of changing specular highlight* and *proportion of occlusion*. The specular highlights in concern are those that changes gradually from one image to the next across the baseline. Multi-baseline Stereo has been described and tested in literature as a method for improving the handling of occlusion and ambiguity across the scan lines Okutomi and Kanade [1993]Jeon et al. [2001] by using the sum of the energy measures across the camera array; e.g. Sum of Sums of Squared Difference (SSSD). Attempts have also been made at dealing with specular highlights by actively detecting specular highlights within the algorithm Li et al. [2002] and treating them as occlusions. However, the problems related to nearby objects are overlooked as the algorithms assume that the area looks the same in all cameras. This chapter presents an alternative measure that utilizes the fact that a multi baseline array consists of subsets of smaller baselines. A large baseline improves depth resolution but it also makes the correspondence more difficult Okutomi and Kanade [1993]. Three factors increase this effect: Being close to the observed object, window correlation size, and orientation of object surfaces.

Precision agriculture is a field with rising interest in 3D computer vision, which is becoming tangible as new high dynamic range cameras and precalibrated multi-view cameras are being developed. These cameras satisfy the epipolar geometry constraints and the intrinsic-

[1] This chapter was based on paper Nielsen et al. [2005a]

and extrinsic calibration can be skipped. Close-up 3D reconstruction of plants is an excellent example where the leaves can be pointing steeply toward the cameras and it needs high depth resolution because the leaves overlap closely to each other. Excellent depth maps has potential to aid the segmentation of individual leaves Lee et al. [1996], if the disparity maps have trustworthy discontinuity edges. This is useful in precision agriculture for segmenting individual leaves for autonomous weed identification, fruit picking, branch thinning, and for finding sampling points on specific locations of a plant Christensen and Jørgensen [2003]Nielsen et al. [2004b]. The image acquisition is expected to be done from a moving platform in an outdoor environment, so reconstruction must be done from a single time slice.

In general terms plants belong to the class of objects that are: semitransparent, biological, non-rigid structures. Disparities are often non-planar and can get very *steep* toward the cameras. Textures are non-existent or highly detailed, and having more or less specular highlights. Fortunately, they are segments of smooth surfaces, but intertwining and overlapping. It is very difficult to get dense ground truth. The Vision based depth map reconstruction is usually confined to fronto-planar depth scenes, where the depth maps can be described as regions of near-equal disparities. These scenes are viewed from a distance and have small finite disparity spaces, where it is reasonable to manually acquire ground truth. As an alternative, structured light can be used. It uses multiple images so that the objects must be rigid in time Scharstein and Szeliski [2003].

5.1 Methods and Material

The stereo correspondence algorithms were all based on a basic Sum of Squared Difference (SSD) dissimilarity (energy) function (eq. 3.4). *The presented methods assumes precalibrated images satisfying epipolar geometry constraints, equal baseline, and zero rotation.*

In the classical multi baseline SSSD the Sum of Squared Difference between the reference camera and the ith camera is computed for N cameras. See equation 5.1.

$$S(x, y, d) = \arg\min_{d} \sum_{c=2}^{N} (E_{1,c}(x, y, \frac{d(c-1)}{N-1})) \tag{5.1}$$

It shows that the binocular case ($N = 2$) is a special case of this equation. $E_{a,b}$ is the SSD measure between camera a and b, such that the function basically sums up the SSD similarity measures between all combinations of the first camera and the others. The d that minimizes this sum is chosen for pixel x,y.

5.1.1 Introducing SISSD

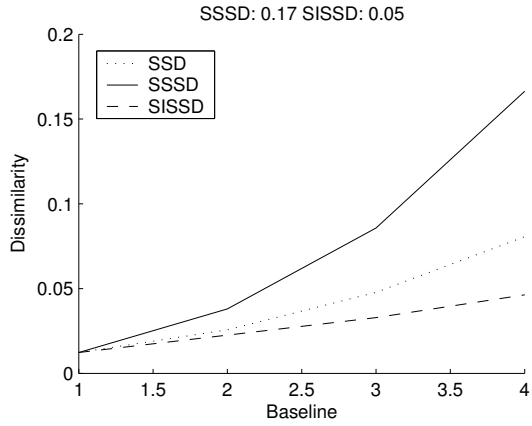
A new measure Sum of Individual Sums of Squared Differences is defined as SISSD (see equation 5.2). This measure was supposed to learn the graduate change in the feature window across the baseline. This could be a problem with occlusions as it would learn the feature of the occluding object, which was countered by including the weighted dissimilarity in regard to the reference camera. In the new measure the Sum of Squared Difference was computed between the $i-1$th and the ith camera, and between the 1st and the ith camera

to ensure that it does not adapt to a completely different object.

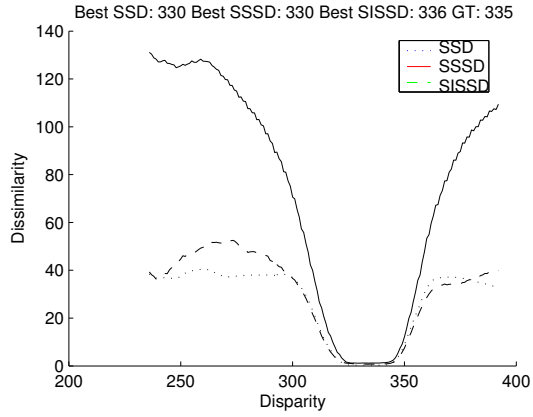
$$S(x,y,d) = \arg\min_{d} \sum_{c=2}^{N} [\alpha(E_{c-1,c}(x,y,\frac{d(c-1)}{N-1})) + (1-\alpha)(E_{1,c}(x,y,\frac{d(c-1)}{N-1}))] \quad (5.2)$$

It shows see that SSSD is a special case of SISSD, where $\alpha = 0.0$. Figure 5.1 shows an example of the case with steep object where the projection distorts the orientation of the leaf. The top shows parts of images of a five camera array. The middle plot the development of the dissimilarity (energy) across increasing baseline. It is obvious that SSSD increases exponentially, while SISSD is even less than SSD. The bottom plot shows the dissimilarity for the three measures across the scan line and prints the best match for SSD, SSSD, SISSD and Ground Truth (GT). This trait should also be an advantage in the presence of specular highlights that travel across the baseline. An example is shown in figure 5.2. Based on these preliminary results, a benchmark experiment was performed. The goal was to validate that SISSD performed better than SSSD on steep-leaved objects and in areas where the specular highlight state changes, and whether the reference similarity constraint could counter the occlusion problem.

(a)

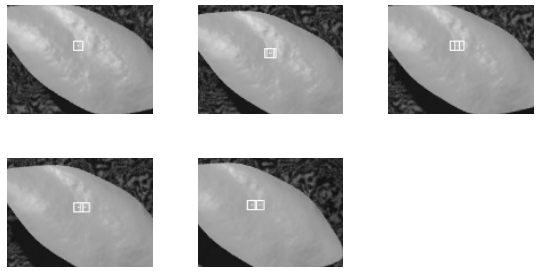


(b)

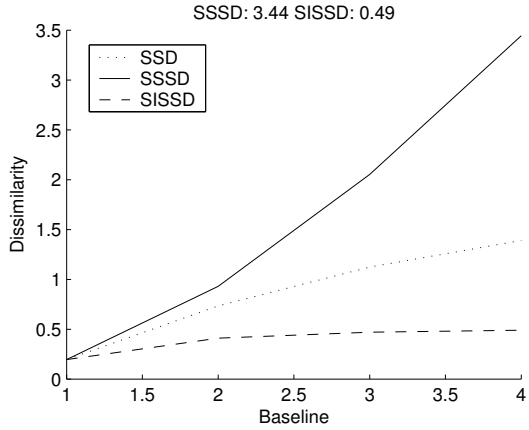


(c)

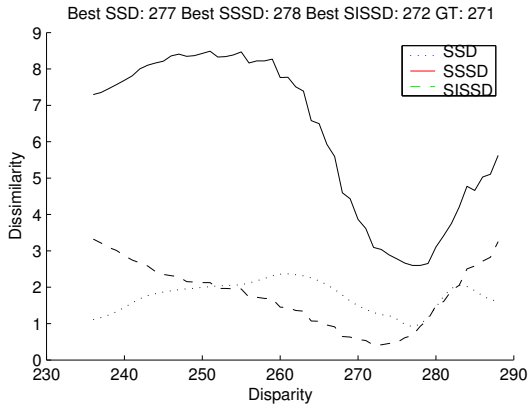
Figure 5.1: The case of steep leaves where projection changes orientation across the baseline. (a) five views of the location on the steep leaf. (b) The development of the dissimilarity across the baseline. (c) The dissimilarity/energy function across the scan line in the image. The best match for SSD, SSSD, SISSD ($\alpha = 1$), and ground truth (GT) is given over the graph.



(a)



(b)



(c)

Figure 5.2: The case of flat leaves where the highlight changing across the baseline. The potential weakness of SISSD is that the dissimilarity difference between the correct match and its surroundings is not very pronounced. This makes the global minimum sensitive to jitter.

5.1.2 Comparable Methods

One of the best 3D reconstruction algorithms available uses a graph cut energy minimization, which yields similar results to the slower simulated annealing. The difference is that graph cuts preserves depth discontinuity Kolmogorov and Zabih [2002]. It does not rely on window sizes which tend to dilate the depth regions and are sensitive to perspective distortion. The main adjustable parameter is the impact of the smoothness constraint, λ. The lambda smoothness constant from Kolmogorov and Zabih is such that high values force more disparity constancy thus it is set to be lower than the automatic settings would to better allow slopes. Since it assumes regions of equal depth, it excels at fronto-planar scenes, but may have trouble when it comes to steep leaves on plant structures. It was interesting to see how it performed in this new context. Kolmogorov's implementation of the graph cut algorithm was used Kolmogorov and Zabih [2002] that is referred to as *kz1*. This is only a binocular algorithm which used the 1st and the Nth camera. λ was given a small value (half of the automatic setting).

There are three common quality metrics root-mean-square, reprojection/prediction of a novel viewSzeliski and Zabih [1999], and percentage of bad matching pixels. The latter is chosen because the focus is to generate correct disparity maps. Root-mean-square error does not ensure that the structure and discontinuities are preserved. Reprojection error does not measure the actual disparity error, but *whether the reprojection of one green pixel happen to hit a matching green pixel* in the novel view. However, in a scene full of green plants that is very likely even if the disparity is very wrong.

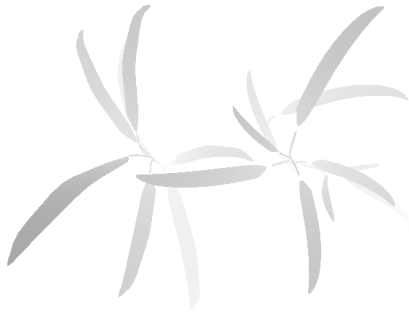
The estimated disparity maps d_E were compared to ground truth (d_{GT}) using the Percentage of Bad Matching Pixels metrics as in Scharstein and Szeliski [2002]:

5.1.3 Experimental Setup

The experimental tests were conducted in order to learn more about the algorithms in the complex context of close-up reconstruction of complex structures. Hence, near-photo realistic ray traced scenes of plants were used in order to control the scene parameters and get valid ground truth disparity maps, occlusion masks, and highlight masks. The scenes had natural outdoor lighting and focal blur, which is a natural problem with plants with steep leaves. Blur is unavoidable, because the aperture cannot be very small and the shutter must be fast when capturing images from a moving platform and the plants are waving in the wind.

Two main classes of plants, long leaf (grass-like, e.g. cereal) and broad leaf (e.g. beet and tomato) were generated. This relates to *surface shape*. For each of these there were plants with steep leaves and flat leaves, respectively. This relates to *surface orientation*. Steep leaves compared to flat leaves have less highlight, more occlusion, and vice versa. A natural case with two grassy plants with flat and steep leaves and a lot of occlusion were used, too. Each scene was generated with textured (spotted) and no texture (glossy), both having bump maps. This relates to *presence of texture*. Finally, all images very generated with and without specularity. This served two purposes; 1. it was required to find the highlight masks (where highlights exist in one frame and not the other), and 2. in order to test overall performance of the algorithms and the same geometrical structure with and without the presence of highlights. There were 18 image sets in total. See figure 5.3 for an example with ground truth.

Reference frame



Ground truth disparity map

Figure 5.3: A natural case, where two grass-like plants are close together and leaves are occluded. The proportion of occluded pixels is 5% and the proportion of changing highlights are 5%.

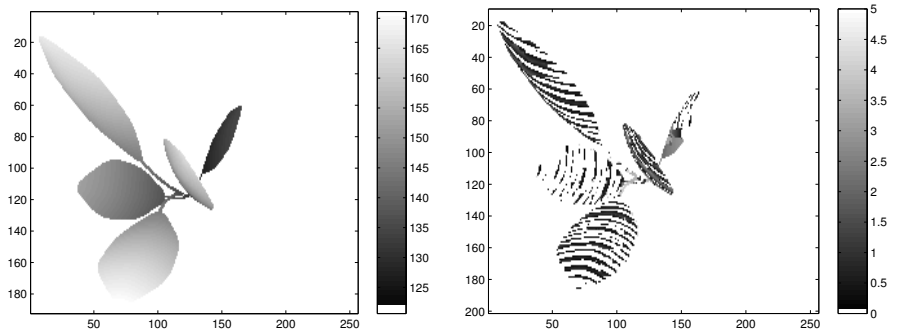
5.2 Results and Discussion

The overall results are shown in table 5.1. It is the mean and spread of performance over all plant types. Note that the ground truth maps were calculated in floating points as to represent the (scaled) inverse of the real height. The disparity maps were integer pixels. If the ground truth had been rounded, the values would have been 10-20% lower. $Multi_{3cam}$ used the same cameras as $Multi_{5cam}$, but skipped camera 2 and 4.

Table 5.1: Comparison of Stereo setups. Mean PBMP (%) and their standard deviations calculated from all pixels (all), pixels with different specularity state (high), and occluded pixels (occ).

Stereo Setup	All	High	Occ
$Multi_3SSSD$	8.9(5.9)	22.1(14.6)	50.3(30.9)
$Multi_3\alpha 0.25$	8.9(5.6)	20.9(13.7)	55.4(28.7)
$Multi_3\alpha 0.50$	9.9(5.6)	20.6(12.1)	64.6(23.8)
$Multi_3\alpha 0.75$	13.5(6.6)	23.0(12.2)	69.1(24.0)
$Multi_5SSSD$	8.3(5.5)	20.3(13.9)	46.1(28.3)
$Multi_5\alpha 0.25$	8.2(5.3)	19.4(13.3)	49.9(24.5)
$Multi_5\alpha 0.50$	8.8(5.4)	19.1(12.7)	55.3(22.5)
$Multi_5\alpha 0.75$	11.6(6.0)	21.0(12.5)	69.1(20.4)
$GraphCut$	14.6(8.7)	19.6(16.3)	73.9(24.3)
$TrinoMin$	10.2(6.5)	23.1(12.5)	30.6(22.3)
$TrinoSum$	9.8(6.9)	23.6(15.8)	40.3(25.0)

The table shows that having those two extra cameras in between the three cameras did improve the result by 11% in average for all pixels, 8% for highlighted pixels, and 8% for occluded pixels. Meanwhile, their spread was approximately equal or slightly narrower (for occluded pixels). The significance of 8.9% versus 8.2% is up to the application to decide. The development within $multi_5$ by increasing $alpha$ was devastating for occluded pixels by 50%, while overall and highlight pixels reach a local minima between $\alpha = 0.25$ and $\alpha = 0.5$. The



(b) (b)

Figure 5.4: (a) Ground truth and (b) Graph Cuts Log(disparity error) for steep spotted broad leaf without highlights. The banding characteristics were caused by the attempt to impose fronto-planar regions on the steep leaves.

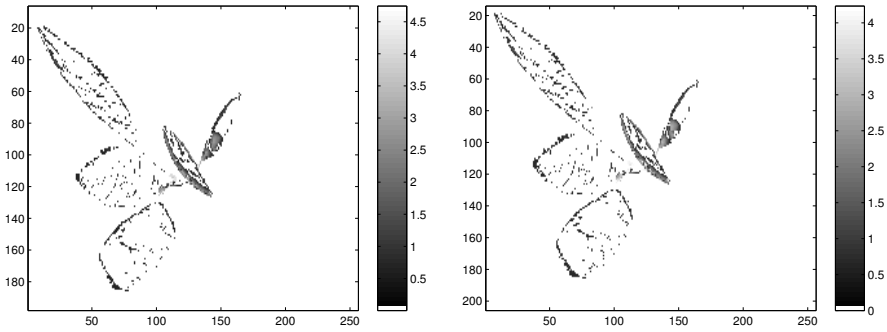
benefit was rather small, though; 1% for all pixels and 5% for highlight pixels. The SISSD measure may be a improvement when using larger window sizes, which tend to be the case when using real images. The trinocular measures did well and they excel at occluded pixels, especially T_m. Graph cuts did the worst, except at correcting highlight pixels by smoothing those areas. Figure 5.4 shows why graph cuts did not do very well. The disparity map was banded, ie. staircase shaped, instead of smooth.

Figure 5.5 shows the errors from the multi-baseline reconstruction of the same plant. The errors were more recognizable as noisy jitter, which could be removed by an energy minimizing sloped smooth surface technique.

Figure 5.6 shows the errors from trinocular results for the same plant. The very steep leaf in the middle and the one to the right of it are difficult for all the algorithms except trinocular minimum (T_m). It is so steep that it is almost a self-occlusion. In the second camera the leaf would be extended along orientation of the baseline, thus occluding the other leaf. T_m simply reconstructed it from the Y direction. The lesson is that it is not only the orientation toward the camera that affects the result, but if the orientation of a leaf aligns with the baseline it can be difficult to reconstruct it. This is especially a problem with textureless grass-like leaves that aligns with the baseline Nielsen et al. [2004b]. In comparison, SISSD was able to reconstruct the steep leaf nearly as good, but the leaf to the right of it was as bad as Trinocular sum (T_s).

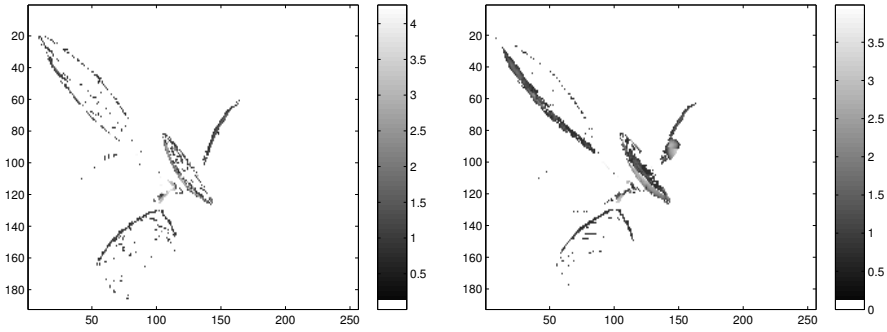
Figure 5.7 plots the all-pixel results grouped by descriptive object parameters, i.e. leaf shape, leaf orientation (flat or steep leaves), texture, and highlights and occlusion. Horizontal axis is the setup: M 0.0 (SSSD), M 0.25 (SISSD $\alpha = 0.25$), M 0.5, M 0.75, Binocular Graph Cut, Trinocular Minimum T_m, and Trinocular sum T_s. The vertical axis is the mean pbmp for window sizes ranging from 4-12. The same goes for figures 5.8 and 5.9 that show the pbmp of highlight pixels and occlusion pixels, respectively.

Figure 5.7 plot (a)(plants without specular highlights) clearly pins down the sources of error for reconstruction in general. The flat-leafed plants (since they had no specular highlights on this plot) all score very well. The errors were large when the leaves were steep or occluding (the model called *two grassy* is 5% occluded in comparison to the steep broad leaf which is



(a) (b)

Figure 5.5: (a) Log(disparity error) Multi-baseline SSSD and (b) SISSD $\alpha = 0.5$. These results did not have any banding, but the difference between the SSSD and SISSD was very small. The result would be excellent if it were combined with a slope- and discontinuity preserving graph cut minimization.



(a) (b)

Figure 5.6: [a] Log(disparity error) trinocular minimum (T_m) and [b] trinocular sum (T_s).

only 1%).

The interesting aspect on plot (a) on figure 5.7 is that it was the steep leaves that best improved slightly from SISSD, while the flat leaves are reconstructed best through SSSD. However, taking a look at plot (b) reveals that when there were highlight on those flat leaves, SISSD was an improvement, too, especially for broad leaf plants.

Note also the fact that the steep leaves were troublesome for graph cuts on plot (a) and (c), especially the glossy steep broad leaf, which was easier for the others compared to grassy plants. Plot (a) to (d) shows consistently that T_s reconstructed grass-like plants better than T_m, but T_m reconstructed broad leaf plant best. This trend is revisited in figure 5.8.

Figure 5.7 Plot (d) shows that in the more natural case, SSSD and T_s were best, even though T_m was best in most occluded parts (figure 5.9 plot (a) and (b)). Maybe the algorithm could dynamically choose T_m by detecting occlusion with left-right consistency Fusiello et al. [2000].

Figure 5.8 plot (a) and (b) shows the subtle strength of SISSD in the highlighted areas.

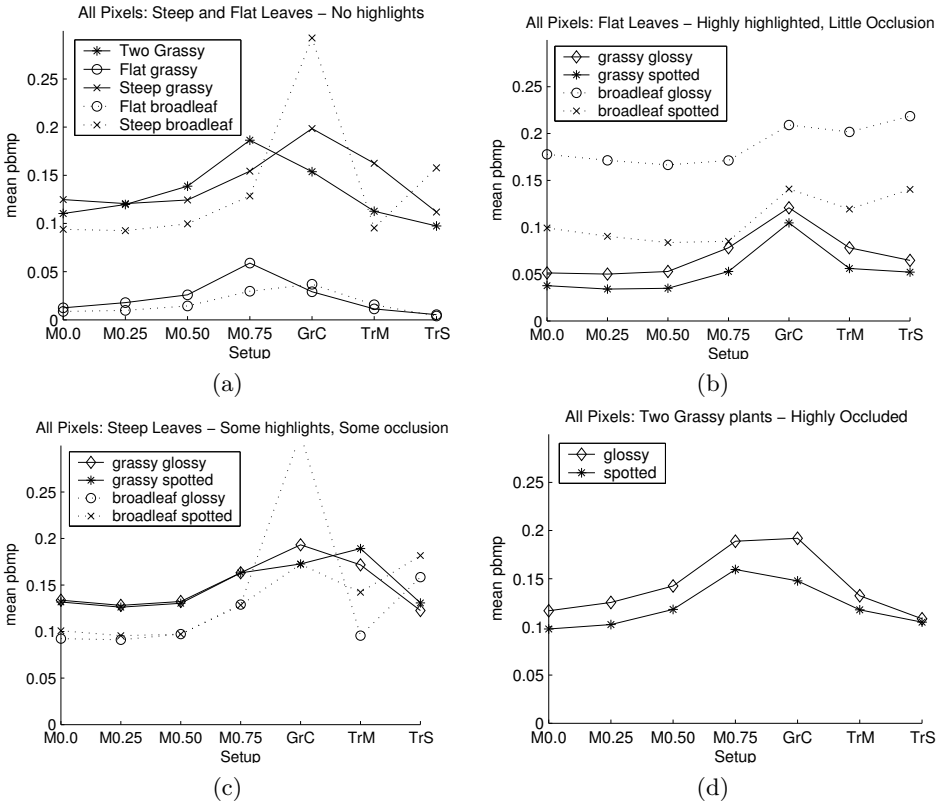


Figure 5.7: PBMP from all pixels results by object type and leaf orientation. The worst case occlusion is the *Two Grassy Plants* model being 5% occluded. The worst case of highlights were the flat grass-like and flat broad-leaf. 20% of their area suffered from changing specular highlights.

The flat glossy broad leaf was the most difficult to reconstruct. Note that this is the plant type that was 50% highlighted, and there were no texture other than shading and bumps to correlate. The graph cut algorithm were particularly bad in this case, because it created non existant surfaces in over the plant from the errors of the highlights.

5.3 Conclusions

The relationship between the performances of the algorithms and the descriptive parameters of the plant objects were investigated. A new multi-baseline Sum of Squared Difference based correlation was defined (SISSD) in order to minimize the effect of perspective distortion within the windows. The results showed that there was a relationship between the performance and the descriptive parameters of the objects. However, SISSD was only a marginal improvement on images with steep leaves (slopes), but more so in the presence of highlights. It was mainly an improvement at the actual highlight areas, especially on shiny

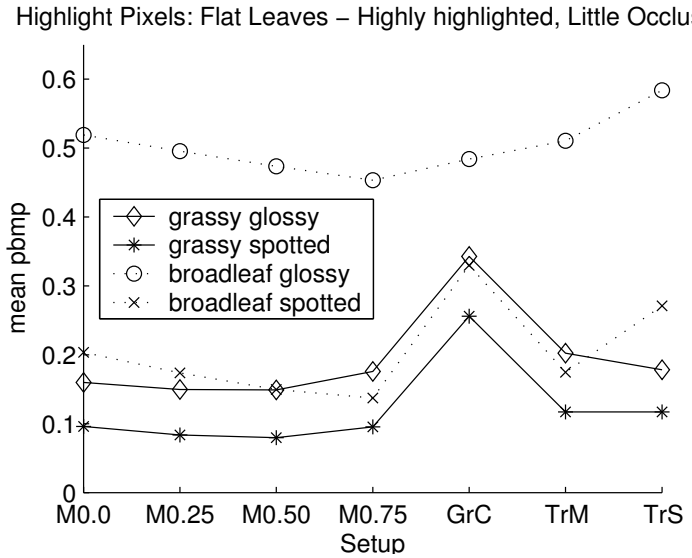
broad leaf plants. On the other hand SSSD was better at matching the occluded areas. The best algorithm for occluded areas was the trinocular T_m algorithm. Binocular Graph cuts were not able to reconstruct the slopes in steep leaves, but the smoothness optimization seemed to smoothen over the errors from highlights, when the highlight areas were not too large. The results showed a complicated relationship of trade-offs that points toward further development combining the strengths of the individual configurations.

5.3.1 Perspectives on Future Work

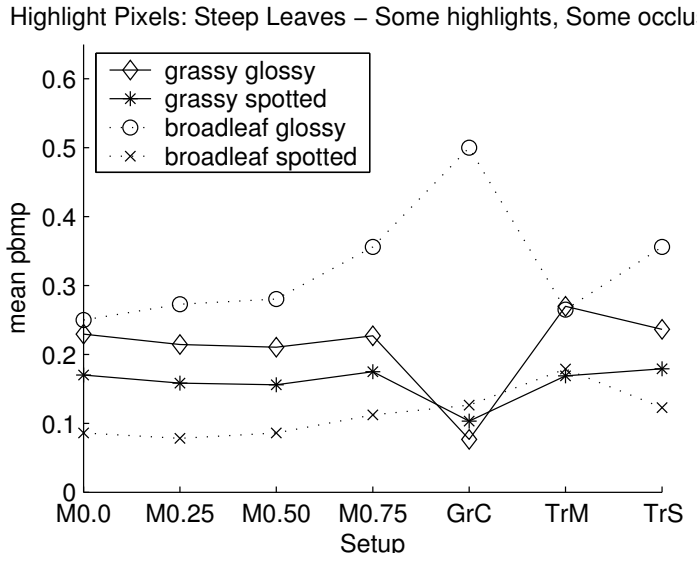
An improvement to the SISSD measure could be to have α depend on the distance from reference image. Another interesting aspect would be to place the 5 cameras in a trinocular setup. The five cameras would then complete two systems of three-camera multi-baseline systems in each direction.

Furthermore, a multi-baseline or trinocular algorithm in combination with graph cuts would be interesting to pursue, and to improve its ability to reconstruct steep slopes. There are other works on these aspects to pay special attention to Buehler et al. [2002]Lin and Tomasi [2004]. Buehler's trinocular algorithm does not handle the situation where occlusion only exist in one camera pair. This was the strength of the trinocular minimum algorithm in this thesis. Lin and Tomasi's algorithm for sloped surfaces relies too strongly on large smooth surfaces. This may be a problem for natural leaves that can be curled and there might only be small segments showing of each leaf, while the surface boundaries are only vaguely defined by intensity edges (sometimes not at all).

The final step is to create a mesh that is able to treat intertwining and overlapping leaves as individual surfaces.

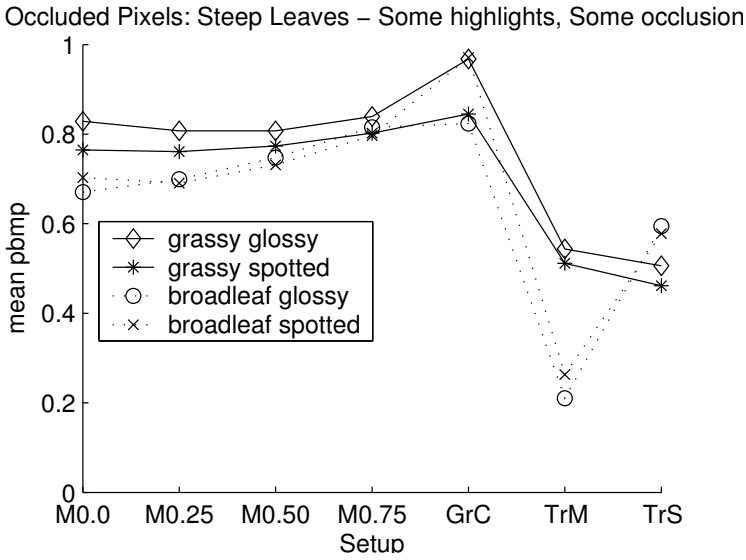


(a)

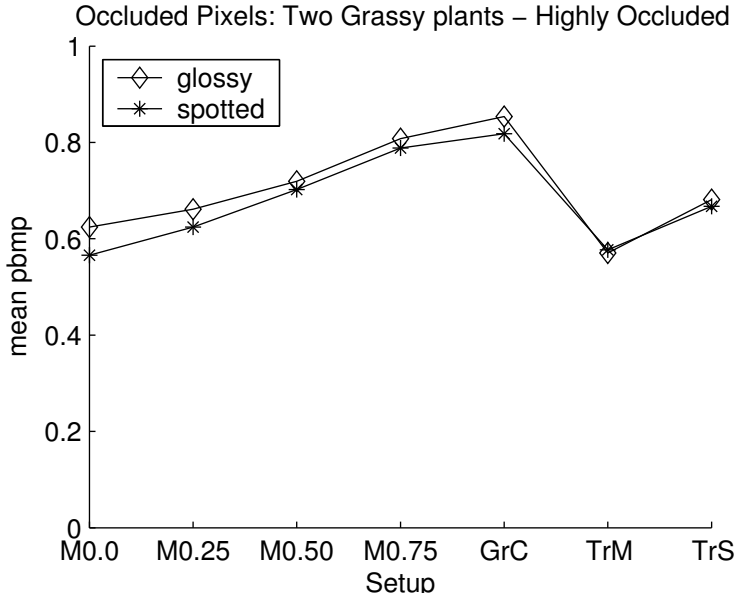


(b)

Figure 5.8: PBMP from specular changing highlight pixels results by object type and leaf orientation. SISSD (M0.25-M0.75) improves performance.



(a)



(b)

Figure 5.9: PBMP from occluded pixels results by object type. Trinocular minimum T_m is the best algorithm for occluded areas.

Chapter 6

Combined SMW

[1]

Summary:

It turned out that smw 1 and smw 5/9 was good at different areas of especially textureless broad leaves in chapter 4. In this chapter it will be attempted to combine the best from both worlds. The idea is that the error map that is output from the SSD matching can be used to toggle between the maps..

Problems with steep slopes in the structures that need to be reconstructed was a major theme in this research. Early testing using the developed test framework revealed that the symmetric multiple window algorithm ([Fusiello et al., 2000]) also have problems with this. That ha never been documented before. The problem might be excessive in the case of plants, where all pixels generally are green so that the impact of perspective distortion is large compared to pixel window uniqueness[2].

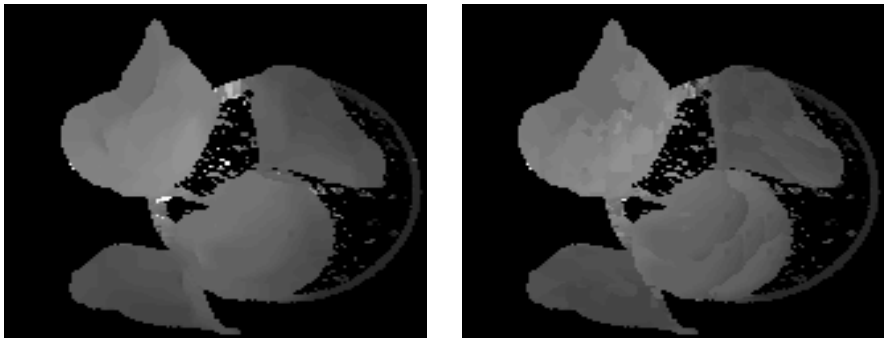
The problem that was found in chapter 4 is shown in figure 6.1. It shows that SMW 1 makes smooth maps but the overlapping leaves blend together. SMV 5 [3] could clearly separate the overlapping leaves, but the smooth surfaces on each leaf was also separated with staircase patterns.

How this is possible might be explained in figure 6.2. It shows a steep leaf seen from 2 viewpoints. Notice how distorted the shape is in view 2. When using multiple windows the SSD is computer for centered as well as off-centered windows (figure 6.2(a) shows 2 such off-centered windows that it is trying to match for the same pixel). Normally the window SSD would be calculated for centered windows on those corresponding windows on the vertical axis (figure 6.2(b) shows the correct match if those windows were centered). Notice how the horizontal offset between the windows is skewed in view 2 compared to view 1. The steeper the slope is, the greater the off-set will be. If it were flat, there would be no offset. The point being that SMW 5-9 is testing windows at different depths for the same pixel. If in scanline Y the smallest SSD is found in a window off-centered above the scanline and the

[1] Results in this chapter was presented in paper Nielsen and Andersen [2008]

[2] i.e. the dissimilarity energy for a wrong correspondence is not larger than the dissimilarity for the correct correspondence.

[3] Clarification: and also SMW 9



(a) (b)

Figure 6.1: Disparity maps of real cotton. Brighter pixels are closer to the camera. Configuration: trinocular sum, window size 16, Colour images. (a) SMW 1. (b) SMW 5.

next scanline $Y + 1$ has a best match in a window off-centered below, those windows most likely find disparities further apart than they should. And that is how the staircase pattern emerges.

6.1 Method

This is an extension of the symmetric multiple windows in [Fusiello et al., 2000]. It combines the results using only one centered window and using five or nine windows centered around the corners (smw 5) and the sides (smw 9).

The smw algorithm is easy to adapt such that it returns a disparity map and SSD error map using the centered window, and a disparity map and SSD error map using smw 5 or smw 9 (depending on the choice).

Figure 6.3 shows the error maps corresponding the the disparity maps in figure 6.1.

The combination of the disparity maps can be found using the SSD error ratios, equation 6.1. Figure 6.4 shows the corresponding error ratio map. The ratio is espeically large at the edges (discontinuities). The disparity for the combined map is chosen by thresholding, equation 6.2.

$$\eta = e_1/e_{59} \tag{6.1}$$

$$d_{comb} = \begin{bmatrix} d_{59} & \text{, if } \eta > T \\ d_1 & \text{, otherwise} \end{bmatrix} \tag{6.2}$$

where T is the error ratio threshold, d_{comb} is the combined disparity map, d_1 is the disparity map from the centered window, and d_{59} is the disparity map from the smw 5 or smw 9.

In practice the centered window is used by default, except where the correlation with the centered window is T times worse than the multiple windows. Note that d_1 is never better

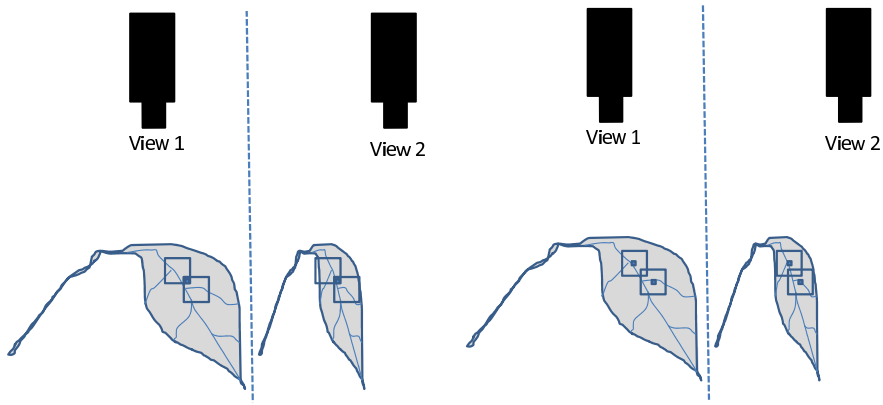


Figure 6.2: A steep leaf seen from two viewpoints. (a) Window correlation is computed for two corresponding pixels using multiple windows off centered above to the left and below to the right. (b) If those windows were computed using centered windows over the scanlines of their centers, they would really be at different depths and thus find best matches at different disparities. This is true only if the object is steep. A flat object does not give this problem.

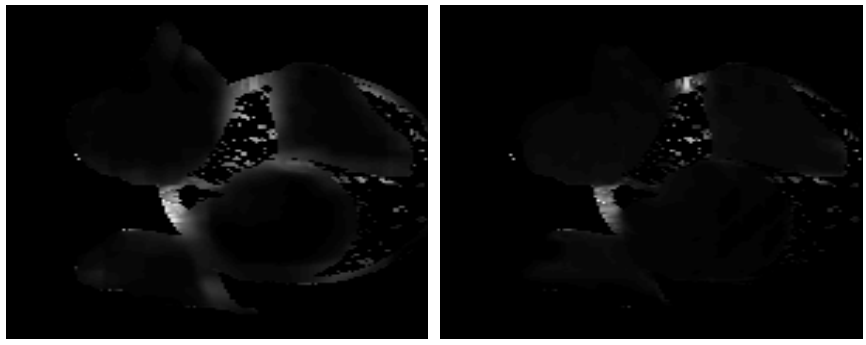


Figure 6.3: Error maps of real cotton. Brighter pixels are larger errors. (a) SMW 1. (b) SMW 5.

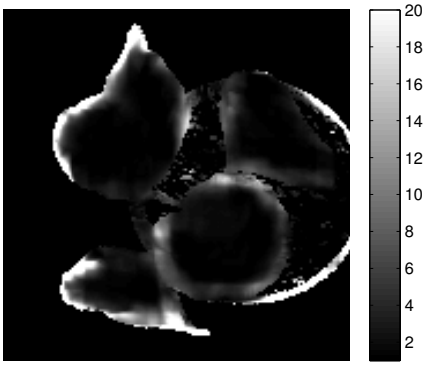


Figure 6.4: SSD error ratio map. Bright pixels are large ratios.

than d_{59}, because if the centered window is better than any of the other windows, then the centered window would be been the best match anyway. Thus $d_{comb} \in [1..\infty[$

6.2 Results

The test was performed on images of the 11 real plant images used earlier in this thesis (chapter 4.2). Figure 6.5 finds the optimal error ratio threshold, $T = 4$. The ratio η does not depend on the window size (w), as the sums of all the window elements ($N = w^2$) is simply the mean times N (μN). Doing the ratio divides N out, see equation 6.3.

$$\eta = \frac{\sum_{i=1}^{w}\sum_{j=1}^{w}(a_{ij}-b_{ij})^2}{\sum_{i=1}^{w}\sum_{j=1}^{w}(a_{ij}-c_{ij})^2} \tag{6.3}$$

$$= \frac{N\sum_{i=1}^{w}\sum_{j=1}^{w}(a_{ij}-b_{ij})^2}{N\sum_{i=1}^{w}\sum_{j=1}^{w}(a_{ij}-c_{ij})^2} \tag{6.4}$$

$$= \frac{\mu_{(a-b)^2}}{\mu_{(a-c)^2}} \tag{6.5}$$

Figure 6.6 shows the result from the cotton plant used in figure 6.1. It shows that the smooth inside of the leaf surface is smooth like smw 1, but the edges between the overlapping leaves are in fact reconstructed as in smw 9.

Note that the boundaries between those areas that are taking from SMW1 and those taken from SMW9 sometimes are not smooth and some work should be dedicated to solving that issue.

6.3 Conclusion

A simple method was developed that combines the strength of using multiple windows with the strength of using a centered window. The improvement in the disparity map was only

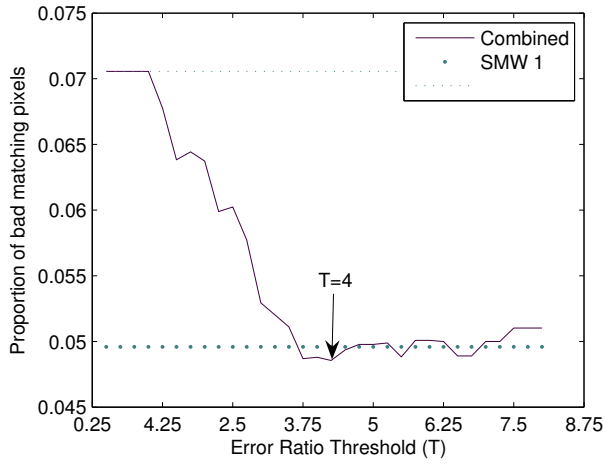


Figure 6.5: PBMP as a function of error ratio threshold.

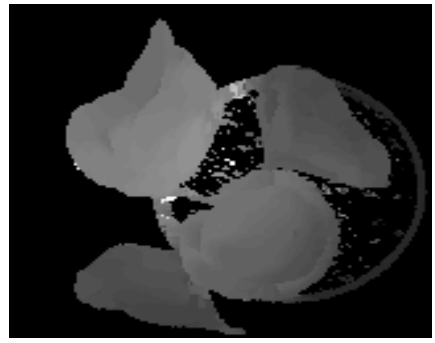


Figure 6.6: Disparity maps of real cotton. Brighter pixels are closer to the camera. Configuration: trinocular sum, window size 16, Colour images. Combined SMW, $T = 4$.

a small improvement, but the effect of this improvement will be examined later in chapter 10.1.

It is worth researching a combination of trinocular sum and trinocular min as well, perhaps by double checking disparity map for all views simultaneously to detect occlusions (where the maps do not correspond to each other). Trinocular min should then be used in those areas. Unfortunately, I cannot do this within the span of my project.

Chapter 7

Trinocular Graph Cuts for Piecewise Smooth and Sloped Surfaces

Summary:

Based on the problems found in the test it was attempted to improve the graph cut algorithm with a third camera and allow sloped surfaces. There is some work in this area [Lin and Tomasi, 2004] where a simple surface b-spline fitting is wrapped around graph cut segmentation. This adds to the complexity of the algorithm but is good if surface fitting is planned. The weaknesses of the sloped surface graph cut algorithm in [Lin and Tomasi, 2004] is that it assumes rather large connected smooth surfaces and it cannot get into the tight spots just like the window based methods. It would be interesting to test it on the images, but that is not possible within the scopes of this thesis. It also makes the complexity of the algorithm much worse. It would be an advantage to retain the graph cut ability to model the small areas while preserving sloped connectedness modeling the natural roughness and sudden bends in the structures. The aim was to find an improvement that does not affect the complexity of the algorithm and keep a modular approach. Some applications may want to process the disparity map directly, while other may need to do the surface fitting that may include other information. By keeping the disparity estimation and surface fitting modular, it is possible to choose separate paths for various applications.

7.1 Theory

Energy minimization is valuable in computer vision problems when there is a number of assumptions that can be used as constraints to the results. Typical situations are when there is a noisy data measure and the assumption that it is piecewise constant or piecewise smooth. See figure 7.1.

An energy function must be designed to smoothen the data without blurring it. This is called discontinuity preserving smoothing. The most general energy function is equation 7.1.

Piecewise constant profile Piecewise smooth profile

Figure 7.1: Different disparity/depth map profiles. Most research has been limited to the piecewise constant profile.

$$E(f) \quad = \quad D(f) + V(f) \tag{7.1}$$

where $E(f)$ is the energy of the configuration f (in general the labels of its variables $x_p = f(p)$). In computer vision the configuration f could be the set of values for its pixels, e.g. intensities (image restoration), object labels (segmentation), or disparities (3d reconstruction). D(f) is the total static cost of assigning the given configuration to the given pixels. V(f) is the total cost of relationships between neighboring variables (e.g. neighboring pixels). The neighborhood scheme can be defined freely. The energies from each neighborhood are simply added and optionally weighted.

Graph cuts is the leading energy minimizing algorithm for computer vision. Simulated annealing is very slow and does not always find a strong local minimum, because it makes random guesses. For certain energy functions graph cuts of binary variables can do that very fast (theoretically within 0.15% of the global minimum), becuase it always makes a qualified suggestion that takes state-changes in the neighbour into account. Disparity maps are computed symmetrically for all the views in the same iteration and thus handles occlusion.

7.1.1 Energy Formulation

Graph cut minimization treats the disparity estimation as a labeling problem. It is efficient to minimize certain energy functions using graph cuts for binary variables if the energy terms are regular metrics [Kolmogorov and Zabih, 2004]. Note that it does not find the global minimum, but it computes a local minimum in a strong sense if the energy terms satisfy the constraints in [Kolmogorov and Zabih, 2004].

Two common procedures that converts multi label problems into binary problems are called α-expansion and β-swap. Both methods formulate local energy functions of binary variables, repeatedly, in order to minimize non-binary variables. In each iteration the area of all available labels are attempted to expand. α-expansion considers all pixels to be changed into α except those that are already α. β-swap considers only those pixels that are already β to be changed into α. Hence, the number of steps in each iteration are quadratic to the number of labels in β-swap, while α-expansion has only steps equal to the number of labels.

The energy function of the stereo algorithm has three energy terms of the labeling f:

$$E(f) = E_{data}(f) + E_{smoothness}(f) + E_{visibility}(f) \quad (7.2)$$

$$E_{data}(f) = \sum D_{((p,f(p),q,f(q))\in I_{data})}(p,q) \quad (7.3)$$

$$E_{smoothness}(f) = \sum V_{p,q\in N}(f(p),f(q)) \quad (7.4)$$

$$E_{visibility}(f) = \sum_{(p,f(p),q,f(q))\in I_{vis}} \inf \quad (7.5)$$

D ensures photo consistency between the multiple views where $((p, f(p), q, f(q)) \in I_{data})$ when $f(p) = f(q)$ and can be chosen arbitrarily.

V ensures smoothness within a neighborhood N and must be chosen carefully to be robust and regular. A common regular and robust energy term is Pott's energy, which is 0 for $f(p) = f(q)$ and λ otherwise. Pott's energy is suitable for piecewise constant data sources. The size of λ is a weighting between the data term and the smoothness term.

Examples of piecewise smooth energy terms are the truncated L1- and L2 norms. Truncation makes them robust and discontinuity preserving. It means that the small changes in the labeling are penalized slightly in the energy, until a certain maximum "roof" is reached. Then it is free to assigned greater disparity jumps. The width of V-shape before it levels out determines how certain the data term is expected to be that there is a discontinuity and how noisy the data measurement is expected to be. The L1 norm is regular but L2 is not.

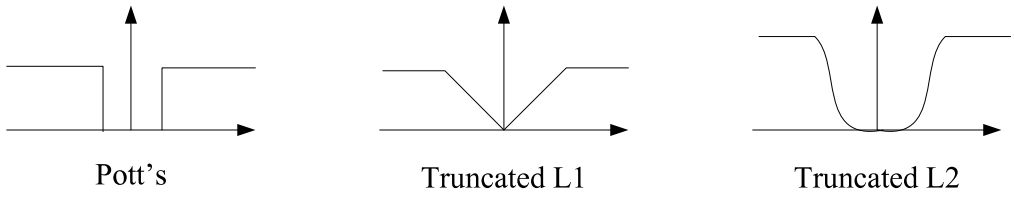


Figure 7.2: Examples of smoothness energy terms V. x axis is $abs(f(p) - f(q))$ and y axis is E(f).

There is a simple test for regularity. In the binary case there is the labels 0 and 1 (source and sink). If $a = V(0,0)$, $b = V(0,1)$, $c = V(1,0)$ and $d = V(1,1)$ then:

$$a + d \leq b + c \quad (7.6)$$

It imposes the demand on the energy that the sum of energies of equal neighbor labels must be less or equal to the sum of neighboring different labels. While this seems easy to uphold, it may not be so, when the sources represents different labels in the multi label problem. Consider an α expansion for $\alpha = 5$ where $f(p) = 0$ and $f(q) = 10$. See table 7.1.

However, in β-swap $V(0,0) = 0$ in all cases and then there is not problem with regularity.

The visibility term handles pixels that blocks the views of others and is infinity when $(p, f(p), q, f(q)) \in I_{vis}$. In the rectified trinocular camera setup p blocks the visibility of q when p is closer to the camera $f(p) > f(q)$.

Table 7.1: Example testing regularity. It shows that L2 is not regular.

Energy term	a	b	c	d	b+c-a-d
Pott's	λ	λ	λ	0	λ
Truncated L1	10	5	5	0	0
Truncated L2	100	25	25	0	-50

7.1.2 Graph Construction

The graph $G = (V, E)$ consists of nodes and edges with weights. There are two terminals source (s) and sink t.

An s-t cut (C)of the graph is a partitioning of the nodes V the sets $s \in S$ and $t \in T$. The algorithm finds the cut that separates the source from the sink that has the lowest cost defined by the sum of all edges that are cut. The procedure is based on the duality between minimum cut and maximum flow. The minimum cut equals the maximum flow that goes into the sink. Flow is defines as the difference between input flow and output flow.

If the graph is constructed in a clever way, the minimum cut and maximum flow equals the energy of the corresponding configuration of the variables plus a predictable constant (K).

$$E(f) \quad = \quad C(f) + E_K \tag{7.7}$$

where E(f) is the energy of the configuration f, C(f) is minimum cut that assigned the configuration f, and E_K is the sum of constant energy terms that does not depend on any variable. It is also usable for compacting the graph (see how below). Note that E_K does not affect the optimal configuration, but is necessary when comparing the energy of a new configuration to the previous.

The thesis bases the work on graph cuts on Kolmogorov's implementation of Kolmogorov and Zabih [2004] for binocular stereo. It is easily adapted for Trinocular stereo by adding energy terms for data penalty, smoothness, and visibility for the third camera on the second baseline.

The data term is based on subpixel sampling:

$$D(p,q) \quad = \quad \min \left(\begin{array}{c} \min((I_1^+(p) - I_2(q))^2, (I_1^-(p) - I_2(q))^2) \\ \min((I_2^+(q) - I_1(p))^2, (I_2^-(q) - I_1(p))^2) \end{array} \right) - K \tag{7.8}$$

where $K > 0$ and I^+ and I^- are maximum and minimum mean pixel values of the 4-neighborhood ($n1 - n4$):

$$I_{n1}(x,y) = \frac{I(x,y)+I(x-1,y)}{2} \tag{7.9}$$

$$I_{n2}(x,y) = \frac{I(x,y)+I(x+1,y)}{2} \tag{7.10}$$

$$I_{n3}(x,y) = \frac{I(x,y)+I(x,y-1)}{2} \tag{7.11}$$

$$I_{n4}(x,y) = \frac{I(x,y)+I(x,y+1)}{2} \tag{7.12}$$

The smoothness term is an adaption of Potts energy that uses edge information as static clues:

$$V_{p,q\in N}(f(p),f(q)) = \begin{cases} 0, \text{if} f(p)=f(q) \\ 3\lambda, \text{if} f(p)\neq f(q) \text{and} \Delta I(p,q) < I_{thresh} \\ \lambda, \text{otherwise} \end{cases} \tag{7.13}$$

It uses information about edges in the images $\Delta I(p,q)$ is the maximum difference between the neighboring pixels in the R, G, and B channels.

The visibility term is the same as above but will be optional. Its benefit will be tested in the test.

In order to build the graph, an energy matrix is built and broken down into a sum of terms that each can be represented by single edges, see figure 7.3. The AB/CD square is broken into terms, where each term depending on zero variables like the constant A, one variable like the 2 middle terms, or a fixed configuration such as the last term B+C-A-D if i=0 and j=1. Those terms can be placed as edges directly into the graph as shown in figure 7.4. $A = A+0+0+0$, $B = A+0+D-C+B+C-A-D$, $C = A+C-A+0+0$, and $D = A+C-A+D-C+0$.

A is a constant that is added to E_K. The second term depends on variable i, the third term depends on variable j, and the last term depends on both $i=0, j=1$. Note that this is just one separation out of many possible ways (see appendix C for another one used in the implementation and a practical example comparing the two graphs.).

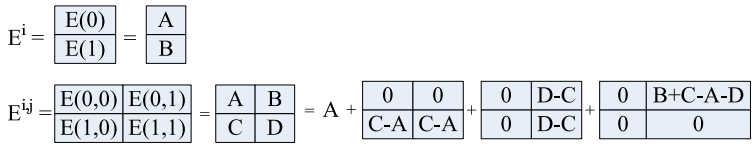


Figure 7.3: In order to build the graph, an energy matrix is built and broken down into a sum of terms that each can be represented by single edges. A is a constant that is added to E_K.

Figure 7.4 shows how to construct the subgraphs between one or two variables once the energies are computed. Consider the data term depending only on one variable.

Note that the data cost for keeping the label as it was (A) is not assigned to the edge that goes from the source to the node, but the opposite side. The same goes for (the data cost for assigning the new label). For example if A is larger than B, then the node should be connected to the sink after the cut. Thus, the larger energy should be assigned to the edge going to the sink. This makes it cheaper to cut the edge to source. In order to spare the number of edges in the graph, the only largest edge is kept with the smaller capacity subtracted (and added to the constant.).

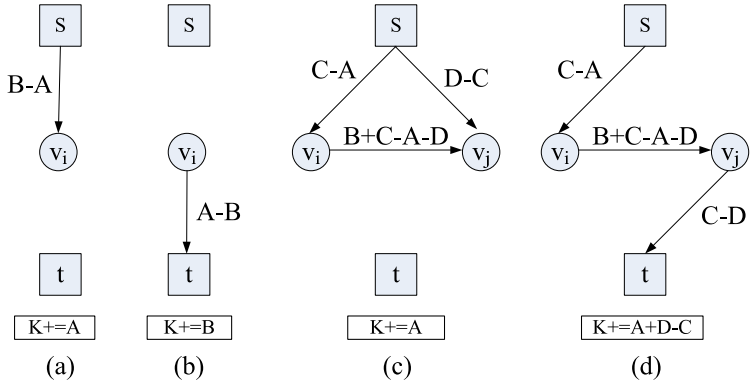


Figure 7.4: Graph construction for E^i and $E^{i,j}$. (a) Single variable where $E^i(0) < E^i(1)$, else (b) $E^i(0) \geq E^i(1)$ on the other side. The smallest energy is added to the constant (K). (c) Two variables where $E^{i,j}(0,0) + E^{i,j}(1,1) \leq E^i(1,0) - E^i(0,1)$, and $C - A \geq 0$ and $D - C \geq 0$. (d) if $C - A < 0$ or $D - C < 0$ then the corresponding edge is subtracted on both sides and added to the constant.

It is easy to test whether the graph representation is correct. Figure 7.5. The sum of the edges that are cut plus the constant should equal the energy of assigning the configuration to 0 (source) and 1 (sink). The fat arrows are those edges that are cut (dashed lines). Note that directional edge between the nodes is only active when the cut assigns node i to source and j to sink. Alternatively, there is no flow between them and it is not necessary to cut the edge to separate the sink from the source.

The complete graph (figure 7.6) is constructed by adding all the edges that goes from the source to the same node or from the same node to the sink.

Summary of Designing a graph cut algorithm

The following is a quick review how easy it is to design a graph cut algorithm from [Kolmogorov and Zabih, 2004].

1. Consider using Graph Cuts if:

- You have a trade-off between at least two factors.
- You can divide these factors into data costs and neighborhood costs
- You assume piecewise constancy or piecewise smoothness.

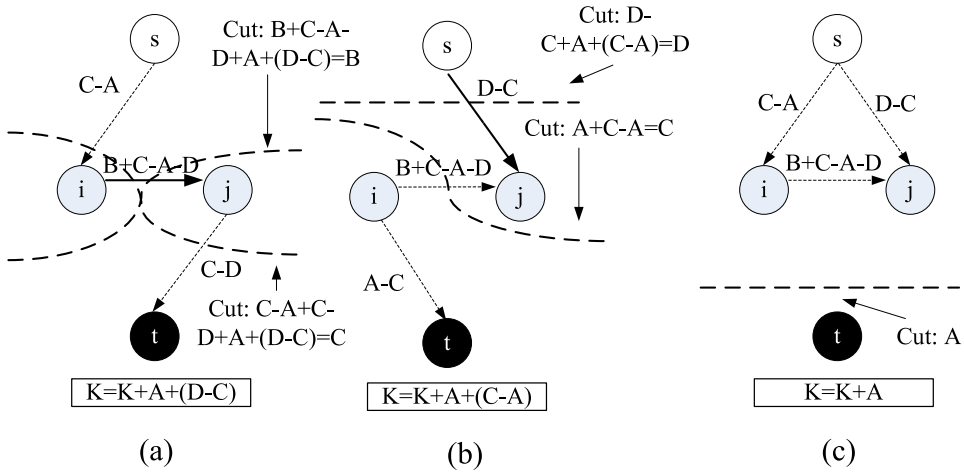


Figure 7.5: The sum of the edges that are cut plus the constant should equal the energy of assigning the configuration to 0 (source) and 1 (sink). The fat arrows are those edges that are cut (dashed lines). Note that directional edge between the nodes is only active when the cut assigns node i to source and j to sink. Alternatively, there is no flow between them and it is not necessary to cut the edge to separate the sink from the source.

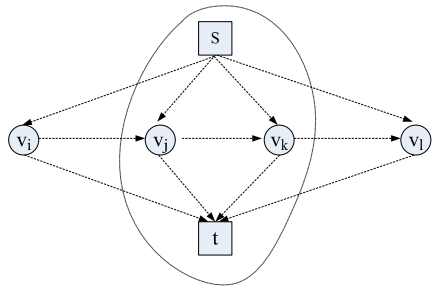


Figure 7.6: Merging of subgraphs by adding the edges together.

- You do not have strict real-time demands.

2. Define your data cost ($d(f(p))$) and neighborhood (smoothness) costs ($V(f(p), f(q))$). The data cost relates to how certain your measurement is that the label ($f(p)$) is true. The neighborhood cost relates to how plausible neighboring assignments ($f(p)$ and $f(q)$) are, thus imposing structural constraints.

4. Demonstrate that $V()$ is in fact convex, robust, and regular in either a α-expansion or a β-swap.

5. Choose a neighborhood scheme (such as 4 or 8 connectivity) based on your assumptions of the structural form.

Appendix C shows a practical example.

7.2 Sloped Extension

The only change that needs to be reconsidered is the smoothness term. Piecewise smoothness rather than constancy is assumed for sloped surfaces. Therefore, Pott's term is not suitable as it would produce staircase results for sloped surfaces. If the truncated L1 norm is adapted to use image edges as static clues like the Pott's term, then the convex part of the function depends on the weighting between data term and smoothness term, and it will always penalize a large gradient more the a small gradient. It is difficult to control to actual weighting between data term and smoothness term. I will investigate adapting Pott's term to allow sloped surfaces.

Pott's term is exactly zero if $|f(p)-f(q)| = 0$. In a 45 deg surface gradient $|f(p)-f(q)| = 1$ and it is penalized in the energy term as much as $|f(p) - f(q)| = 100$. An adaption could be to open up the range where the energy is zero: $|f(p) - f(q)| \leq S$. This allows a gradient of S. The trade-off includes a tolerance for noise in the same range S. Equation 7.14 shows the slope extension for pott's term.

$$V_{p,q\in N}(f(p), f(q)) = \begin{cases} 0, \text{if} f(p) = f(q) \\ \lambda_s, \text{if} |f(p) - f(q)| \leq S \\ 3\lambda, \text{if} f(p) \neq f(q) \text{and} \Delta I(p,q) < I_{thresh} \\ \lambda, \text{otherwise} \end{cases} \tag{7.14}$$

Let us take a closer look at the slope extension term, figure 7.7, it is still robust and convex but not necessarily regular.

Figure 7.8 shows an example where $S = 1$, $\lambda = 6$, and there is no edge.

Regularity can be imposed online by checking that $b+c-a-d \geq 0$. If it is not regular then change b (the energy for assigning the new label to the neighbor) such that $b+c-a-d = 0$, i.e. $b = a+d-c$. Changing b only affects the edge between the neighbor nodes, which would have been negative. This new energy is close to the real energy.

Comparing to adapted truncated L1, where $a = 2$, $b = 1$, $c = 1$ and $d = 0$. The difference is that the energy for keeping the old configuration is much lower, so that is becomes more difficult to insert a gradient between them.

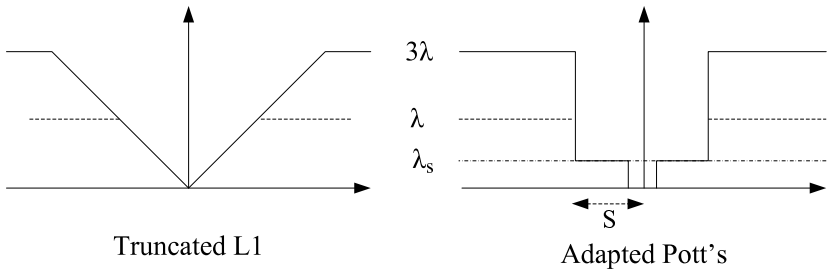


Figure 7.7: Comparison of truncated L1 and adapted pott's for slopes

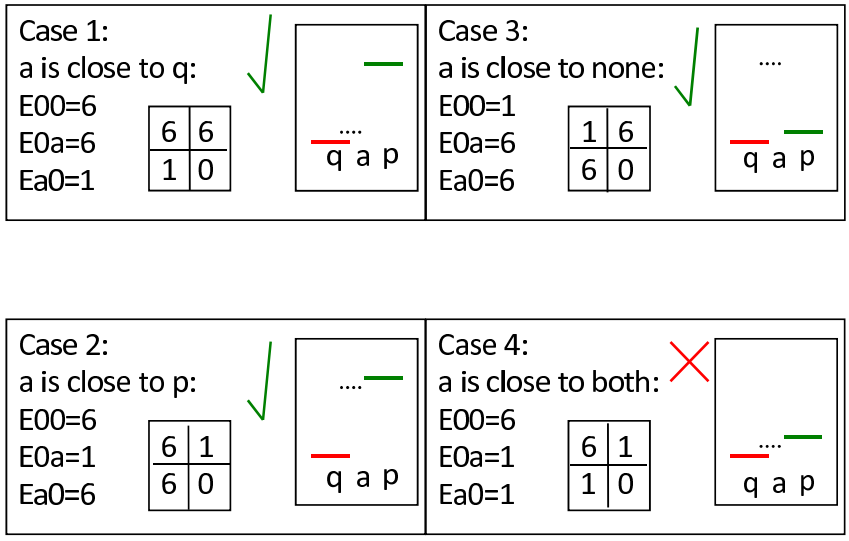
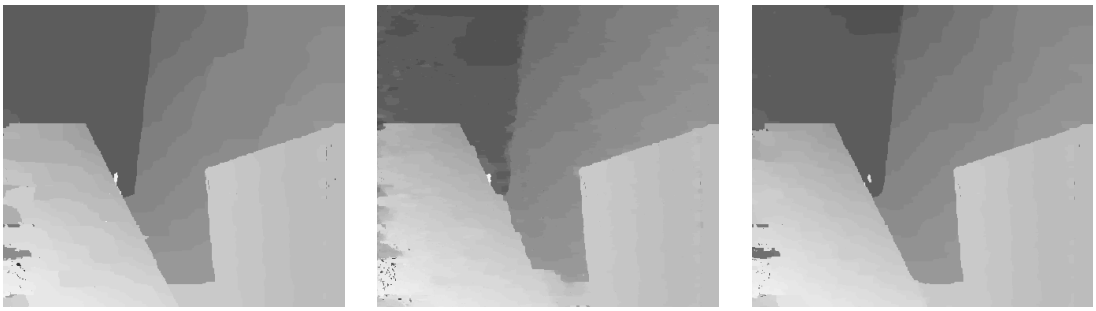


Figure 7.8: Testing regularity

Figure 7.9 shows the comparison between the linear and the adapted slope Pott's. It shows that weak discontinuities are more difficult for the truncated linear term. The original fronto planar term is best at the edges. There are gross quantization in the depth axis resulting in tilted disparity "bands". It is interesting to note that the normal Pott's term (a) produces unrealistic (unevenly distributed) disparity "bands". The sloped extension for Pott's term produces the best results (where λ_s is about $\frac{1}{3} - \frac{1}{2}$) of λ).



[a] [b] [c]

Figure 7.9: Comparison of kz1 with fronto-planar Pott's term (a), sloped extension with truncated linear term (b) sloped extension with adapted pott's term (c). Results after 3 iterations, $\lambda = 5$ (from kz1's automatic heuristic).

7.3 Results

The following is a series of tests on images from the middlebury test set and some plant scenes.

Figure 7.10 and 7.11 shows examples of estimated disparity maps for steep leaves using the standard graph cut (kz1) and my slope extension with $\lambda_s = 0$ and $\lambda_s = 1$. It shows that kz1 gives a stair case pattern, which is reduced with the sloped extension. When $\lambda_s = 0$ there's a little noise, which can be attenuated with $\lambda_s = 1$ without introducing much stair case pattern.

The middlebury test images Tsukuba, Sawtooth and Venus are included to make sure the sloped extension does not significantly affect the results on these well known images. Performance scores are proportion of bad matching pixels. See figures 7.12

The middlebury images can be roughly described at frontoplanar structures. Even through Venus and Sawtooth are not, the quantization of the disparities approximates the slopes as large constant planes. It is found that the visibility constraint helps the disparity estimation a little and that using the sloped extension with $\lambda_s > 0$ performs as well as the normal Pott's term.

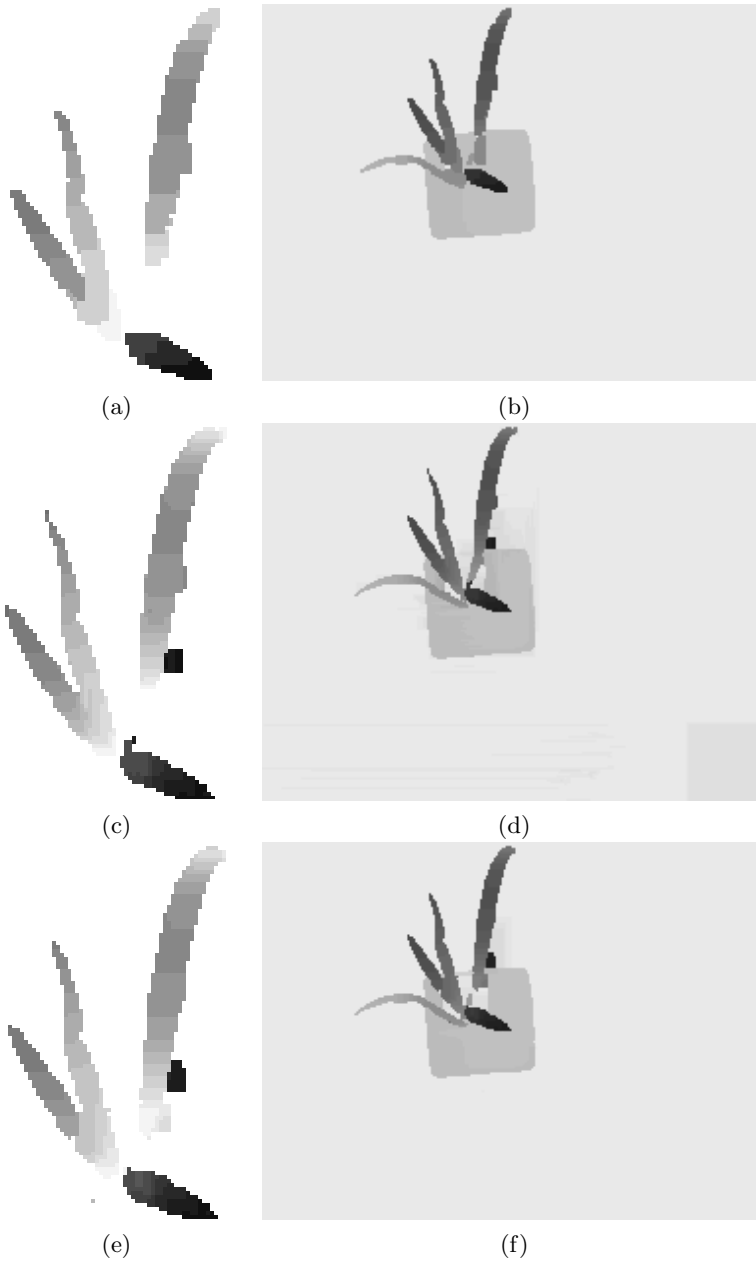


Figure 7.10: Comparison of kz1 to sloped extension with adapted pott's term. a-b: kz1. c-d: sloped extension where $S = 1$ and $\lambda_s = 0$. e-f: sloped extension where $S = 1$ and $\lambda_s = 1$. To the right is full disparity maps and to the left is zoomed in on the plants and improved contrast.

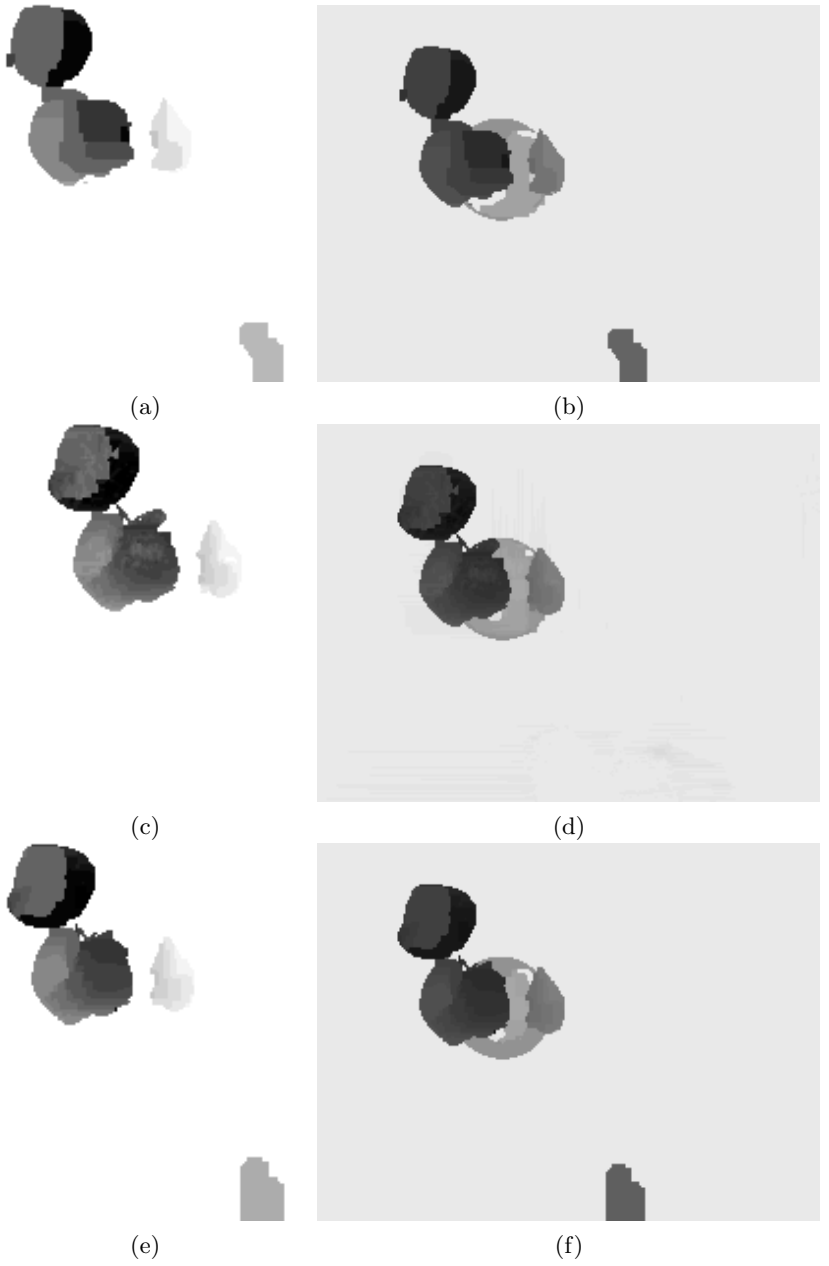


Figure 7.11: a-b: kz1. c-d: sloped extension where $S = 1$ and $\lambda_s = 0$. e-f: sloped extension where $S = 1$ and $\lambda_s = 1$. To the right is full disparity maps and to the left is zoomed in on the plants and improved contrast.

The comparison is also done with virtual plant images in figure 7.13. With these structures the visibility constraint makes the performance slightly worse. The best performance is achieved by choosing the sloped extension with λ_s = approximately $\frac{1}{4}$-$\frac{1}{3}$ of λ.

The fact that the energy measure in the sloped extension to Pott's energy term is more complex and not regular might seriously affect the energy minimization time (the number of iterations). The number of iterations and compared in figure 7.14. The absolute values cannot be compared but note how fast they converge. Using the sloped extension roughly requires 1-2 extra iterations.

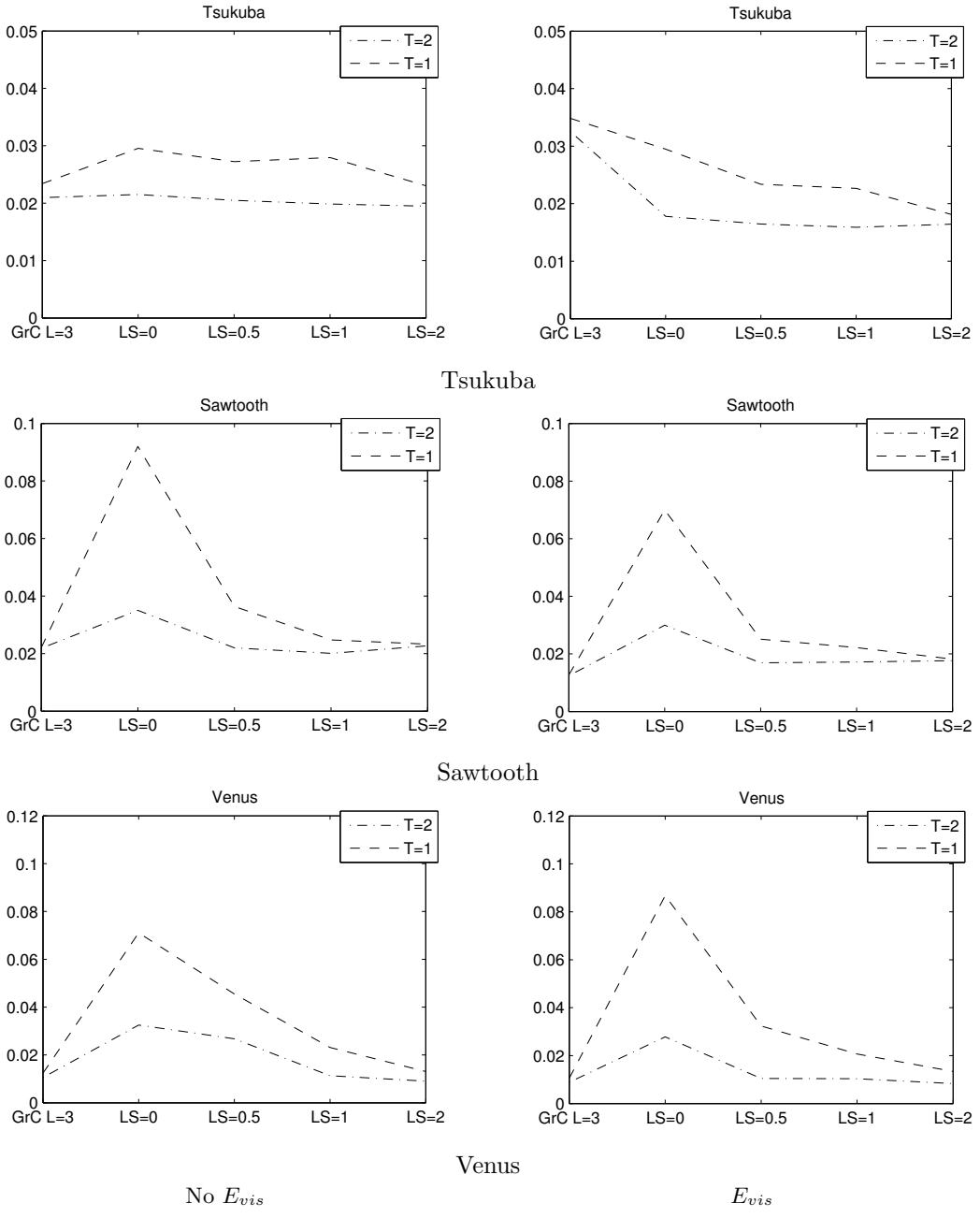


Figure 7.12: Comparison on real images with and without visibility constraint. Results with Pott's term when $\lambda = 3$ (L) and for sloped extension with different λ_s (LS). T is the tolerance in the proportion of bad matching pixels.

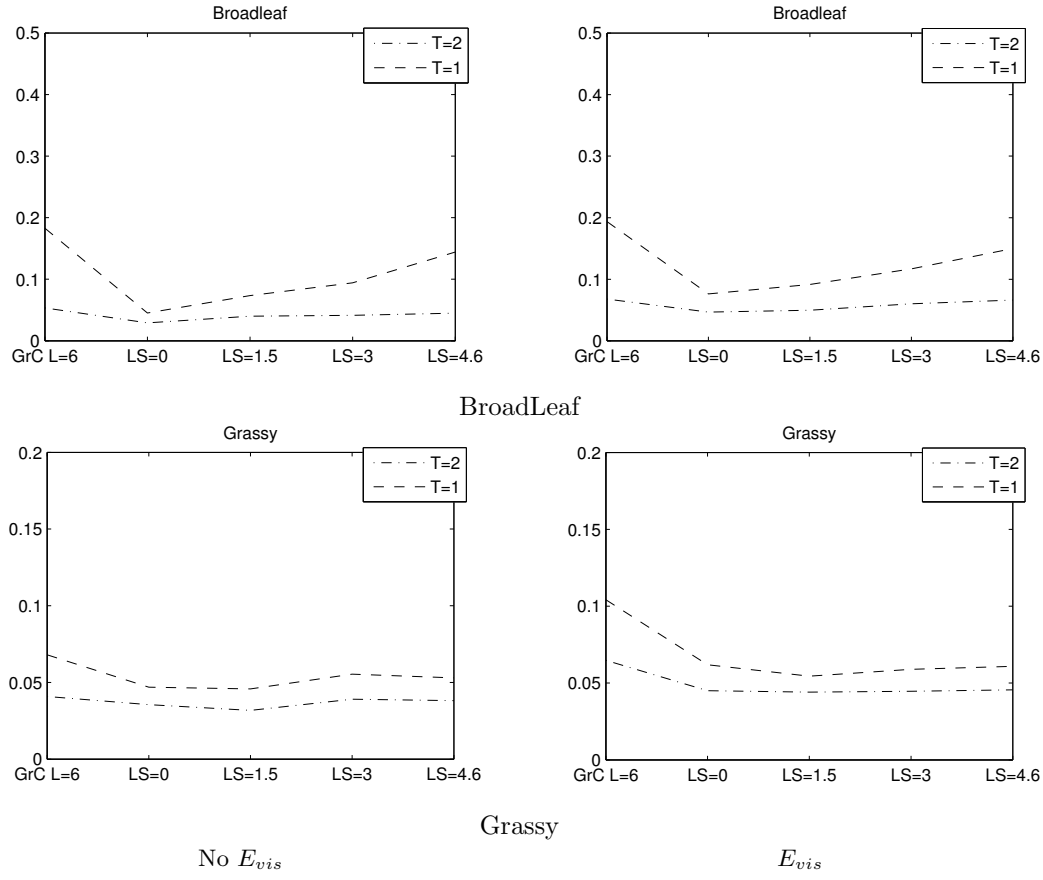


Figure 7.13: Comparison on virtual images with and without visibility constraint. Results with Pott's term when $\lambda = 6$ (L) and for sloped extension with different λ_s (LS). T is the tolerance in the proportion of bad matching pixels.

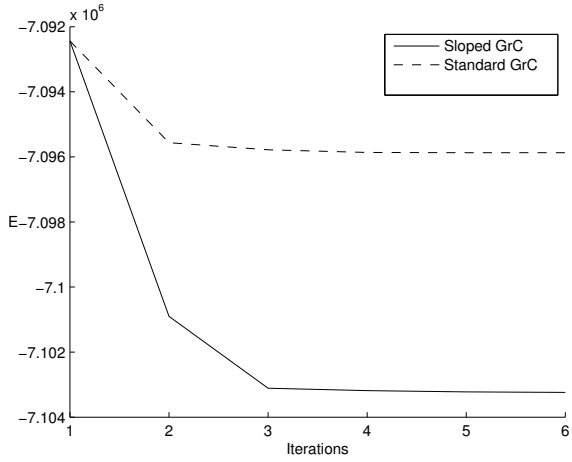


Figure 7.14: Energy minimization time by iterations.

Figure 7.15 shows a full comparison of algorithms and settings. The comparisons are done in pairs. Each pair uses the overall best setting from the former pairs, e.g. the λ test uses only the Trinocular results and the visibility test uses only results from Trinocular and $\lambda = 6$, etc. The new *sloped* extension is an advantage for all images except image 8, which is a very flat grass leaf plant. Trinocular is better than binocular. Sloped extension is better especially when lambda is higher. The visibility constraint turned out to be a disadvantage because entire leaves not reconstructed when a strong constant soil region was chosen as a stable label first.

Figure 7.16 shows an example of estimated disparity maps for steep leaves using the standard graph cut (kz1) and my slope extension.

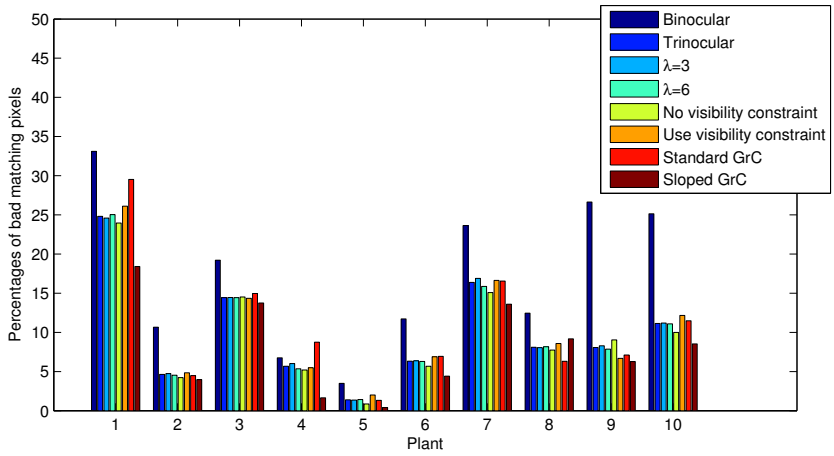
7.4 Conclusion

The goal with this chapter was to adapt graph cut energy minimization to reconstructing plants which are sloped, piecewise smooth surfaces rather than piecewise constant. I wanted to achieve this without making the algorithm more complicated. I also wanted to keep the disparity map estimation independent of a further surface fitting because not all applications actually requires this. This way it is as general as possible.

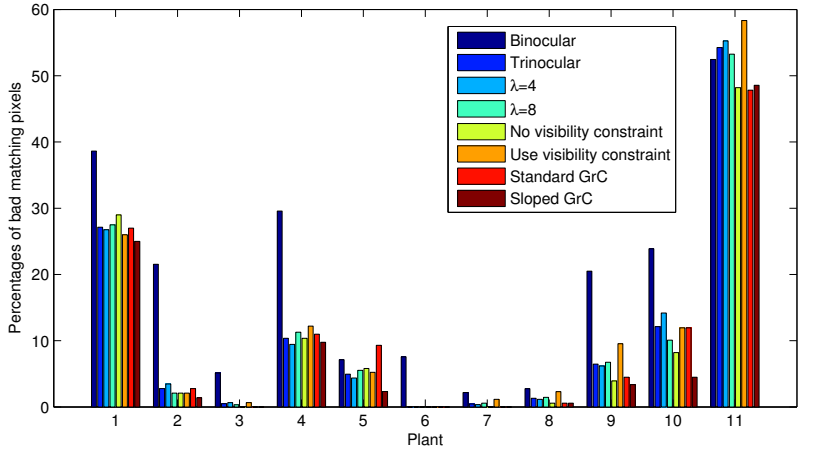
This chapter contained an explanation where and how to use graph cuts and how to make the energy given by the cuts match the real energy of the chosen configuration by keeping track of a constant E_K.

The standard graph cut algorithm [Kolmogorov and Zabih, 2002] performed badly on my plant images, because it created staircase patterns on the slopes. The third camera in the trinocular setup was integrated into the graph simply by adding the energies. It is equivalent to the *trinocular sum SMW* algorithm used earlier.

A new adaption of Pott's energy was designed and found a valuable improvement. While it is still not perfect, it makes for a better a-priori disparity map for a following surface fitting. It introduced 2 parameters S and λ_s. $S = 1$ should be usable in most cases. If it is higher, there's a risk that the noise will be too high. $\lambda_s = \frac{\lambda}{4}$ is generally a good choice.

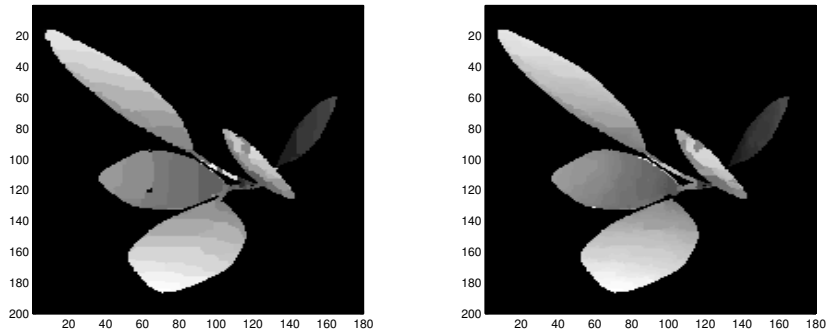


(a)



(b)

Figure 7.15: Graph Cut comparison of settings. The comparisons are done in pairs. Each pair uses the overall best setting from the former pairs, e.g. the λ test uses only the Trinocular results and the visibility test uses only results from Trinocular and $\lambda = 6$, etc. The new *sloped* extension is an advantage for all images except image 8, which is a very flat grass leaf plant. (a) Virtual Plants. (b) Real plants.



(a) (b)

Figure 7.16: Steep plants (a) standard graph cut and (b) slope extension.

Part III

Structural Representations

Chapter 8

Surface Reconstruction

Summary:

Plant structures consists of individual smooth segments. The disparity maps can be separated into individual connected surfaces that can be described individually. Surface reconstruction based on triangulation of segmented discrete disparity maps has staircase patterns which gives a higher area and an unnatural look. Fitting Non-Uniform-Rational-B-Spline yields much more natural structures.

8.1 Segmenting Disparity Maps

The individual objects in the disparity map is segmented using a typical mask based labeling with depth step separation. It is an addaption of the old Rosenfeld and Pfaltz unionfind algorithm.

Refer to figure 8.1. The image is searched from point (0,0). When the first object pixel is found in (8,1), the first region is created with label 1. The next pixel sees the first pixel under the mask and the pixel is close enough in depth ($\Delta Z < maxz$) and is set as the same region. The next pixel that has no other regions under the mask is (12,1) and region 2 is created. The next interesting case is pixel (8,2). Here there are two different regions under the mask; region 1 and 4.

The top pixel has priority and the pixel is set as region 1, and region 4 is marked for replacement with region 1. The same thing occurs in pixel (4,3), but note that the pixel and the replacement is set as the region that is pointed back to by the region under the mask.

Always recurse back to the non-replaced region. Another example is in pixel (12,4) where region 2 has been set to be replaced by region 3, but region 3 is to be replaced by region 5. Thus, the pixel is set as region 5. Pixel (17,5) is not part of Region 5, because it is a 4 connected labeling. In the second pass region 4 and 6 will be added to region 1, and 2 and 3 will be added to region 5.

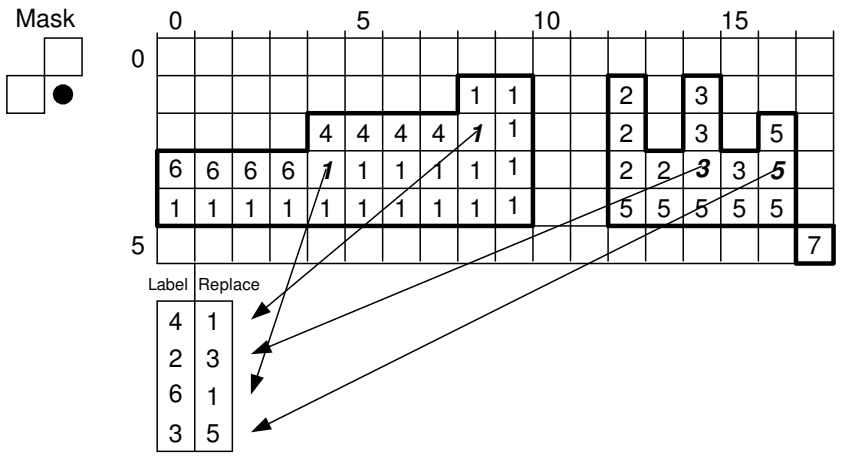


Figure 8.1: In the first pass of the labeling a list of regions are created. Sometimes the regions must be merged so the list also contains information on which unique region a given region is to be merged with in the 2nd pass.

8.2 Non-Uniform Rational B-Splines

Using Non-Uniform Rational B-Splines (NURBS) for geometric models is the industry standard because of their ability to efficiently evaluate models of complex structures and mathematical primitives. Compared to B-spline curves and Bezier curves they are projection invariant (instead of just affine variant) because they use homogeneous coordinates in R^4, which also adds the feature to weight the control points. This means that any projection of the evaluated curve/surface points can be obtained by doing the projection of the control points (incl. weights). That is a fact that will be used later in this chapter.

The NURBS is defined by a set of control points $P \in R^4$ and a set of basis functions N of pth degree corresponding to a set of knots U. The smoothness of the curve depends on the degree.

$$C(u) = \sum_{i=0}^{n} N_{i,p}(u) P_i \qquad u \in [0,1] \tag{8.1}$$

$$U = \{0_0, ..., 0_p, u_{p+1}, ..., u_{m-p-1}, 1_{m-p}, .., 1_m\} \qquad m = n + p + 1 \tag{8.2}$$

$$\tag{8.3}$$

Surfaces are a set of curves distributed along a V-direction. Hence, the (pth, qth) degree NURBS surface is defined by:

$$S(u,v) = \sum_{i=0}^{n}\sum_{j=0}^{m} N_{i,p}(u)N_{j,q}(v)P_{i,j} \qquad u \in [0,1] v \in [0,1] \tag{8.4}$$

$$U = \{0_0, ..., 0_p, u_{p+1}, ..., u_{r-p-1}, 1_{r-p}, .., 1_r\} \qquad r = n+p+1 \tag{8.5}$$

$$V = \{0_0, ..., 0_q, u_{q+1}, ..., u_{s-q-1}, 1_{s-q}, .., 1_s\} \qquad s = m+q+1 \tag{8.6}$$

The *ith* knot basis function of *pth* degree is defined by:

$$N_{i,0}(u) = \begin{cases} 1 & \text{if } u_i \leq u < u_{i+1} \\ 0 & \text{otherwise} \end{cases} \tag{8.7}$$

$$N_{i,p}(u) = \frac{u-u_i}{u_{i+p}-u_i}N_{i,p-1}(u) + \frac{u_{i+p+1}-u}{u_{i+p+1}-u_{i+1}}N_{i+i,p-1}(u) \tag{8.8}$$

8.3 Fitting NURBS

There are many algorithms for fitting NURBS. The simplest use least squares approximation for a fixed number of control points and a uniformly distributed knot vector. Alternatively, an acceptable error is chosen and the algorithm finds the placement and number of control points either by starting with a low number and increasing them until the error tolerance is fulfilled, or by starting with many points and decreasing their numbers. This way corners and sudden extrusions on the surface are modeled better. For segmented, labeled leaves I will assume that a fixed number is sufficient, because there are no corners and extrusions.

The least squares approximation is formulated in the following way. Q $(m \times dim)$, and N $(m \times n)$are known variables and P $(n \times dim)$ is the unknown. The number of control points must be less than the number of data points. N^T is multiplied to the equation because N is not invertible, while N^TN is.

$$NP + \epsilon = Q \tag{8.9}$$

$$N^TNP + \epsilon = N^TQ \tag{8.10}$$

$$P + \epsilon = (N^TN)^{-1}N^TQ \tag{8.11}$$

The data points can be weighted by including a diagonal weight matrix where $w(m,m) = w_m$:

$$wNP + w\epsilon = wQ \tag{8.12}$$

$$N^TwNP + w\epsilon = N^TwQ \tag{8.13}$$

$$P + w\epsilon = (N^TwN)^{-1}N^TwQ \tag{8.14}$$

It is also possible to constrain some of the points if it is very certain. Those points are simply not considered for change in the fitting process. Often this is done to the end points. It could also be a detected feature such as leaf tips or leaf curls.

8.3.1 Fitting Multiple Views

The NURBS can make some wavy shapes when the weights are low at some points. Low weights loosen the grip over the shape and it can deviate much more at these locations. This allows a better approximation at the higher weighted points. Looking at the result from the some view looks correct, but it will look strange from another view (for example fig. 8.2). If other views are known, they can be taking into account while approximating the control points.

a b c

Figure 8.2: A bad reconstruction of a day lily. Looking at a reconstruction using a single NURBS from the same view as the reference camera looks correct (a), but it looks strange and wavy from another view (b). The other view of a better reconstruction using multiple NURBS is shown (c).

The following can be used if multiple disparity maps are known and that objects are segmented and matched in all the views.

First, consider the 2 view case. Assuming that data set Q_1 and Q_2 are related to each other:

$$Q_2 = Q_1 T \tag{8.15}$$

And knowing that NURBS are projection invariant the same transformation T can be applied to P. Thus, the P that minimises least squared error for the two equations is desirable:

$$w_1 N P T_1 = w_1 Q_1 \tag{8.16}$$
$$w_2 N P T_2 = w_2 Q_2 \tag{8.17}$$

where T_1 is the transformation for dataset 1 (here identity matrix is used) and T_2 is the transformation from Q_1 to Q_2. It is easy to add more views up to (Q_I).

It can be written as:

$$w_1 NP = w_1 Q_1 T_1^{-1} \quad (8.18)$$

$$w_2 NP = w_2 Q_2 T_2^{-1} \quad (8.19)$$

$$\begin{bmatrix} w_1 N \\ w_2 N \end{bmatrix} P = \begin{bmatrix} w_1 Q_1 T_1^{-1} \\ w_2 Q_2 T_2^{-1} \end{bmatrix} \quad (8.20)$$

$$\begin{bmatrix} N^T & N^T \end{bmatrix} \begin{bmatrix} w_1 N \\ w_2 N \end{bmatrix} P = \begin{bmatrix} N^T & N^T \end{bmatrix} \begin{bmatrix} w_1 Q_1 T_1^{-1} \\ w_2 Q_2 T_2^{-1} \end{bmatrix} \quad (8.21)$$

And finally:

$$P = (N^T w_1 N + N^T w_2 N)^{-1} N^T (w_1 Q_1 T_1^{-1} + w_2 Q_2 T_2^{-1}) \quad (8.22)$$

$$(8.23)$$

This can easily be extended to I views:

$$P = (\sum_{v=1}^{I} N^T w_v N)^{-1} N^T (\sum_{v=1}^{I} w_v Q_v T_v^{-1}) \quad (8.24)$$

Note that if the weights for all other than Q_1 are zero, then the equation is identical to the normal least squares approximation for one Q.

If the NURBS are fitted to the disparity maps rather than the 3D model T_2 is simply:

$$T_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad T_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$Q_1 = \begin{bmatrix} x_{R1} & y_{R1} & d_{R1} & 1 \\ x_{R2} & y_{R2} & d_{R2} & 1 \\ . & . & . & . \\ x_{RM} & y_{RM} & d_{RM} & 1 \end{bmatrix} \quad Q_2 = \begin{bmatrix} x_{L1} & y_{L1} & d_{L1} & 1 \\ x_{L2} & y_{L2} & d_{L2} & 1 \\ . & . & . & . \\ x_{LM} & y_{LM} & d_{LM} & 1 \end{bmatrix}$$

But how is it possible to match the labels of multiple labeled disparity maps?

8.3.2 Matching the labels in multiple views

Given two labeled disparity maps from two views and a known projection from one view to the other it is easy to find out if a labeled segment (Q_{i1}) in one view matches a labeled segment in the other view (Q_{j2}). See equation 8.25. $Q_i 1$ is a $M_i \times 3$ matrix containing M_i disparity points $[x, y, d]^T$ (d is disparity for point $[x, y]$) from the label i. $Q_j 2$ is a $N_j \times 3$ matrix containing N_j disparity points $[x, y, d]^T$ from the label j.

$$e_m(i, j) = \sum Q_{j2} = Q_{i1} * T_2^T \quad (8.25)$$

where $e_m(i, j)$ is the estimated match between label i in view 1 and label j in view 2.

In principle the method projects the disparity segment i onto view 2, and checks how well the segment overlaps (in x,y, and d) a segment j in view 2.

8.3.3 Choosing the Weights

There are a number of ways to set the weights. A number of approaches were considered:

- Error map from the SSD matching. However, it turned out to yield bad results, because the error map not a good estimate for correctness. Steepness and edges affect the error gravely. Graph cuts do not have the error map. The results were rather wavy at the boundaries, see explanation in figure 8.3.
- Deviation from local mean disparity and assign half weight to large deviations. This filters out the noise from single pixels.
- Occlusion map: Assign half weight to occluded pixels when such detection has taken place.

The two last methods were inconclusive because there was not performed enough testing to actually conclude how well they performed. At first glance the results were identical to using fixed weights.

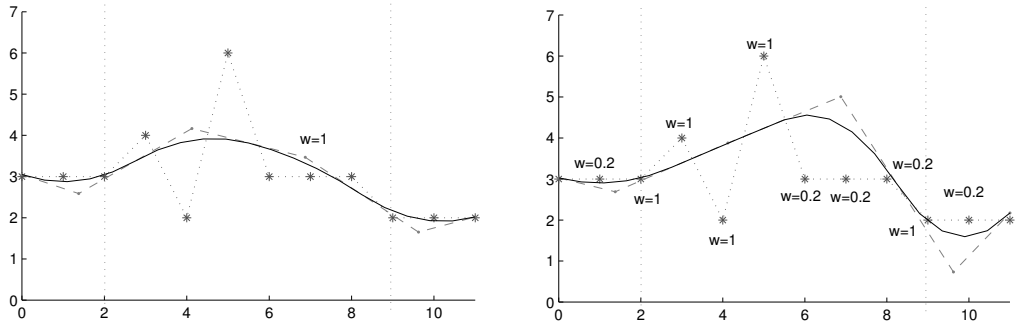


Figure 8.3: The dotted line shows the Q points. The dashed line shows the control points (P). The solid line was the resulting NURBS (C). (a) A fit with fixed weights. (b) A fit with variable weights. Note how the end is thrown into a large curvature because the weights were small. This is a typical case if the weights are chosen from the SSD error map, where the errors are high at the boundaries of the objects.

8.4 Generation of the Mesh

The mesh was generated from the disparity maps. The maps were labeled and a NURBS was fitted to each object. Then a sub pixel disparity map was reconstructed for each NURBS

object. Triangles for the mesh were extracted from neighboring pixels by the principle of figure 8.4. This construction made sure that there are no overlapping redundant triangles. It was done for each object so that the mesh could have overlapping objects.

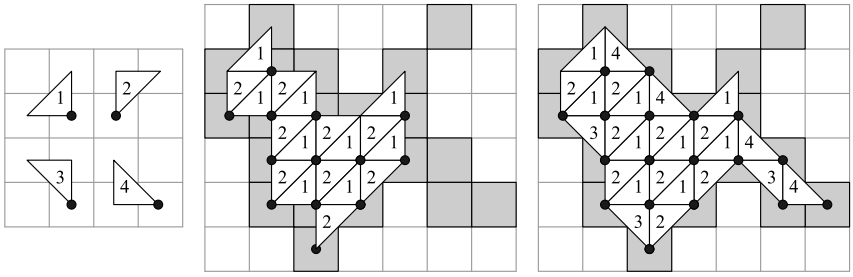


Figure 8.4: There are 4 types of triangles to consider for each pixel. Types 1 and 2 are always placed when all their corners hit the object. Types 3 and 4 are only placed if type 1 was not already placed for the active pixel (the one under the dot).

8.4.1 Looking behind the disparity map

The relationship between disparity maps (equation 8.16) can also be used for merging disparity maps. This makes it possible to generate layered disparity maps where each layer corresponds to an object. Where a layered is overlapped by another layer (occluded), the missing parts can be transfered from other views (assuming that the information is correct).

The procedure is quite simple from section 8.3.2 is being applied.

When a match has been found between Q_i and Q_j, the parts of the disparity segments that are not overlapping are appended to each other. See figure 8.5. The result is that separate objects can coexist at the same $[x, y]$ at different disparities. The limitation is that the same object cannot such as a pig's tail.

The benefits is an increased chance of guessing invisible connections between visible object parts, which allows the algorithm to recreate those connection and describe the object more accurately regarding especially surface area and for example the number of leaves on a plant.

The weakness in using this method is that it is an additional error source from label mismatches and disparity errors. It works only if the disparity was estimated accurately in the alternative view(s), but it is especially those disparities that are occluded in one view that are difficult to estimate.

Figure 8.6 shows a reconstruction of a real hypoestes plant. The leaftmost leaf is outside the reference frame, so that it cannot be reconstructed, but with the occlusion gap filling method developed here, it is possible to append the missing region.

Figure 8.7 shows a reconstruction of the Middlebury sawtooth images. Normally there would be holes behind the disparity planes (8.7a) but with the method developed here, it is possible to look behind the front layer (8.7b).

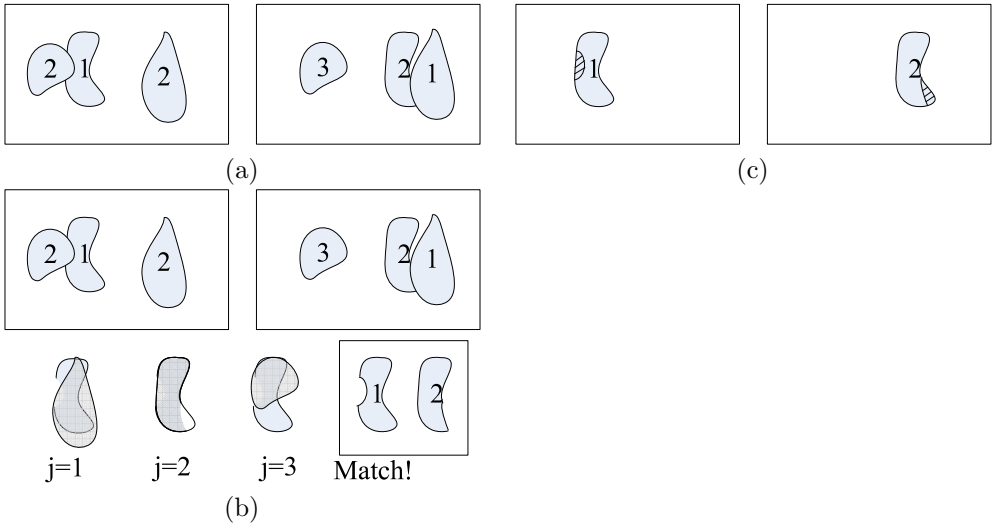
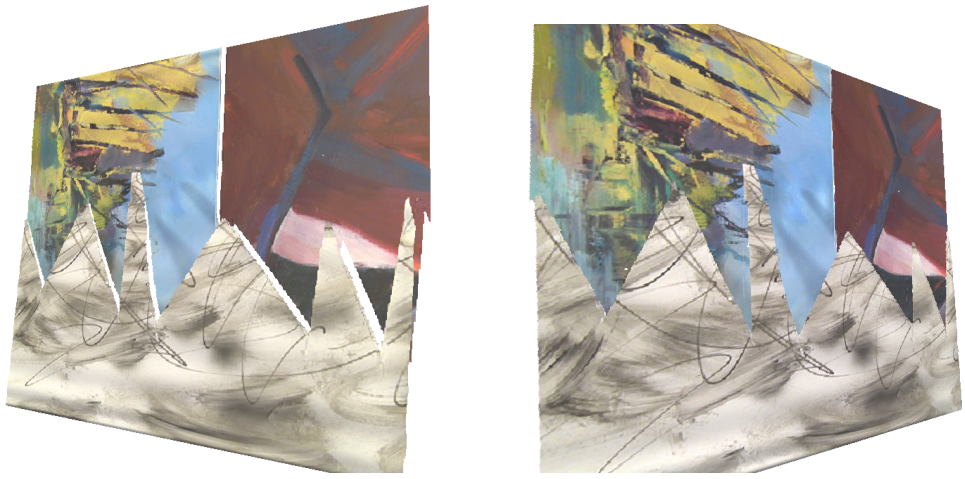


Figure 8.5: Left and right views with disparity segments for three objects. (a) The label numbers do not match. Segment 1 is partially occluded. (b) Object 1 in the left view must be matched with the objects in the right view. Object 1 is projected onto the right view and the object with most disparities (not only the spatial overlap) overlapping is the best match. Here it is object 2. (c) Parts from object 2(right) that does not overlap object 1(left) is appended to object 1 and vice versa.

Figure 8.6: (a) Reconstruction without filling in the gaps from an alternative view. (b) Reconstructed from two disparity maps.



(a) (b)

Figure 8.7: a) Normally there is a gap behind a disparity discontinuity. b) But small gaps can be filled using the occlusion filling technique.

8.4.2 Reconstruction Performance

The effect on the disparity maps is tested by reconstructing the virtual plant images using single view NURBS fitting and dual view fitting. Figure 8.8 shows the percentage of bad matching pixels of the virtual plant image set using grahp cuts with sloped extension, single view NURBS fitting (appendix B shows the disparity maps and where the errors are), and dual view with NURBS fitting.

The effect is positive for most images. Spotted broad leaf 1 and smooth grass leaf 1 (plant 4 and 7 in appendix) were worse off being NURBS fitted with a single view, but dual view fitting was less destructive. Smooth grass leaf 4 (plant 10 in appendix) shows the drawback when the leaves are not matched correctly.

8.5 Extraction of Information

There is valuable information in the mesh that can explain something about the canopy structure.

- Steepness
- Normal image (showing the surface normals at each pixel of the image).
- Area
- Height
- Bounding box

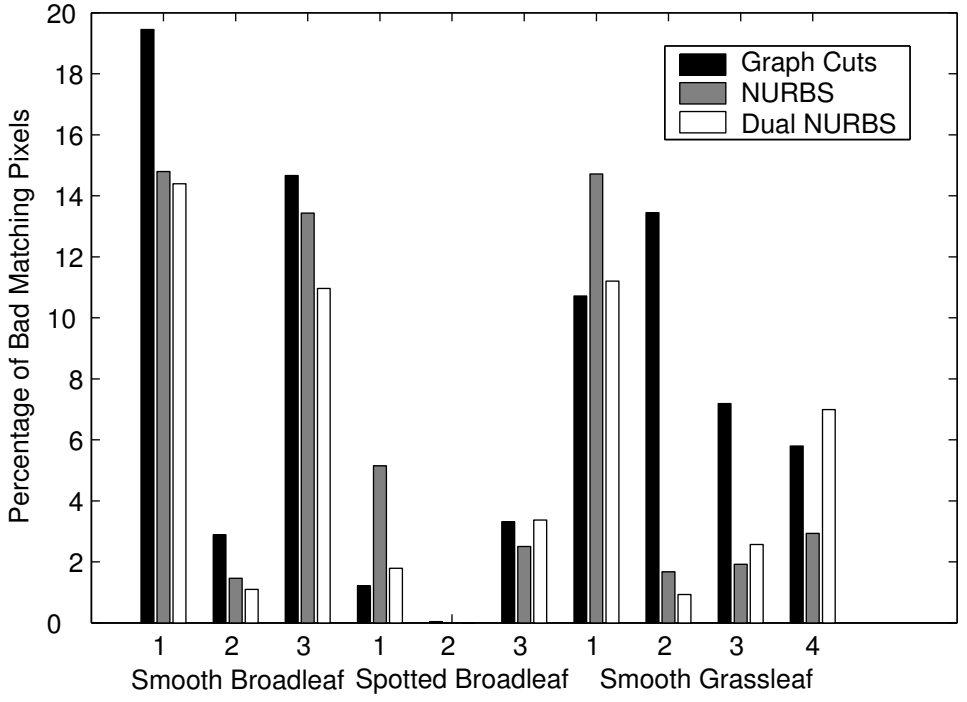


Figure 8.8: Comparison of graph cut disparity images and their respective NURBS approximation using only the right disparity map and both the right and left disparities.

The steepness of an object is the average tilt of the triangles. Basically, it is the average angle to the Z of the surface normals. The normal is found using the cross product between two vectors connecting the points in the triangle.

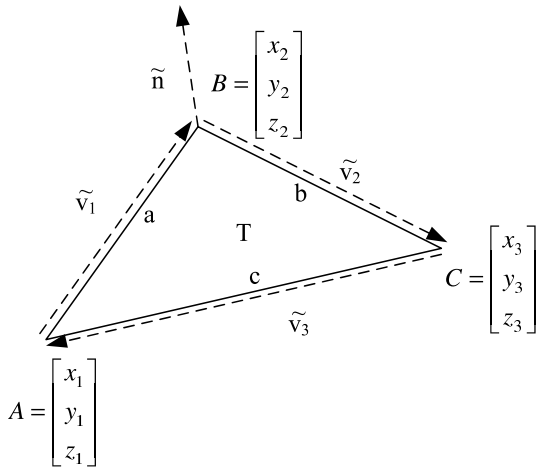


Figure 8.9: Notation of triangles in this section.

$$\vec{v}_1 = \begin{pmatrix} x_2 - x_1 \\ y_2 - y_1 \\ z_2 - z_1 \end{pmatrix} \tag{8.26}$$

$$\vec{v}_2 = \begin{pmatrix} x_3 - x_2 \\ y_3 - y_2 \\ z_3 - z_2 \end{pmatrix} \tag{8.27}$$

$$\vec{n} = \frac{\vec{v}_1 \times \vec{v}_2}{|\vec{v}_1 \times \vec{v}_2|} \tag{8.28}$$

$$s = \frac{180}{N\pi} \sum \arccos\left(\frac{n_z}{1}\right) \tag{8.29}$$

An example of usage could be to use the steepness of each individual leaf to calculate an basic plant steepness by computing the average steepness of all the leaves weighted by the size of the leaf.

Each pixel in the normal image is the average of the normals of the triangle the pixel is a part of. The steepness of the object is the average of all its normals. The normal image is very smooth because of the NURBS fitting. This makes it very useful in future perspectives. For example, it is possible to evaluate leaf angle distributions and orientations, and to follow the slopes toward the tips and bases.

The total area of the object is the sum of the areas of all triangles. The area of a triangle is found using the lengths of all three vectors connecting the points:

$$\vec{v}_3 = \begin{pmatrix} x_1 - x_3 \\ y_1 - y_3 \\ z_1 - z_3 \end{pmatrix} \tag{8.30}$$

$$a = |\vec{v}_1| \tag{8.31}$$

$$b = |\vec{v}_2| \tag{8.32}$$

$$c = |\vec{v}_3| \tag{8.33}$$

$$p = \frac{a+b+c}{2} \tag{8.34}$$

$$a = \sqrt{p(p-c)(p-b)(p-a)} \tag{8.35}$$

where a, b, and c are the lengths of the sides, and p is the semi-perimeter.

The average height is straight forward to compute as the mean of all Z values.

The bounding box can be the absolute bounding box as the minimum and maximum x, y, and z coordinates, respectively. Noisy outlying points can be filtered by using the mean plus-minus three times the standard deviation as the bounding box.

$$h = \frac{1}{3N} \sum z_1 + z_2 + z_3 \tag{8.36}$$

$$B_a = \begin{pmatrix} \min(x_1, x_2, x_3) & \max(x_1, x_2, x_3) \\ \min(y_1, y_2, y_3) & \max(y_1, y_2, y_3) \\ \min(z_1, z_2, z_3) & \max(z_1, z_2, z_3) \end{pmatrix} \tag{8.37}$$

$$B_f = \begin{pmatrix} \mu_x - 3\sigma_x^2 & \mu_x + 3\sigma_x^2 \\ \mu_y - 3\sigma_y^2 & \mu_y + 3\sigma_y^2 \\ \mu_z - 3\sigma_z^2 & \mu_z + 3\sigma_z^2 \end{pmatrix} \tag{8.38}$$

There was no ground truth steepness for the real plant images, so a set of ray traced images were generated for the test of steepness. The steepness was extracted using a simple mesh and a NURBS mesh. The scene consisted of a wooden box with the surface area $7.5 * 5.5 = 41.25$. It was placed in 5 different orientations; 0, 30, 45, 60, and 90 degrees. See result summary in table 8.1.

Table 8.1: Results from steepness test.

	Simple mesh		NURBS mesh	
Steepness	Angle	Area	Angle	Area
0^o	0	41.6	0	41.6
30^o	15.3	48.2	28.6	40.7
45^o	30.1	46.4	43.4	39.8
60^o	55.7	41	56.6	38.6
90^o	74.5	45.2	86.7	37.5

The results using a NURBS mesh was superior, because it was predictable. Steepness correlation $R^2 = 0.9997, p < 0.001$. Residuals (absolute errors after regression): $\mu = 0.38, \sigma^2 = 0.32$. See figure 8.10

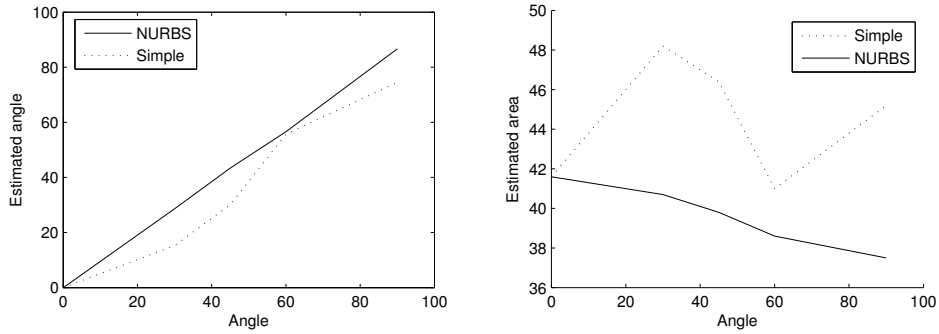


Figure 8.10: [Left] Correlation between real steepness and estimated steepness. [Right] Estimated area in relation to the steepness angle.

Its inaccuracy underestimating the steepness and area at large angles could be due to a small camera matrix focal length discrepancy. The results from the simple mesh has no apparent pattern. Steepness correlation $R^2 = 0.9604$. Residuals : $\mu = 6.65, \sigma^2 = 2.16$. The area estimations were chaotic.

The real plants had their leaf heights measured at three locations across the leaves. The cotton and hypoestes leaves were averaged to get a leaf height to compare with the surface reconstruction height parameter. When leaves were connected through their stems in the surface reconstruction their heights were compared to the average height of the real leaves.

Figure 8.12 shows a correlation plot between them. The mean absolute error before regression was 0.27 with a standard deviation of 0.23. The regression resulted in $y = 0.9672x + 0.8297$, where y is the ground truth heights and x is the estimated heights. The correlation was $R^2 = 0.9959, p < 0.001$. Residuals (absolute errors after regression): $\mu = 0.18, \sigma^2 = 0.18$.

Compared to the resolution in the disparities this was a very good result. The resolution in height was $1 - 4mm/pixel$ within the range of the plants. The average error was within the same range as the disparity resolution.

Figure 8.13 shows snapshots of the VRML models generated from the algorithm.

8.6 Conclusion

It was possible to split the disparity maps into individual objects using simple mask based segmentation. Each object was described with NURBS and thus the disparity maps were improved. It was possible to fit the NURBS using a single view or combining multiple views if such disparity maps were available. It was also possible to reconstruct overlapping objects because there was one NURBS per object. Information about the objects were also extracted and the accuracy of the information was acceptable.

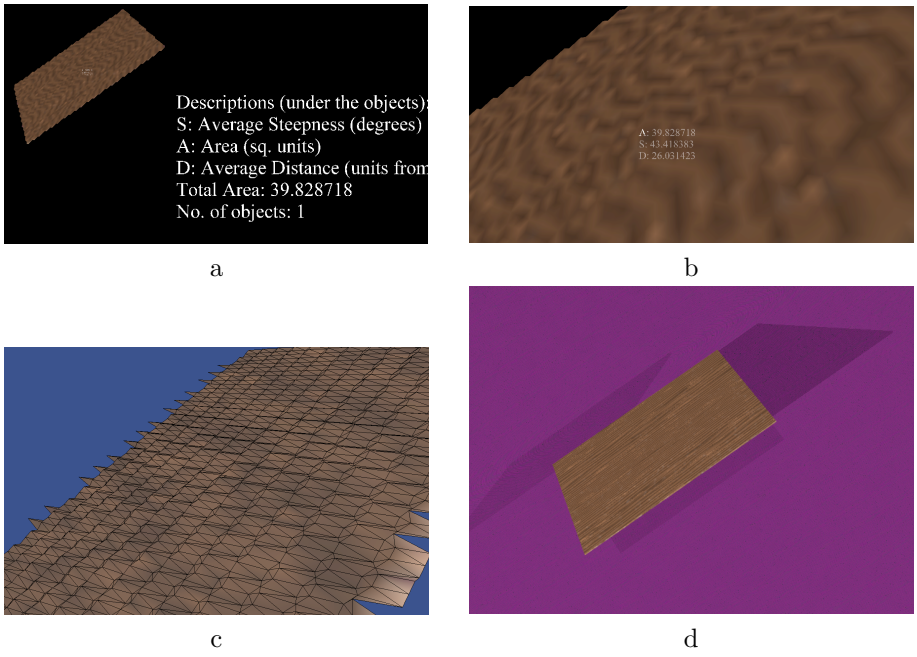


Figure 8.11: a) VRML model of the NURBS mesh 45^o wood box. b) a close-up reveals area and steepness of the large single object. c) the problem with the simple mesh is that even after smoothing, it is still step-like.

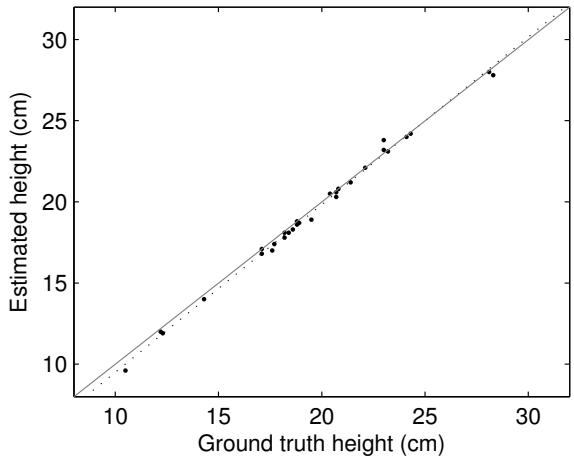


Figure 8.12: Leaf heights correlation. Dashed regression line. Grey solid unit line. $R^2 = 0.9959, p < 0.001$. Residuals: $\mu = 0.18, \sigma^2 = 0.18$

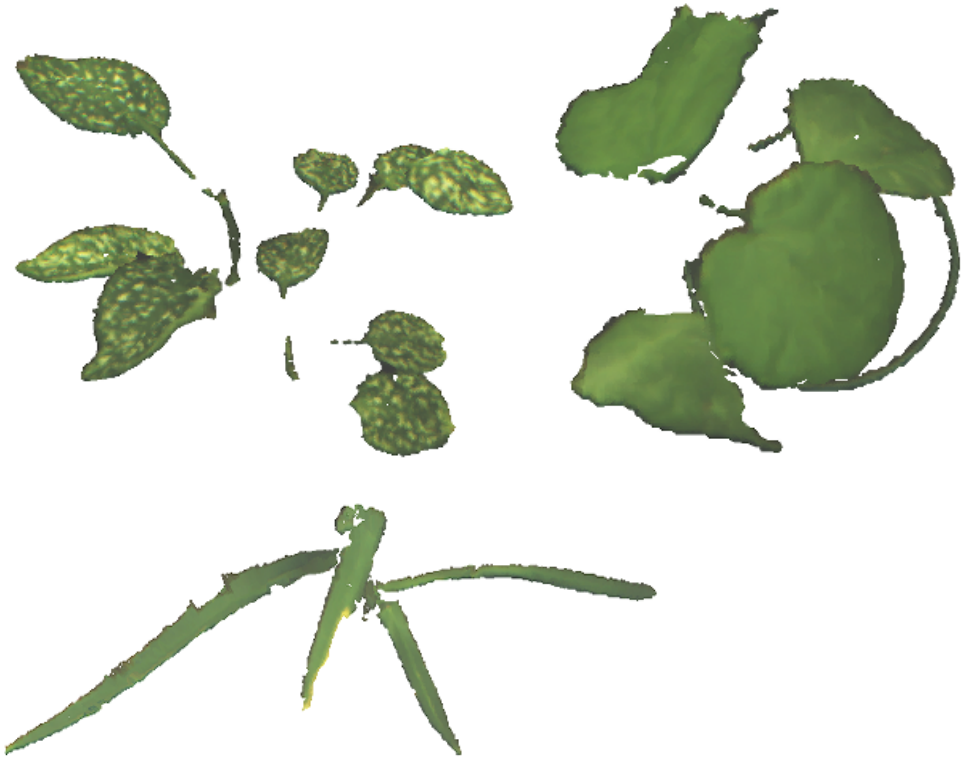


Figure 8.13: Reconstructions of the real plants from figure 4.3 seen from a slightly different angle.

Chapter 9

Automatic Annotation of Structures

Summary:

This chapter presents the more practically minded research where two cases were investigated. Experiments with nutrient deficiency in barley and tomato-weed classification were conducted in conjunction with agricultural engineering facilities. From pictures of barley the plant leaf area from 2D and 3D of the plants was found. Novel methods for extraction of sample spots, i.e. tips and bases were also found. The tip finder was very accurate, while the base was a rough estimate that could be further pinpointed using the surface normal maps from chapter 8[1]*. Leaf angle was not determined at the time of the experiment but the method described in chapter 8 would make it possible. In the tomato field experiment the weeds and tomatoes were distinguished using leaf height. Previous work in weed detection had problems with overlapping leaves. Different approaches such as watershed was tested for the purpose of separating the leaves before generating actual spray maps. An adaption of watershed and the image labeling technique in section 8.1 combined with a unique depth histogram based probability measure were best for detecting weed cells in the grid.*

[1]Note: These experiments were done before the improved disparities and surface reconstruction was made.

9.1 Automatic Annotation of Barley for Classification of NPK Deficiencies

The application of nutrients to field crops is of critical importance to optimize crop yield and product quality. Farmers must balance the competing goals of supplying enough nutrients to their fields, and minimize input in order to avoid adverse environmental effects and economic penalties from reduced yields [Blackmore et al., 2002a].

Farm managers do in most cases carry out agricultural monitoring for nutrient deficiencies detection in a given field manually. The only tools they use are their eyes detecting spatial variability from an area of the field and to analysis of a single leaf.

Spatial variability of nutrients and variations in soil features affect nutrient utilization by the crop. Non-destructive plant nutrient status identification by use of optical sensors is a way to determine a field's nutrition distribution. Thus a variable rate application system is essential in order to reduce the outputs of fertilizers in agriculture. However, great challenges lies within automatic spectral discrimination of the different deficiency symptoms.

Nitrogen, phosphorus and potassium are all mobile nutrients within the plant which is due to their deficiency symptoms appears on the older leaves first. The deficiency symptom for nitrogen (N) occurs as an overall chlorosis and inhibits biomass production significantly. Phosphorus (P) deficiency symptoms appear as deep green, purple coloration of petioles as for potassium (K) deficiency occur as interveinal chlorosis and bronze coloration[Marschner, 1995].

Recent research shows that mutually information about the plants spectral reflection and the spatial location of it may open new possibilities for discrimination between N, P, or K stress symptoms in Barley [Christensen et al., 2004][Christensen and Bennedsen, 2004][Christensen and Jørgensen, 2003], i.e. the reflectance comes from the tip, middle or base of the leaf or from the bottom or top leaf. It may be expected that these characteristics may include other cereal crops and hence facilitate a new way for characterization of crops growth status and potential.

However, to implement a methodology taking advantage of these new results in an operational context calls for development of automatic methods for description of the plants structural characteristics. This challenge will be addressed in the following study by investigation of computer vision potential for estimation and characterization of geometrical structure of the crops. The research will involve estimation of the following features, leaf area for estimation of the leaf area index (LAI), and position of the plant base and leaf tips for spatial location of reflections from the crop.

The first feature LAI is an important parameter for correction of the spectral reflection according to the current canopy growth [Filella and Peuelas, 1994][Friedl et al., 1994], while second features are essential for localization of the spectral reflection from the canopy necessary for utilization of the new approach for plant analysis.

Estimation of the plants leaf area, base and leaf tips was done on Barley plants with 4 to 12 leaves isolated in pots. The test plant were part of an experiment with varying fertilization of N, P, and K, hence the large spread in development.

Both 2D and 3D computer vision techniques were used for estimation of the plants leaf area. There was no significant difference between the area estimate obtain with either of the

methods. Actually the estimate from the 2D analysis was better than the estimate by 3D analysis. However, an important feature of the 3D estimation is that it gives the leaf area in world coordinates compared to the 2D estimation techniques that due to the projection of the plants onto the image plane reduces the dimension to 2D and as a result the area estimate will be related to a specific location and orientation of the camera, e.g. height above the crop. To be independent of a precise location and orientation of the camera will be very desirable in an operational context and hence the result is promising for future development of LAI sensors.

For estimation of the plant base a new method using the structural information that leaves from a cereal crop emerge from the plant base is introduced. The base is found by counting the intersection between the perimeter of a circle with its center located at the base and the plant leaves segmented from the background. The base is determined by the location where the circle's perimeter has maximum intersections with the plant.

The leaf tips was found by first calculating the edge image using Canny edge detector[Canny, 1986]. At every edge pixel a circle is placed at the number of intersection between the circles perimeter and edge pixels was calculated giving a signature vector. At the leaf tips the signature vector will give an unique pattern which is recognized.

9.1.1 Materials and Methods

This section will present the materials used in the experiment, and the methods developed for the feature extraction.

Test Plants

Spring barley plants (Optic) were collected in an 8.5 ha field of The Royal Veterinary and Agricultural University research farm at the campus in Taastrup, Denmark (about 18 km west of central Copenhagen). The field had been producing cereals over a period of 32 years, without addition of phosphorus (P) and potassium (K), thus the soil content of these minerals was very low. Different plots were established added with various amounts of nutrients (Table 9.1) in order to establish a variation of the nutrient content in the spring barley plants creating appearance variation.

Table 9.1: Fertilization Scheme. In $kg\ ha^{-1}\ year^{-1}$ and type

Plot	N	P	K	S	Fertilization
A	0	0	0	0	
B	60	0	60	25	Fertilizer
C	60	10	0	25	Fertilizer
D	60	10	60	25	Fertilizer
E	120	20	120	50	Fertilizer
F	75	10	75		Manure
G	150	20	150		Manure

Approximately five weeks after sowing (BBCH[2] 21[Lancashire et al., 1991]) the plants were transferred into laboratory environment, where there were transplanted individually into sphagnum for then to be measured. The resulting amount of different plant representative of the different fertilization treatments was 35 plants.

Materials

The experiment was carried out in a laboratory environment. Sunlight was simulated with flicker free lamp with a color temperature of 5500 Kelvin.

Dhond [Dhond and Aggarwal, 1990] shows the benefits of trinocular stereo vision giving a more accurate 3D estimate at the expense of processing complexity. A frame was built to accommodate three monochrome JAI M4 camera link cameras approximately 1 meter above the ground, as shown in figure 3.1. Their orientations are fixed to optimize the common field-of-view of the plants by pointing at the plant base, when the plant is in the center of each image plane.

Images of all plants were taken using red and green filters for each camera. These wavebands can be used to distinguish soil and leaf matter [Brivot and Marchant, 1996][Filella and Peuelas, 1994].

The 2D based Leaf Area is calculated from the largest segmentation mask of the three cameras by the following equation:

$$A = K(max(\sum\sum S_c(x,y))) \tag{9.1}$$

where K is a constant which approximates the conversion from pixel area to real world area, the error caused by not knowing the orientation of the leaves projected onto the image plane, and the average occlusion. A large but almost vertical leaf will look small on the image plane. The heuristic K is found by minimizing the error on a small test data set, and it assumes a fixed distance to the crop, as opposed to using the 3D method.

Leaf Tips The leaf tips are found in the segmented images using edge signatures. The idea is similar to the method used in [Chapron et al., 1997] with corn leaves.

The segmentation mask is first morphologically closed to remove gaps inside the leaves. The edges in the segmentation mask is found with the canny edge detector. A response signature is generated for each edge pixel by counting when the perimeter of a circle hits the edges. The signature is generated from 0 to 900 degrees, 2.5 turns, in order to ensure detection of signature B on figure 9.1, which shows four example signatures.

A pointy edge will respond with a low angle between the intersections (See signature A on figure 9.1). However, it is possible that the perimeter starts inside the leaf (signature B). Therefore, detection depends on the first 4-5 hits. First measure the angles between hit no. 1 and 2 (δ_1), and no. 2. and 3 (δ_2). A straight line will respond with a $\delta_1 = \delta_2 = 180\,\text{deg}$.

[2]The abbreviation BBCH derives from the institutions that jointly developed this scale: BBA, Biologische Bundeanstalt fr Land- und Forstwirtschaft (German Federal Biological Research Center for Agriculture and Forestry); BSA, Bundessortenamt (German Federal Variety Authority); CHemical Industry, Industrieverband Agrar, IVA (German Association of Manufactures of Agrochemical Products).

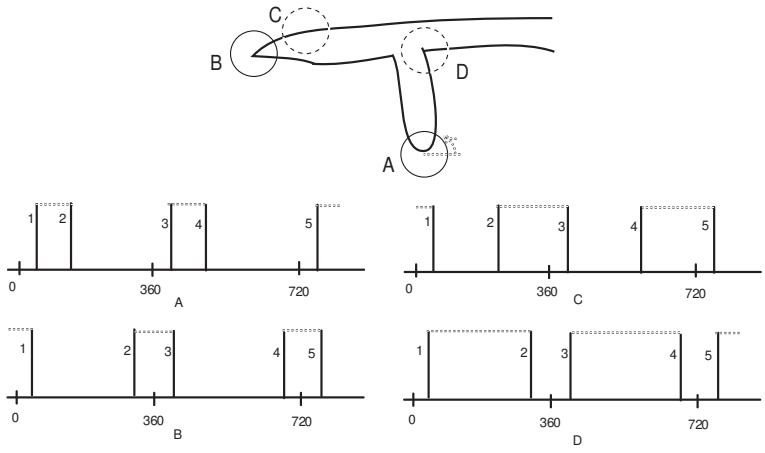


Figure 9.1: Four examples of edge signatures. The enumerated vertical bars are where the perimeter intersects the edge at a certain angle. The dotted horizontal line denotes that the perimeter at this angle is inside the leaf.

For a leaf tip, the smallest of the two will be below a certain threshold (T) and the largest must be above the remaining perimeter ($360 - T$), or there are other edges present.

$$0 < min(\delta_1, \delta_2) < T \tag{9.2}$$

$$360 - T < max(\delta_1, \delta_2) < 360 \tag{9.3}$$

The angles between hit no. 3 and 4 (δ_3), and no. 4. and 5 (δ_4) are then used to confirm the match. If another edge is hit, it will change the order of hits. A pointy indentation will produce the correct response. This is avoided by checking that the leaf pixels are between the hits of the smallest δ_i.

The radius of the circle reflects how broad a leaf tip can be, but a large radius increases the chance for disruption in the signature by a nearby leaf.

Plant base The philosophy used in the detection of the plant base, is that the plant base is the origin of leaves. The leaves grow from this origin into arbitrary directions. If a circle is drawn from this origin, the perimeter will intersect all the leaves. If the leaf are spread evenly on the perimeter, a circle will intersect the maximum number of leaves near the base of the plant. The smaller the circle, the closer to the base its center will be. Figure 9.2 shows examples with circles of different sizes on two plants.

A base measure ($B(x, y)$) is calculated for each plant pixel in the image by this equation:

$$B(x, y) = w_1 h_1 + w_2 h_2 + w_3 h_3 \tag{9.4}$$

where h_i is the number of hits on the circle of radius r_i and w_i is the weight. If two locations have the same B, the one where the smallest radius has the most hits is chosen.

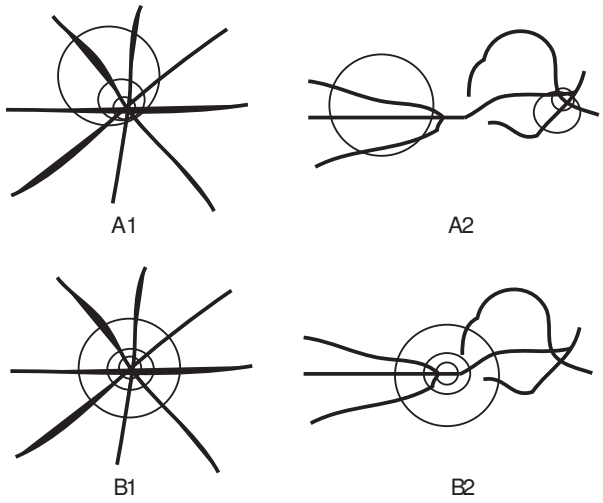


Figure 9.2: Finding the Leaf base. A1-2: Examples using large or small circles, and an ideal symmetric plant and a natural plant situation, respectively. B1-2: Combining the usage of circles of different sizes at the same center.

The radiuses should be weighted inverse to their size, ie. the smallest is weighted highest. The largest radius (r_1) is chosen such that the diameter of the circle is approximately a half plant wide. $r_2 = r_1/2$ and $r_3 = r_2/2$. Thus, the weights are $w_1 = 1$, $w_2 = 2$, and $w_1 = 4$. The plant base is found on all three camera views and the average position is chosen.

9.1.2 Results

This section presents the results from the parts of the image processing algorithms.

Leaf Area

Ground truth areas were obtained by scanning the leaves in a flatbed scanner and counting the plant pixels.

Figure 9.3 a-b, illustrates the relationship between estimated and scanned leaf area (regarded as ground truth). Both models shows a clear significant linear relationship (p¡0.001) and a correlation coefficient at approximately 0.90.

An example of the 3D reconstruction of a plant is illustrated in figure 9.4 using binocular and trinocular stereo, respectively. It is the plant which is shown in figure A.1. The figure illustrates how the introduction of third camera can fix the occlusion problem. This is obvious in the two upper leftmost leaves in the disparity maps which are lower rightmost in the 3D illustrations. Figure A.1 shows that one of the leaves occludes the other leaf, which results in a large disparity jump. The trinocular algorithm counteracts this jump for most of the pixels, while the relaxation scheme in the binocular algorithm cannot. The same problem is seen near the base. It also shows that the leaves on the binocular 3D model are more waving

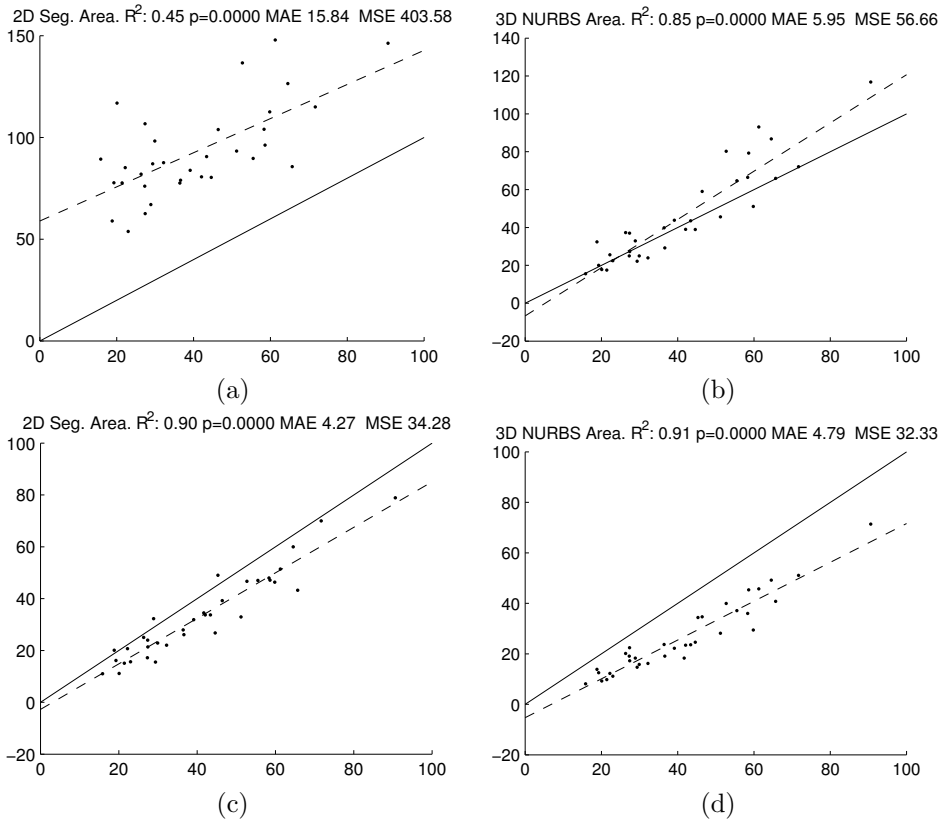


Figure 9.3: Relationship between estimated and scanned leaf area. a) Low threshold. Areas (in cm^2) found by 2D segmentation: $R^2 = 0.45$ and $p < 0.001$. b) Low threshold. Trinocular reconstruction using combined SMW: $R^2 = 0.85$ and $p < 0.001$. c) High threshold. Areas (in cm^2) found by 2D segmentation: $R^2 = 0.90$ and $p < 0.001$. d) High threshold. Trinocular reconstruction using combined SMW: $R^2 = 0.91$ and $p < 0.001$.

up and down compared to the trinocular model. It is also obvious that the NURBS fitting creates a nice smooth mesh, but it also evens the height of the leaves at the bases.

Plant base and leaf tips

The method introduced for localization of the plant base was in average able to find the location of the base with an error of 66 pixels with a standard deviation of 44.3 pixels. Figure 9.5 shows an example of the localization method with an error of approximately 66 pixels, which corresponds to approximately 1-2 leaf widths.

The results from the leaf tip localization method is illustrated in figure 9.5. The histogram shows that 79% of the leaves are automatically located correctly. The major reason to failure of the method is due to occlusion by other leaves or the tip is outside the boundary of the

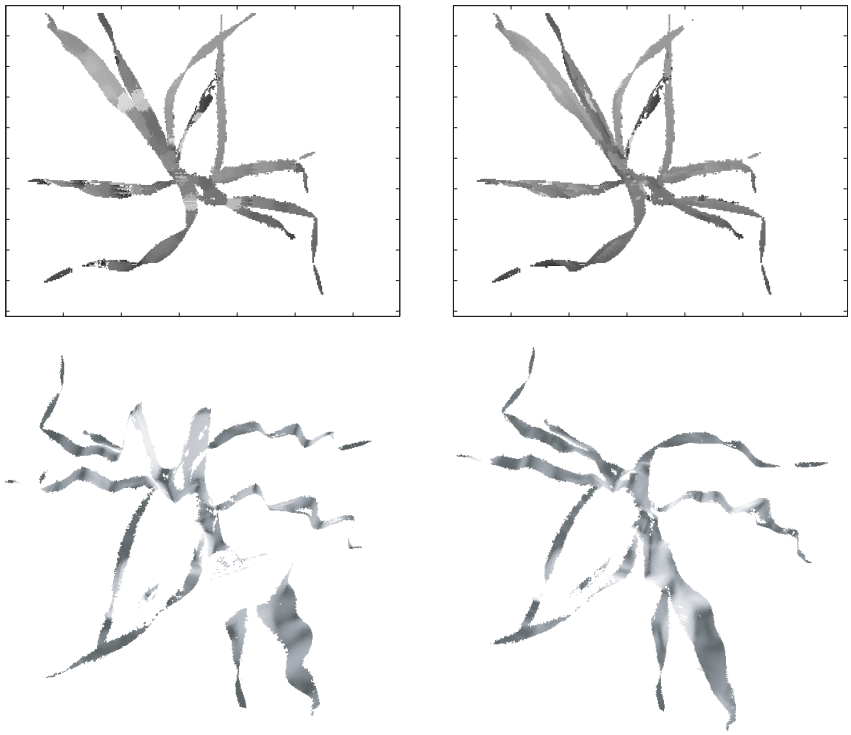


Figure 9.4: Disparity map (top) and 3D reconstruction by NURBS surface (bottom) of the plant in figure A.1. Left: Based on binocular stereo with multiple windows and simulated annealing. Right: Based on trinocular stereo with a single centered window and without simulated annealing.

image. These two sources account for 13% of the errors so strictly only in 8% the method fails. Looking closer at the reason for failure 5% is due to errors in the segmentation and only 3% is due to direct failure of the method.

An example of the success of the tip detection is shown in figure 9.5.

9.1.3 Discussion

The technical solutions

The trinocular solution for the 3D reconstruction is fast and can successfully extract leaf area. The area calculation from the three 2D segmentation images yields similar results to the 3D areas, but the method depends on the scale, given by distance to the crop, and it is more sensitive to the segmentation (large degrading performance, when the segmentation threshold changes). However, the 3D model is not detailed enough for individual leaf features. Experiments making a reconstruction based on each camera pair shows that it is random which of them makes the best (visual) 3D reconstruction. This has three reasons:

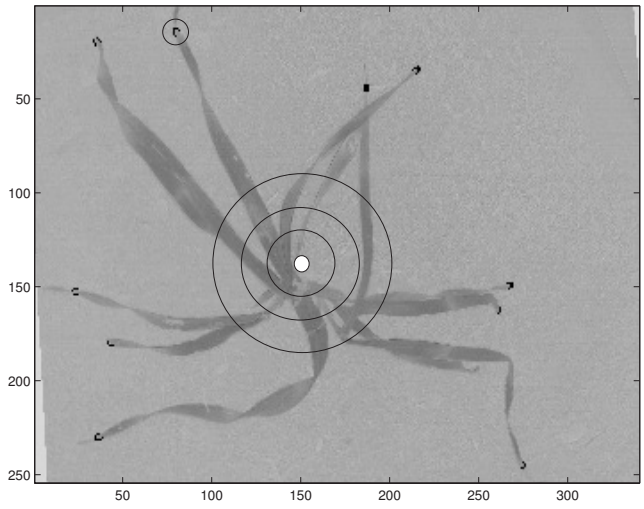


Figure 9.5: Example of plant base and leaf tip detection. The white dot denotes the estimated position of the plant base. The error in this example is close to the mean base error. The intersections of the three circles are candidates for individual leaf bases. The black spots on the tips are the detected tip pixels.

1. Upright leaves
2. Orientation of leaves
3. Occluded leaves

Problem 1: Using an area of course gives a problem with steep leaves, whose orientations from two different views are significantly different. The SSD measure will not favor the match between these areas.

Problem 2: Leaves that are aligned with the epipolar line between the camera pair in question are difficult to reconstruct. The search space in the 2nd camera is very large, because the length of the leaf is on the epipolar line. This problem is successfully made less apparent by the projection test onto the third camera view.

Problem 3: Choosing a reference image, deselects all the leaves that are occluded in that image. Those that are not occluded might be occluded in one or two of the other cameras.

The success of the reconstruction depends on how well the leaf can be seen in the given camera pair, whether it is aligned with the epipolar line, and how similar the light is on the entire leaf.

The method presented for the leaf tip detection in the individual segmented images is successful. It detects most of the tips that are found and introduces few false positives. The test does not reveal how many of the physical leaf tips are not seen in any camera. This would be interesting to look into in an experiment, when the tips are going to be projected into 3D space.

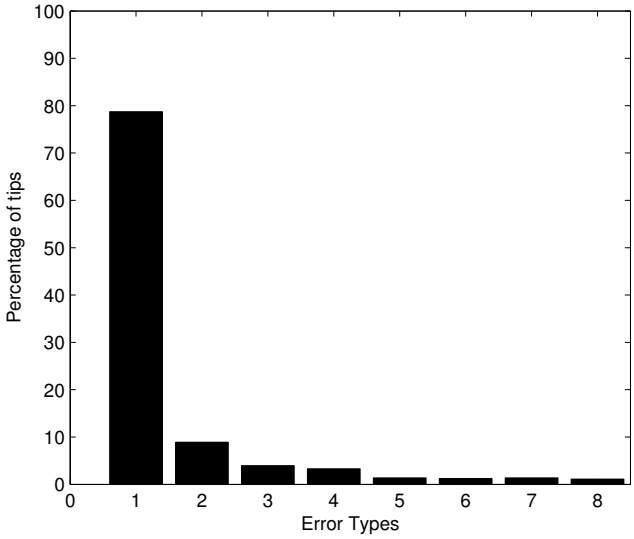


Figure 9.6: Leaf Tip Results. Bars denote the percentage of tips are 1. Success. 2. Occluded. 3. Out of bound. 4. False positives because the segmentation cut up the leaf. 5. Not segmented correctly. 6. broken/blunt. 7. False positives. 8. Failed.

The plant base is found with satisfactory precision. The leaf base candidates given by the method needs further work. This further development could combine this leaf base selection with leaf tip error filtering. This could be based on heuristics on the location of tips in accordance to the skeletons of the leaves.

Once the individual leaves are found, they can be used to extract individual leaves, which can aid the 3D reconstruction. The leaves can be extracted from the images in which they are viewed optimally. Consequently, the 3D reconstruction can be optimized by choosing the camera pair whose epipolar line is as perpendicular as possible to the leaf, and the tips can be used as stable control points. Using one NURBS Surface per leaf will also support overlapping leaves and discontinuity at the plant base and when the leaves are close to each other. Each leaf can then be syntactically described as: Leaf age given by height at the base, base-, middle-, and tip position.

Significance in Precision Agriculture

Naturally, at this stage a spectral data registration on individual plants throughout a field in order to establish a robust and optimal NPK fertiliser strategy is costly. However, combining it with other monitoring systems at higher scales such as satellite- and/or airborne information systems, a sampling strategy can be created. This way, the problem regions could be identified and thus representative plant level sampling could be carried out in these identified regions. Moreover, these results do not only relate to the fertilization issue, but has great potential for both fungicide and pesticide application strategies.

The autonomous field vehicle (API) developed as a collaborative effort between nine part-

ners in Denmark[3]. This vehicle can potentially map the parameters necessary for the retrieval of the relevant crop bio-physical parameters and leaf optical properties, if equipped with the proper instrumentation. The API is ideal for collecting high resolution point measurements ($< 0.25m^2$) at hot spots in the field, i.e. areas that represent extremes. However, in order to secure a representative up-scaling based on a very limited number of e.g. 3D stereo vision points it is crucial to select points in an rational way. If the focus is on crop diagnostic a measurement of e.g. a weed patch could introduce severe bias to the up-scaling procedure. Therefore a reasoning algorithm must be developed identifying the optimal sampling points at the time the robot enters the area of interest [Hardwick and Stout, 1998].

There is an abundance of remote sensing technology available to measure variability in plants and soils. Also, there is a shortage of information about the causes of plant condition variability and the management solutions needed manage variability to improve crop production. The lack of knowledge needed to answer these variability questions is restricting the development of precision farming management decision support systems.

The challenge ahead is thus to register, manage and combine the right information sources temporally and spatially in order to obtain an optimal crop diagnostic information system.

9.1.4 Summary

The trinocular stereo camera setup has successfully extracted leaf features of individual young barley plants. Leaf area was estimated from 3D with a significant correlation, $R^2 = 0.89$. A relaxed binocular with smoothness optimization and a simpler trinocular 3D reconstruction were tested, and the trinocular vision was superior, despite being faster in its execution. Candidates for individual leaf bases were given based on an estimate of the basic plant base. This plant base was found with a mean 66 pixels precision, corresponding to 1-2 leaf widths. 79% of the leaf tips were found on the 2D images, while 13% were occluded.

Combining the knowledge of found tips and the base candidates from the individual 2D images is expected to lead to a better 3D reconstruction and more information about individual leaves, such as individual leaf base, middle, and tip, in future work.

The sub-leaf scale remote sensor will be a tangible solution combined with autonomous data collection and sampling strategy designed from satellite or airborne imaging.

[3]DIAS Bygholm, DIAS Flakkebjerg, AAU Dept. of Computer Science, AAU Dept. of Control Engineering, KVL Dept. of Agricultural Sciences - AgroTechnology, Dronningborg Industries, Hardi International, Sauer-Danfoss and ECO-DAN.

9.2 Classification of Weeds in Tomato fields

[4]

Tomato continues to be an important crop in areas like California, which produced 9.25 million tons of processing tomatoes on 280 thousand acres in 2003 and had about 19.5 thousand acres of fresh market tomatoes in production in 2003. (PTAB, CASS, 2003).

Weed control is a critical factor in maximizing crop yields and studies have shown (e.g. Vargas et al. 1996) that weed control during the early part of the season has the greatest impact on crop yield. Yield reductions of 48% to 88% have been reported (McGiffen et al., 1992; Miyama, 1999; Monaco et al., 1981) in tomato when weeds were present at seed line densities ranging from 4.8 plants/m^2 to 11 plants/m^2.

Chemical inputs (e.g. herbicides, insecticides, and fertilizers) in agricultural operations continue to raise environmental and economic concerns. For example in 1994, 5.9 million kg of agricultural chemicals (herbicides, insecticides, fungicides, and other chemicals) were used to produce processing tomatoes in California alone (USDA, NASS and ERS, 1995). The main alternative to chemical weed control is hand hoeing; however labor costs and availability issues make this alternative less attractive to farmers.

Automated weed control methods have been investigated by many researchers. For example, Lee et al. (1999) and Lamm et al. (2002) developed machine vision-based robotic weed control systems for precision weed control in row crop applications. However the use of machine vision as a weed sensor in an automatic weed control system for use in an agricultural field can be challenging due to the variability of object appearance and viewing conditions in a natural environment. Tang et al. (2001) developed a camera-based system that could estimate weed density and observed that variability in outdoor lighting resulted in variations in system performance. The most common machine vision techniques for weed detection have been based upon the analysis of leaf or plant shape, texture or leaf color using a 2D image. Under ideal conditions, these techniques can produce accurate techniques for weed detection, however ideal viewing conditions are not common in a typical agricultural field and issues like visual occlusion due to leaf overlap, can degrade their performance.

Some machine vision techniques have been developed to mitigate the occlusion problem. The Hough transform has been used with some success by some researchers to detect partially occluded objects of regular shape in 2D images; however it is computationally intensive, requiring specialized hardware for many real-time applications. Manh, et al. (2001) used deformable templates in order to adapt classical morphological recognition techniques to the variable environment of an agricultural field and showed success in recognizing partially occluded 2D images of one weed species. Lee (In press) used the Watershed method to separate occluded objects in 2D images and reported a 2 to 3 fold improvement in the recognition rate of occluded tomato leaves and weeds in natural outdoor images.

Three methods using leaf height information as the feature used in the crop/weed classification of leaf area will be presented. These will be tested on 34 images of tomato with occluded weeds. The methods will be compared by generating spray maps for each method. An error grid is defined in which the cells containing sprayed tomato and sprayed-, and unsprayed weeds are counted. The number of targeted tomatoes and weeds are also counted. Finally, the results are discussed regarding perspectives and future work.

[4] This section was based on paper Nielsen et al. [2004a]

9.2.1 Materials and Limitations

This section will present the material setup, limitations, methods, and how to test them.

Materials

The images were recorded in tomato transplant rows with weeds. A 1024x768 color Point Grey Research Digiclops 3D system was used for the image acquisition. It was mounted in a light tunnel with shielding from sunlight. Eight halogen lamps were mounted in the tunnel. The camera was placed 533 mm above the seed line, and its lens focal length was 4 mm. The field of view at the soil plane is then 640 mm by 480 mm.

The tunnel was mounted on the tractor, which traveled 1.2 km per hour, and the image grabbing was 3 fps. The vertical axis was in the driving direction. Figure 9.7 shows the camera and light tunnel. The heights of the tomato plants were measured manually and their mean height was 191 mm with a standard deviation 31 mm.

The rows contained a substantial number of weeds of the types: Black Nightshade, Hairy Nightshade, Pigweed, Purslane, and grass weeds.

Limitations

The basic assumption in the presented methods is that the tomatoes are significantly taller than the weeds. This is assumed to be true for tomato transplant fields. Tomato transplants tend to have a long naked stem near the soil. The stereo vision is simplified by the assumption that:

1. The images do not contain intrinsic distortion. They have been rectified. This is included in the Point Grey Digiclops Software Development Kit.

2. There is no rotation between the cameras.

3. The translations between the two camera pairs are equal, i.e. 10 cm.

These assumptions were verified using Bouquet's Camera Calibration Toolbox for Matlab which is based on Heikkilla and Silven 1997.

9.2.2 Methods

The procedure is divided into three steps:

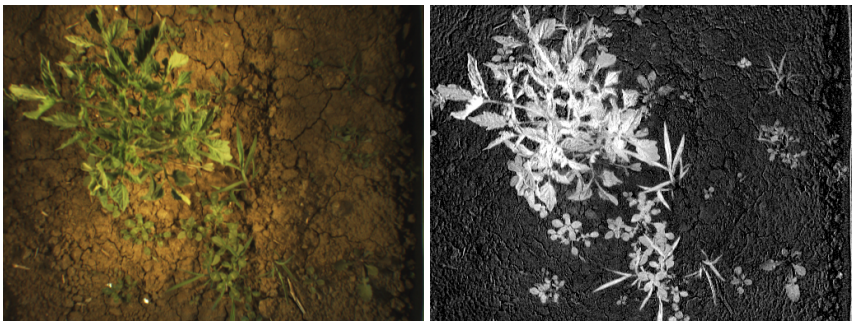
1. Calculate leaf height

2. Classification of tomato versus weed

3. Generate spray map

(a) (b)

Figure 9.7: (a) Digiclops (b) Tomato under light tunnel.



(a) (b)

Figure 9.8: (a) Original color image of tomatoes and weeds. It is difficult for human eyes to distinguish leaf pixels from soil pixels (b) Green chromacity. Leaf pixels stand out clearly.

Calculate leaf height

The tomato hieght is estimated from the dense disparity maps. These were estimated using the SMW 5 algorithm with window size 20x20. The resulting disparity map of the leaf pixels are shown in figure 9.10.

In order to limit the search space for pixel correspondences a leaf mask was generated by thresholding the green chromacity images. The result is shown in figure 9.8.

Further observations of the images show that the disparity ranges from approx. 140 to 280 pixels. Each disparity relates to a certain height, which can be found in the look up table plotted in figure 9.9.

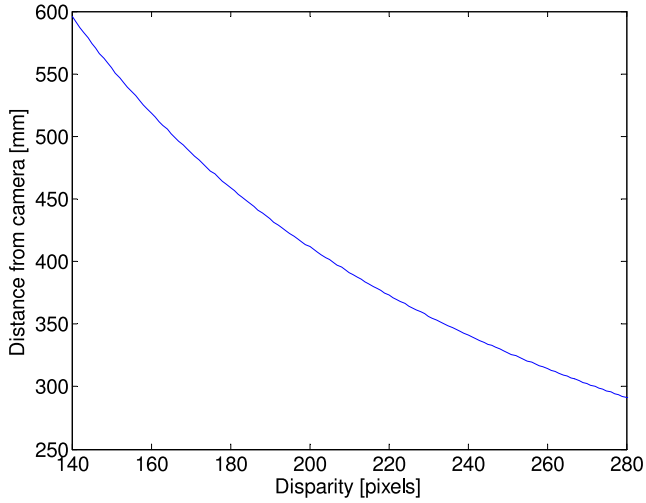


Figure 9.9: Disparity-to-height look-up table. The table actually shows the distance from the reference camera to the object. It requires knowledge of the distance to the soil to get the actual height of an object.

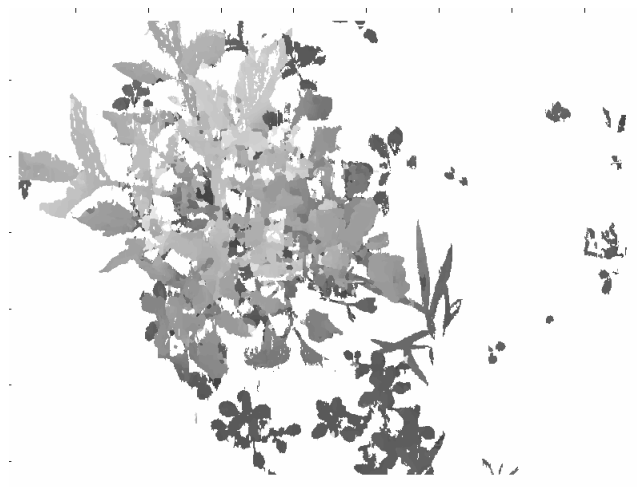


Figure 9.10: Disparity map of the plant previously shown. The disparity jump is easily seen on the tomato occluding a grass weed. Lighter pixels mean taller areas. White is the soil and is not considered.

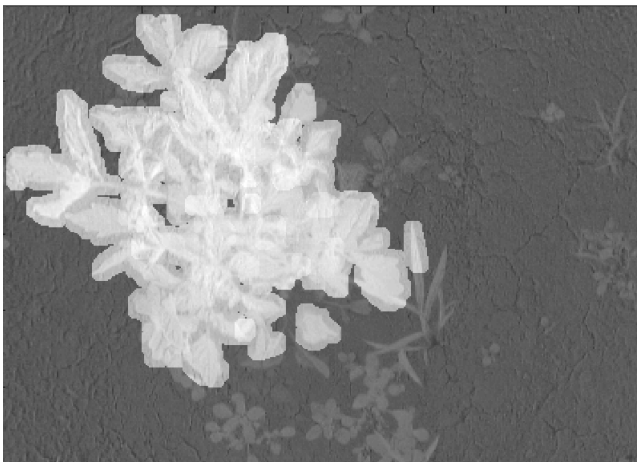


Figure 9.11: Classification with height thresholding. Highlighted tomato region. Parts of tomato leaves are sticking out of the region. Increasing the dilation will enlarge the region and that will bring more weeds into safety.

Classification using disparity map threshold

The simplest method is to define leaf pixels with a disparity over a certain threshold to be tomato. A morphological open operation (Gonzalez and Woods 1992, section 8.4) removes very small noise particles caused by wrong correspondence on weeds or soil in the disparity map. A dilation operation is used to expand the tall areas on top of the low areas. This closes the gaps in the tomato and enlarges the edge of the tomato to cover the low leaves sticking out at the bottom.

A problem with this method is that it is difficult to tell if some branches that are low towards the ground will stick out farther than the dilation will expand the tomato region. Some weeds, especially grass weeds, point upwards. The tallest part of these leaves can be mistaken for tomato. It is necessary to use some more information about the objects such as pixel connectivity. An example is shown in figure 9.11.

Classification using connected blob histogram

This method will look at the histograms of entire blobs (connected pixels), i.e. connected pixels in the disparity image. This allows the usage of the knowledge about the heights and standard deviation of the tomatoes in the rows. Classification will be based in the histogram distribution. Equation 3 shows how to extract the feature from the histograms. Figure 9.12 shows the principle with example tomato and weed histograms.

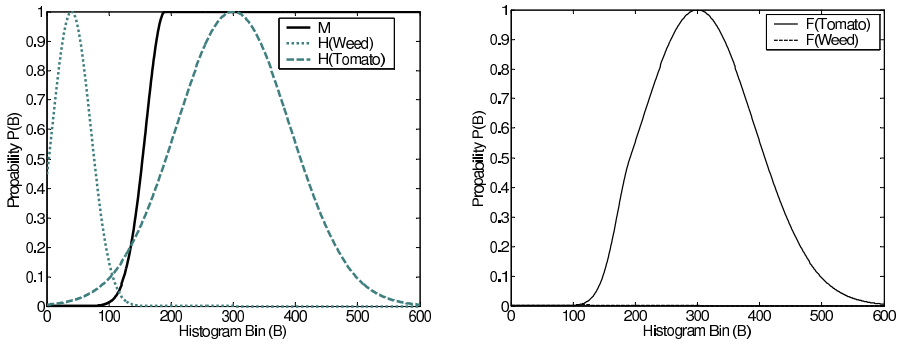
Equation 3. Tomato feature (F) is the sum of resulting histogram (H) bins (B) multiplied by the modified normal distribution (M) of tomato heights.

$$F = \sum_B H_B M_B \tag{9.5}$$

$$M_B = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(B-\mu)^2}{2\sigma^2}} \quad , B < \mu \tag{9.6}$$

$$M_B = \frac{1}{\sqrt{2\pi\sigma^2}} \quad , B < \mu \tag{9.7}$$

$$\tag{9.8}$$



(a) (b)

Figure 9.12: (a) the histogram plots of the modified normal distribution (M) and a weed (H_{weed}) and a tomato (H_{tomato}) (b) The results of the bin-wise multiplication $H_B M_B$. F is basically the area underneath the resulting curve. The area under the weed response is 0.14 and the tomato response is 217.49.

F will be close to zero if the histogram H has mostly low bin counts. F will be largest when the histogram has only counts in bins higher than the mean. A blob is classified as tomato, when F is higher than a certain threshold.

Furthermore, tomato blobs whose centers of gravity are placed near the middle can be verified by their sizes. If it is near the edges, it might be a small part of a tomato and the size cannot be used as a feature.

The problem with this method is that objects merge as one. In some of the images there are tomato branches that are close to the ground and are disconnected from the large region. This is because the stems can be brown rather than green, so they are not segmented as being part of a plant.

Watershed introducing disparity utilization

This method will use the same method for classification as the previous, but the blobs will be divided into smaller blobs using the watershed algorithm (Vincent and Soille, 1991). Previous work with the watershed algorithm has shown a morphological operation on the distance map can avoid excessive local minima creating catchment basins (Lee and Slaughter, In Press).

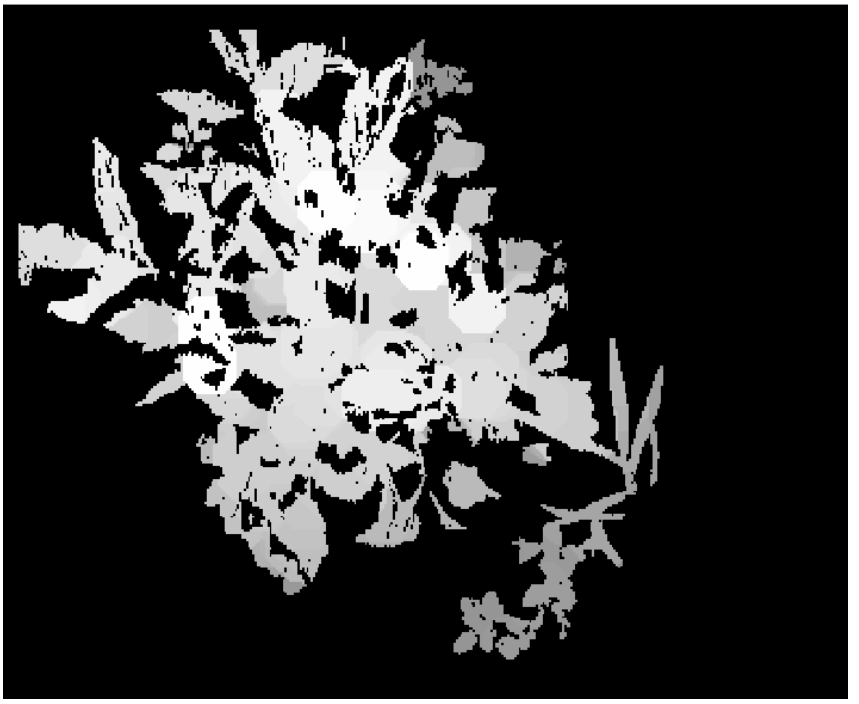


Figure 9.13: Classification with the histograms of whole blobs. It shows the region classified as tomato. It contains a bunch of weeds as well because the weeds are connected to a tomato region.

Disparity discontinuities also provide information about how to divide the blobs. This can be incorporated into the distance function.

Equation 4. Distance function (dist). x and y are two points in the image. P is the path completely contained in image A between the points. l is the length of P.

$$dist_A(x, y) \quad = \quad -\min(l(P)) \tag{9.9}$$

The length of P is often defined by the Euclidean-, chessboard-, or city block distance measures. The two latter measures are very efficient measures. The calculation of the distances will become very slow if the algorithm must search the path for disparity changes. Instead, the chessboard distance function is used with "dams" added to the distance map. Dams are added where the edges with a magnitude over a threshold in the disparity image are located. The edge strengths are calculated as the sums of the disparity convoluted with the vertical and horizontal Sobel filters (Gonzalez and Woods 1992, pp 418-419).

Equation 5. Edge strength (Ed) of disparity image(d) is given by convolution with horizontal and vertical Sobel filters (S).

$$E_d \quad = \quad |S_v * d| + |S_h * d| \tag{9.10}$$

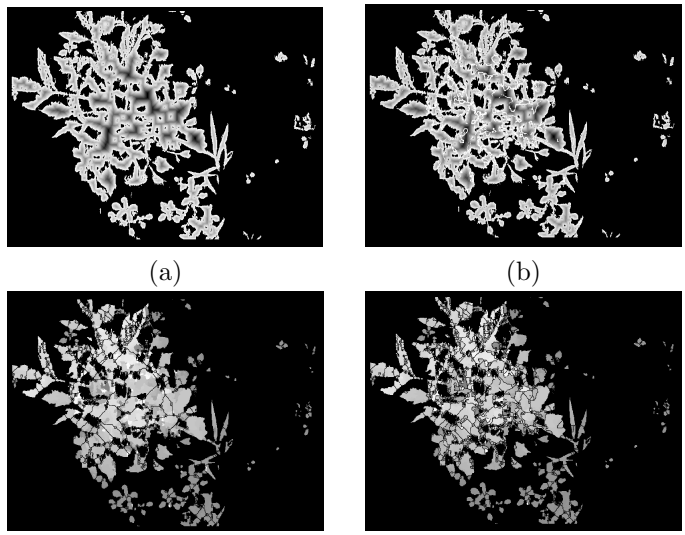


Figure 9.14: (a) Distance map without edge modification. (b) Distance map with edge modification. (c) Watershed result without edge modification. (d) Watershed result with edge modification.

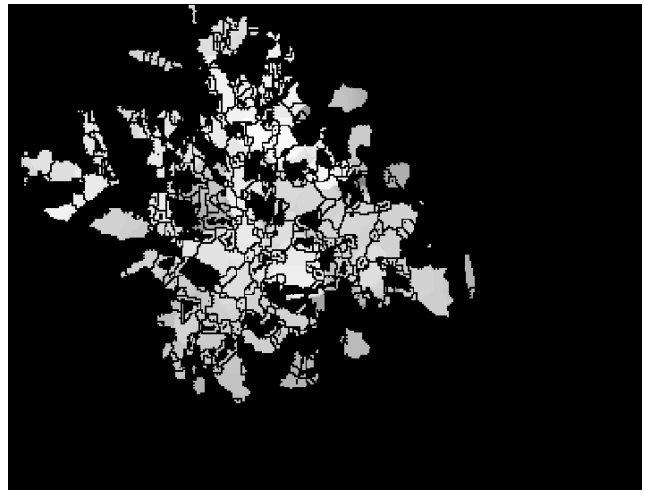
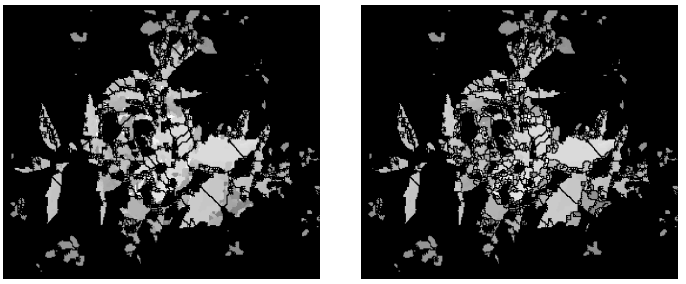


Figure 9.15: Classification of blobs separated by watershed. It shows the region classified as tomato. Only one leaf of the grass weed on the right side of the tomato is classified as tomato.

Figure 9.16: Another disparity image with occluded leaves.

See figure 9.14 for examples with and without this modification, and see the resulting classification in figure 9.15. Figure 9.16 and figure 9.17 shows another good example using the disparity edge. It is easy to see that the edge dams make sure that the watershed does not create blobs that belong to multiple overlapping leaves. Unfortunately, it creates a lot of tiny blobs. It will still be a problem that leaves near the edge of the tomato region can be mistaken for weeds, as well as tomato leaves on long brown branches sticking out from the tomato region.



(a) (b)

Figure 9.17: (a) Watershed result without edge modification. (b) Result with edge modification. From classification to spray maps.

The micro sprayers have a resolution of 12.7 mm (Lee, et al. 1999). A cell grid is defined where each cell is 12.7 x 12.7 mm. The visual area can be covered by 42 nozzles, and there can be 30 cells along the driving direction. The estimated tomato region is dilated by the size of 2 cells to be safe. If a cell contains over 10% leaf pixels and not a single tomato pixel it is sprayed.

9.2.3 Test

A ground truth (GT) cell array is generated manually, where values 0 = soil, 1 = tomato, and 2 = weed. Evaluating the spray grid cell array is done using an error grid cell array (EG). See Equation 6.

Equation 6. Calculating the error grid cell array. EG is error grid cell array, GT is ground truth cell array, SG is spray array cell array, and I is image number.

$$EG(x,y) \quad = \quad SG(x,y) + 2GT(x,y) \tag{9.11}$$

The meaning of the values in EG is shown in table 9.2.

Table 9.2: The meaning of the values in the Error Grid (EG) cell array.

EG	cell value Meaning Goal
0	Soil not spray Maximize
1	Soil spray Minimize
2	Tomato not spray Maximize
3	Tomato spray Minimize (critical)
4	Weed not spray Minimize
5	Weed spray Maximize (important)

An error vector (E) can be calculated from EG as the sum of cells that have a certain value (Equation 7). This E vector shows the overall performance of the method used to generate the spray grid cell array.

Equation 7. E is a vector that shows the total no. of cells of the value v.

$$E(v) \quad = \quad \sum_{y} \sum_{x} (EG(x,y) == v) \tag{9.12}$$

A manual count is necessary to detect how many actual plants are sprayed at least in one cell.

9.2.4 Results

This section presents and discusses the results that relates to the spray maps. First the disparity map is verified by measuring the distance from the camera to the soil.

Disparity map

The disparity map was verified by calculating the average distance to the soil, which was 528 mm (The ground truth starting height was 533 mm). The disparity maps showed correctly that the seed line was slightly indented. Table 9.3 shows the average distances in the seed line and on both sides.

A surface fit to the ground soil surface would help tracking real plant height because the camera measures the distance to the plant and a fixed height is assumed. This may not be the case so tracking the soil to camera distance will help. It could also be used for other purposes such as seed line guidance and obstacle tracking.

Table 9.3: Measured average distance to the soil.

Left of seed line	In seed line	Right of seed line
521 mm	543 mm	515 mm

Spray Maps

Table 9.4 shows the result totals for the 34 images. Results are produced twice for each method. "Unsafe tomato" is produced with dilation/threshold settings that are unfavorable for tomato edges. "Safe tomato" is produced with settings that keep the tomato edges safer, i.e. larger dilation and height thresholds. The most important numbers are bold.

Table 9.4: Spray results for the surface methods. The sums of Error Grid (EG) 1-6 cells and the number of killed plants are sprayed at least in one cell.

	1. Height Threshold		2. Blobs Histogram		3. Watershed Blobs	
Error vector	unsafe tomato	safe tomato	unsafe tomato	safe tomato	unsafe tomato	safe tomato
E(0)	26047	26056	26044	26056	26036	26047
E(1)	89	80	92	80	100	89
E(2)	10921	11238	11227	11240	11043	11137
E(3)Tom. Spray	**327**	**10**	**21**	**8**	**205**	**112**
E(4)	2692	3505	2990	3472	2302	2928
E(5)Weed Spray	**2764**	**1951**	**2468**	**1984**	**3154**	**2528**
Killed Tomatoes	21	3	4	3	15	8
Killed Weeds		239	325			351

Table 9.5: Spray results for the new results using the disparity labeling technique for the surface reconstruction.

	4. Surface segmentation	
Error vector	unsafe tomato	safe tomato
E(0)	26001	26027
E(1)	114	109
E(2)	11229	11236
E(3)Tom. Spray	**19**	**12**
E(4)	2501	2861
E(5)Weed Spray	**2955**	**2595**
Killed Tomatoes	4	3
Killed Weeds	413	363

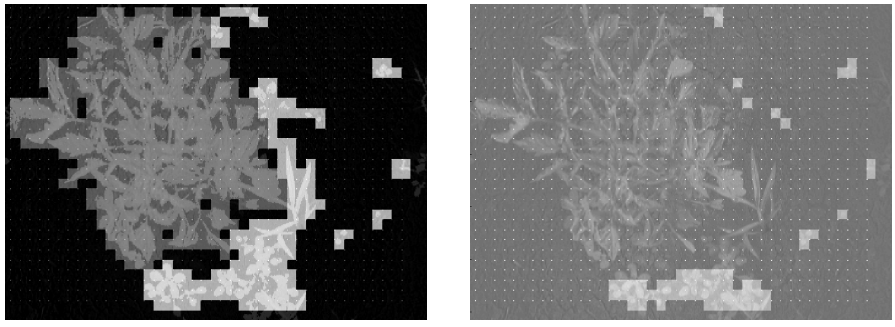
It shows that the first method was very bipolar; either it killed many tomatoes or few weeds. It used no knowledge of the connected pixels and was thus only guessing where the edge of the tomato region was. When assuming the region was very large, the weeds close to the tomato could not be detected. The second method successfully used the knowledge of connected pixels to define a tight boundary around the tomato region. Thus, weeds close

to the tomato could be sprayed, as long they were not connected to the tomato. It does not increase the number of tomato hits much to allow more weeds to be sprayed. The third method was a combination of the properties of the first and second methods. This method was the best at hitting occluded weeds, but it also hit more tomato leaves located at the edge of the tomato region. The watershed algorithm cropped the tomato region where the outer leaves were close to the ground. It is interesting to compare the result with the first method. The sprayed weed cells increase more from "safe tomato" to "unsafe tomato" with the third method, while it is the sprayed tomato that increases drastically in the first method. Figure 9.18 and figure 9.19 shows the spray maps and ground truth for the image example used throughout this section. It shows the "tighter fit" with the blob methods, and the benefits of dividing the blobs into smaller pieces. The occluded grass weed can be sprayed.

Table 9.5 shows new results using the disparity labeling technique for the surface reconstruction developed in chapter 8. Comparing the number of sprayed tomato versus sprayed weed cells to the other techniques is evidence of a much better separation between weed and tomato. But it is still a problem that some weeds are taller than some tomato leaves.

9.2.5 Discussion

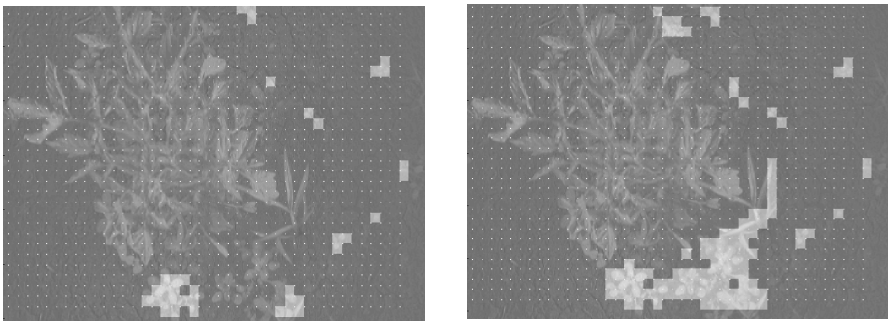
The seed line indentation was not constant through the entire sequence. When the indentation is present, a weed outside the seed line appears taller than the same weed in the seed line. The result of soil distance calculation makes it possible to update the soil levels on the go. The classification methods are not ready for controlling a real micro sprayer array. The blob histogram method can be used with a large safety zone around the estimated tomato region if some weed population is acceptable. It does not solve the occlusion problem, though. The watershed blob histogram method was promising but it needs completion before it is usable. Further work should improve the watershed segmentation to avoid the very small blobs. It would be an advantage if the blobs resemble whole leaves. Incorporation of disparity information into the distance function needs to be tested. It is also possible that blobs can be reconnected afterward.



(a) (b)

Figure 9.18: (a) Ground truth. Light grey cells are weeds. Darker cells are tomato. (b) Spray map using disparity map threshold.

The supposed weeds that are close to the tomato must be double-checked using other features than height. This can be done using leaf shape and color features. If a significant



(a) (b)

Figure 9.19: (a) Spray map using blob histogram. (b) Spray map using modified watershed and blob histogram.

amount of certain tomato pixels and certain weed pixels are present, classes can be created for a Bayesian classification of the weeds that might be tomato. The effect of viewpoint also needs investigation. It was notable in the images that the viewpoint was important for the visibility of occluded weeds. Some weeds would be visible in the top camera or left camera but not in the reference camera. It is especially an interesting aspect when the micro sprayers are shooting from different viewpoints. Finally, utilizing the temporal information needs to be investigated. This means using information from the previous image and taking advantage of the alternative viewpoint again. A weed may not be visible on the current image, but will be in the next or vice versa.

Part IV

Results

Chapter 10

Contributions

[1]

Summary:

In order to wrap up the contents in this thesis a final test compared the algorithms pursued during the work in this dissertation. The new algorithms and the NURBS surfaces were compared. Furthermore, the number of leaves per plant in the final 3D models were tested. The analysis using only percentage of bad matching pixels are found to be misleading, because the structures may be better represented in a disparity map that is equal or slightly worse than another. The final discussion and conclusion of the entire dissertation follows.

The developed algorithms were tested on the virtual and real plant image sets using the best settings found in the previous chapters. Each algorithm was tested against *pbmp* and its ability to segment the correct number of leaves. Test were performed before and after the NURBS surface fitting.

10.1 Results at Disparity Level

This test is the same as performed on existing algorithms in the beginning but now includes the new algorithms before and after NURBS fitting. The methods have been tested in previous chapters with different settings so this time the settings are fixed based on the findings in their respective chapters. Besides the pbmp test that tests only the disparity maps, their usage for further plant reconstruction was also tested in a leaf counting algorithm.

Figures 10.1 and 10.2 show the performance of the different algorithm using the best settings (found in the previous chapters). The figure included percentage of bad matching pixels before and after NURBS surface fitting. The results after doing dual view NURBS fitting for the graph cuts results is also shown. For some plants it is a small advantage but for others it is performing worse. This is because of the over fitting that was seen in section 8.3.2.

The means of the result are listed in table 10.1. Plant 11 is left out in the third column,

[1] Results in this chapter was presented in paper Nielsen and Andersen [2008]

Table 10.1: Final Results. Mean percentage of bad matching pixels for good parameters. The upper part of the table are SSD based algorithms and the lower part are graph cut algorithms. Within these two groups the best results are emphasized.

Algorithm	Virtual	Real	Real (-plant 11)
smw1	12.30	***4.96***	***4.36***
nurbs smw1	10.20	***4.56***	***4.06***
smw9	14.13	7.06	6.45
nurbs smw9	10.20	6.46	6.21
comb smw	11.48	***4.87***	***4.43***
nurbs comb smw	***8.30***	***4.78***	***4.37***
GrC kz1	10.27	10.58	6.63
nurbs GrC kz1	***3.92***	11.78	7.86
dual nurbs GrC kz1	5.51	19.66	15.31
Sloped GrC	8.36	***8.48***	***4.9***
nurbs Sloped GrC	***3.45***	***8.41***	***4.8***
dual nurbs Sloped GrC	***3.55***	11.22	6.33

because it is was very destructive on the graph cut results.

The results on virtual images shows that the developed algorithms combined smw and sloped graph cut are improvements over their old counterparts, especially after fitting NURBS surfaces, even though the positive effect was lessened. It clearly shows that graph cuts are better than SSD correlation methods.

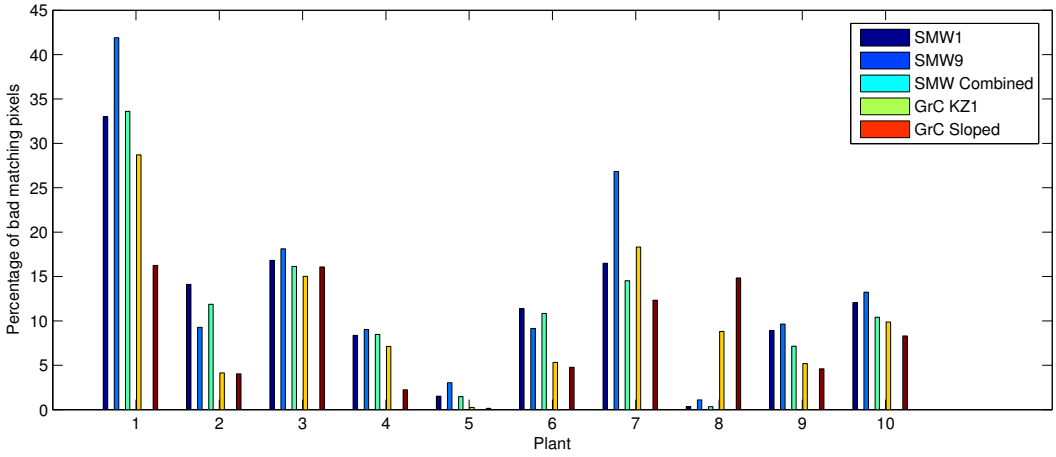
Results with real images were less obvious than the promising virtual results. Plant number 11 was distractive because an entire leaf was covered in the left image and outside the image in the third image, which made it more difficult for the graph cut algorithm to reconstruct it (it relies on reconstructing disparity maps for all three views simultaneously).

Smw 1 and combined smw performed similarly after NURB fitting. Not accounting for plant 11 the sloped graph cut algorithm perform almost the same. The NURBS fitting did not always make the disparity map better. Results after NURBS fitting reflects the structural quality of the disparity maps. The NURBS fitting worked best with Combined SMW and Sloped Graph Cut, especially because the segmentation functioned well.

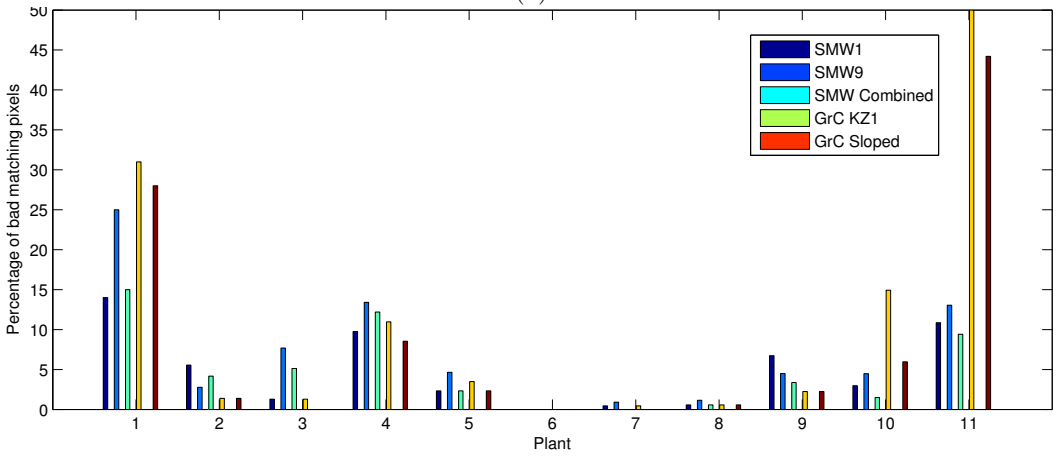
10.2 Results at Plant Level

The following test was done by counting the leaves that were reconstructed in the VRML output files and annotated. The leaf counting test results are shown in tables 10.2 and 10.3. The results confirm that combined SMW and Sloped Graph Cut worked best.

Despite the fact that smw 1 was slightly better than combined smw, and that sloped graph cut was marginally better than the standard graph cut (kz1), the disparity maps was a much better quality for the surface descriptions, because the leaves were separated much better. Note the leaf counts for virtual and real images that the standard algorithms perform much worse on real images compared to virtual images. The new algorithms perform almost the same for virtual and real images. The number sequences (10 9 21 9) and (17 11 36 11) follow

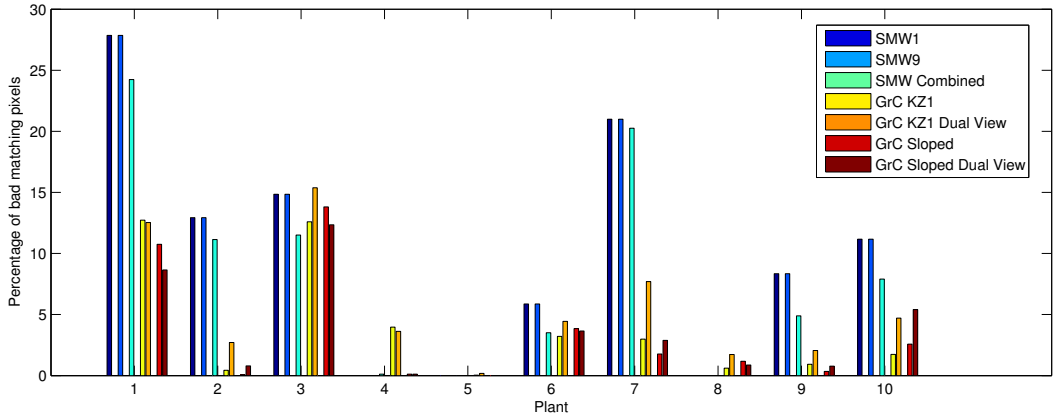


(a)

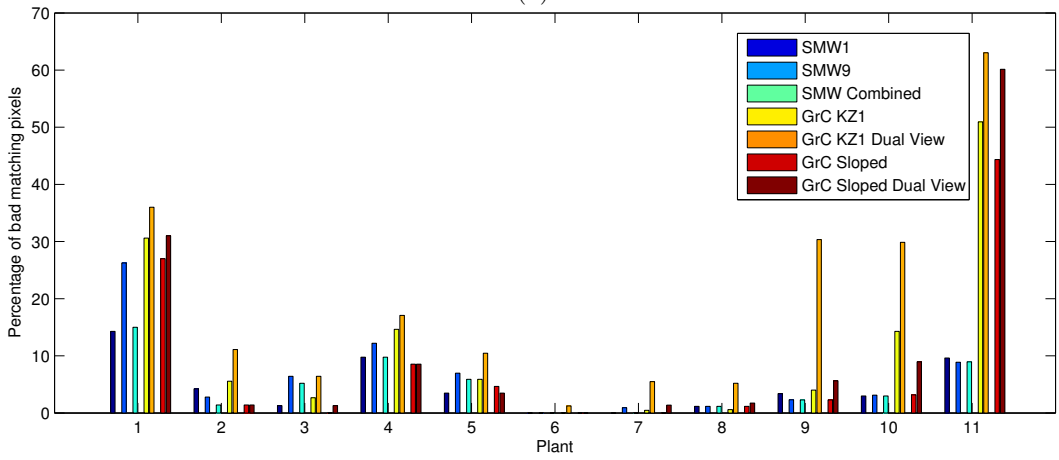


(b)

Figure 10.1: Final test results before NURBS surface fitting. [a] Virtual plants. [b] Real plants.



(a)



(b)

Figure 10.2: Final test results after NURBS surface fitting. [a] Virtual plants. [b] Real plants.

Table 10.2: Leaf Count Results for Virtual Test Set. After NURBS fitting.

Plant	GT No. Leaves	SMW1	SMW Comb	GrC	GrC Sloped
1	5	4	4	7	5
2	6	6	6	8	6
3	6	4	4	9	4
4	5	4	4	3	2
5	6	6	6	6	4
6	12	13	13	15	12
7	5	3	3	8	4
8	5	5	5	6	5
9	10	12	11	12	11
10	16	15	15	19	16
Total error:		10	9	21	9

Table 10.3: Leaf Count Results for Real Test Set. After NURBS fitting.

Plant	GT No. Leaves	SMW1	SMW Comb	GrC	GrC Sloped
1	3	4	4	7	5
2	3	3	3	9	3
3	4	3	3	4	3
4	3	3	3	9	3
5	4	1	4	11	3
6	5	5	5	5	5
7	9	7	8	10	7
8	5	3	3	3	4
9	5	3	5	7	5
10	4	6	6	10	4
11	6	10	10	8	10
Total error:		17	11	36	11

same pattern. But if seeing the average error 0.9 compared to an average of 7.6 leaves per plant for virtual plants and average Error of 1.0 compared to 4.6 leaves per plant there is an offset. However, this is good news, because it shows that the virtual plant test framework still produces valid results for higher abstraction data. The actual error is offset by a certain degree but the test parameter (e.g. algorithm or setting) that produces the best or worst results is the same. This matches the findings from testing the pbmp in chapter 4.2.

10.3 Discussion

10.3.1 Test Framework

The test framework (chapter 4) is expected to serve as a catalyst in the research of more robust 3d computer vision based sensors for precision agriculture. It can be expensive to perform experiments in this application field. Hence, our test framework is a valuable tool that makes it possible to test performance before investing in expensive equipment and field experiments. Generating simulated images with ray-tracing is a versatile approach, which can simulate different camera setups, indoor- and outdoor lighting conditions, focal blur, motion and glossiness. The test framework allowed statistical- and graphical analysis for benchmarking 3D reconstruction algorithms on synthesized crop scenes and for finding the relationships between their many combinations of parameter configurations and the crop features. The performance of the synthesized images was comparable with real images, when the structure was the same and all the important error sources were simulated.

Wilcoxon Signed Rank Test vs T-Test

In this work the t-test was used to test null hypotheses. The standard deviations were rather large. This warrants the consideration whether the t-test is usable for comparing different setups and experiments. After all, the variability is not caused by natural phenomena. Instead, Wilcoxon's signed rank test should be considered. However, using the signed rank test to analyze the results showed the same results as the t-test in chapter 4. But it did find significant improvement in the final result in section 10.1. It will need further examination to find the situations where the t-test and the singed rank test shows different results.

10.3.2 Disparity Estimation and Surface Reconstruction

It was chosen to do surface reconstruction in a modular structure with individual disparity map estimation and then NURBS Fitting. This was because focus had been on finding good disparity maps. The benefit from this modular approach was that disparity estimation and surface fitting are independent of each other and can be solved as separate problems. Furthermore, it was shown that some applications could be solved using only disparity maps and for such applications it is a lightweight approach to eliminate the surface reconstruction module.

Disparity Map Estimation

A number of methods was investigated in this dissertation and they are discussed below.

Trinocular Symmetric Multiple Windows

The relationship between the performances of the algorithms and the descriptive parameters of the plant objects were investigated. Two trinocular formulaes and a new multi-baseline Sum of Squared Difference based correlation was also defined (in chapter 5) and were analyzed in relation to occlusion and specular highlights and parameters describing the objects in the image. They all had their pros and cons. Multibaseline was a marginal improvement

with steep slopes but was very good at highlights. The trinocular minimum measure was good at occlusions. Trinocular sum was best overall.

An interesting experiment would be to place the 5 cameras in a trinocular setup. The five cameras would then complete two systems of three-camera multi-baseline systems in each direction. It would also be a further development to base a selection of trinocular sum and min on occlusion detection.

In section 6 a combination of correlation based reconstruction using a centered window and using symmetric multiple windows was developed. The result was a minor pbmp improvement, but a major improvement when the disparity maps had to be segmented into individual leaves.

Trinocular Graph Cuts with Sloped Extension

The standard graph cut algorithm [Kolmogorov and Zabih, 2002] performed badly on the plant images, because it created staircase patterns on the slopes. The third camera in the trinocular setup was integrated into the graph simply by adding the energies. It is equivalent to the *trinocular sum SMW* algorithm used earlier.

A new adaption of Pott's energy was designed and found a valuable improvement. While it is still not perfect, it makes for a better disparity map for a following surface fitting. It introduced 2 parameters S and λ_s that were easy to set automatically like the other parameters in the graph cut algorithm. The new term was not regular, but a novel method was applied to handle this by detecting situations, where the energy term is not regular. In case of problems the energy is made regular by altering the energy for the neighboring vertex.

Using energy minimization was not a huge advantage to the combined SMW algorithm. It introduced some additional problems with segmenting individual leaves.

Surface Fitting

It was possible to split the disparity maps into individual objects (chapter 8) using simple mask based segmentation. Each object was described with NURBS and thus the disparity maps were improved. It was possible to fit the NURBS using a single view or combining multiple views if such disparity maps were available. It was also possible to reconstruct overlapping objects because there was one NURBS per object. Information about the objects were also extracted and the accuracy of the information was acceptable.

It was difficult to choose the threshold at which an object should be separated, especially when the disparity map was reconstructed using graph cuts.

Non-Modular Approach

The modular approach is very sensitive toward the z threshold in the individual object segmentation. It results in cut leaves, while stems are attached to leaves.

Especially the graph cut disparity maps has the risk of being cut in the middle of the leaves. If the disparity estimation and surface fitting is wrapped into an energy minimization (as in [Lin and Tomasi, 2004]) this could be avoided.

In future research it would be interesting to do it differently using the more flexible multiple

NURBS representation and the developed dual view fitting formulas. Graph cuts would segment the leaves, using disparity discontinuities combined with shape parameters to cut leaves from stems (perhaps using watershed cuts). The result would be overlapping NURBS surfaces. Given detected feature information such at tips and curls it was possible to constrain or add more weight to such locations.

10.3.3 Structural Representations

Extraction of syntactic data from plants using computer vision (chapter 9) was demonstrated. Some information was found in 2D and some in 3D. 2D analysis led to distinct feature descriptions such as tips, a rough estimate of leaf area and the base of the plant. It was also possible to get an estimate of the leaf area, but the 3D estimation was slightly more accurate and with more generality.

By analyzing the 3D reconstructions it was possible to extract information on individual leaves such as their height from the ground, steepness, number of leaves, and more. Combined with the 2D annotated data it can result in a novel sensor for use in precision agriculture. In section 9.2 it was demonstrated that the height parameter is usable for weed detection. The work was done directly on the disparity maps so here the surface fitting was not necessary.

Chapter 11

Summary

This thesis presents a series of computer vision based methods designed for plant images using 3 color cameras. Plant structures are naturally variable and complex, wave in the wind, can point straight up toward the camera, or even broken or eaten partly away. Their surfaces can be purely green and specular reflective.

The purpose was to extract information for usage in precision agriculture for diagnostics of local crop health. The scope of the problem was focused on spatial 3D structural information in the work. This was a scarcely ventured area because plants are non-rigid biological objects. Most work in biological domains have been medical imagery where the image sources are different (e.g. MRI slices) and where image formation take a long time.

A key point in this work is that image capture is instant which means it can be captured on-the-fly while driving through the field regardless of plants waving in the wind. Of course, light conditions must allow for a fast shutter to minimize motion blur. In chapter 9.2 this was shown to be possible using a mounted metal canopy with 4 light sources. The light inside was almost diffuse to eliminate shadow problems and specular reflections from leaves. The light should not be ideally diffuse, because a little shading on the leaves was an advantage for determining photo consistency between camera views (just like the idea behind structured light).

During a test performed with a ray-tracing test framework (chapter 4) it became clear that a major problem for 3D reconstruction was the fact that the structures point toward the camera. Typical methods for disparity map generation assume disparity constancy. Window based methods that does not assume constancy around the neighborhood also cannot handle these steep slopes, because a window of an object with a steep slope looks very different from two viewpoints. It is very difficult to correlate such windows when all the object pixels are almost the same green color.

Below is a summary of the contributions made in the thesis:

- In chapter 4 an extensive test framework was made. Instead of the usual geometric forms, randomized plant models were made in Plant Studio for testing purposes in ray traced images. It was possible to control the structures, lighting, and camera setups. In order to make the ray traced image sets as realistic as possible they were made to be slightly misaligned, brightness and contrast was adjusted randomly, added poison

noise, and rendering using depth of field. Depth maps were converted into disparity maps, and occlusion maps and specular highlight maps were also generated.

- In chapter 5 A novel trinocular method (trinocular min) and a new multi-baseline Sum of Squared Difference based correlation were developed and tested with the test framework. Multibaseline was a marginal improvement with steep slopes but was very good at specular highlights. The trinocular minimum measure was good at occlusions.

- In chapter 6 a simple but novel method was developed that combines the strength of using multiple windows with the strength of using a centered window. The improvement in the disparity map was only a small improvement, but the effect of this improvement was positive when the disparity map had to be separated into smaller parts.

- In chapter 7 a famous graph cut algorithm was adapted to use trinocular image sets (using the trinocular sum principle). Furthermore, a novel smoothness term was developed to reconstruct sloped surfaces better than the original graph cut algorithm without adding to its complexity. The new term was not regular, but a novel method was applied to handle this. However, using energy minimization was not a huge advantage to the combined SMW algorithm. It introduced some additional problems with segmenting individual leaves.

- A novel method for detecting leaf area, leaf tips, overall bases, leaf heights, leaf steepness was developed. These descriptors are novel contributions for sensors in precision agriculture. Weed classification and NPK deficiency was used as practical cases. The methods in this dissertation is a tangible step toward a practical solution to these application. Real-time demands are still problematic, though.

Tangible solutions have been presented for 3D reconstruction based plant annotation. The reconstructed 3D models did not have a photo realistic quality, but it was demonstrated that the quality was enough to extract some information. Some of the contributions are not complete algorithms, but rather extensions (or add-ons) that could be taken out of their algorithmic context and placed within a new algorithm. Each method has its own merit and future research should consider to put some of the pieces into their algorithms. Below are some examples for future work:

- The trinocular minimum term could be implemented as part of a graph cut algorithm and applied where the graph cut algorithm finds occlusion.

- The multibaseline SISSD term could be implemented in a trinocular multibaseline setup.

- The Graph cut algortihm could be wapped around the labelling and NURBS surface fitting.

- The combined SMW aproach could be used in other window correlation based algorithms.

- Combinations of the above...

11.1 Perspectives

Information that is given by the crop canopies combined with spectral sampling should allow for better assessment of crops. Of course, the limitation of computer vision based sensors is that the symptoms must show visible signs, or at least visible in the near-infrared spectral response. It also means that it will work best at early growth stages where the canopies do not occlude each other. Luckily, this is also when fertilizer adjustment has the most impact. A 3D computer vision based sensor will be usable as either sensor-based variable rate application or map-based variable rate application with targeted sampling schemes.

The computer vision based sensor could be part of a multiple scale decision support system; Large scale (very distant) sensors, such as satellite/aerial images, can be used to target sampling strategies to be used by the small scale (ground level) sensors collecting crop data (e.g. computer vision based) and soil data (e.g. EM-38). The sensor data could be merged in a decision support system, which would be used directly in any type of variable rate application or to aid the farmer's management. The approach is highly modular and other data can be incorporated such as weather data and related news reports.

11.2 Acknowledgments

I would like to thank the Danish Technical Research Council, the Danish Agricultural and Veterinary Research Council and the Danish Ministry of Food, Agriculture and Fisheries for funding through the National Research program "Sustainable Technology in Agriculture". I would also like to thank the people at BAESIL at University of California for a pleasant stay and facilities.

Bibliography

H. J. Andersen. *Outdoor computer vision and weed control.* PhD thesis, Laboratory of Computer Vision and Media Technology, Aalborg University, 2001.

H. J. Andersen, K. Kirk, and L. Reng. Geometric plant properties by relax stereo vision using simulated annealing. *Computers and Electronics in Agriculture*, pages 219–232, 2005.

L. Assémat and M. Chapron. The detection of relative height of weeds as a main determinant of an early competition index estimate in the field. In *Proc. of 4th European Conference on Precision Agriculture*, Berlin, 2003.

B. S. Blackmore, H. W. Griepentrog, and L. K. Christensen. *Visioner for fremtidens jordbrug (Visions for Future Agriculture)*, pages 127–140. Gads Forlag, Kbenhavn, 2002a.

S. Blackmore, R. J. Godwin, and S. Fontas. The analysis of spatial and temporal trends in yield map data over six years. *Precision Agriculture*, pages 169–182, 2002b.

S. Blackmore, R. J. Godwin, and S. Fontas. The analysis of spatial and temporal trends in yield map data over six years. *Biosystems Engineering*, pages 455–466, 2003.

R. Brivot and J. A. Marchant. Segmentation of plants and weeds for a precision crop protection robot using infrared images. In *IEE Proceedings: Vision, Image and Signal Processing*, volume 143, pages 118–124, 1996.

C. Buehler, S. J. Gortler, M. F. Cohen, and L. McMillan. Minimal surfaces for stereo. In *ECCV (3)*, pages 885–899, 2002. URL `citeseer.ist.psu.edu/buehler02minimal.html`.

J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6), 1986.

M. Chapron and P. Huet. Characterization of stone densities of soils by image processing. In *Proc. of 4th European Conference on Precision Agriculture*, Berlin, 2003.

M. Chapron, K. Khalfi, E. Monton, P. Boissard, and L. Assemat. Weed recognition by color image processing. In *The 10th Scandinavian Conference on Image Analysis*, 1997.

L. K. Christensen. *NPK deficiencies discrimination by use of spectral and spatial response.* PhD thesis, The Royal Veterinary and Agricultural University, AgroTechnology Section, Department of Agricultural Sciences, April 2004.

L. K. Christensen and B. S. Bennedsen. NPK stress discrimination in spring barley using hyper spectral and spatial information. *Submitted*, 2004.

L. K. Christensen and R. N. Jørgensen. Spatial reflectance at sub-leaf scale discriminating NPK stress characteristics in barley using multiway regression (N-PLS). In *2003 ASAE Annual International Meeting, Las Vegas, Nevada, USA, July 27-30, Paper no. 031138*, 2003.

L. K. Christensen, B. S. Bennedsen, R. N. Jørgensen, and H. Nielsen. Modeling nitrogen and phosphorous content at early growth stages in spring barley using spectral line scanning. *Biosystem Engineering*, 88 (1):19–24, 2004.

C. Demoulin and M. Van Droogenbroeck. A method based on multiple adaptive windows to improve the determination of disparity maps. In *ProRISC/IEEE Workshop on Circuit, Systems and Signal Processing*, pages 615–618, Veldhoven, The Netherlands, November 2005.

U. R. Dhond and J. K. Aggarwal. Structure from stereo: A review. *SMC*, 19(6):1489–1510, November 1989.

U. R. Dhond and J. K. Aggarwal. Binocular versus trinocular stereo. In *IEEE Intl. Conf. on Robotics and Automation*, pages 2045–2050, 1990.

I. Filella and J. Peuelas. The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. *International Journal of Remote Sensing 15(7)*, pages 1459–1470, 1994.

M. A. Friedl, J. Michaelsen, F. W. Davis, H. Walker, and D. S. Schimel. Estimating grassland biomass and leaf area index using ground and satellite data. *International Journal of Remote Sensing 15(7)*, pages 1401–1420, 1994.

A. Fusiello, V. Roberto, and A. Verri. Symmetric stereo with multiple windowing. *International Journal of Pattern recognition and Artifical Intellingence*, 14(8):1053–1066, 2000.

R. Gangloff and K. Westfall. Spatial dependency of soil samples and precision farming applications. In J. Stafford, editor, *Proceeding for 4th ECPA - 1st ECPLF*, June 2003.

J. P. Hardwick and Q. F. Stout. Flexible algorithms for creating and analysing adaptive sampling procedures. In *New Developments and Applications in Experimental Design*, volume 34, pages 91–105, 1998.

J. Heikkila and O. Silven. A four-step camera calibration procedure with implicit image correction. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1106–1112, June 1997.

A. Jensen and Z. Young. Capturing low-cost aerial images to use in site-specific agriculture. In J. Stafford, editor, *Proceeding for 4th ECPA - 1st ECPLF*, June 2003.

J. Jeon, K. Kim, C. Kim, and Y. Ho. Robust stereo matching algorithm using multiple-baseline cameras. In *IEEE Pacific Rim Conference on Communications, Computers and signal Processing*, volume 1, pages 263–266, 2001.

T. Kanade and M. Okutomi. A stereo matching algorithm with an adaptive window: Theory and experiment. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16(9):920–932, 1994. ISSN 0162-8828. doi: http://dx.doi.org/10.1109/34.310690.

V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. In *IEEE European Conference on Computer Vision*, May 2002.

V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(2):147–159, 2004. ISSN 0162-8828.

P. D. Lancashire, H. Bleiholder, T. van der Boom, P. Langeluddeke, R. Stauss, E. Weber, and A. Witzenberger. A uniform decimal code for growth stages of crops and weeds. *Annals of Applied Biology 119*, pages 561–601, 1991.

W. S. Lee, D. C. Slaughter, and D. K. Giles. Development of a machine vision system for weed control using precision chemical application. In *International Conference on Agricultural Machinery Engineering '96, Seoul, Korea.*, pages 802–811, 1996.

W. S. Lee, D. C. Slaughter, and D. K. Giles. *Robotic weed control system for tomatoes*, volume 1, pages 95–113. Wageningen Academic, 1999.

Y. Li, S. Lin, H. Lu, S. Kang, and H.-Y. Shum. Multibaseline stereo in the presence of specular reflections. In *IEEE Intl Conf. on Pattern Recognition*, volume 3, pages 573–576, Aug. 2002.

M. Lin and C. Tomasi. Surfaces with occlusions from layered stereo. In *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, volume 26(8), pages 1073–1078, August 2004.

D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60 (2):91–110, 2004.

H. Marschner. *"Mineral nutrition of higher plants"*. "U.S. Edition Published by Academic Press Inc., San Diego, CA 92101", 1995. ISBN 0-12-473542-8.

M. Morgan and D. Ess. *The Precision Farming Guide for Agriculturists.* John Deere Publishing, 1997.

D. J. Mulla, A. C. Sekely, and B. M. Evaluation of remote sensing and targeted soil sampling for variable rate application of lime. In P. Robert, editor, *Proc. 5th Intl. Conf. on Precision Agriculture*, pages 16–19, July 2000.

J. Müller, J. Smit, J. Hofstee, and D. Goense. Student design contests promote hands-on learning and innovation in precision agriculture. *Agro Informatica*, 2003.

J. Mulligan and K. Daniilidis. Trinocular stereo: A real-time algorithm and its evaluation. *International Journal for Computer Vision*, 47:51–61, April 2002.

M. Nielsen and H. J. Andersen. Plant and leaf analysis based on 3d reconstruction. In *Agricultural and Biosystems Engineering for a Sustainable World, International Conference on Agricultural Engineering and Industry Exhibition, Hersonissos, Crete, 23-25 June*, 2008.

M. Nielsen, H. J. Andersen, D. C. Slaughter, and D. K. Giles. Detecting leaf features for automatic weed control using trinocular stereo vision. In *International Conference on Precision Agriculture, Minneapolis, MN, USA*, July 2004a.

M. Nielsen, L. K. Christensen, and H. J. Andersen. Sub-leaf scale remote sensor for npk discrimination using stereo vision. In *Engineering the Future, International Conference on Agricultural Engineering, Leuven, Belgium, 12-14 September, Session 10, no. 327*, 2004b.

M. Nielsen, H. J. Andersen, and E. Granum. Comparative study of disparity estimations with multi-camera configurations in relation to descriptive parameters of complex biological objects. In O. Hellwich, I. Niini, C. Ressl, V. Rodehorst, D. Scharstein, and P. Sturm, editors, *BenCOS - Towards Benchmarking Automated Calibration, Orientation and Surface Reconstruction from Images. ISPRS Workshop in conjunction with ICCV 2005*, pages 63–68. ISPRS Working Groups - WG III/1 Automatic Calibration and Orientation of Optical Cameras - WG III/2 Surface Reconstruction, October 2005a.

M. Nielsen, H. J. Andersen, D. C. Slaughter, and E. Granum. Ground truth evaluation of 3d computer vision on non-rigid biological structures. In J. Stafford, editor, *Precision Agriculture 05*, pages 549–556. Wageningen Academic Publishers, The Netherlands, June 2005b.

M. Nielsen, H. J. Andersen, D. Slaughter, and E. Granum. Ground truth evaluation of computer vision based 3d reconstruction of synthesized and real plant images. *Precision Agriculture*, 8(1-2):49–62, 2007.

M. Okutomi and T. Kanade. A multi-baseline stereo. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 15(4), page 353363, 1993.

D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1/2/3):7–42, 2002.

D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003), Madison, WI, USA*, pages 195–202, 2003.

D. L. Snyder, C. W. Helstrom, A. D. Lanterman, M. Faisal, and R. L. White. Compensation for readout noise in ccd images. *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, 12(2), 1995.

R. Szeliski and R. Zabih. An experimental comparison of stereo algorithms. In *Springer, International Workshop on Vision Algorithms, Corfu, Greece*, pages 1–19, Aug. 1999.

R. Taylor and D. Whitney. Using yield monitor data to direct soil sampling, 2001. URL `http://www.oznet.ksu.edu/pr_prcag/yield-monitor.shtml`.

D. M. Woebbecke, G. E. Meyer, K. Von Bargen, and D. A. Mortensen. Shape features for identifying young weeds using image analysis. In *Transactions of the ASAE 38(1)*, pages 271–281, 1995.

N. Zhang and C. Chaisattapagon. Effective criteria for weed identification in wheat fields using machine vision. In *Transactions of the ASAE 38(1)*, pages 965–974, 1995.

List of Figures

1.1 Overview of the relevant parts and the focus of this thesis. Emphasis is on generating high quality disparity maps for extraction of individual leaves and the automatic annotation of them. 6

2.1 Leaf Reflectance at different N-application levels. NDVI is the ratio between 2 wavelengths, e.g. at 800 nm compared to 660 nm. 15

2.2 (a) A given symptom can come from many factors such as water, soil compaction, and pest. (b) The limiting factor is the weakest link in crop health. . 16

2.3 Growth Stages. Initial aim of the sensor is that it will be functional in the Tillering period. This is when fertilizer adjustment has most impact. Adapted from Agronomy Guide For Field Crops, Ontario Ministry of Agriculture, Food and Rural Affairs, 2003. 17

2.4 Grid sampling: The field is divided into a dense grid of cells, which can be sampled. Multiple samples from each cell can be averaged for better reliability. Targeted sampling: Sparse sample locations are selected intelligently. 18

2.5 Combining resources. The 3D based computer vision sensor as part of a decision support system. 19

3.1 Camera Setup looking at winter barley. (a) As seen from the side: Three JAI M4+CL cameras at fixed positions, denoted Left, Right and Bottom Camera, respectively. (b) Seen from above. 23

3.2 Example tomato image (a) Excessive green and (b) Green chromacity. The latter is les prone to shading. 25

3.3 (a) The SMW algorithm uses 9 windows, and the best similarity score is chosen. (b) In this thesis SMW 1 refers to using the centered window alone. SMW 5 refers to using those windows marked with a 5. SMW 9 uses all windows. 26

4.1 An aligned binocular setup where the second camera is transposed [b 0 0] and rotated [0 0 0] from the first. The third camera in a trinocular setup is transposed [0 b 0]. 32

4.2 Illustration of when occlusion occurs. The dots represent objects or pixels. Their positions are shifted to the right (how far is determined by their disparities) on the X axes from image R to image L. Equation 4.4 perceives an occlusion at point (x,y) if pixels to the left of (x,y) would be shifted onto the same position as itself. 34

4.3 [Top] Examples of real spotted broadleaf, smooth/glossy broadleaf- and grassy plants [Bottom] Examples of the rendered counterparts. 36

4.4 Disparity maps of a real cotton plant. Symmetric multiple windows cannot handle the steep slope on the leaves. 38

4.5 Comparison of window sizes in relation to leaf types and texture types. . . . 38

4.6 Comparison of Window Sizes in relation SMW. 39

4.7 Comparison of Window Sizes in relation to camera setup. 39

5.1 The case of steep leaves where projection changes orientation across the baseline. (a) five views of the location on the steep leaf. (b) The development of the dissimilarity across the baseline. (c) The dissimilarity/energy function across the scan line in the image. The best match for SSD, SSSD, SISSD ($\alpha = 1$), and ground truth (GT) is given over the graph. 48

5.2 The case of flat leaves where the highlight changing across the baseline. The potential weakness of SISSD is that the dissimilarity difference between the correct match and its surroundings is not very pronounced. This makes the global minimum sensitive to jitter. 49

5.3 A natural case, where two grass-like plants are close together and leaves are occluded. The proportion of occluded pixels is 5% and the proportion of changing highlights are 5%. 51

5.4 (a) Ground truth and (b) Graph Cuts Log(disparity error) for steep spotted broad leaf without highlights. The banding characteristics were caused by the attempt to impose fronto-planar regions on the steep leaves. 52

5.5 (a) Log(disparity error) Multi-baseline SSSD and (b) SISSD $\alpha = 0.5$. These results did not have any banding, but the difference between the SSSD and SISSD was very small. The result would be excellent if it were combined with a slope- and discontinuity preserving graph cut minimization. 53

5.6 [a] Log(disparity error) trinocular minimum (T_m) and [b] trinocular sum (T_s). 53

5.7 PBMP from all pixels results by object type and leaf orientation. The worst case occlusion is the *Two Grassy Plants* model being 5% occluded. The worst case of highlights were the flat grass-like and flat broad-leaf. 20% of their area suffered from changing specular highlights. 54

5.8 PBMP from specular changing highlight pixels results by object type and leaf orientation. SISSD (M0.25-M0.75) improves performance. 56

5.9 PBMP from occluded pixels results by object type. Trinocular minimum T_m is the best algorithm for occluded areas. 57

6.1 Disparity maps of real cotton. Brighter pixels are closer to the camera. Configuration: trinocular sum, window size 16, Colour images. (a) SMW 1. (b) SMW 5. 60

6.2 A steep leaf seen from two viewpoints. (a) Window correlation is computed for two corresponding pixels using multiple windows off centered above to the left and below to the right. (b) If those windows were computed using centered windows over the scanlines of their centers, they would really be at different depths and thus find best matches at different disparities. This is true only if the object is steep. A flat object does not give this problem. 61

6.3 Error maps of real cotton. Brighter pixels are larger errors. (a) SMW 1. (b) SMW 5. 61

6.4 SSD error ratio map. Bright pixels are large ratios. 62

6.5 PBMP as a function of error ratio threshold. 63

6.6 Disparity maps of real cotton. Brighter pixels are closer to the camera. Configuration: trinocular sum, window size 16, Colour images. Combined SMW, $T = 4$. 63

7.1 Different disparity/depth map profiles. Most research has been limited to the piecewise constant profile. 66

7.2 Examples of smoothness energy terms V. x axis is $abs(f(p) - f(q))$ and y axis is E(f). 67

7.3 In order to build the graph, an energy matrix is built and broken down into a sum of terms that each can be represented by single edges. A is a constant that is added to E_K. 69

7.4 Graph construction for E^i and $E^{i,j}$. (a) Single variable where $E^i(0) < E^i(1)$, else (b) $E^i(0) \geq E^i(1)$ on the other side. The smallest energy is added to the constant (K). (c) Two variables where $E^{i,j}(0,0) + E^{i,j}(1,1) \leq E^i(1,0) - E^i(0,1)$, and $C - A \geq 0$ and $D - C \geq 0$. (d) if $C - A < 0$ or $D - C < 0$ then the corresponding edge is subtracted on both sides and added to the constant. 70

7.5 The sum of the edges that are cut plus the constant should equal the energy of assigning the configuration to 0 (source) and 1 (sink). The fat arrows are those edges that are cut (dashed lines). Note that directional edge between the nodes is only active when the cut assigns node i to source and j to sink. Alternatively, there is no flow between them and it is not necessary to cut the edge to separate the sink from the source. 71

7.6 Merging of subgraphs by adding the edges together. 71

7.7 Comparison of truncated L1 and adapted pott's for slopes 73

7.8 Testing regularity 73

7.9 Comparison of kz1 with fronto-planar Pott's term (a), sloped extension with truncated linear term (b) sloped extension with adapted pott's term (c). Results after 3 iterations, $\lambda = 5$ (from kz1's automatic heuristic). 74

7.10 Comparison of kz1 to sloped extension with adapted pott's term. a-b: kz1. c-d: sloped extension where $S = 1$ and $\lambda_s = 0$. e-f: sloped extension where $S = 1$ and $\lambda_s = 1$. To the right is full disparity maps and to the left is zoomed in on the plants and improved contrast. 75

7.11 a-b: kz1. c-d: sloped extension where $S = 1$ and $\lambda_s = 0$. e-f: sloped extension where $S = 1$ and $\lambda_s = 1$. To the right is full disparity maps and to the left is zoomed in on the plants and improved contrast. 76

7.12 Comparison on real images with and without visibility constraint. Results with Pott's term when $\lambda = 3$ (L) and for sloped extension with different λ_s (LS). T is the tolerance in the proportion of bad matching pixels. 78

7.13 Comparison on virtual images with and without visibility constraint. Results with Pott's term when $\lambda = 6$ (L) and for sloped extension with different λ_s (LS). T is the tolerance in the proportion of bad matching pixels. 79

7.14 Energy minimization time by iterations. 80

7.15 Graph Cut comparison of settings. The comparisons are done in pairs. Each pair uses the overall best setting from the former pairs, e.g. the λ test uses only the Trinocular results and the visibility test uses only results from Trinocular and $\lambda = 6$, etc. The new *sloped* extension is an advantage for all images except image 8, which is a very flat grass leaf plant. (a) Virtual Plants. (b) Real plants. 82

7.16 Steep plants (a) standard graph cut and (b) slope extension. 83

8.1 In the first pass of the labeling a list of regions are created. Sometimes the regions must be merged so the list also contains information on which unique region a given region is to be merged with in the 2nd pass. 88

8.2 A bad reconstruction of a day lily. Looking at a reconstruction using a single NURBS from the same view as the reference camera looks correct (a), but it looks strange and wavy from another view (b). The other view of a better reconstruction using multiple NURBS is shown (c). 90

8.3 The dotted line shows the Q points. The dashed line shows the control points (P). The solid line was the resulting NURBS (C). (a) A fit with fixed weights. (b) A fit with variable weights. Note how the end is thrown into a large curvature because the weights were small. This is a typical case if the weights are chosen from the SSD error map, where the errors are high at the boundaries of the objects. 92

8.4 There are 4 types of triangles to consider for each pixel. Types 1 and 2 are always placed when all their corners hit the object. Types 3 and 4 are only placed if type 1 was not already placed for the active pixel (the one under the dot). 93

8.5 Left and right views with disparity segments for three objects. (a) The label numbers do not match. Segment 1 is partially occluded. (b) Object 1 in the left view must be matched with the objects in the right view. Object 1 is projected onto the right view and the object with most disparities (not only the spatial overlap) overlapping is the best match. Here it is object 2. (c) Parts from object 2(right) that does not overlap object 1(left) is appended to object 1 and vice versa. 94

8.6 (a) Reconstruction without filling in the gaps from an alternative view. (b) Reconstructed from two disparity maps. 94

8.7 a) Normally there is a gap behind a disparity discontinuity. b) But small gaps can be filled using the occlusion filling technique. 95

8.8 Comparison of graph cut disparity images and their respective NURBS approximation using only the right disparity map and both the right and left disparities. 96

8.9 Notation of triangles in this section. 97

8.10 [Left] Correlation between real steepness and estimated steepness. [Right] Estimated area in relation to the steepness angle. 99

8.11 a) VRML model of the NURBS mesh 45^o wood box. b) a close-up reveals area and steepness of the large single object. c) the problem with the simple mesh is that even after smoothing, it is still step-like. 100

8.12 Leaf heights correlation. Dashed regression line. Grey solid unit line. $R^2 = 0.9959, p < 0.001$. Residuals: $\mu = 0.18, \sigma^2 = 0.18$ 100

8.13 Reconstructions of the real plants from figure 4.3 seen from a slightly different angle. 101

9.1 Four examples of edge signatures. The enumerated vertical bars are where the perimeter intersects the edge at a certain angle. The dotted horizontal line denotes that the perimeter at this angle is inside the leaf. 107

9.2 Finding the Leaf base. A1-2: Examples using large or small circles, and an ideal symmetric plant and a natural plant situation, respectively. B1-2: Combining the usage of circles of different sizes at the same center. 108

9.3 Relationship between estimated and scanned leaf area. a) Low threshold. Areas (in cm^2) found by 2D segmentation: $R^2 = 0.45$ and $p < 0.001$. b) Low threshold. Trinocular reconstruction using combined SMW: $R^2 = 0.85$ and $p < 0.001$. c) High threshold. Areas (in cm^2) found by 2D segmentation: $R^2 = 0.90$ and $p < 0.001$. d) High threshold. Trinocular reconstruction using combined SMW: $R^2 = 0.91$ and $p < 0.001$. 109

9.4 Disparity map (top) and 3D reconstruction by NURBS surface (bottom) of the plant in figure A.1. Left: Based on binocular stereo with multiple windows and simulated annealing. Right: Based on trinocular stereo with a single centered window and without simulated annealing. 110

9.5 Example of plant base and leaf tip detection. The white dot denotes the estimated position of the plant base. The error in this example is close to the mean base error. The intersections of the three circles are candidates for individual leaf bases. The black spots on the tips are the detected tip pixels. 111

9.6 Leaf Tip Results. Bars denote the percentage of tips are 1. Success. 2. Occluded. 3. Out of bound. 4. False positives because the segmentation cut up the leaf. 5. Not segmented correctly. 6. broken/blunt. 7. False positives. 8. Failed. 112

9.7 (a) Digiclops (b) Tomato under light tunnel. 116

9.8 (a) Original color image of tomatoes and weeds. It is difficult for human eyes to distinguish leaf pixels from soil pixels (b) Green chromacity. Leaf pixels stand out clearly. 116

9.9 Disparity-to-height look-up table. The table actually shows the distance from the reference camera to the object. It requires knowledge of the distance to the soil to get the actual height of an object. 117

9.10 Disparity map of the plant previously shown. The disparity jump is easily seen on the tomato occluding a grass weed. Lighter pixels mean taller areas. White is the soil and is not considered. 117

9.11 Classification with height thresholding. Highlighted tomato region. Parts of tomato leaves are sticking out of the region. Increasing the dilation will enlarge the region and that will bring more weeds into safety. 118

9.12 (a) the histogram plots of the modified normal distribution (M) and a weed (H_{weed}) and a tomato (H_{tomato}) (b) The results of the bin-wise multiplication $H_B M_B$. F is basically the area underneath the resulting curve. The area under the weed response is 0.14 and the tomato response is 217.49. 119

9.13 Classification with the histograms of whole blobs. It shows the region classified as tomato. It contains a bunch of weeds as well because the weeds are connected to a tomato region. 120

9.14 (a) Distance map without edge modification. (b) Distance map with edge modification. (c) Watershed result without edge modification. (d) Watershed result with edge modification. 121

9.15 Classification of blobs separated by watershed. It shows the region classified as tomato. Only one leaf of the grass weed on the right side of the tomato is classified as tomato. 121

9.16 Another disparity image with occluded leaves. 122

9.17 (a) Watershed result without edge modification. (b) Result with edge modification. From classification to spray maps. 122

9.18 (a) Ground truth. Light grey cells are weeds. Darker cells are tomato. (b) Spray map using disparity map threshold. 125

9.19 (a) Spray map using blob histogram. (b) Spray map using modified watershed and blob histogram. 126

10.1 Final test results before NURBS surface fitting. [a] Virtual plants. [b] Real plants. 131

10.2 Final test results after NURBS surface fitting. [a] Virtual plants. [b] Real plants. 132

A.1 Rectification principle. 156

A.2 Rectification. Stipled boxes mark matching points with equal disparity (d). . 157

A.3 (a) The difference between the depths achieved by triangulating different disparities individually from each camera pair. (b) The difference between the disparities achieved by reprojecting world coordinates at different depths onto the three images. 157

B.1 Virtual Plant No. 1-2. 160

B.2 Virtual Plant No. 3-4. 161

B.3 Virtual Plant No. 5-6. 162

B.4 Virtual Plant No. 7-8. 163

B.5 Virtual Plant No. 9-10. 164

B.6 Virtual Plant No. 10 masks. 165

B.7 Real Plant No. 1-2. Cotton. 167

B.8 Real Plant No. 3-4. Cotton. 168

B.9 Real Plant No. 5-6. Cotton/Hypoestes. 169

B.10 Real Plant No. 7-8. Hypoestes. 170

B.11 Real Plant No. 9-10. Day lily. 171

B.12 Real Plant No. 11. Day lily. 172

C.1 In order to build the graph, an energy matrix is built and broken down into a sum of terms that each can be represented by single edges. A is a constant that is added to E_K. 173

C.2 Graph construction for E^i and $E^{i,j}$. (a) Single variable where $E^i(0) < E^i(1)$, else (b) $E^i(0) \geq E^i(1)$ on the other side. The smallest energy is added to the constant (K). (c) Two variables where $E^{i,j}(0,0) + E^{i,j}(1,1) \leq E^i(1,0) - E^i(0,1)$, and $C - A \geq 0$ and $D - C \geq 0$. (d) if $C - A < 0$ or $D - C < 0$ then the corresponding edge is subtracted on both sides and added to the constant. 174

C.3 The sum of the edges that are cut plus the constant should equal the energy of assigning the configuration to 0 (source) and 1 (sink). The fat arrows are those edges that are cut (dashed lines). Note that directional edge between the nodes is only active when the cut assigns node i to source and j to sink. Alternatively, there is no flow between them and it is not necessary to cut the edge to separate the sink from the source. 175

C.4 Practical example with two label 0 and 1. The initialization is all zeros. The data term is given on the lines from the nodes to the labels 0 and 1. The initial energy is 12. 175

C.5 A graph is constructed from the data term given the rules in figure 7.4A and 7.4B. 176

C.6 A smoothness term is added to remove noise. Pott's energy is used with static edge clues, $\lambda = 3$. $E(0,0) = E(1,1) = 0, E(1,0) = E(0,1) = 3$ or 9. There's an edge between nodes 2 and 3. [Left] using rules from figure C.2. This method does not add anything to the constant. [Right] Rules from figure 7.4. 176

C.7 The minimal cuts are at the same place but with different flows, but the total energies are the same. 176

C.8 The best configuration is now 0,0,1,1 with the energy 11. 177

Appendix A

Trinocular Rectification

An adaption of Bouquet's Camera Calibration Toolbox for Matlab which is based on [Heikkila and Silven, 1997] was used to rectify images from a trinocular L-setup. The intrinsic parameters was calibrated for each camera individually and extrinsic parameters was calibrated each camera pair (using the same image set of the chess board).

$$KK = \begin{bmatrix} f_x & \alpha & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{A.1}$$

$$P = KK \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix} \tag{A.2}$$

$$\begin{bmatrix} x_c \\ y_c \\ 1 \end{bmatrix} = P \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{A.3}$$

The following is an overview of how the rectification works using basic transformations (see figure A.1):

- Consider the relative translation and rotation of camera 2 and 3 in relation to the first (reference).
- Rotate the reference camera to align the translation of camera 2 with the x axis perfectly and the translation of camera 3 with the y axis almost as good as possible (fig. A.1A-B).
- Align the rotation of camera 2 and 3 with the reference camera (fig. A.1B-C).
- Adjust the affine distortion skew (α) of the camera matrices to align the y axis perfectly.
- Adjust the length of the vertical baseline so that it is equal to the horizontal baseline.

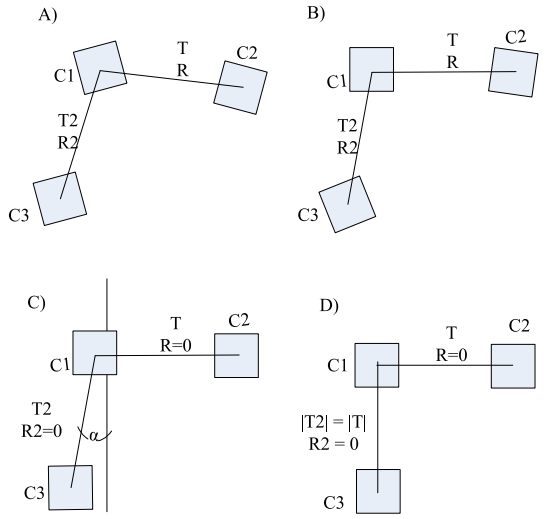


Figure A.1: Rectification principle.

- Select a common focal length for the cameras (fig. A.1C-D).
- Now the projected images might end up outside the image views:
- Center the principal point of each camera so that as much of the original images are inside the image while the translation is equal, or adjust the image sizes accordingly.
- The principle points must satisfy the following: $c_y1 = c_y2$, $c_x1 = c_x3$, $c_x2 = c_y3$
- Store the new ppm's, t's, and R's.

After the rectification the pixels are aligned along the horizontal and vertical epipolar lines in each camera pair, see figure A.2.

Of course, the results are not ideal. Figure A.3 shows an example of disparity imprecisions for the camera setup used in section 9.1. The figures show that when treating the 3 cameras as 2 pairs there is a discrepancy between doing the disparity-depth and vice versa translation the horizontal pair or the vertical pair.

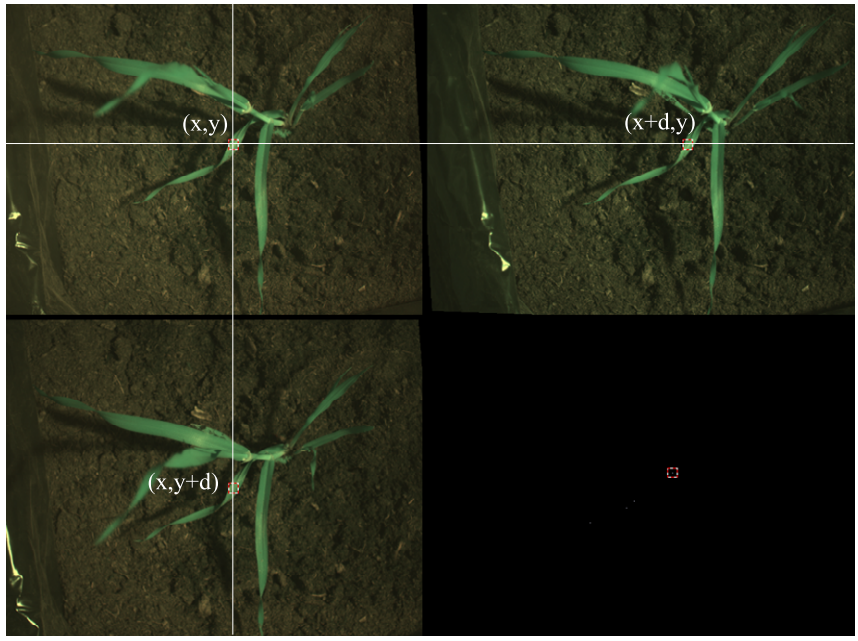
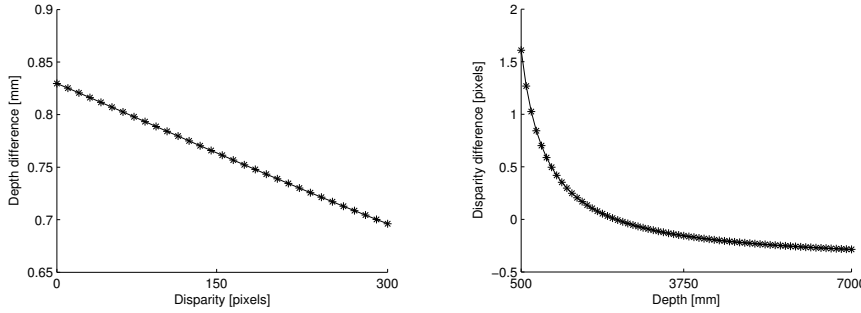


Figure A.2: Rectification. Stipled boxes mark matching points with equal disparity (d).



(a) (b)

Figure A.3: (a) The difference between the depths achieved by triangulating different disparities individually from each camera pair. (b) The difference between the disparities achieved by reprojecting world coordinates at different depths onto the three images.

Appendix B

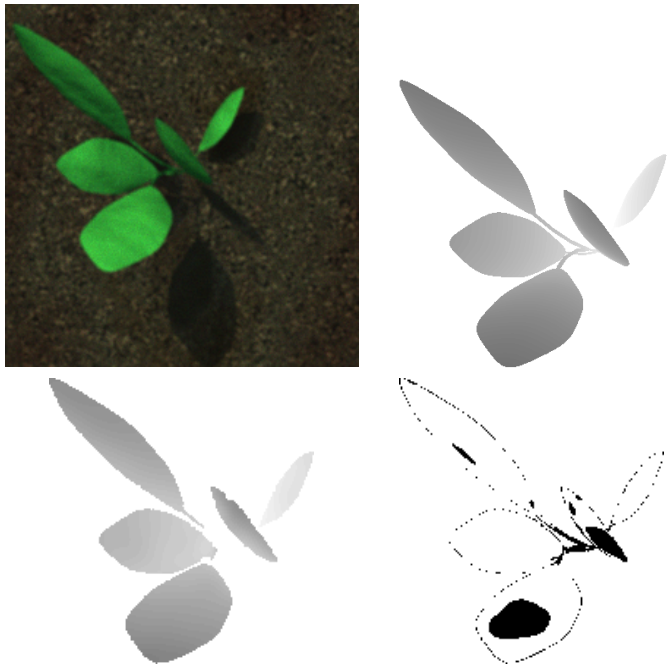
Reconstructed Plant Models

This appendix shows some of the plants that were used in the tests in this thesis. The most interesting plants have been chosen for display. Each plant is shown in 4 ways: TOP A B, BOTTON C D. A and B are right and left views of the plant, C is the result from the combined SMV with NURBS algorithm and D shows where points were annotated and whether they were found correctly on the result shown in C. Correct is denoted o and wrong is denoted x.

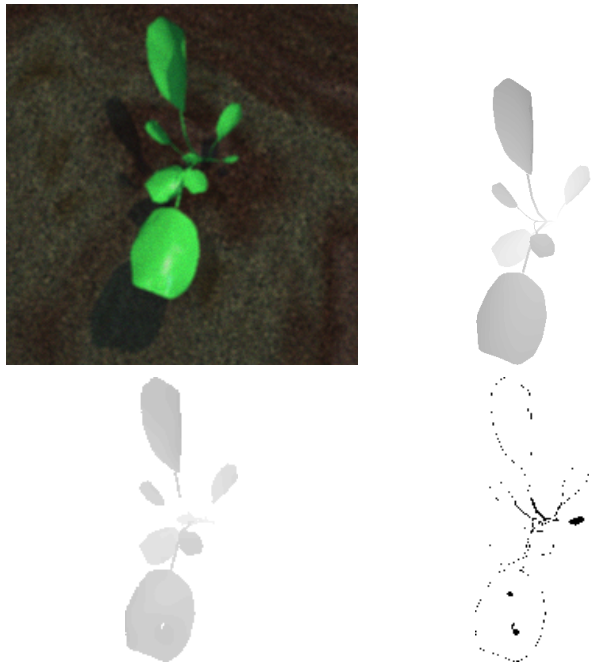
B.1 Virtual Plant Set

The following figures show the reference images, ground truth disparity maps, the reconstructed disparity maps(*), and the error maps for the most commonly used virtual plant images used in the thesis.

(*) Reconstructed using graph cuts with sloped extension and NURBS fitting.

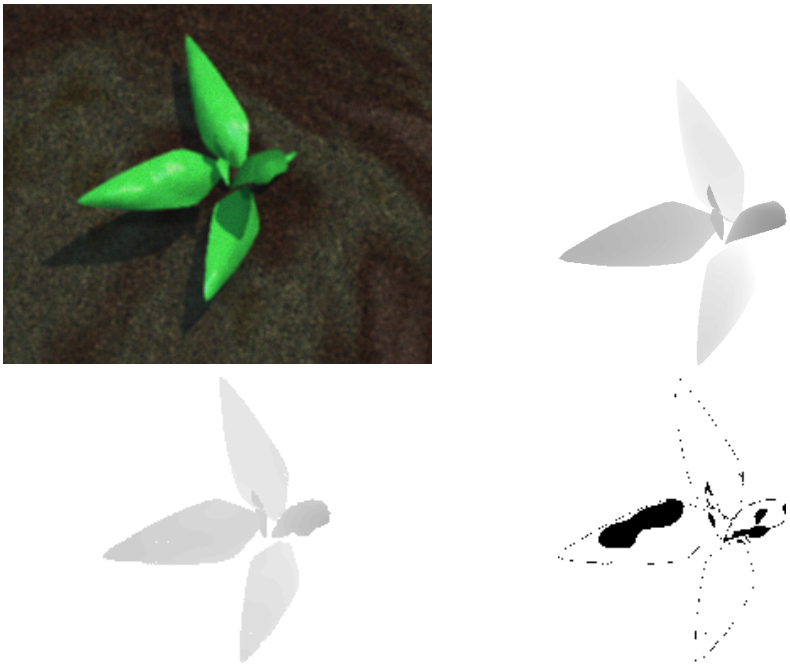


Virtual Plant No. 1

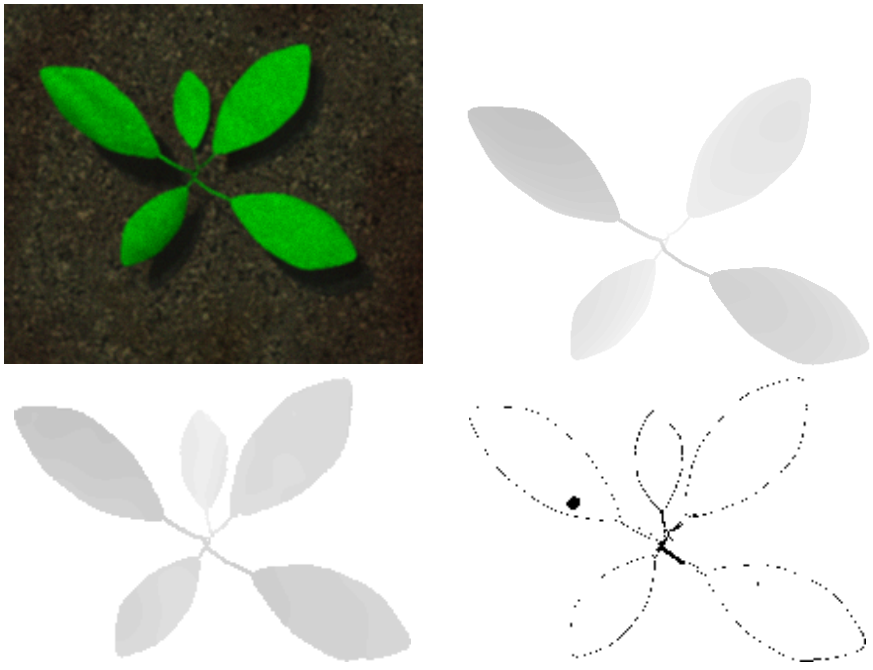


Virtual Plant No. 2

Figure B.1: Virtual Plant No. 1-2.

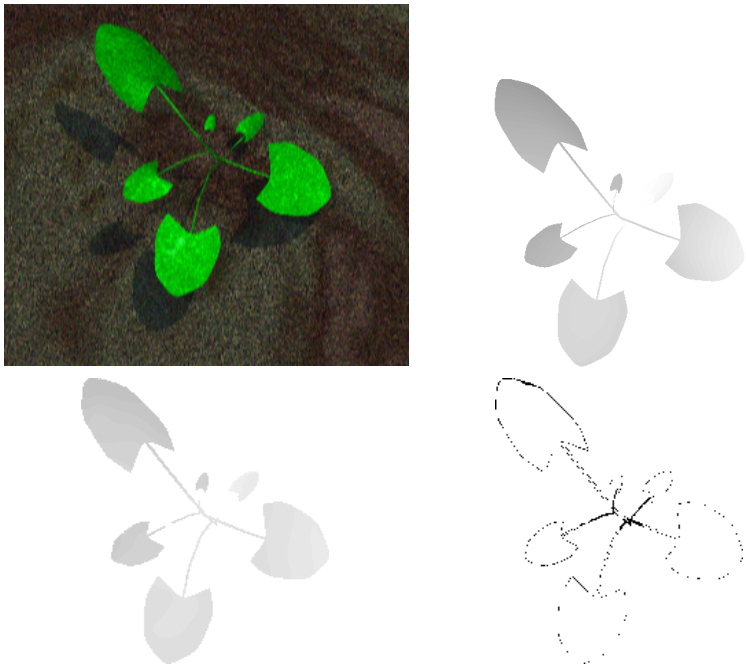


Virtual Plant No. 3

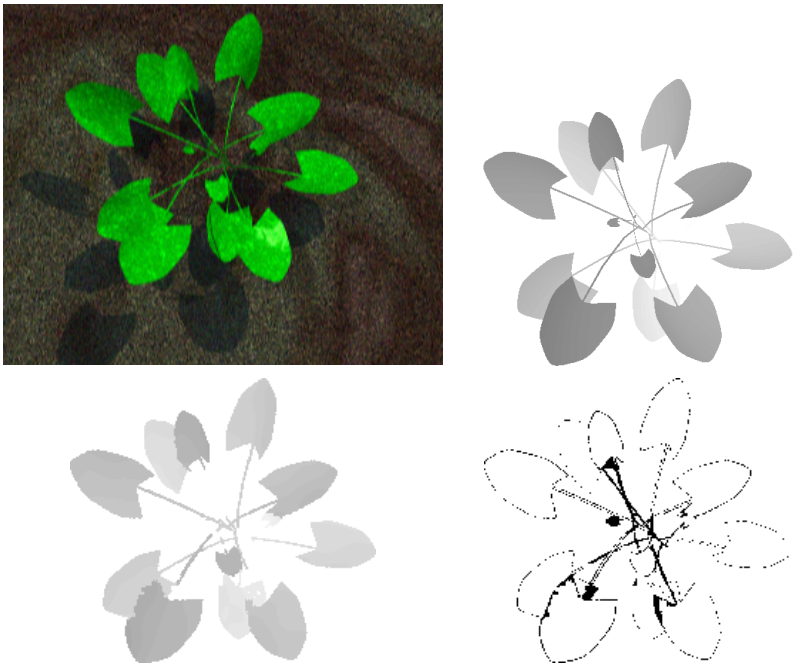


Virtual Plant No. 4

Figure B.2: Virtual Plant No. 3-4.

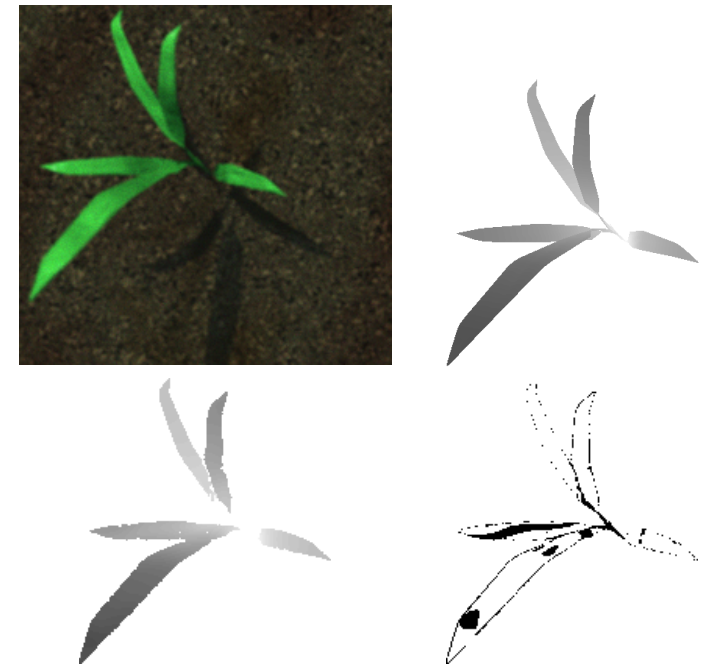


Virtual Plant No. 5

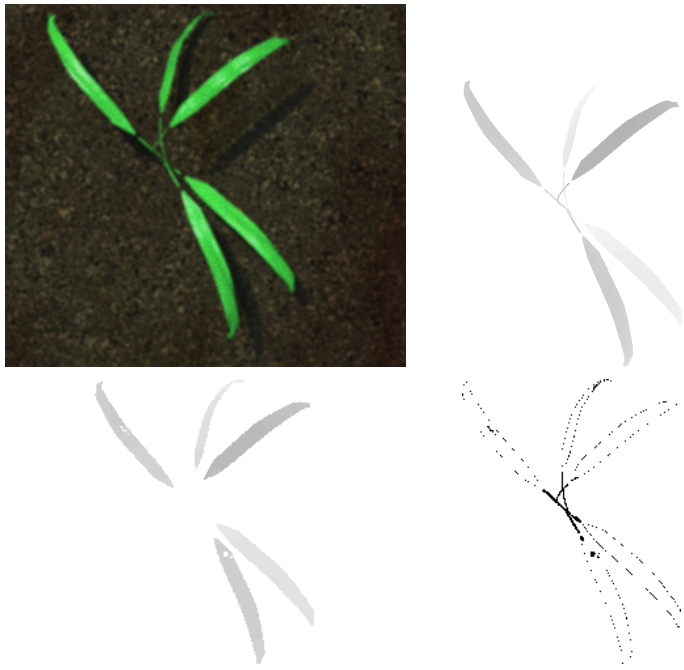


Virtual Plant No. 6

Figure B.3: Virtual Plant No. 5-6.



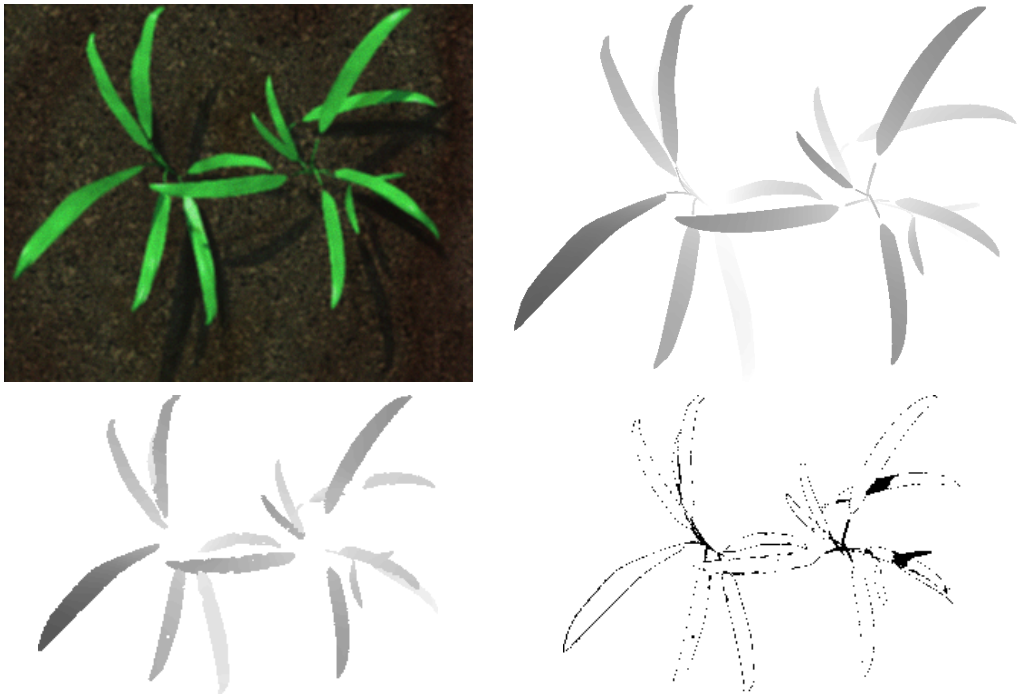
Virtual Plant No. 7



Virtual Plant No. 8

Figure B.4: Virtual Plant No. 7-8.

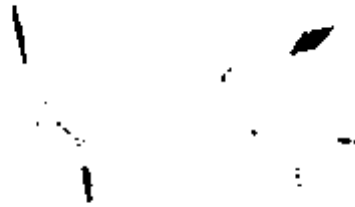
Virtual Plant No. 9



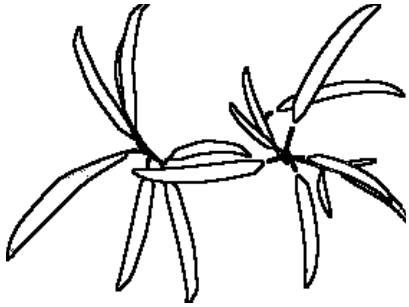
Virtual Plant No. 10

Figure B.5: Virtual Plant No. 9-10.

Occlusion mask t=0



Occlusion mask t=1



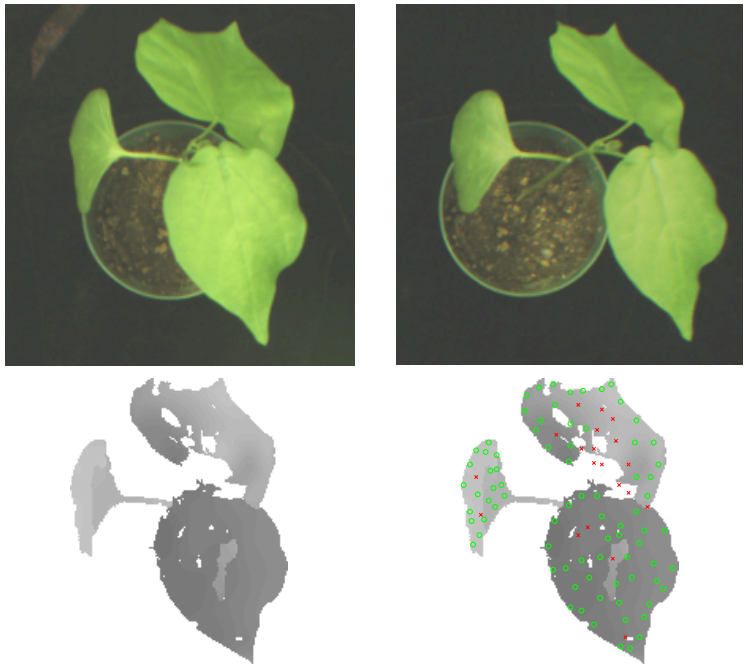
Discontinuity mask t=2

Mut. Ex. Highlight mask t=5

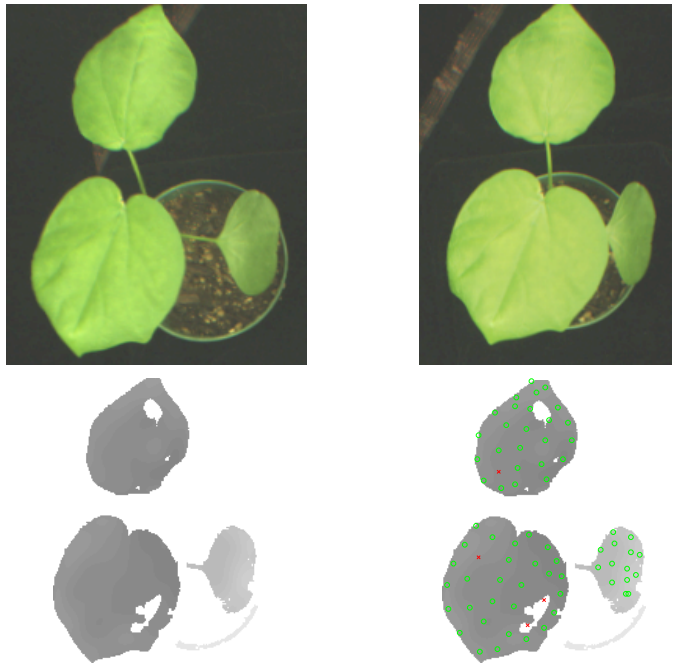
Figure B.6: Virtual Plant No. 10 masks.

B.2 Real Plant Set

The real image set consists of cotton, day lily, and hypoestes plants. The cells in this section are right and left images, the disparity maps calculated using the combined SMW algorithm with NURBS, and the same showing which points was annotated. The annotation points are marked with an x if the pixel is an error and a o if it is correct.

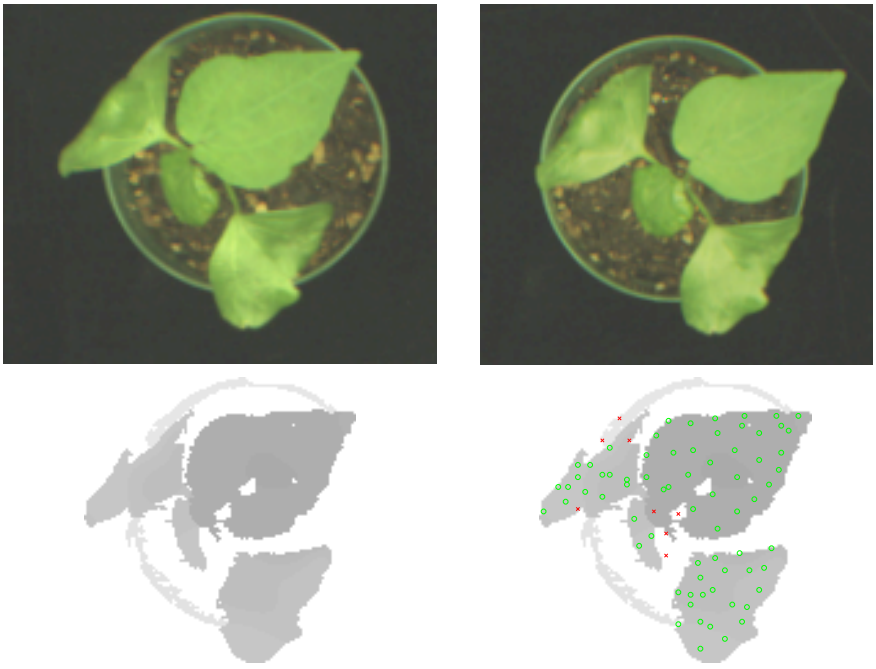


Real Plant No. 1

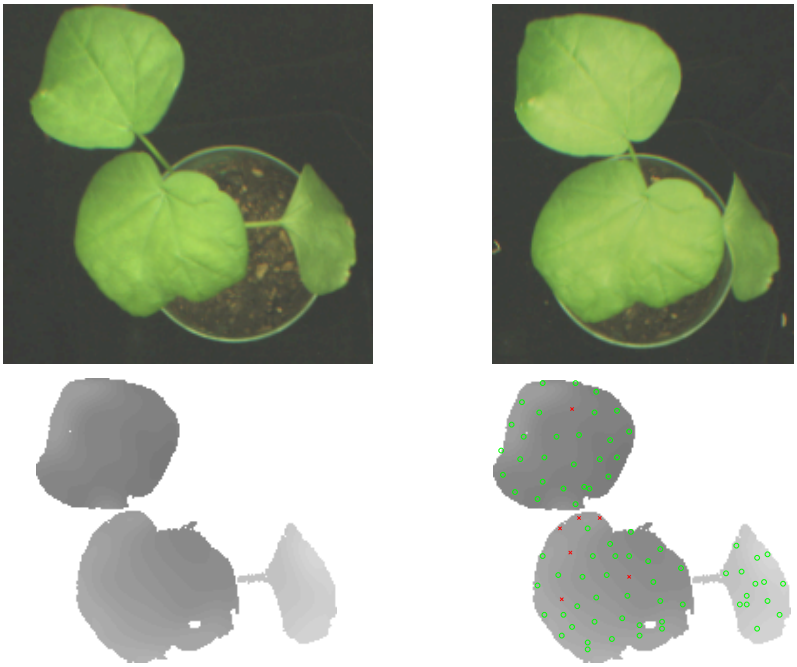


Real Plant No. 2

Figure B.7: Real Plant No. 1-2. Cotton.

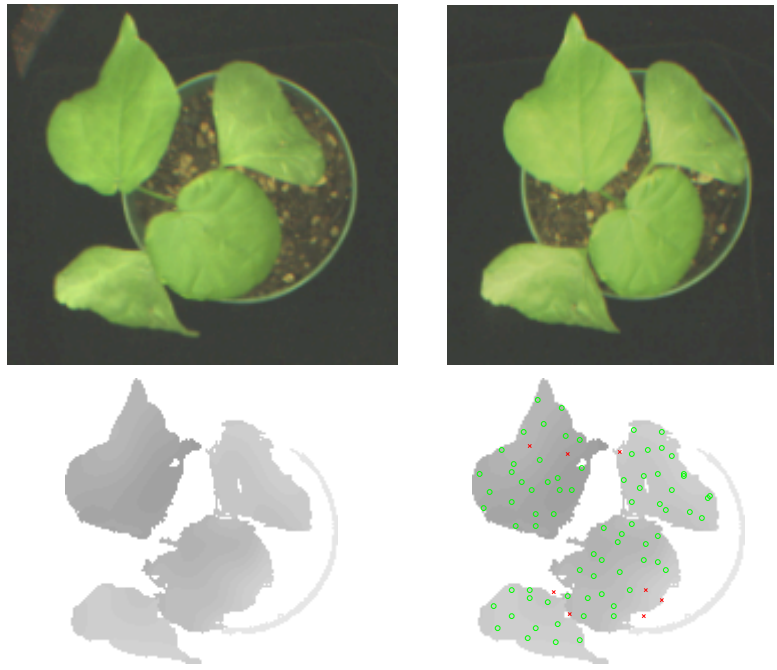


Real Plant No. 3

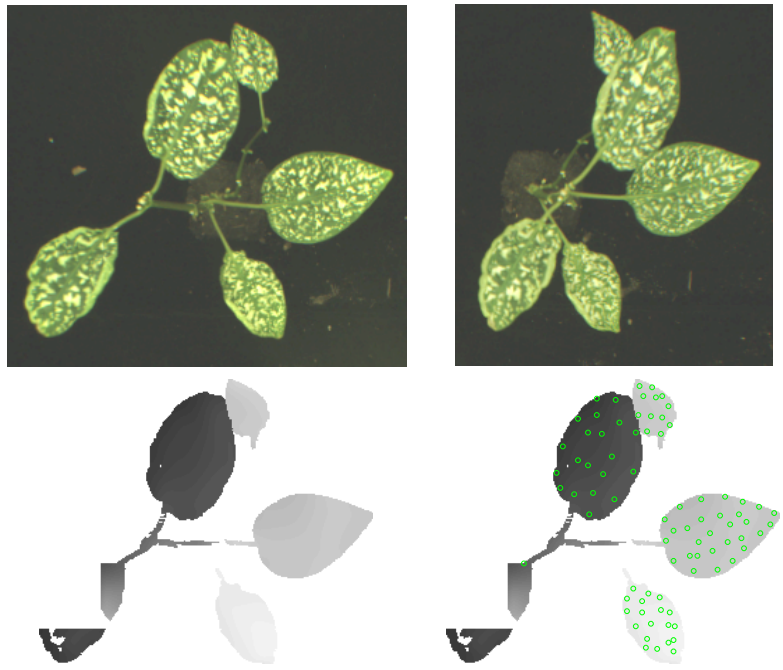


Real Plant No. 4

Figure B.8: Real Plant No. 3-4. Cotton.

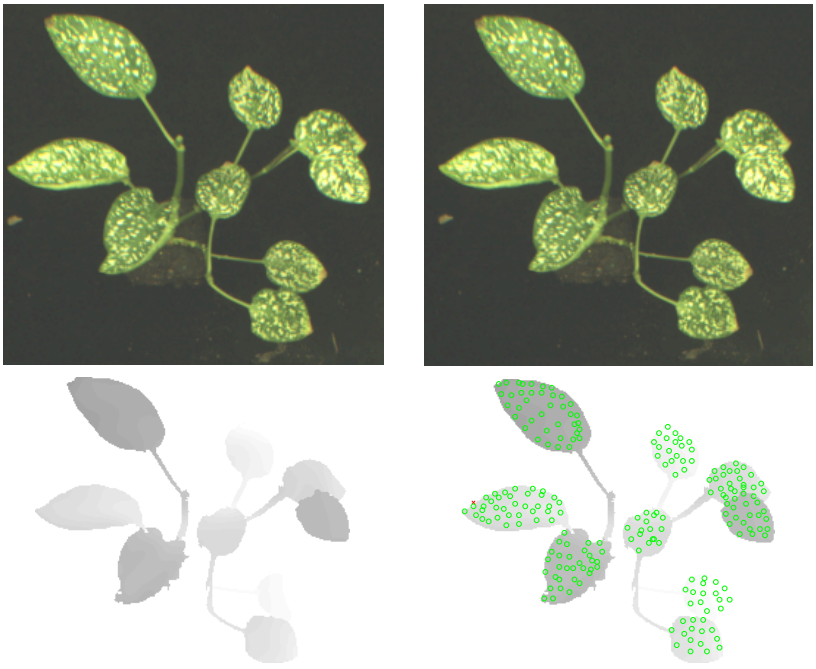


Real Plant No. 5

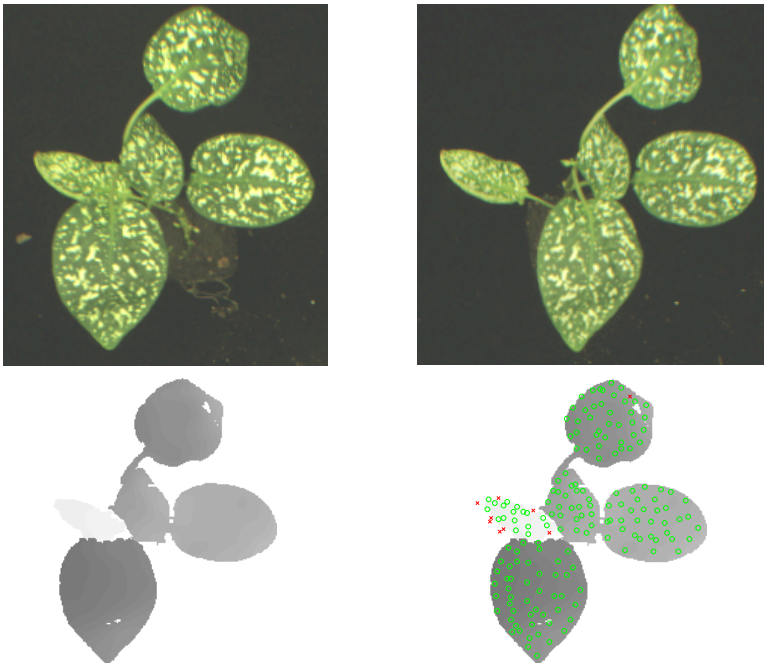


Real Plant No. 6

Figure B.9: Real Plant No. 5-6. Cotton/Hypoestes.

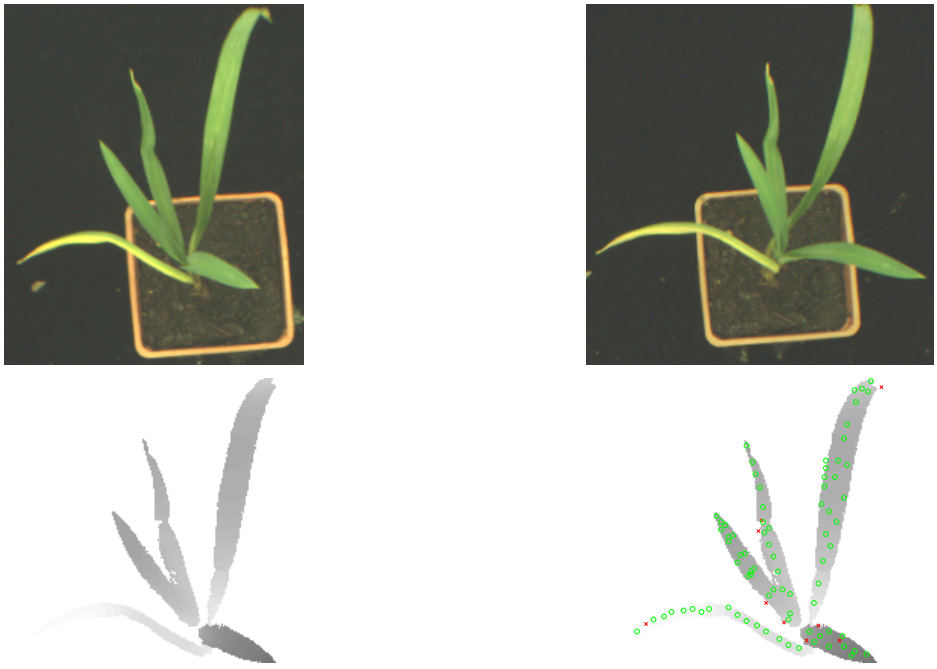


Real Plant No. 7

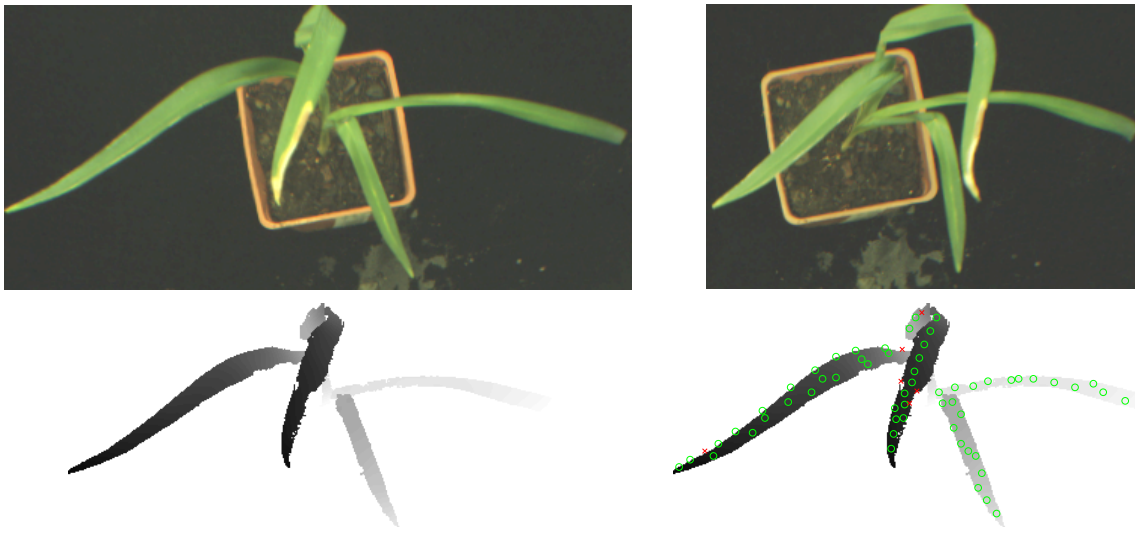


Real Plant No. 8

Figure B.10: Real Plant No. 7-8. Hypoestes.

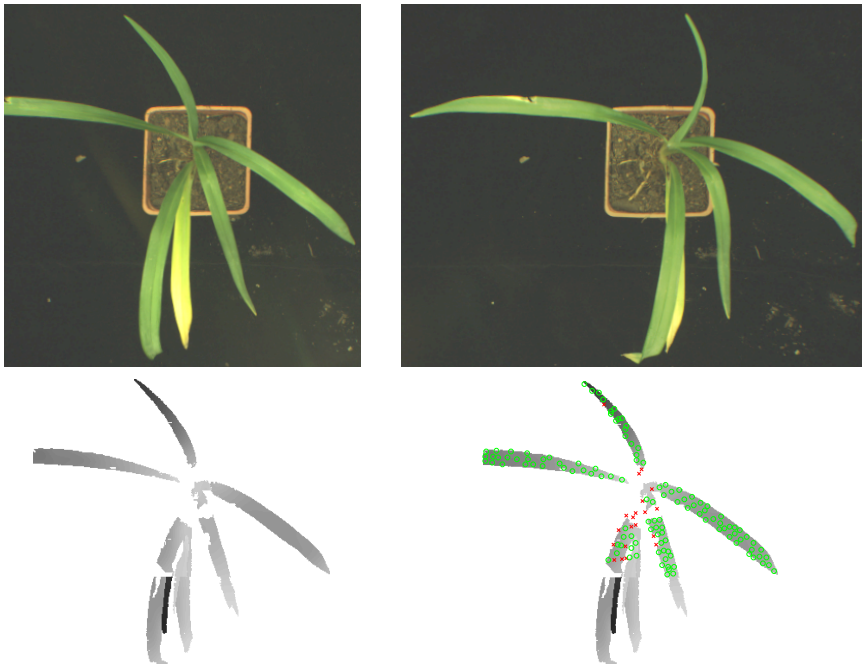


Real Plant No. 9



Real Plant No. 10

Figure B.11: Real Plant No. 9-10. Day lily.



Real Plant No. 11

Figure B.12: Real Plant No. 11. Day lily.

Appendix C

Another Construction of a Graph

This appendix shows another example how to construct a graph for $E^{i,j}$ in an energy minimization. This construction was used on Kolmogorov's graph cut implementation of the method in [Kolmogorov and Zabih, 2004].

In order to build the graph, an energy matrix is built and broken down into a sum of terms that each can be represented by single edges. A is a constant that is added to E_K. The second term depends on variable i, the third term depends on variable j, and the last term depends on both $i = 0, j = 1$. Note that this is just one separation out of many possible ways.

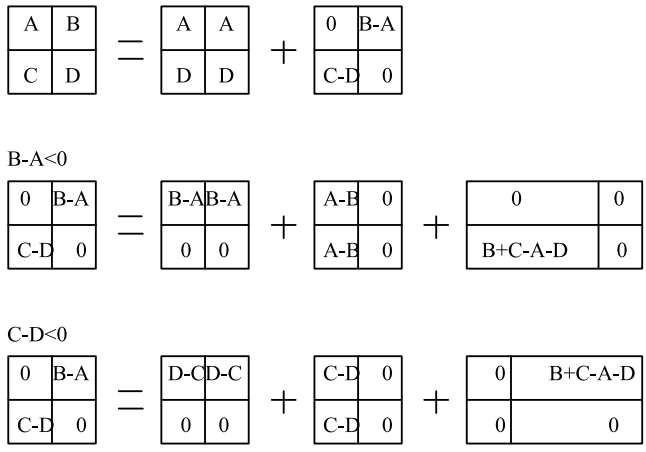


Figure C.1: In order to build the graph, an energy matrix is built and broken down into a sum of terms that each can be represented by single edges. A is a constant that is added to E_K.

Figure C.2 shows how to construct the subgraphs between one or two variables once the

energies are computed. Consider the data term depending only on one variable.

Note that the data cost for keeping the label as it was (A) is not assigned to the edge that goes from the source to the node, but the opposite side. The same goes for (the data cost for assigning the new label). For example if A is larger than B, then the node should be connected to the sink after the cut. Thus, the larger energy should be assigned to the edge going to the sink. This makes it cheaper to cut the edge to source. In order to spare the number of edges in the graph, the only largest edge is kept with the smaller capacity subtracted (and added to the constant).

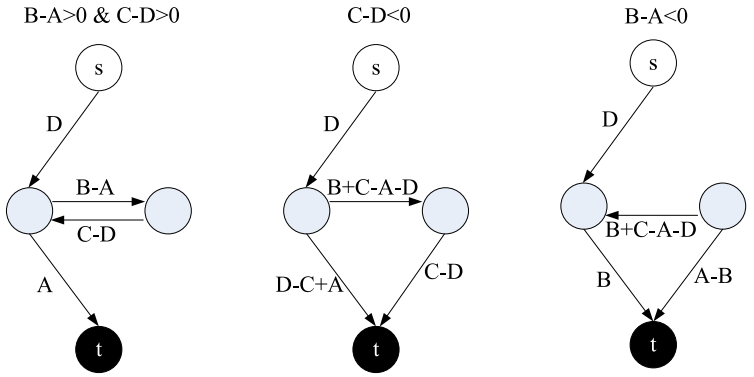


Figure C.2: Graph construction for E^i and $E^{i,j}$. (a) Single variable where $E^i(0) < E^i(1)$, else (b) $E^i(0) \geq E^i(1)$ on the other side. The smallest energy is added to the constant (K). (c) Two variables where $E^{i,j}(0,0) + E^{i,j}(1,1) \leq E^i(1,0) - E^i(0,1)$, and $C - A \geq 0$ and $D - C \geq 0$. (d) if $C - A < 0$ or $D - C < 0$ then the corresponding edge is subtracted on both sides and added to the constant.

It is easy to test whether the graph representation is correct. Figure C.3. The sum of the edges that are cut plus the constant should equal the energy of assigning the configuration to 0 (source) and 1 (sink). The fat arrows are those edges that are cut (dashed lines). Note that directional edge between the nodes is only active when the cut assigns node i to source and j to sink. Alternatively, there is no flow between them and it is not necessary to cut the edge to separate the sink from the source.

C.1 Practical example

Figures C.4 through C.8 will give an example of a practical case with 4 nodes and two labels (a single α-expansion).

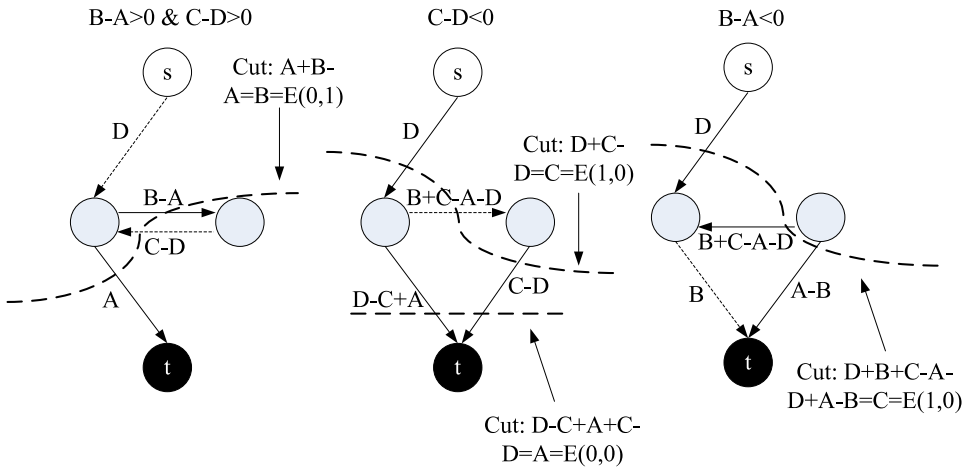


Figure C.3: The sum of the edges that are cut plus the constant should equal the energy of assigning the configuration to 0 (source) and 1 (sink). The fat arrows are those edges that are cut (dashed lines). Note that directional edge between the nodes is only active when the cut assigns node i to source and j to sink. Alternatively, there is no flow between them and it is not necessary to cut the edge to separate the sink from the source.

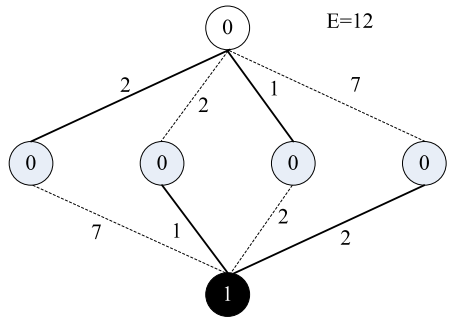


Figure C.4: Practical example with two label 0 and 1. The initialization is all zeros. The data term is given on the lines from the nodes to the labels 0 and 1. The initial energy is 12.

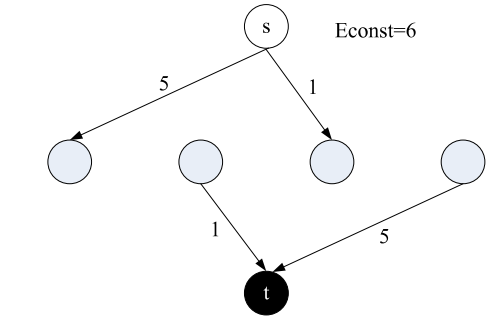


Figure C.5: A graph is constructed from the data term given the rules in figure 7.4A and 7.4B.

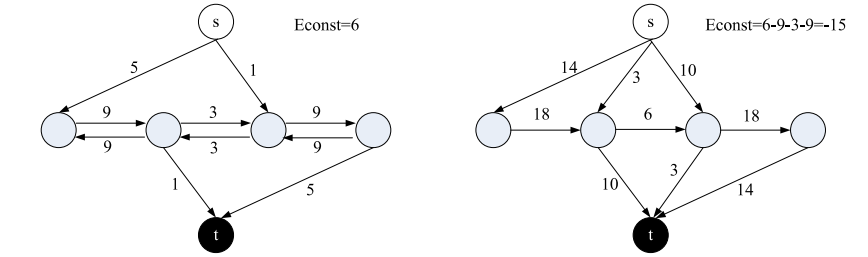


Figure C.6: A smoothness term is added to remove noise. Pott's energy is used with static edge clues, $\lambda = 3$. $E(0,0) = E(1,1) = 0, E(1,0) = E(0,1) = 3$ or 9. There's an edge between nodes 2 and 3. [Left] using rules from figure C.2. This method does not add anything to the constant. [Right] Rules from figure 7.4.

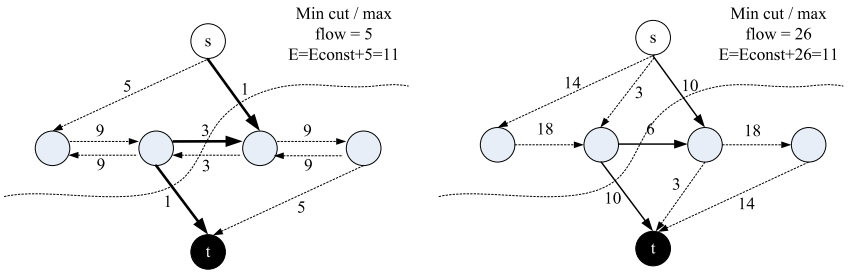


Figure C.7: The minimal cuts are at the same place but with different flows, but the total energies are the same.

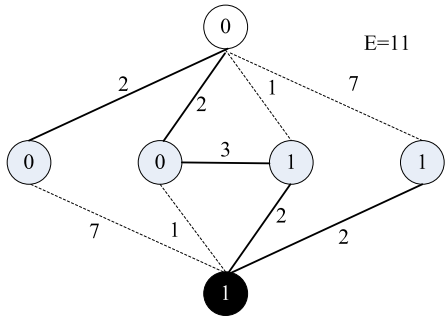


Figure C.8: The best configuration is now 0,0,1,1 with the energy 11.

Media Technology
Aalborg University
Niels Jernes Vej 14
DK-9220 Aalborg
Denmark

Telephone: +45 9940 8793
TeleFax: +45 9940 9788

URL: http://www.create.aau.dk

ISBN 978-87-992732-3-2