Master's Thesis

Rigid Body Reconstruction for Motion Analysis of Giant Honeybees Using Stereo Vision

Michael Maurer, BSc
maurer@icg.tugraz.at

Institute for Computer Graphics and Vision (ICG)
Graz University of Technology
Inffeldgasse 16
8010 Graz, Austria

Supervisor I: Univ.-Prof. Dipl.-Ing. Dr.techn. Horst Bischof
Supervisor II: Dipl.-Ing. Dr.techn. Matthias Rüther

February, 2010

TU Graz
Graz University of Technology
Senat

Deutsche Fassung:
Beschluss der Curricula-Kommission für Bachelor-, Master- und Diplomstudien vom 10.11.2008
Genehmigung des Senates am 1.12.2008

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommene Stellen als solche kenntlich gemacht habe.

Graz, am …………………………… ………………………………………………..
(Unterschrift)

Englische Fassung:

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

…………………………… ………………………………………………..
date (signature)

Abstract

Zoologists are interested in the rapid movements of giant honeybees. Especially the movement of all single bees during the defense behavior is of interest. Currently they are only able to measure the movement of a single bee using a laser vibrometer. A single measurement dose not provide any information on speed, intensity and the starting point of a wave. They are interested in a sensor that enables a 3D reconstruction of the individuals while performing a defense wave. In order to solve this problem, a vision based measurement system is proposed. A portable stereo setup using two high resolution cameras with high frame rates is designed in this thesis to acquire the image sequences of the defense wave in an outdoor environment. The functionality of the acquisition setup has also been proven at an expedition to Nepal. Additionally, a framework to segment and reconstruct the single bees is presented. For the segmentation three different methods are proposed and evaluated. The correspondence problem is faced using reduced graph cuts to get accurate matches in the presence of repetitive patterns. The evaluation has been done by comparison to manually labeled data.

Keywords: stereo reconstruction, NCC, reduced graph cuts, maximum flow, correspondence problem, MSER, shape prior segmentation, rigid body, giant honeybees, apis dorsata, shimering, defense waving

Kurzfassung

Zoologen wollen das Abwehrverhalten der Riesenhonigbienen untersuchen. Dabei soll das Bewegungsmuster der einzelnen Bienen während einer Verteidigungswelle analysiert werden. Derzeit ist es nur möglich die Bewegung einer Biene mittels eines Laservibrometers zu messen. Aus einer Einzelmessung ist jedoch keine Aussage über die Geschwindigkeit, die Intensität und den Startpunkt möglich. Daher ist man an einem Aufnahmesystem interessiert, das eine 3D Rekonstruktion der einzelnen Bienen während der Verteidigungswelle liefert. Um diese Aufgabe zu lösen, wird ein Vision basiertes Messsystem vorgestellt. Zur Aufnahme der Verteidigungswellen im Freien wurde ein portables Stereo-System entwickelt, das zwei hochauflösende Kameras mit hoher Bildwiederholrate besitzt. Die Funktionalität des Aufnahmesystems wurde bei einer Nepal Expedition gezeigt. Zusätzlich wird ein Softwarepaket zur Segmentation und Rekonstruktion der Einzelbienen vorgestellt. Für die Segmentierung werden drei verschiedene Methoden präsentiert und evaluiert. Zum Auffinden von Korrespondenzen wird ein reduzierter Graph eingesetzt, dessen Verhalten mittels manuell annotierten Korrespondenzen überprüft wird.

Stichwörter: Stereorekonstruktion, normalisierte Kreuzkorrelation, Graphentheorie, maximaler Fluss, Korrespondenzen, MSER, Rigide Körper, Riesenhonigbienen, apis dorsata, Verteidigungswellen

Acknowledgments

First of all I would like to thank Prof. Dipl.-Ing. Dr.techn. Horst Bischof who arouse my interest on computer vision in his lectures and his support. Special thanks go to Dipl.-Ing. Dr.techn. Matthias Rüther who gave advice to my thesis, supported me with suggestions to reach the goal and proofreading my master's thesis. Further on I want to thank Ao.Univ.-Prof. Mag. Dr.phil. Gerald Kastberger for enabling this project, the good collaboration and the pleasant atmosphere during the experiments in Nepal.

I would like to thank my family for their ever-present love, mental and financial support and their trust.

Special thanks go to the robot vision community at the ICG, especially Christian, Jakob, Katrin, Markus, Martin and Matthias for the pleasant working environment, the rich in content discussions and them becoming friends.

Further, I want to mention the team of the Nepal expedition: Frank, Gerald, Helmut, Ilse, Julia, Klaus, Madhu and Thomas. Many thanks for the fun we had in-between the work and the enjoyable time.

Finally, I would like to give thanks to all my friends for accompanying me on my years of study, the social gatherings, the shared leisure time activities and their friendship.

Contents

List of Figures

List of Tables

Chapter 1

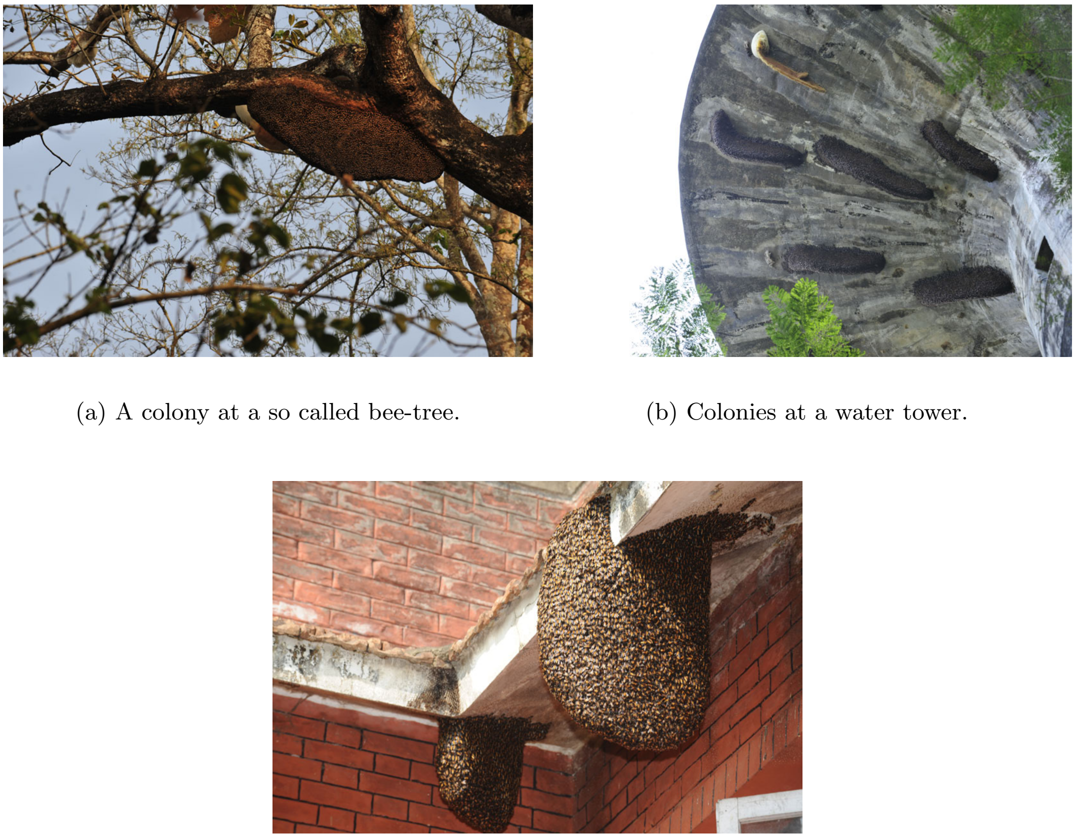
Introduction

Contents

This thesis proposes a hardware setup for acquiring image sequences in outdoor environments and a framework to reconstruct the three-dimensional structure of the defense waving of giant honeybees for motion analysis.

1.1 Motivation

Giant honeybees (lat. apis dorsata) belong to the species of honeybees and live in the mainly forested areas of southern and southeastern Asia like Nepal. The subspecies apis dorsata dorsata that was observed in this project has the second largest individuals among all honeybees. The typical worker bees are about 17 to 20mm long. A colony of giant honeybees consists of up to 100,000 individuals that live in an open-nest. This means that the nest is made up of a single comb with a dimension of up to two meters in horizontal span and about one meter in vertical span. The nests are usually built in exposed places far off the ground. These places are for example overhanging cliffs, trees or buildings (see Figure 1.1). The adult bees cover the comb in multiple layers. This covering is done loosely fixed to the comb and in literature it is terminated as bee curtain. The bee curtain is parted in two regions. One shows locomotion and the bees there are responsible for food, water and the larvae. The other one shows almost no locomotion. There the bees are mostly uniformly oriented with the heads up. They form the integument that is responsible for nest climate, humidity and for protection against wind, rain and sun. Additionally to these functions they can be recruited for defense purpose. Giant honeybees have several methods of defense including an organized mass attack where hundreds of bees attack the intruder and the defense waving also called shimmering behavior. It can be described as social waves that slide over the bee curtain at a speed of about one meter per second. The waves are generated by coordinated rising of the



(a) A colony at a so called bee-tree.

(b) Colonies at a water tower.

(c) Colony at a building.

Figure 1.1: Places used by giant honeybees to build their nests.

abdomen of neighboring bees and can be observed as dark waves on a video (see Figure 1.2). The reasons for defense waving are wasps, birds or honey buzzards that approach the nest. [Kas], [Wik].

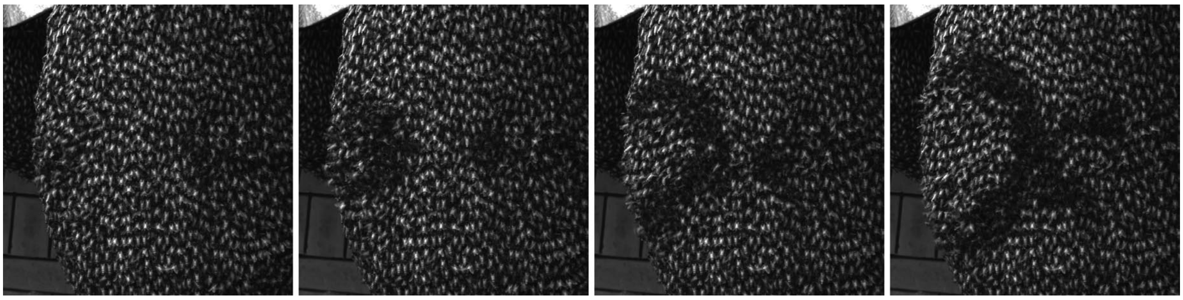


Figure 1.2: Bee cluster performing a wave by synchronized rising of their abdomen. The frames were captured with a rate of $12fps$.

From the viewpoint of zoologists research on the global structure of the nest are quite interesting. These are:

- measurements of the mesh width,
- the movement of the wave in space,
- the speed in the different directions and
- the participation of second layer bees that are bees under the surface of the curtain.

Additionally they are looking for the number of active and passive bees. Active bees are those that rise their abdomen to generate the wave. Also the age of the participating bees is of interest and can be obtained from the coloring of the abdomen. Technically to be able to solve these requirements a segmentation of the single bees and a three-dimensional (3D) reconstruction of the bee curtain structure is required. Having the segmentation of the bees the age determination is possible by analyzing the coloring of the abdomen. Also the number of single bees is solved by the number of segmented regions. Having a 3D reconstruction of the bees over time the movement of the bees can be evaluated. This evaluation includes the participation of second layer bees that can be observed by global movements of the curtain. Further, the mesh width can be calculated having the 3D reconstruction.

The challenge to be able to generate the 3D reconstruction are the field of view defined by the dimensions of the curtain, the speed the waves are sliding on the curtain and the fact that the bees can only be observed in a passive way. An active observation would disturb the bees and force a defense attack. So stereo reconstruction is a suitable method. Reasons are the field of view that can be observed with stereo cameras, the one shot principle that enables the treatment of the wave speed and the passiveness of the stereo setup. It only monitors the bees without any extra illumination or pattern projections. The challenge for the construction of such a stereo setup are the environmental conditions, the location of the nests, speed and the occurrence of repetitive patterns. A stereo setup satisfying these requirements is constructed and evaluated in this master's thesis. Further, a method to reconstruct the bees has been developed and evaluated.

1.2 Overview

The master's thesis is structured as follows. In Chapter 2 an overview of the related work is given containing methods used to track insects, followed by segmentation methods for rigid objects and stereo reconstruction methods focused on finding correspondences. Chapter 3 describes the problem of data acquisition which contains the sensor design and calibration. Possible methods to segment the bees are summarized in Chapter 4. There the focus is on maximally stable extremal regions, shape prior segmentation and template matching. Chapter 5 deals with stereo reconstruction and can be parted in stereo matching using reduced graph cuts and triangulation. The theoretical part is followed by experiments and evaluations of the methods in Chapter 6 and discussion and outlook in Chapter 7.

Chapter 2

Related Work

Contents

The focus of this thesis lies on the 3D reconstruction of the defense waving of giant honeybees. To perform the reconstruction the bees have to be segmented, matched and triangulated. First methods dealing with tracking of insects are summarized in Section 2.1. Then a selection of segmentation methods for rigid objects is provided in Section 2.2. Followed in Section 2.3 with a selection of stereo reconstruction methods.

2.1 Tracking of Insects

As this thesis leads to track the single bees performing defense waves previous papers dealing with tracking of insects are of interest.

Veeraraghavan et al. [VCS08] presented a shape and behavior encoded tracking framework to track and simultaneously analyze the movement of a bee. With this framework it is possible to analyze complex behaviors that are modeled using a three-tier hierarchical motion model. The first tier models the local motions and they act as a vocabulary for behavior modeling. This behavior is modeled in the second tier by a Markov motion model. The third tier models the switching between single behavior using a Markov model. Figure 2.1 shows a bee performing a waggle dance and the according behavioral model.

Maitra et al. [MSS09] extended the framework of [VCS08] and created a robust bee tracker with adaptive appearance template and geometry constrained resampling. They use static and adaptive appearance templates. The static appearance template is responsible to prevent from mis-tracking and the adaptive one handles appearance changes. Further, they

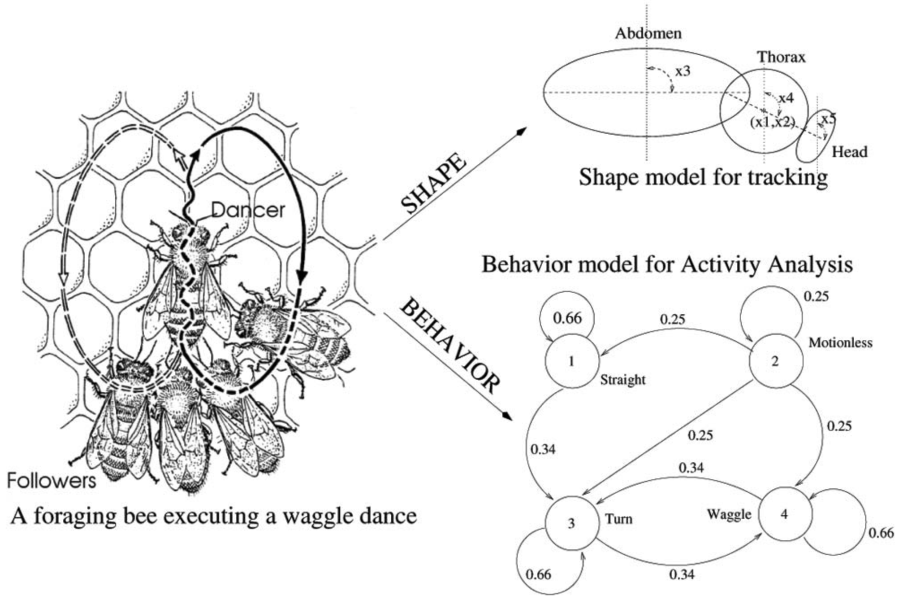


Figure 2.1: A bee performing a waggle dance and the behavioral model for the waggle dance [VCS08].

added to the resampling step of the particle filter a constraint to the geometry to improve the prediction of unreliable parameters of head and thorax orientations. Their benefit is a more stable tracking as shown in Figure 2.2. There they compare the method of [VCS08] (top row) with their own one (bottom row).

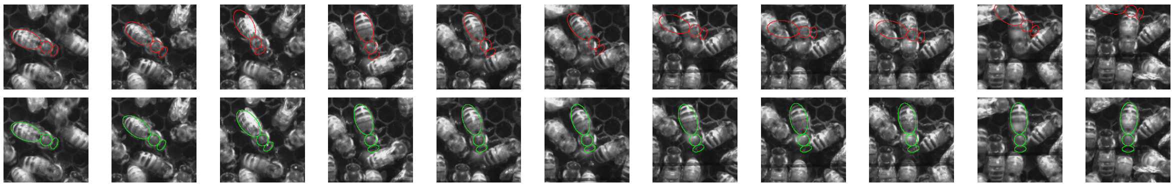


Figure 2.2: Comparison of the PF-Gaussian (top row) and the proposed method (bottom row) taken from [MSS09]. Note the improvement in tracking under appearance change (3rd column) and the unreliable head feature (5th column). Then the error quickly accumulates.

In general biologists use image processing and analysis software as Image Pro [Med], Optimas [IG] or ImageJ [oH] to extract visual information of the acquired image.

2.2 Segmentation of Rigid Objects

In computer vision segmentation is a widely discussed topic and so there exist numerous different methods. A summary of all segmentation methods would go beyond the scope of this thesis. Here, methods suitable for the input data as shown in Figure 2.3 are discussed.

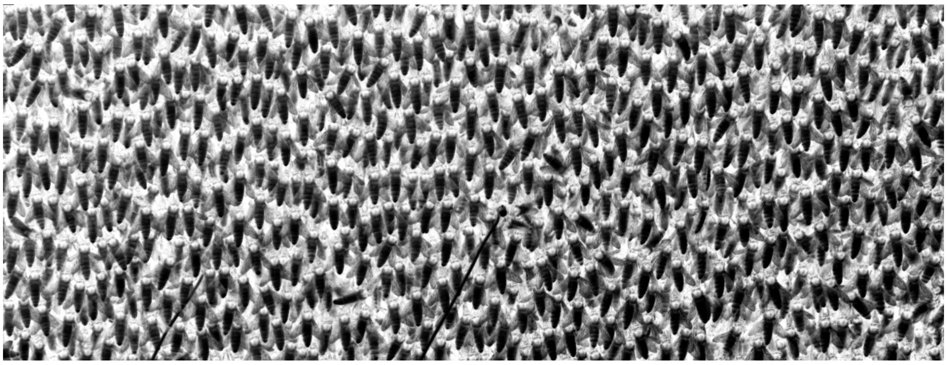


Figure 2.3: Cutout of a giant honeybee nest.

These are maximally stable extremal regions (MSER), shape prior segmentation (SPSeg) and normalized cross correlation (NCC). The motivation of taking MSERs is the dark regions that represent the abdomen of the bees. The fact that the abdomen of the bees are almost the same shape leads to the selection of shape priors to get an accurate segmentation of the abdomen's boundary. Finally taking the pattern of a bee into account the bees can be segmented using an image patch and calculating the correlation. So NCC is selected because of its invariance to illumination.

An introduction to image segmentation can be found in Gonzalez et al. [GW02]. There the detection of discontinuities, thresholding, region-based segmentation, segmentation by morphological watersheds and using motion in segmentation are described. Further, Sonka et al. [SHB99a] gives an overview of segmentation methods.

Maximally Stable Extremal Regions (MSER) are introduced by Matas et al. [MCUP02]. They show that this regions are closed under continuous transformation of image coordinates and that this regions are closed under monotonic transformation of image intensities. So MSERs can be used for wide baseline stereo as demonstrated in their experimental section.

Donoser et al. [DB06] show that using a component tree it is possible to calculate MSERs in quasi-linear time. Further, they show that using MSER tracking it is possible to improve the detection of single MSERs by a factor of 4 to 10. The tracking stability is also improved by weighted features and by using backward tracking that is possible under the use of component trees. So a novel MSER tracking algorithm is formulated by them.

In [Ved07] by Vedaldi an implementation of multi-dimensional MSER is presented that can be used for tracking in video sequences by directly extracting the 3D region from a stack of frames.

Werlberger et al. [WPUB09] present a variational model for interactive shape prior segmentation. Starting with a manually adjusted position of the object they perform a local optimization routine by transforming the shape prior. Therefor the variational formulation of the Geodesic Active Contour energy is used and minimization is done with a fast primal-dual approach. The framework they present performs a local optimization of the shape prior to get a accurate segmentation of the object. This approach can also be used for tracking objects in videos or segment multiple objects with a single shape prior as it can be seen in Figure 2.4.

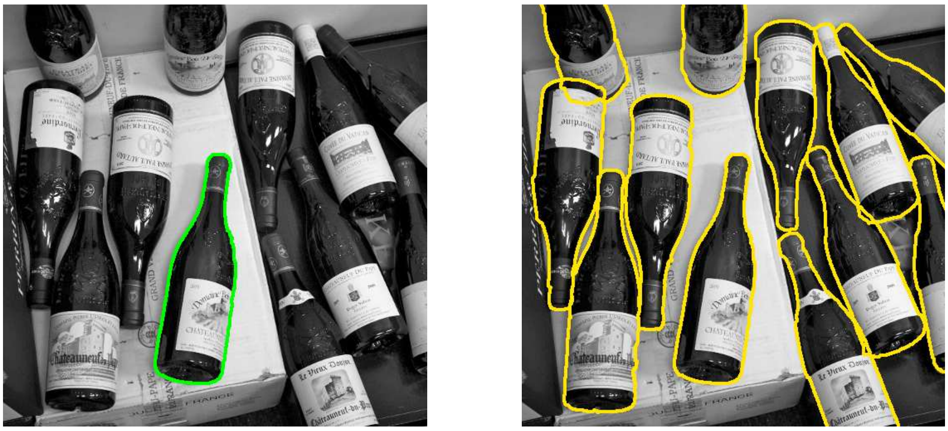


Figure 2.4: Segmentation of bottles with a single shape prior [WPUB09]

Matching is another method to segment a known object in an image. Using a criterion of optimality the best match can be figured out. Matching is often based on directly comparing gray-level properties as criterion. Further, correlations up to complex approaches of graph matching can be used. [SHB99b] So normalized cross correlation can be used as optimality criterion.

Lewis [Lew95] shows a method to efficiently calculate normalized cross correlation by using precomputed integral image and image^2 over the search window. The normalized cross correlation is obtained from transform domain convolution and provides a speedup of an order of magnitude compared to spatial domain computation.

2.3 Stereo Reconstruction

Before a reconstruction can be calculated using triangulation the problem of correspondences has to be solved. This is a very extensively studied problem in computer vision and can be solved in many different ways.

A detailed evaluation of dense two-frame stereo correspondence algorithms can be found in Scharstein et al. [SS01].

To find corresponding bees in the stereo frames of the giant honeybee nest the algorithm has to deal with repetitive patterns. Methods that deal with about the same problem are mentioned.

Du et al. [DZC07] developed a method to reconstruct 3D scenes made up of large numbers of dynamic particles that have to be track able. They call their approach Relative Epipolar

Motion (REM) and solve the problem of correspondences in stereopsis by utilizing the motion clue. They are matching feature trajectories instead of the features themselves and so they are able to reconstruct dynamic 3D scenes of large number of indistinguishable drifting particles and are able to establish correspondences for dynamic surfaces made up of repetitive textures. They also offer a method to project structured light in active mode for deforming surface reconstruction.

Zhang et al. [ZCS03] extends the binocular stereo problem into the space-time domain using active illumination. They reduce the ambiguity and increase the accuracy by utilizing both spatial and temporal appearance variation. This is done by simultaneously matching intensities in multiple frames by minimizing a sum of SSD (SSSD). Their framework serves as general structured light framework, can handle natural scenes with repetitive textures and chaotic behaviors, such as waving trees and flowing water and it handles over time moving and deforming objects too.

Kamiya and Kanazawa [KK08] created a method to match scenes with repetitive patterns as buildings or walls. The matching is done in two phases. The first phase deals with the repetitive patterns. There they detect the elements of repetitive patterns and divide them into regions whose correspondences are coarsely estimated by RANSAC. Afterwards features in the repetitive regions are matched using the obtained rough correspondences. The second phase consists of the matching in the remained regions by feature matching using the epipolar constraint calculated out of the point matches of the repetitive pattern regions. Results of their method can be seen in Figure 2.5.

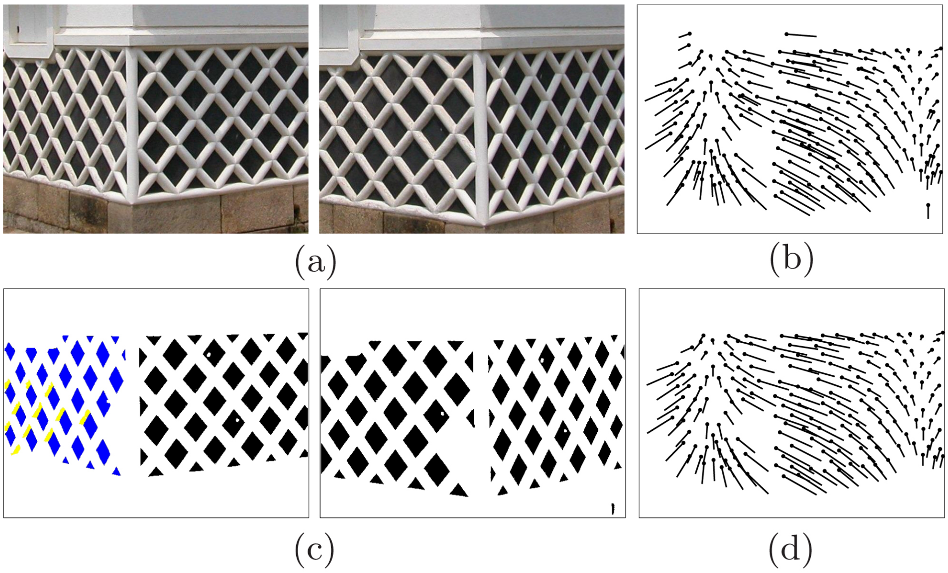


Figure 2.5: Real image example taken from [KK08]. (a) Original images. (b) Result. (c) Region grouping results. (d) Obtained matches from repetitive pattern regions.

As described by Matas et al. [MCUP02] it is possible to use MSER as a feature for robust wide baseline stereo. To get an accurate estimation of the epipolar geometry of the scene, first the detected MSERs are robustly matched by voting, followed by tentative correspondences using correlation. First, a coarse approximation of the epipolar geometry is calculated by the centers of gravity of distinctive regions. These regions have a higher correspondence at their affine normalized image patch than a threshold. The final accurate estimation of the epipolar geometry is robustly calculated with the centers of the convex hulls of the distinctive regions that were inliers of the rough estimation. The resulting epipolar geometry is depicted in Figure 2.6.

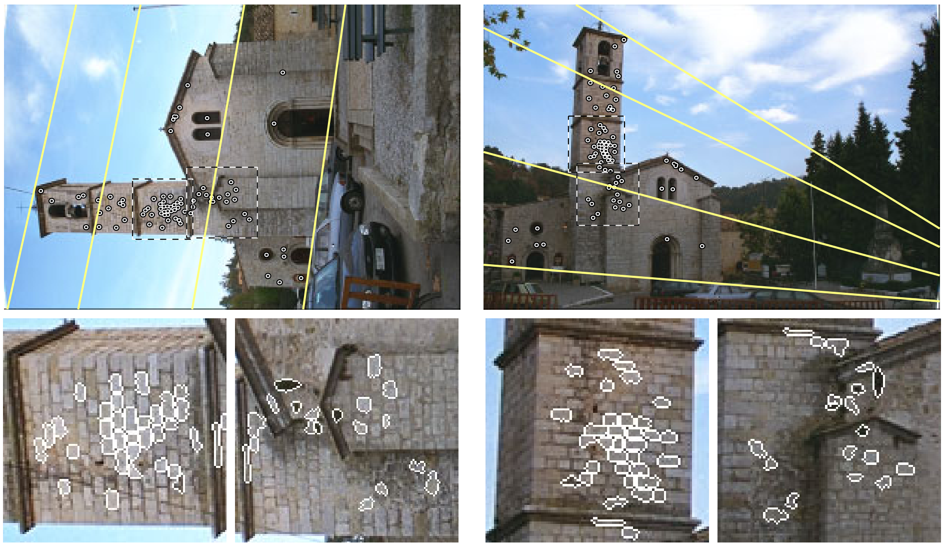
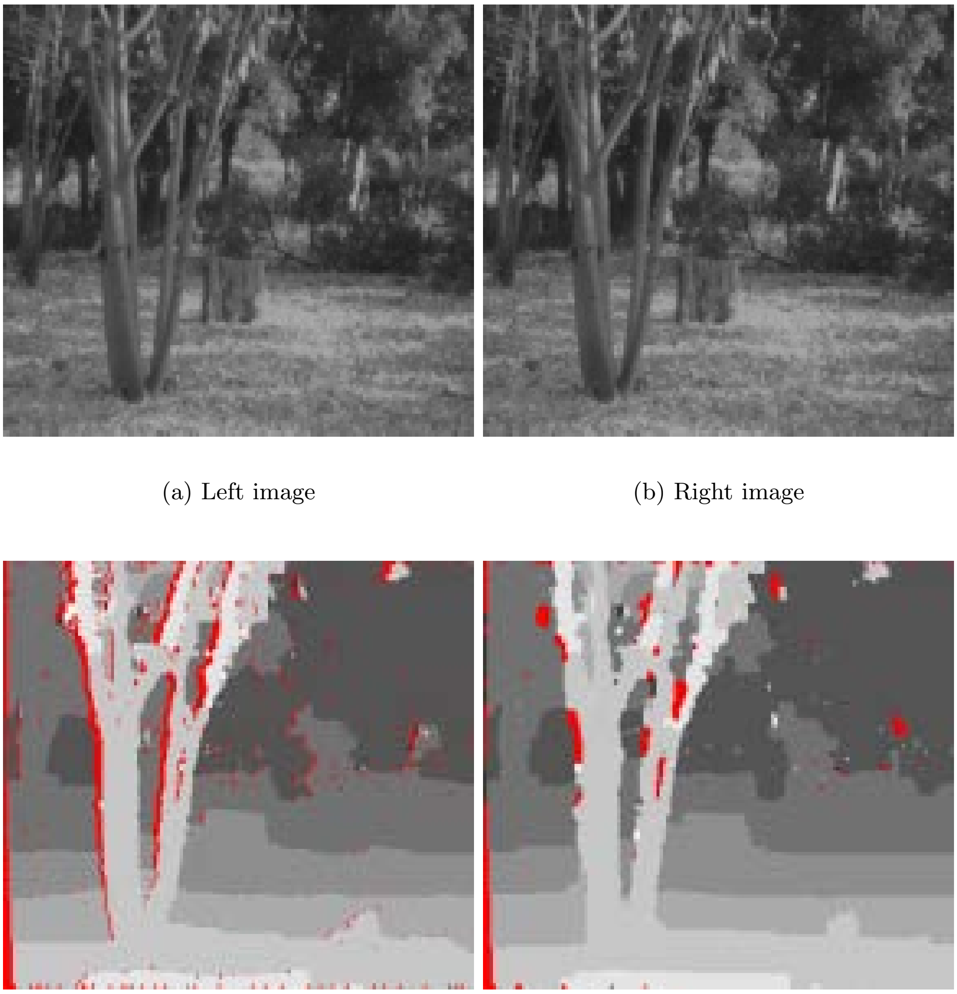


Figure 2.6: VALBONNE: Estimated epipolar geometry and points associated to the matched regions. The image is taken from [MCUP02].

The correspondance problem can also be solved using graph cuts. There, many different versions of energy functions exist that are minimized to get a global minimum or at least an approximation by a strong local one. Such energy functions and the algorithms to solve the minimization are as follows.

Boykov et al. [BVZ99] formulated two algorithms based on graph cuts to efficiently approximate the global minimization of energies which is NP-hard. They introduced two large moves, namely expansion moves and swap moves to find a local minimum. The advantage of these moves compared to many standard algorithms are the opportunity to change arbitrary large sets of pixels simultaneously. They show that the expansion algorithm finds a labeling within a known factor of the global minimum and the swap algorithm is capable to handle more general energies.

In [KZ01a] and [KZ01b] Kolmogorov and Zabih present an algorithm that properly deals with occlusion while computing visual correspondences. The algorithm is based on the expansion move algorithm of [BVZ99] and satisfies the uniqueness constraint. A pixel in one image should correspond to at most one pixel in the other image and pixels corresponding to no other pixels should be labeled as occluded. The algorithm finds a strong local minimum of an energy function and pixels that violate uniqueness are punished with an infinite energy. They gained promising results as it can be seen in Figure 2.7.



(a) Left image

(b) Right image

(c) Horizontal motion (method [KZ01a])

(d) Horizontal motion (method [BVZ99])

Figure 2.7: Stereo results on the SRI tree sequence [KZ01b]. Occluded pixels are shown in red.

Kim et al. [KKZ03] extended the visual correspondence problem using graph cut by adding mutual information to the energy function. Their method does not suffer from the problem of fixed windows, namely poor performance in low textured areas and at discontinuities. It combines the tolerance for intensity changes that comes from the mutual information with the graph cut based energy minimization that preserves discontinuities and handle regions with low texture. The results of their algorithm are comparable to other energy minimizing approaches and outperforms standard correlation-based methods.

Hong and Chen [HC04] formulate the stereo matching problem as an energy minimization problem in the segment domain instead of the pixel domain. Using graph cuts the corresponding disparity planes for the segments will be estimated. They segmented the input images by color. Their method gains performance in textureless regions, disparity discontinuous boundaries and occluded portions. Further, the disparity map is modeled as segmentations and plain models and so it is much compact. But the algorithm can not handle disparity boundaries appearing inside the initial color segments.

Zureiki et al. [ZDC07] present a graph cut based method of finding stereo correspondences using only potential values in the disparity range. This leads to a reduced graph and gives the ability to make the disparity range wider. The values for the graph can be found by local analysis of stereo matching using local similarity measurements like SAD. So the algorithm sensibly ameliorates the quality of the disparity image resulting from only local methods and avoids the combinatorial explosion of graph cuts.

Chapter 3

Data Acquisition

Contents

The aim of the complete measurement process is to get the 3D reconstructions of the movements of the single bees that are placed in the non locomotion area of the bee curtain. Focus lies on the movements during defense waves and there especially on the z-movement (towards the cameras). So the positions of all bees should be measured in 3D at the same time. This measurement should be repeated at a frame rate of 40fps to be able to reconstruct the defense wave over time.

Difficulties arise from the location the giant honeybees build their nests. These are at exposed positions far off the ground as shown in Figure 1.1. Further, the size of an individual compared to the nest is quite small. Next the rapid sliding of the waves over the nest and the even faster flipping of the abdomen put high demands to the frame rate of the capturing device.

Possible non contact optical shape measurement approaches are summarized in Figure 3.1.

A possible acquisition setup to deal with the difficulties just presented is a passive stereo setup as presented in Section 3.1 and 3.2.

The selection of a stereo setup and not structured light or any laser scanning method is motivated by the passive way of the image acquisition. Taking a structured light setup it is required to project patterns onto the nest and a multi shot variant has to be considered for accurate results. This variant requires a still observation object that is contradictory to the

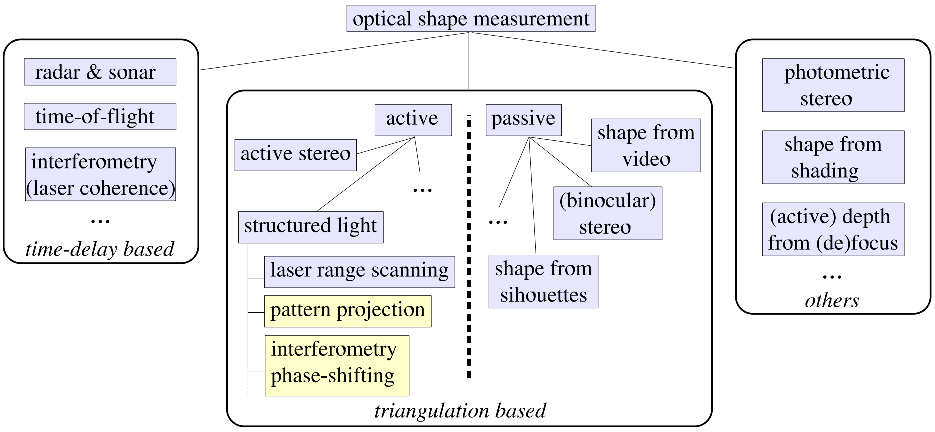


Figure 3.1: Possible non contact optical shape measurement approaches [Kon05]

rapid moving waves that should be observed. Further, by projecting patterns onto the nest the bees will be disturbed in their daily routine and the observation results will be useless. Also a laser scanning method is not able to measure the distances to the single bees at a time and so it is unusable for this acquisition scenario too. A time of flight camera does not have the required resolution to measure each bee. Further, it does not have the required accuracy and the active illumination would disturb the bees.

The sensor design, sensor geometry and the sensor calibration will be described in the following sections of this chapter.

3.1 Sensor Design

The following restrictions govern the design of a suitable stereo setup:

- The bee cluster has a size of about $700 \times 700 \times 100\text{mm}^3$ (height × width × depth), that should be in the measurement range.
- The cameras cannot be mounted arbitrarily close to the bee cluster because the bees will be disturbed by the setup and will attack the intruder. A clearance distance of two meters has to be maintained.
- The size of a single giant honeybee is about 20mm × 6mm (length × diameter). In a single image the diameter of a bee should be at least 10px.
- The reconstruction error in z-direction should not exceed one millimeter.
- An abdomen flip of a bee lasts for about 0.2s and should be resolved with 10 frames.
- The recorded sequence should contain at least one wave that requires about one second to slide over the nest, because it is not possible to predict the start of a wave a longer recording time is required.

- The cameras have to be triggered synchronously.
- The stereo setup has to be placed outdoors, so it has to be portable, flexible and stand outside influences as dust.
- Further, because of the outdoor environment the setup has to be self powered and should be able to operate for one working day.

For capturing the images 4Mpx cameras were selected. The cameras have a resolution of 2352×1728px. According to the constraints 700mm will fit to 1728px one pixel will represent roughly 0.405mm in real world. Because of this a bee with a diameter of 6mm will be represented by about 15px. To resolve an abdomen flip with 10 frames a frame rate of at least 50fps has to be satisfied. The cameras used are able to capture 60fps and so resolve an abdomen flip with 12 frames. Two cameras at a frame rate of 60fps result in a data stream of 480MB/s that has to be stored. So, because a usual hard disc can not cope with such a data stream, the captured image sequence is buffered to RAM and is later written to disc. This leads to RAM constraining the capture time. A total of 8GB of RAM results in an acquisition time of 15s that is capable to capture several waves at a time. To achieve synchronization the cameras are interconnected as a master-slave system. This means that the frame grabber card of one camera acts as the master and generates a trigger signal that externally triggers the second frame grabber. To have a flexible and portable setup the cameras have been mounted on an aluminum rig. This rig enables a positioning of the cameras at different baselines b with variable camera angles α. The calculation of the baseline and camera angle required for the asked measurement accuracy in z-direction will be described in Section 3.2. Finally the setup has to be self powered. This is solved by a 12V PC power supply and a car battery of 100Ah. So an operation time of seven hours can be reached. Figure 3.2 shows the resulting stereo setup mounted at the place of action even to the second floor's ceiling.

3.2 Sensor Geometry

Referring to the geometric constraints a baseline b can be calculated as follows:

$$b = 2 * d * \tan\left(\frac{\alpha}{2}\right) \tag{3.1}$$

where b is the baseline, d the distance of the setup to the bee cluster and α the angle enclosed by the cameras. Figure 3.3 shows the geometric interrelationship of (3.1).

To be able to select the appropriate object lenses the focal length f can be calculated as:

$$f = \frac{wd * size_{img}}{size_{obj} + size_{img}} \tag{3.2}$$

where wd denotes the working distance, $size_{obj}$ the dimension of the object and $size_{img}$ the size of the image, that is calculated as:

$$size_{img} = size_{px} * numPixels \tag{3.3}$$

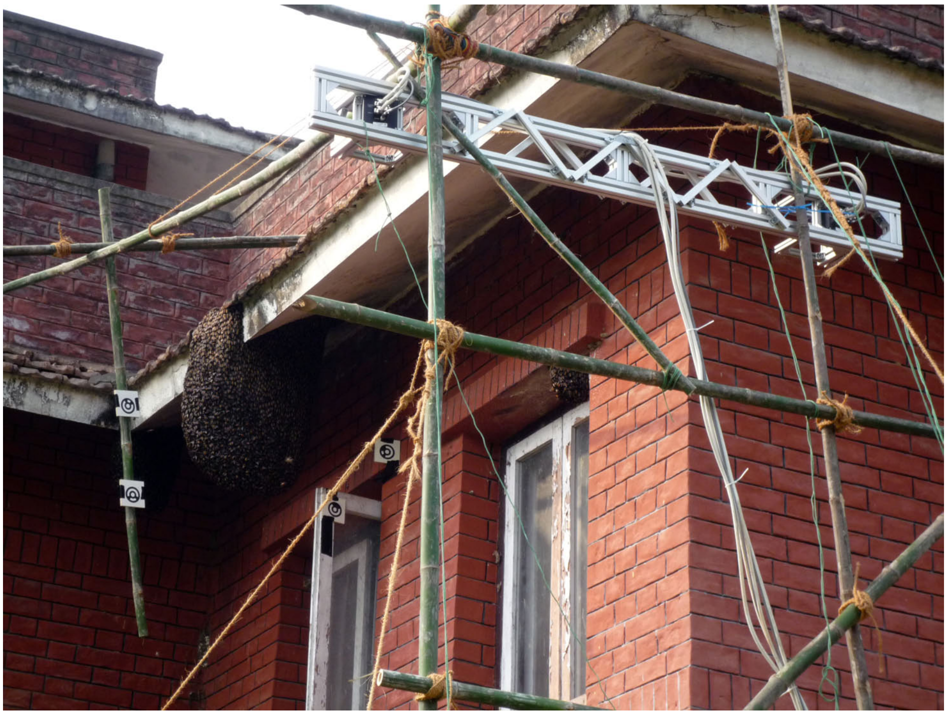


Figure 3.2: The stereo setup placed in front of a bee colony even to the second floor ceiling.

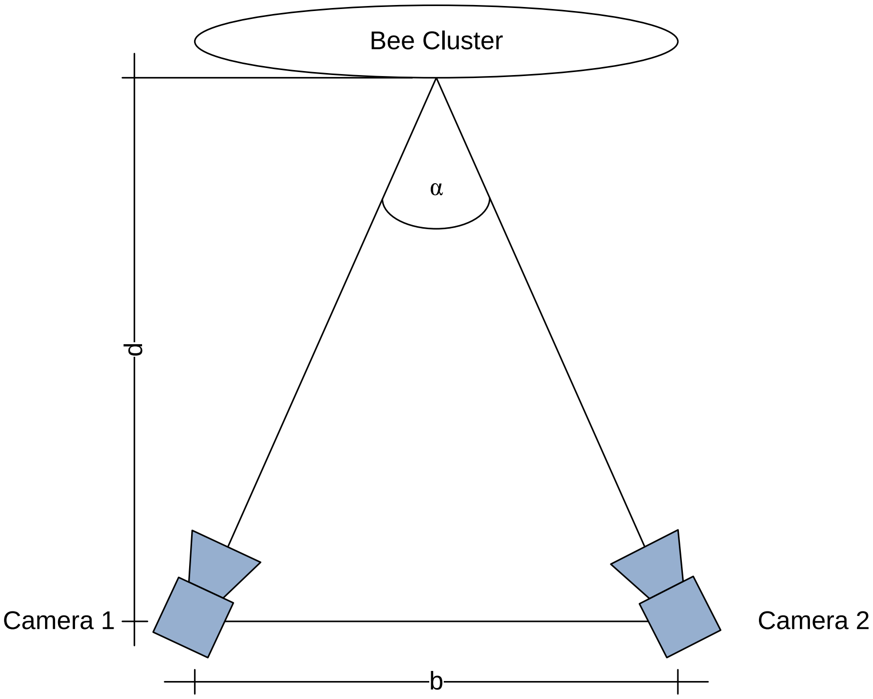


Figure 3.3: Geometric interrelationship for calculating the baseline b as formulated in (3.1). d represents the working distance and α represents the stereo angle.

where $size_{px}$ represents the dimension of a pixel on the camera chip and $numPixels$ the amount of pixels. Figure 3.4 illustrates the required distances.

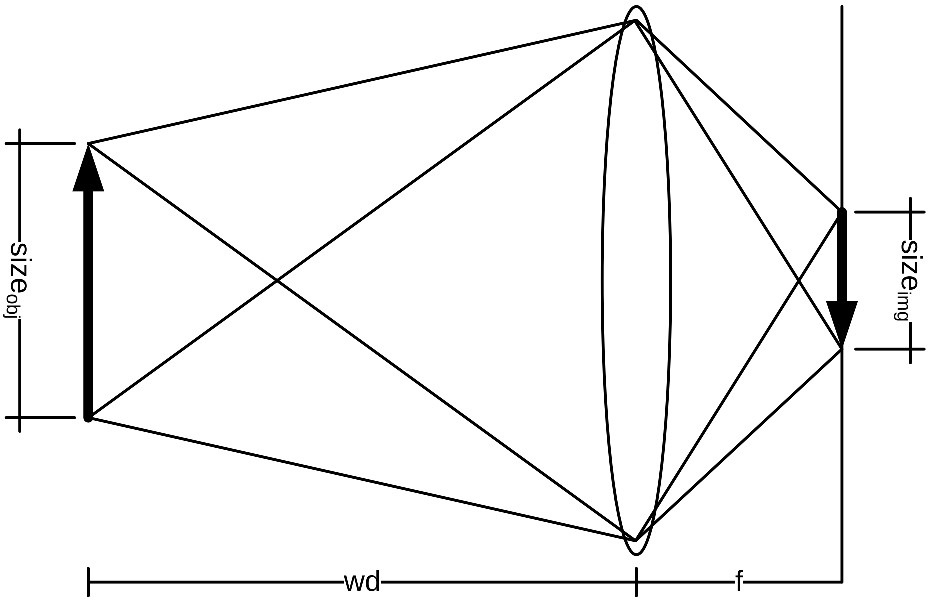


Figure 3.4: Schematic of the required distances for the calculation of the back focal distance f as formulated in (3.2). Where wd represents the working distance, $size_{img}$ the size of the image and $size_{obj}$ the size of the object to display.

The accuracy e_z can be estimated under the assumption of affine cameras as:

$$e_z = e_{px} * m/sin(\frac{\alpha}{2})$$

where e_{px} is the expected pixel error of segmentation α the stereo angle and m the relation of field of view to the number of pixels. To satisfy the required accuracy α has to be at least 23.37°. With the choice of $\alpha = 30°$ an accuracy of $0.78mm$ can be reached.

3.3 Sensor Calibration

To be able to get a proper reconstruction the sensor has to be calibrated. The parameters that result from the calibration process are the intrinsic camera parameters for each camera, the radial distortion parameters κ_1 and κ_2 of the optics and the extrinsic parameters. The intrinsic camera matrix K is composed as:

$$\mathbf{K} = \begin{pmatrix} f & s & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{pmatrix}$$

where f denotes the focus, s the skew and $(p_x \; p_y)^T$ the principal point. The extrinsic parameters describe the transformation between the two camera centers as depicted in Figure 3.5

using a rotation and an translation.

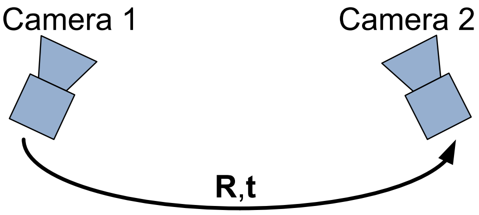


Figure 3.5: Illustration of the pose description of the second camera relative to the first one by a rotation and translation.

The resulting calibrated camera are composed of as:

$$\mathbf{P}_1 = \mathbf{K}_1[\mathbf{I}|\mathbf{0}]$$

$$\mathbf{P}_2 = \mathbf{K}_2[\mathbf{I}|\mathbf{0}][\mathbf{R}|\mathbf{t}]$$

where $\mathbf{K}_i$ denotes the calibration matrix for camera i and $\mathbf{R}$ and $\mathbf{t}$ are the rotation and translation estimated by the stereo calibration.

The method used for calibration follows the flexible camera calibration by Zhang [Zha99]. There few images of a planar target under different orientations are captured. After the feature point detection in the images the five intrinsic parameters and all the extrinsic parameters are estimated by solving a nonlinear minimization problem. Further, with a linear least-squares solution the coefficients of the radial distortion will be estimated and by minimization all parameters are refined. A detailed description of the method can be found in [Zha99]. Further, all parameters are refined by using a bundle adjust and optimizing the motion and structure.

The reason for using the method by Zhang [Zha99] and not a photogrammetric calibration method is that the calibration is required on place. Using a photogrammetric calibration a precise 3D calibration object is required and its operation in outdoor environments won't be easy compared to the simple and robust planar target used by Zhang.

3.4 Summary and Conclusion

In this chapter the design of the acquisition setup has been considered. It emerges that a stereo setup is suitable to capture the defense waves of the giant honeybees. The reasons are the passive way the images are captured and the field of view that is captured at a high resolution. Further, the required frame rate of 50fps is achieved by the cameras. The setup consists of two 4Mpx cameras that are mounted on a stand. Because the setup is used outdoors it is self powered by a rechargeable battery. To be flexible to the environmental constraints the cameras are adjustable in orientation and distance to each other.

The experiments of Section 6.1 show that a 100Ah battery powers the acquisition setup for one working day. Further, the required synchronization of the cameras is solved by a master-slave triggering.

The geometry of the sensor is constrained by the clearance distance, the reconstruction accuracy and the required field of view. To gain a reconstruction error less than one millimeter a camera angle of 30° has to be chosen. The according baseline results in 1600mm under satisfying the clearance distance.

To be able to perform reconstructions using the stereo images the setup has been calibrated in-site. This was done by the calibration method introduced by Zhang [Zha99] followed by a bundle adjust to refine structure and motion.

Concluding this acquisition setup satisfies the requirements listed in Section 3.1 and during the experiments in Nepal the adjustments to the environment were possible. Just the stability could be improved because slight deformations occurred while positioning the setup in front of the bee curtain. But this results of the tradeoff of being light weighted. For the further processing of the data these deformations were corrected by bundle adjust. The selection of the bee colony has to be chosen with intent on the illumination. A bright nest was beneficial to have short exposure times and so more sharp images. All together the acquisition setup captured high quality image sequences and stood the influences of dust in Nepal.

Chapter 4

Image Segmentation of Bees

Contents

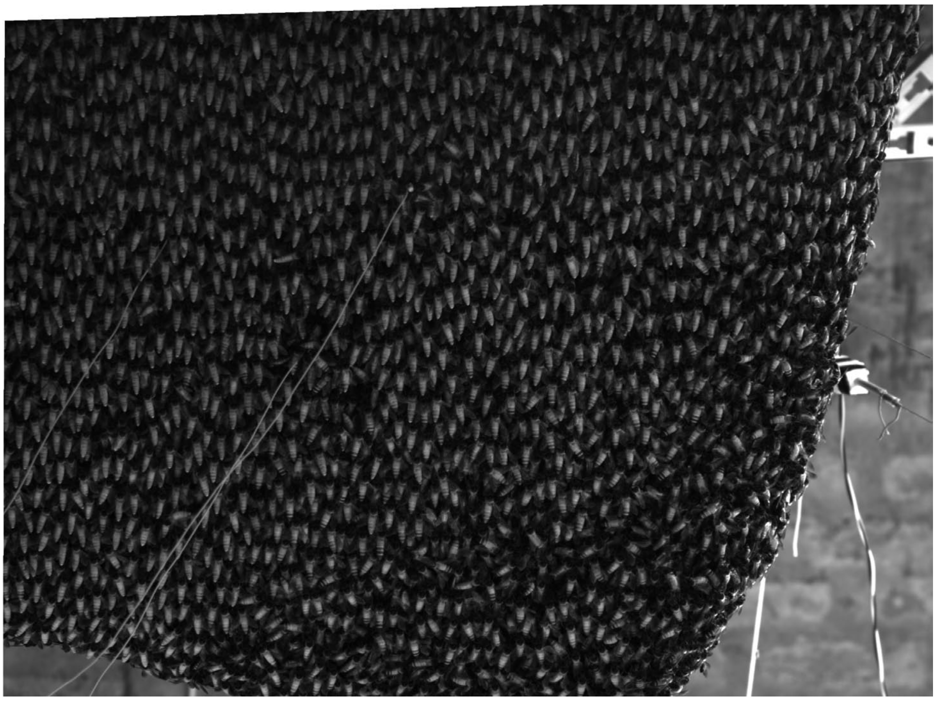


Figure 4.1: Giant honeybee nest as captured using the camera setup presented in Chapter 3.

This chapter deals with the segmentation of individual giant honeybees in rectified stereo images captured with the setup presented in Chapter 3. The aim is to segment the bees and determine their positions, orientations and lengths. Problems arise in separating the bees because they cover the comb in multiple layers and overlap each other. The heads and the thoraxes only contrast weakly from the background and the semi-transparent wings cover parts of neighboring bees. During a defense wave the wings blur the captured images by rapid flaps. Further, the resolution of a bee is limited to 50 × 16 pixels.

A bee with its chitin carapace can be seen as a rigid body. The extent of a bee can be approximated by an ellipsoid in 3D. The abdomen of a bee is bright compared to the head and the wings. The older the bee the darker the end of the abdomen gets and the coloring is used as an index for the age. The abdomen of the bees is textured by stripes.

Generally, segmentation can be done in different ways, categorized by the information used:

- edge based,
- region based,
- shape based,
- template based or
- based on local features and voting maps.

With local features and voting maps it is possible to segment the bees from the background. A segmentation of the single bees is more difficult and may be solved by center voting algorithms.

Having an image of a giant honeybee, it can be approximated by an ellipse. Figure 4.2 shows an image of a giant honeybee located at the bee curtain. It can be seen that the head and the thorax have about the same intensity values as the background.

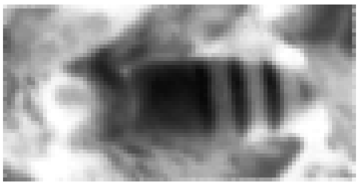


Figure 4.2: A inverted image of single giant honeybee (lat. apis dorsata) with its textured abdomen.

Regarding the whole bee, they show a repetitive pattern. A template containing a representative pattern can be used to segment the bees by template matching.

Having a look at a input image (Figure 4.1) it can be seen that the abdomen of the bees are clearly visible against the background and the head and thorax show weak contrast. Therefore it is also possible to focus on the abdomen and handle the abdomen as stable regions. The method used is based on maximally stable extremal regions (MSER). The ellipse fitted to the boundary of an MSER results in a determination of the orientation and the length of individuals.

Based on the observation that the abdomen all have the same shape and there exists an edge between the abdomen and the background, a shape based segmentation method like shape prior segmentation (SPSeg) is also promising. The segmentation should result in measurements of the position, the length and the orientation of the bee. This enables to compare the segmentation with a manually labeled ground truth.

In this section three segmentation methods are presented and evaluated in Section 6.2. These are the maximally stable extremal regions (MSER), the shape prior segmentation (SPSeg) and the normalized cross correlation (NCC).

4.1 Maximally Stable Extremal Regions (MSER)

Maximally Stable Extremal Regions (MSER) are elements of an image that show stable boundaries under increasing intensity thresholds. Here the regions should represent the abdomen of the bees. Informally the concept of extracting MSERs can be described as follows. For the range of the intensity values of an image I a sequence of binary images I_t is generated by thresholding I with all possible thresholds. The thresholds are all intensity values contained in I. The result is a sequence of binary images I_t, which starts with a totally white image and with increasing t, more and more black spots appear until the whole image is black. An example is shown in Figure 4.3. All connected components of all frames build the all maximal regions set. MSERs are those regions where the local binarization is stable over a large range of thresholds. The properties of such regions are stability, multi-scale detection, the invariance of affine transformation of image intensities and covariance to adjacency preserving transformations on the image domain [MCUP02]. The result of the MSER are regions that are made up of connected pixels. These regions can be approximated by an ellipse fitted on the boundary of the region. For a formal definition of MSER see Matas et al. [MCUP02].

The abdomen of a single bee might represent a stable region because there are visible edges between the abdomen and the background. The stripes of the abdomen will also result in stable regions, but are filtered later. Using MSER on the input image returns maximal regions of different sizes. To ease the selection of the stability threshold and enhance the stability of the regions the image is histogram equalized. This will adjust the contrast of the image according to its histogram. Now MSER returns ellipses of different aspect ratios that approximate the stable regions as shown in Figure 4.5(a). These regions do not only segment abdomens of bees. To get regions that describe abdomens the MSER result can be filtered by constraining the major and minor axis as:

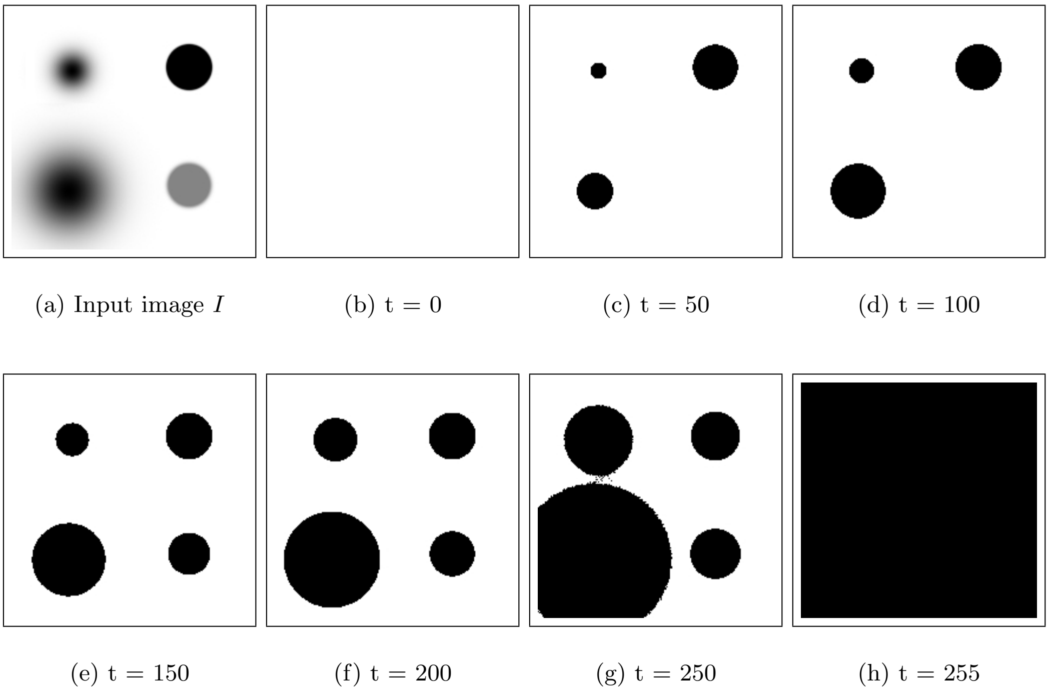


Figure 4.3: Illustration of the thresholding procedure executed while MSER calculation. Figure 4.3(a) shows the input image I, figure 4.3(b) to 4.3(h) show the binary images I_t at different values of t. The region in the upper right corner is highly stable compared to the one in the lower left corner that increases over t. As presented in the lower right corner it is also possible for regions to appear at higher levels of t but a region never disappears or shrinks.

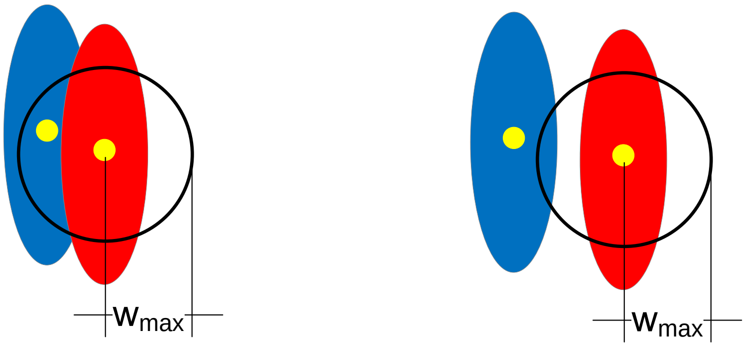
$$l_{min} \leq x_{major} \leq l_{max} \tag{4.1}$$

$$x_{minor} \leq w_{max} \tag{4.2}$$

where x_{major} and x_{minor} are the major and minor axis lengths respectively of the ellipse. l_{min} and l_{max} denote the minimal and maximal length of a bee. w_{max} is the maximal width of a bee.

This results in regions that are of the size of an abdomen as shown in Figure 4.5(b). Filtering by size only will not result in one region for each bee. The reason is that MSER also detects nested regions and some regions containing false matches may have the same aspect ratio.

For further refinements the result regions with a distance smaller than w_{max} are detected and paired. This means that neighboring regions whose distance of the centers go below w_{max} are paired as illustrated in Figure 4.4 and can be formulated as:



(a) Paired regions

(b) Unpaired regions

Figure 4.4: Illustration of the pairing of regions. (a) paired regions whose distance of the centers is smaller than w_{max} and (b) unpaired regions.

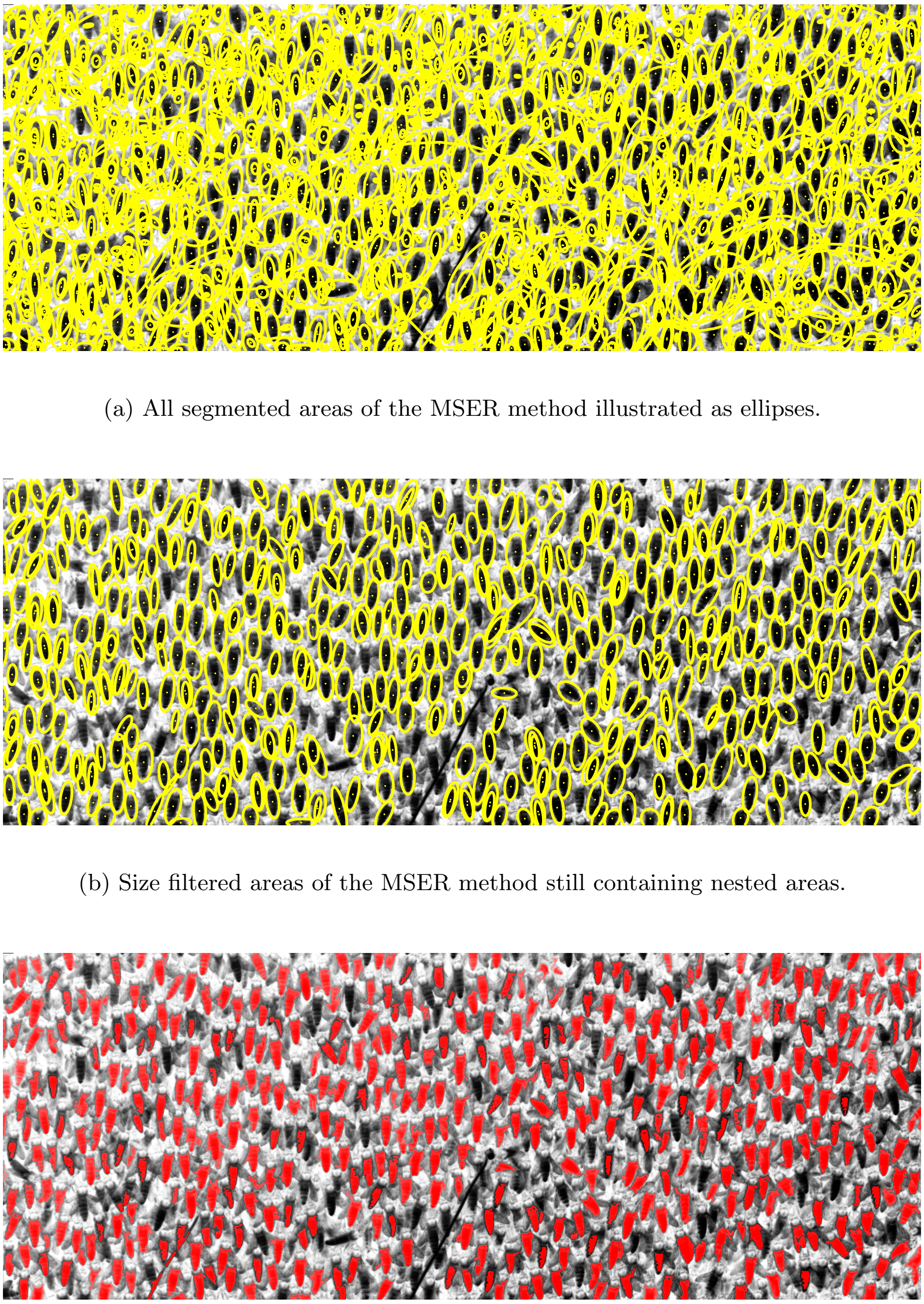
$$paired(p,q) = \begin{cases} 1 & \text{if} dist(p,q) \leq w_{max} \\ 0 & \text{else} \end{cases} \tag{4.3}$$

where p and q represent two regions and $dist(p,q)$ the euclidean distance of the centers of p and q.

The mean length and width of all unpaired regions are calculated and serve as criteria for the paired regions. The region of a pair whose length and width fits the mean length and width best is selected to represent a bee. The final result of the MSER segmentation can be seen in Figure 4.5(c). The pixels representing the regions are overlaid in red. The position of a bee can be described by the center of the fitted ellipse. The orientation can be calculated by the angle between ellipse major axis and the horizontal axis. The length is the major axis length of the ellipse.

4.2 Shape Prior Segmentation (SPSeg)

According to Werlberger et al. [WPUB09] shape prior segmentation is a semi-automated segmentation method that uses the variational formulation of the Geodesic Active Contour (GAC) energy as minimization function and additionally considers a shape prior. The minimization is done with a fast primal-dual approach. Adding high level information in terms of a shape prior enhances the robustness of the segmentation method. To reduce computation time the segmentation is done semi-automated. This means that the shape prior is set by the user to an initial pose and a local optimization routine estimates the transformation parameters $\phi = \{t, R, S\}$, where t represents a translation, R a rotation and S a scale. The segmentation energy can be formulated as:



(a) All segmented areas of the MSER method illustrated as ellipses.

(b) Size filtered areas of the MSER method still containing nested areas.

(c) Result of the MSER segmentation method. Detected regions are plotted in red.

Figure 4.5: Intermediate results of the MSER segmentation method. Raw MSER results (a), size filtered results (b) and final results (c).

$$\min_{u,\Phi}\left\{\int_{\Omega} g\left|\nabla u\right| dx + \lambda\int_{\Omega}\left(\phi\left(t,R,S\right)\circ s\left(x\right)\right) u\ dx\right\} \tag{4.4}$$

where $\int_{\Omega} g\left|\nabla u\right| dx$ represents the GAC as weighted total variation (see [LO05], [BEV$^+$05] and [BEV$^+$07]). In the shape force term, $s\left(x\right)$ describes the shape as a binary function and $\phi\left(t,R,S\right)$ the adaption of the shape prior. Weighting factor λ is required to balance between the regularization and shape force [WPUB09].

The usage of shape prior segmentation is motivated by the fact that the abdomen of the bees have almost the same shape. Using a prior should result in accurate segmentation of the abdomen because the prior should jump on the contour. An over segmentation, for example by additionally selecting the wings is not expected.

Taking the SPSeg framework as presented in [WPUB09] it solves the energy minimization for a single prior located at one position. Constraining the search range for the local optimization is required to reduce calculation time. Because of the limited search area initial poses of each bee are required. These poses can for example be set up using MSER. Now the shape prior segmentation is calculated and returns an optimized position and a mask describing the shape for each initial pose. Wrong initial poses are not handled because the resulting energy is not thresholded and can be arbitrary high. This means that each initial pose results in a position and mask.

Here, SPSeg is applied to refine the segmentation of the abdomen and does not handle wrong initializations. Figure 4.6 shows the result of SPSeg, initialized with MSER and in Figure 4.7 the refinement is demonstrated by a comparison of the MSER segmentation with the results of SPSeg. It can be seen that the masks snap to the boundaries of the abdomen. Afterwards the features describing a region (position, length and orientation) are extracted from the masks. The position is retrieved by the center of gravity. To determine orientation and length an ellipse is fitted on the mask-boundary. The orientation is calculated by the angle between the ellipse major axis and the horizontal axis. The length is twice the ellipse's major radius.

4.3 Template Matching

As presented in [RW00] cross correlation can be used for template matching. The result is a position for a template in an image. The similarity is calculated as:

$$c(u,v)=\sum_{x,y} f(x,y)t(x-u,y-v) \tag{4.5}$$

where c represents the correlation, f is the image and the sum is over x, y under the window containing the template t at the position u, v. Cross correlation has to deal with the problem of illumination changes. To get rid of this fact an illumination invariant version, the normalized

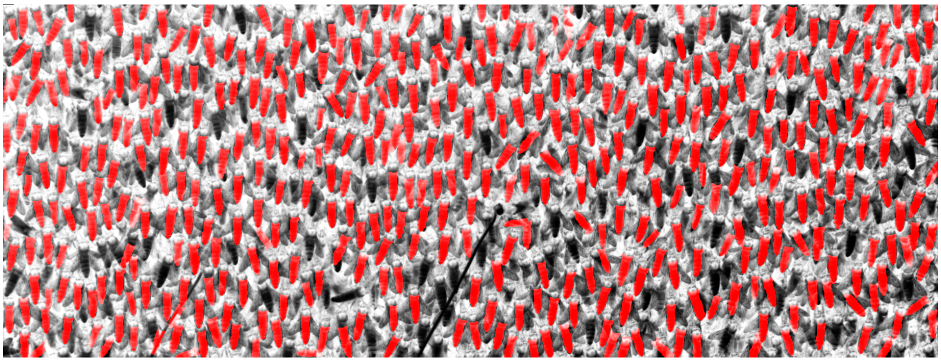
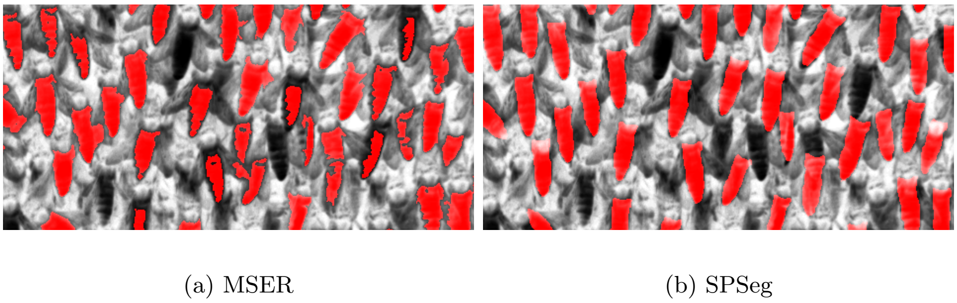


Figure 4.6: Results of a shape prior segmentation of bees initialized using MSER. In most cases the shape prior snaps to the boundary of the abdomen.



(a) MSER

(b) SPSeg

Figure 4.7: Demonstration of the refinements achieved by SPSeg. (a) shows the segmentation results of MSER and (b) the results of SPSeg initialized by MSER.

cross correlation (NCC), can be formulated by normalizing the image and template vectors to unit length (4.6).

$$\gamma(u,v) = \frac{\sum_{x,y}\left[f(x,y) - \bar{f}_{u,v}\right]\;\left[t(x-u,y-v) - \bar{t}\right]}{\sqrt{\sum_{x,y}\left[f(x,y) - \bar{f}_{u,v}\right]^2\;\sum_{x,y}\left[t(x-u,y-v) - \bar{t}\right]^2}} \tag{4.6}$$

where $\bar{t}$ is the mean of the template and $\bar{f}_{u,v}$ is the mean of the image under the template. Here the template is made up of an image of a bee as shown in Figure 4.8(a). Compared to the shape prior used for SPSeg as shown in Figure 4.8(b) the template also contains texture information of the abdomen. A drawback of NCC is the fixed template size and orientation while matching. Further, NCC only returns the position of the template. A more detailed representation of the bee can be achieved by manual annotations in the template. So for each template the length of the abdomen and the position of the center of the abdomen can be defined.

As it can be seen in Figure 4.9 the bees are of different sizes and are not orientated in exactly the same direction. To handle the drawbacks of NCC it has been extended to deal with limited changes in orientation and size. The orientation changes are handled by taking one

(a)

(b)

Figure 4.8: Comparison of the template used for NCC (a) and the shape prior used for SPSeg (b).

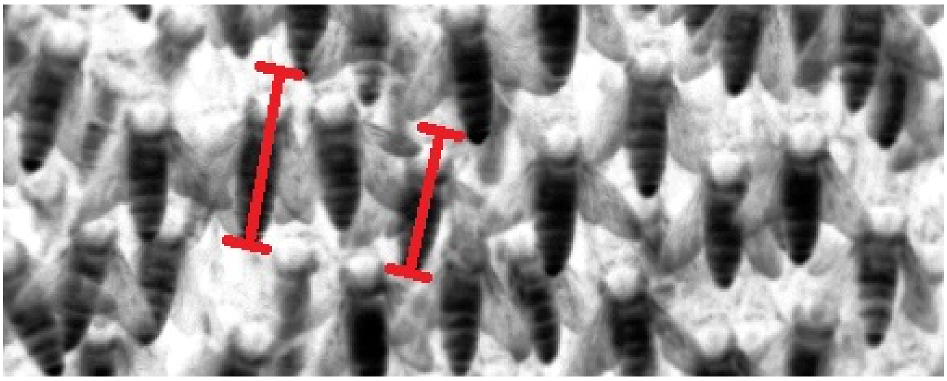


Figure 4.9: Part of an input image to illustrate the different lengths of the bees. The length of two example bees are plotted.

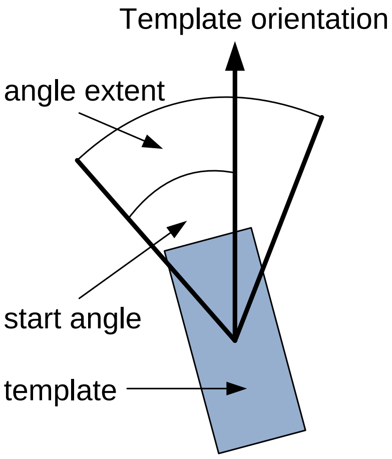


Figure 4.10: Illustration of the angles used for varyation in orientation.

template and generate multiple by rotating the template. The number of resulting templates is constrained by:

$$numTemplates_{orient} = floor(\frac{angle_{Extent}}{angle_{Step}}) \tag{4.7}$$

where $angle_{Extent}$ is the range of variation in orientation and $angle_{Step}$ is the amount the angle is incremented from one orientation to the next. The incrementation of the angle is started by an offset to the main orientation as shown in Figure 4.10. The single orientation angles can be calculated as:

$$angle_i = angle_{Start} + i * angle_{Step} \text{ with } i = 1..numTemplates_{orient} \tag{4.8}$$

where $angle_{Start}$ denotes the start offset to the main orientation.

The variation in size is handled by manually selecting multiple templates of different size. Further, these templates are made invariant to orientation as described. To eliminate multi detections and overlaps a non maxima suppression is performed.

An example result of the NCC segmentation of the abdomen of the bees with multiple templates is shown in Figure 4.11.

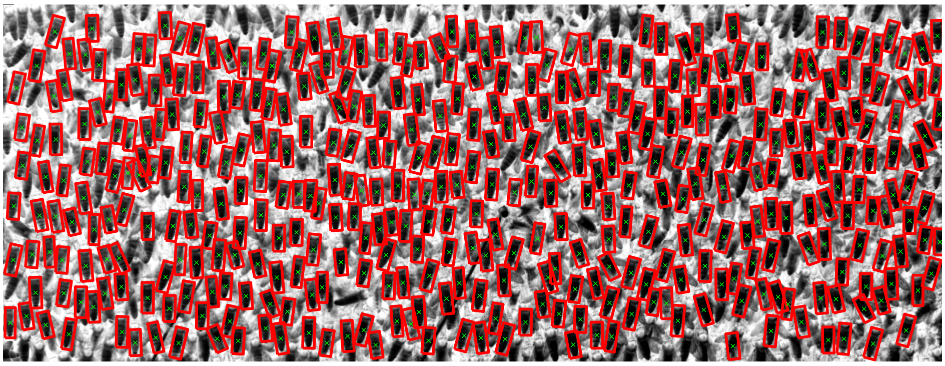


Figure 4.11: Results of a NCC segmentation of bees using five templates with a variation in orientation of 23°.

4.4 Summary and Conclusion

In this Chapter three different methods to segment giant honeybees are presented. These methods are:

- maximally stable extremal regions (MSER),
- shape prior segmentation (SPSeg) and
- template matching using normalized cross correlation (NCC).

The methods are evaluated in Section 6.2 and 6.3. Having a look at the recognition rate, MSER segments most regions, but compared to a ground truth (GT) NCC results in more true positives. The behavior of the segmentation methods concerning the reconstruction ability, SPSeg refines the results but requires a hundred times the computation time compared to MSER. Further, SPSeg is dependent on the initialization and does not handle false initializations. Therefore SPSeg is not suitable to handle the amount of test data in an accurate time and will not be used further. MSER segments the abdomen quite well, but NCC outperforms it in accuracy and time. A segmentation of an image of the experimental sets is done in about three seconds. This leads to template matching using NCC being the appropriate segmentation method for getting the individual bees.

The comparison of the time for the single segmentation methods is quite difficult and has to be considered with care. The methods are implemented at different levels of optimization. NCC for example is a highly optimized library that utilizes multi core CPUs. In contrast SPSeg was an experimental version calculated on the GPU that has further capabilities for parallelization and optimizations in read and write access.

My expectations to the three methods were a dense segmentation of the single bees. Having a look at the statistical evaluation of the recognition rate of the methods (Figure 6.10) it can be seen that all three methods segment the bees. In my opinion NCC is the appropriate method to segment the bees because it results in less false positives and these false positives are most bees that are not labeled in the GT. However MSER also contains regions that do not represent a bee but are of the same size.

The presented MSER method can be improved by changing the pairing. Instead of taking the distance of the centers an overlap of the region pixels should be detected. This would result in accurate filter results. Further improvements in filtering would be received by also constraining the minimal width of the ellipse. SPSeg can be improved by analyzing the energy resulting from the shape optimization. Thresholding of this energy by a maximum value would result in filtering bad initializations. Using this constraint of the resulting energy may lead to use random seeds to initialize SPSeg. An improvement of template matching would be achieved by learning the templates instead of manually selecting them. Template matching using NCC might be improved by making the templates size invariant by using a scale space. This would lead to improve the distribution of the lengths as shown in Figure 6.2.3.2 and approximate the distribution of the GT data more accurate. To get rid of the manual selection of the templates these could be learned.

Chapter 5

Stereo Reconstruction

Contents

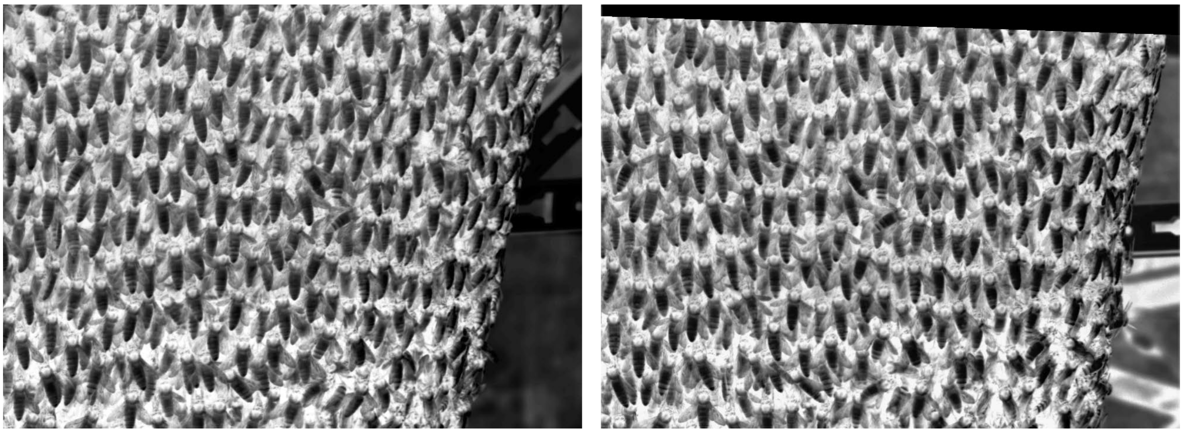


Figure 5.1: Pair of stereo images as they result from the preprocessing step. The images are rectified, histogram equalized and inverted.

This Chapter deals with the 3D reconstruction of the segmented bees using stereo vision. An example stereo image is shown in Figure 5.1. It is assumed that individual bees are already segmented, using methods proposed in Chapter 4. Stereo reconstruction is parted in two main steps. These are the correspondence problem and the triangulation of the correspondences. Problems arise by finding corresponding bees because of the presence of repetitive patterns, as shown in Figure 5.2. Using only a maximum similarity search using a measure like NCC will not solve the problem of repetitive patterns.

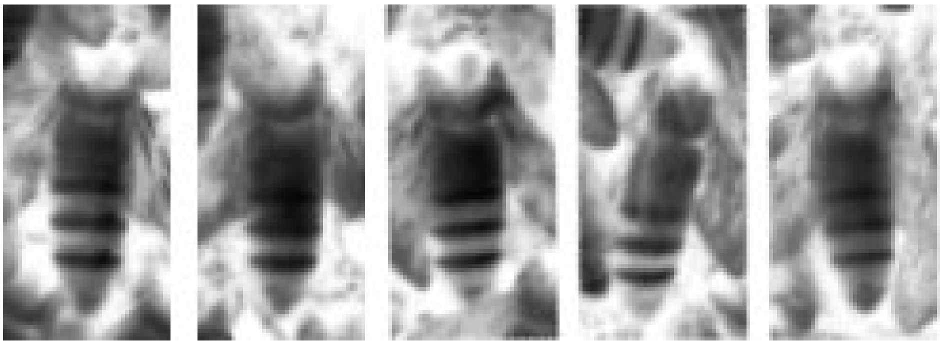


Figure 5.2: Sequence of bees in a single image to show the occurrence of repetitive patterns.

To generate a 3D reconstruction form a pair of stereo images, first the correspondence problem has to be solved. This means that for each bee in the left image its corresponding bee in the right image has to be identified. This problem is equivalent to calculating the disparity map. The map can be calculated using local methods, global methods or dynamic programming. Local methods focus on the matching cost calculation and on the cost aggregation. They use a local winner-take-all optimization at each pixel. Local method are limited by enforcing uniqueness only for the reference image and are not suited to deal with repetitive patterns. Dynamic programming can find the global minimum for independent scan-lines in polynomial time but problems arise by dealing with occlusions and by enforcing inter-scan-line consistency. Additionally dynamic programming enforces a monotonicity or ordering constraint. Global methods are formulated by minimizing the energy of a disparity function d:

$$E(d) = E_{data}(d) + \lambda E_{smooth}(d) \tag{5.1}$$

where the data term, $E_{data}(d)$, measures the match between the disparity function and the input image pair. $E_{smooth}(d)$ represents the smoothness between neighbors for example by measuring disparity differences [SS01].

Because of the possibility to define smoothness between neighbors and preserve neighborhood, graphs have been chosen to solve the correspondence problem. The graph is constructed based on the method of reduced graphs as presented by Gusfield and Irving in [GI89] and minimized using the minimum-cut / maximum-flow algorithm as proposed by Goldberg and Tarjan [GT88]. A minimum-cut results in a strong local minimum and so approximates the global solution.

Having the correspondences, these can be triangulated.

In the following sections first the algorithm to solve the correspondence problem is described followed by the triangulation routine.

5.1 Stereo Matching Using Reduced Graph Cuts

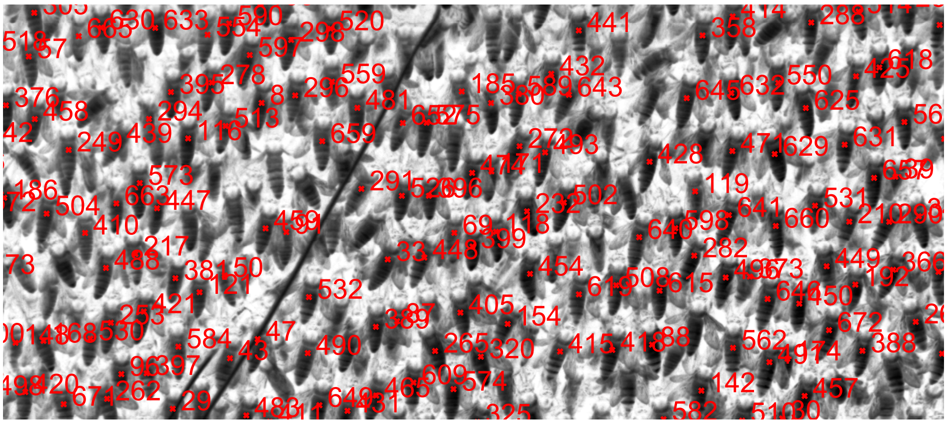


Figure 5.3: Segmented bees resulting from the template matching method using NCC as presented in Section 4.3.

To calculate the minimum-cut of a graph, the graph has to be constructed. Considering the simplest version by associating each bee of the first image with all bees of the second image would result in a computational too expensive problem. This can be solved by implying the epipolar geometry. Only bees that are located in the same row are possible corresponding candidates for rectified stereo images. The fact that bees are segmented with different templates results in loosing the constraint and allowing a vertical deviation d_y. The pre-selection constraint can be formulated as:

$$y_l - d_y \leq y_{r,i} \leq y_l + d_y \tag{5.2}$$

where y_l represents the y-coordinate of the reference bee in the left image and $y_{r,i}$ the y-coordinate of the i^{th} right bee.

To be able to construct a reduced graph as presented in [ZDC07] a similarity measure for the bees is required. This is done by calculating the normalized cross correlation of a patch containing the reference bee at the position of a possible corresponding bee and will further be called $s_{i,p}$, where p represents the index of the reference bee and i the index of the possible candidate. Now, for each bee p a chain of N possible candidates is constructed by selecting the N greatest similarities of $s_{i,p}$ and sort them by their horizontal positions in the image. The cost of matching $D_{\{p,s_{i,p}\}}$ will be defined as:

$$D_{\{p,s_{i,p}\}} = 1 - s_{i,p} \tag{5.3}$$

These chains are connected with the source and the sink of the graph. An illustration of the graph is presented in Figure 5.4. The capacity of the t-links are calculated as:

$$Ct(p,i) = D_{\{p,s_{i,p}\}} + K_p \tag{5.4}$$

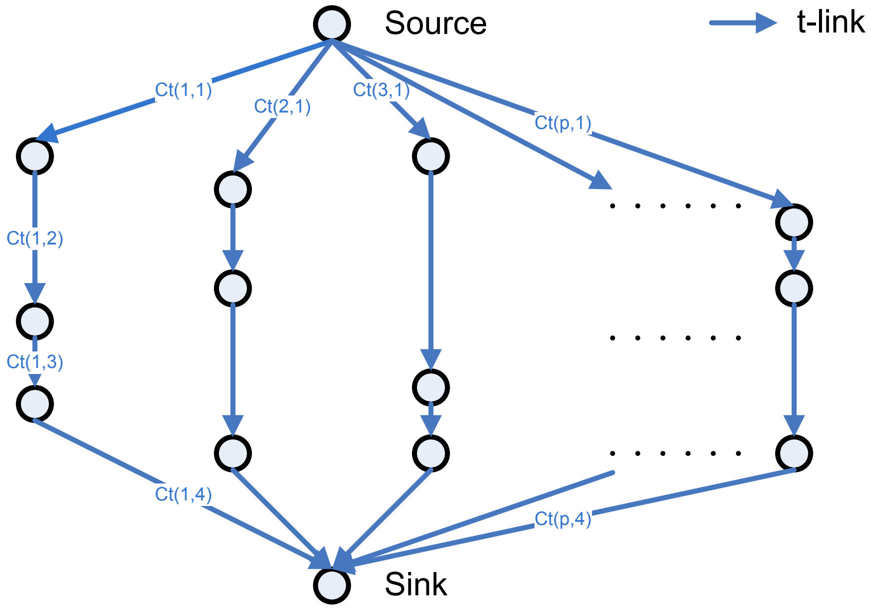


Figure 5.4: Illustration of a reduced graph with only t-links.

where K_p is a constant satisfying:

$$K_p > N * |max(d_{i,p}) - min(d_{i,p})| \text{ with } i = 1..N \tag{5.5}$$

where $d_{i,p}$ is the disparity of the bee p to its possible corresponding one i.

Calculating the minimum-cut of the up to now constructed graph would result the same as performing a maximum search on the similarity $s_{i,p}$ of the possible candidates. A neighboring constraint has to be introduced for the k nearest neighbors. In the graph the neighboring constraint is presented by n-links of a capacity $Cn(p, q, i)$:

$$Cn(p, q, i) = (|d_{i,p} - d_{i,q}| + 1) \text{ with } i = 1..N - 1 \tag{5.6}$$

where p and q denote two neighboring bees in the first image. The graph containing example n-links can be seen in Figure 5.5.

Using this graph to get the correspondences by calculating the minimum cut will result in loosing the neighborhood of the bees in the second image because for example $s_{1,p}$ does not have to be close to $s_{1,q}$. A reorganization of the graph solves this problem. Instead of having a chain of N t-links the chain consists of M slots and the N possible candidates are associated to the slots satisfying:

$$m * \frac{2 * span}{M} \leq d_{i,p} - d_{est}(p) < (m + 1) * \frac{2 * span}{M} \text{ with } m = -\frac{M}{2}..\frac{M}{2} \tag{5.7}$$

where m represents the slot position, $d_{est}(p)$ the estimated disparity of bee p and $span$ the maximal distance of the possible corresponding bee to the estimated disparity in pixel.

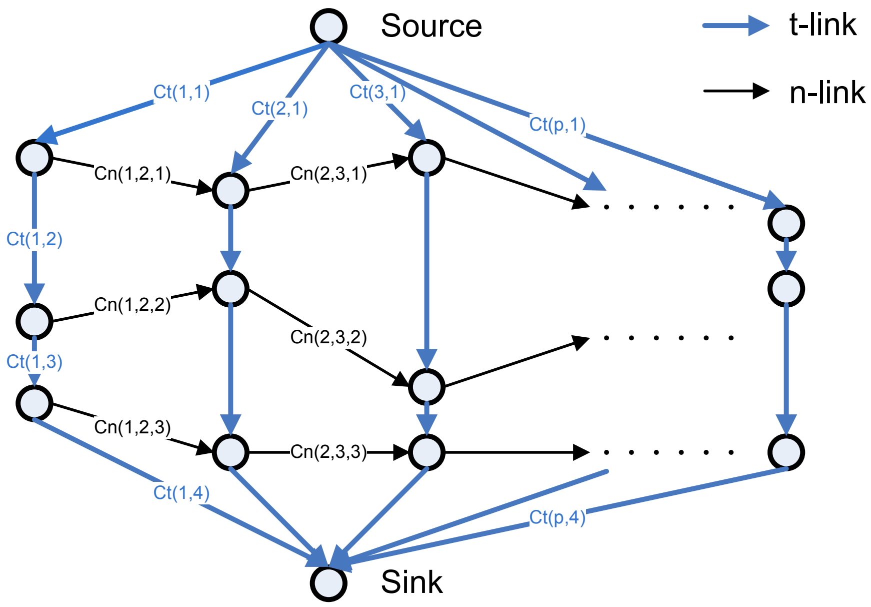


Figure 5.5: Structure of a reduced graph for $(N-1)=3$ levels where Ct and Cn represents the weights the links have.

The number of slots M has to be chosen that only one bee will be associated to one slot. This can be formulated as:

$$\frac{2 * span}{M} < \min_{p,q,i}(d_{i,p} - d_{i,q}) \tag{5.8}$$

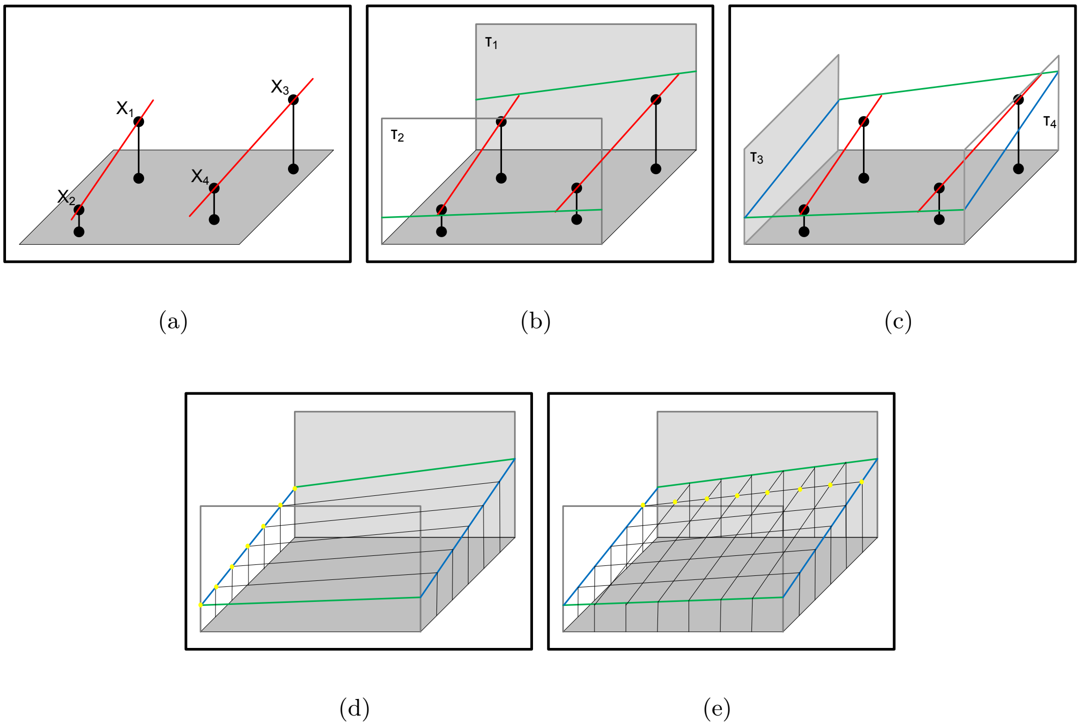
The estimated disparity $d_{est}(p)$ is needed to further constrain the dimension of the tree. The estimation is done by manually selecting four correspondences located at the corner regions of the overlapping part of the stereo image pair. Using the disparities of these four points a surface is fit through the supporting points and so an approximate disparity can be calculated for each pixel of the left image. To calculate the estimated disparities given the four disparity offsets and the coordinates of the supporting points of the left image the following steps have to be performed.

First four three dimensional points X_i have to be defined as:

$$X_i = (x_i \; y_i \; d_i)^T$$

where x_i and y_i represent the coordinates of the i^{th} supporting point of the left image and d_i the manually annotated disparity. Then two lines are set up by using X_1 and X_2 for the first line and X_3 and X_4 for the second line as shown in Figure 5.6(a). These two lines are intersected with the planes τ_1 and τ_2. The planes are defined by:

$$\tau_1 = n_{img} \times v_{row,first}$$



(a) (b) (c)

(d) (e)

Figure 5.6: Illustration of the steps to estimate the disparity.

$$\tau_2 = n_{img} \times v_{row,last}$$

where n_{img} represents the normal vector of the image plane, $v_{row,first}$ the first row vector and $v_{row,last}$ the last row vector of the image. The cutting points at each plane set up two lines (see Figure 5.6(b)). These lines are intersected with τ_3 and τ_4:

$$\tau_3 = n_{img} \times v_{col,first}$$

$$\tau_4 = n_{img} \times v_{col,last}$$

where $v_{col,first}$ denotes the first column vector and $v_{col,last}$ the last column vector of the image. Connecting the points respective the plane they are member of results in two lines as illustrated in Figure 5.6(c) by the blue lines. These two lines have to be parted in as many equal distances as the number of image rows decremented by one. The points that are a member of $\tau 3$ are called $X(0, y_l)$ where y_l represents the row index. These points are connected with the corresponding points located on the opposite line (see Figure 5.6(d)) and a gradient k_{y_l} in y-direction can be calculated for each line. For each pixel the estimated disparity $d_{est}(x_l, y_l)$ can be calculated as:

$$d_{est}(x_l, y_l) = x_l * k(y_l) + x_3(0, y_l)$$

where $x_3(0, y_l)$ is the third coordinate of $X(0, y_l)$.

An example of such an disparity estimation plane can be seen in Figure 5.7.

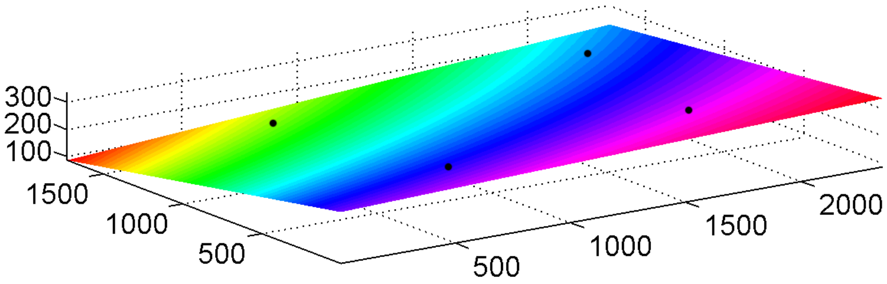


Figure 5.7: Disparity estimation plane with color coded disparities. The black points are the points of support.

To get complete connections between the source and the sink for each bee the remaining empty slots after the association of the N bees with their capacities as formulated in (5.5) have to be assigned a capacity Ct_0 as:

$$Ct_0(p) = 2 * (K_p + 1) \tag{5.9}$$

Also the n-links have to be updated and added to the graph bidirectionally. The n-link does not only aim from bee p to q but also from q to p. For two bees p and q the n-links are calculated as:

$$Cn(p,q,i) = \begin{cases} K_p * \alpha * (1 + \frac{1}{|d_{i,p} - d_{i,q}| + 1}) & \text{if slot } i \text{ contains a bee for } p \text{ and } q. \\ K_p * \alpha & \text{if at least one slot } i \text{ is empty.} \end{cases} \quad \text{with } i = 1..M \tag{5.10}$$

where α is responsible for smoothing the cut. Figure 5.8 shows an example graph where the slots containing a bee are marked in red and the weights are calculated using (5.5), (5.9) and (5.10).

Finally a cut that separates the source from the sink has to be calculated. The summation of the capacities that are cut should be minimal and can be solved by a minimum-cut / maximum-flow algorithm as presented by Goldberg and Tarjan [GT88]. The result of the cut are corresponding pairs of bees. Figure 5.9 shows an example cut by the dashed red line.

Analyzing the corresponding pairs inconsistencies in uniqueness can be figured out. These conflicts arise if more than one bee of the first input image aim to the same bee of the second image. The uniqueness violations result from the asymmetric treatment of the stereo images. Up to now only a forward matching has bee performed. The uniqueness violations can be solved by also applying a backward matching (repeating the whole for switched input images) and merging the two sets of correspondences by creating the intersection.

The stereo matching algorithm using reduced graph cuts is summarized as:

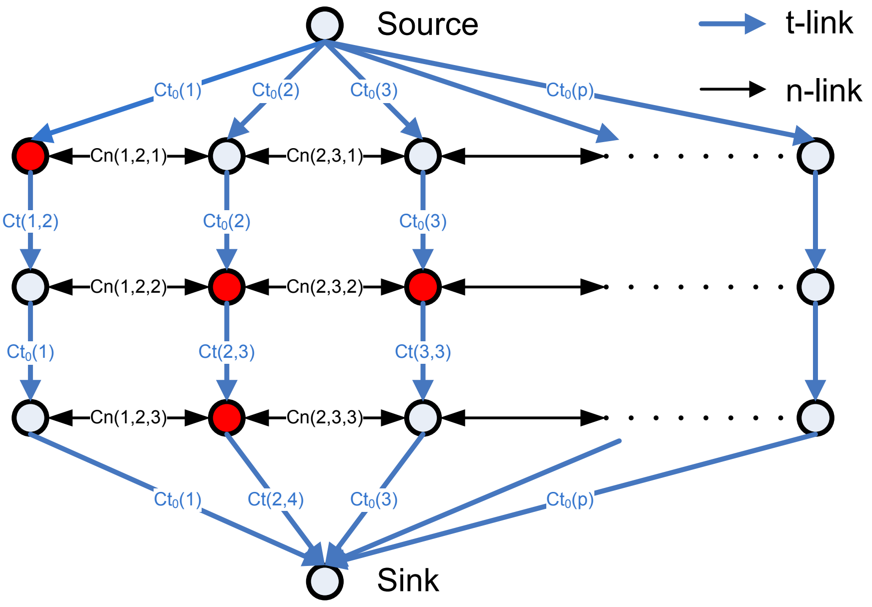


Figure 5.8: Structure of a reduced graph for $M = 3$ levels where bees are located at the red nodes.

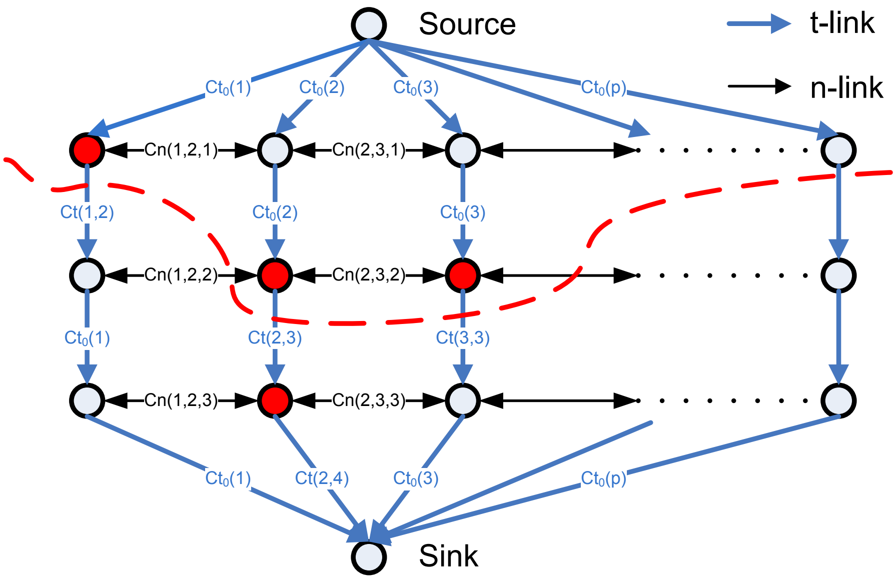


Figure 5.9: Cut through a reduced graph for $M = 3$ levels where bees are located at the red nodes. The dashed red line represents the cut.

1. Estimate the disparity by manually selecting four correspondences and fit a plane.

2. Preselect possible candidates by constraining according the epipolar geometry (5.2).

3. Generate an empty tree that contains a chain of M slots for each bee and connect the top of the chain with the source and the end with the sink. The Capacities are calculated by (5.9).

4. For each bee and its possible candidates locate the slot in the tree by (5.7) and update the capacity of the t-link by (5.5).

5. Add bidirectional n-links with a capacity calculated by (5.10) for the k nearest neighbors at each slot level.

6. Calculate the minimum-cut.

7. Repeat 2 to 6 for switched input images.

8. Merge the two sets of correspondences by intersecting them.

5.2 Triangulation

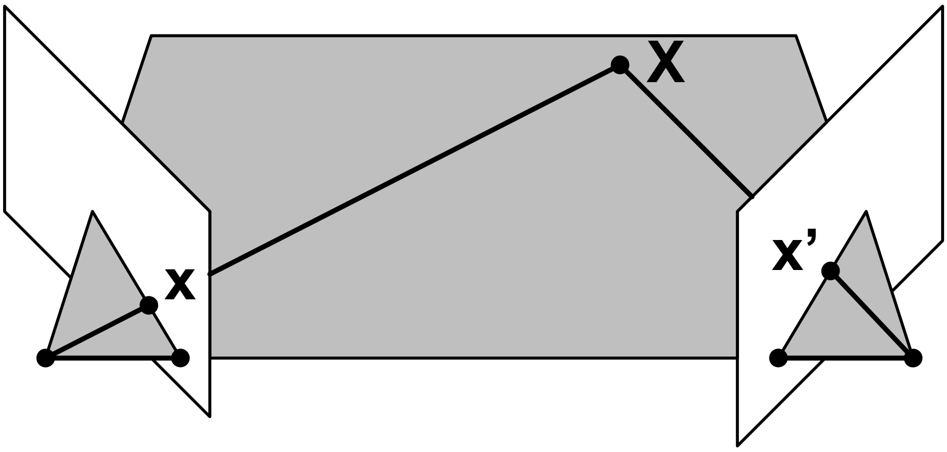


Figure 5.10: Interrelationship between the two 2D image points $\mathbf{x}$ and $\mathbf{x}'$ and the 3D triangulated point $\mathbf{X}$

The goal of the triangulation is to get an approximation of the 3D point $\mathbf{X}$ that is calculated using two corresponding points $\mathbf{x}$ and $\mathbf{x}'$ in the image planes of the cameras as illustrated in Figure 5.10. Having corresponding image points for each matched bee the 3D positions can be calculated.

According to Hartley and Zisserman [HZ03] to triangulate two corresponding points, $\mathbf{x}$ and $\mathbf{x}'$, of the image planes their camera matrices $\mathbf{P}$ and $\mathbf{P}'$ are required. A camera matrix $\mathbf{P}$ is formulated as:

$$\mathbf{P} = \mathbf{K}[\mathbf{I}|\mathbf{0}]\mathbf{H} \tag{5.11}$$

where $\mathbf{K}$ represents the calibration matrix, $\mathbf{I}$ a 3×3 identity matrix and $\mathbf{H}$ a projective transformation in 3-space. $\mathbf{K}$ is made up of:

$$\mathbf{K} = \begin{pmatrix} f & s & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{pmatrix} \tag{5.12}$$

where f is the focus, s the skew and $\mathbf{p} = (p_x p_y)^T$ the principal point. A projective transformation in 3-space is composed as:

$$\mathbf{H} = \left(\begin{array}{c|c} \mathbf{R} & \mathbf{t} \\ \hline \mathbf{0} & 1 \end{array}\right) \tag{5.13}$$

where $\mathbf{R}$ is a 3×3 rotation matrix and $\mathbf{t}$ a 3×1 translation vector.

For a 3D point $\mathbf{X}$ there exists back projections in the image planes that satisfy:

$$\mathbf{x} = P\mathbf{X} \tag{5.14}$$

$$\mathbf{x}' = P'\mathbf{X} \tag{5.15}$$

Analogue to the DLT method these equations can be combined and form:

$$A\mathbf{X} = \mathbf{0} \tag{5.16}$$

This is done by calculating the cross product to get three equations for each image point $\mathbf{x}$ and $\mathbf{x}'$. For the first image it results in:

$$\mathbf{x} \times (P\mathbf{X}) = \mathbf{0} \tag{5.17}$$

This can be written out:

$$x(\mathbf{p}^{3T}\mathbf{X}) - (\mathbf{p}^{1T}\mathbf{X}) = 0 \tag{5.18}$$

$$y(\mathbf{p}^{3T}\mathbf{X}) - (\mathbf{p}^{2T}\mathbf{X}) = 0 \tag{5.19}$$

$$x(\mathbf{p}^{2T}\mathbf{X}) - y(\mathbf{p}^{1T}\mathbf{X}) = 0 \tag{5.20}$$

$$\tag{5.21}$$

where $\mathbf{x} = (x\ y)^T$, $\mathbf{x}' = (x'\ y')^T$ and $\mathbf{p}^{iT}$ are the rows of the projection matrix P.

Repeating it for the second image and taking the two independent equations of the cross product of each image, results in four equations that are linear in the components of $\mathbf{X}$ and form a equation system of $A\mathbf{X} = \mathbf{0}$ with:

$$A = \begin{pmatrix} x\mathbf{p}^{3T} - \mathbf{p}^{1T} \\ y\mathbf{p}^{3T} - \mathbf{p}^{2T} \\ x'\mathbf{p}'^{3T} - \mathbf{p}'^{1T} \\ y'\mathbf{p}'^{3T} - \mathbf{p}'^{2T} \end{pmatrix} \tag{5.22}$$

A detailed description of solving the set of four equations with four homogeneous unknowns by either a homogeneous method or an inhomogeneous method can be found in Hartley and Zisserman [HZ03].

5.3 Summary and Conclusion

In this chapter a method to find corresponding bees is presented. It is done by a reduced graph cut that takes the neighborhood in the left and right rectified stereo image into account. The graph dimension can be reduced by constraining the possible bees to an interval around the estimated disparity. By choosing different smoothing factors α the smoothness of the cut and so the influence of the neighbors is controlled. In Section 6.4 the performance of the stereo matching algorithm using reduced graph cuts is evaluated by experiments. The resulting correspondences are compared with ground truth data and result in 98% true matches. By further visual inspection of the correspondences an accurate result has been gained by the stereo matching using reduced graph cuts. Graph cut exceeded the expections by improving the result of a maximum similarity search by 96.4%. The required matching time of two minutes can be tolerated and is still a very high improvement to manually labeling the correspondences. For an expert manually labeling of one image pair takes about four hours and the presents of errors can not be excluded.

Exemplary other methods to solve the correspondence problem were considered and tested. These were local methods as SIFT and growing correspondence seeds. An exact evaluation was not performed, but a look at the results showed that they are not suitable for finding corresponding bees.

The functionality of the matching algorithm has been tested at images of all three test sets and the results have been evaluated by visual inspection because of the lack of ground truth data. It results that the stereo matching works for the other experimental series and because the images of one series are very similar the matching should work for all 521 sequences of defense waves. Generally there are about 10 ambiguities at one epipolar line. By constraining the horizontal search range to 300px around the estimated disparity the number of possible candidates can be reduced to about 3.5.

Chapter 6

Experiments

Contents

This chapter deals with the experiments done to evaluate the properties of the acquisition setup and the reconstruction framework. The experiments include an evaluation of the hardware according the synchronization of the cameras, the required power consumption and the acquisition buffer. Further, the reconstruction framework is evaluated. The first part to evaluate is the segmentation of the single bees using the three methods discussed in Chapter 4. There the number of segmented bees and the reconstruction ability give hints of the usability of the segmentation method. Focus lies on the template matching method using NCC because it can be shown that it is the appropriate method to segment the bees. A detailed evaluation of its parameters has been carried out. To achieve good triangulation results a proper solution of the correspondence problem is required. The stereo matching algorithm using reduced graph cuts as described in Chapter 5 will be evaluated against a manually labeled ground truth to quantize its results.

In the first part of this chapter the experimental setup will be described. Afterwards the setup will be evaluated to ensure that it fulfills the environmental constraints. The next experiments deal with the evaluation of the reconstruction framework and can be parted in the evaluation of the segmentation methods and the stereo matching.

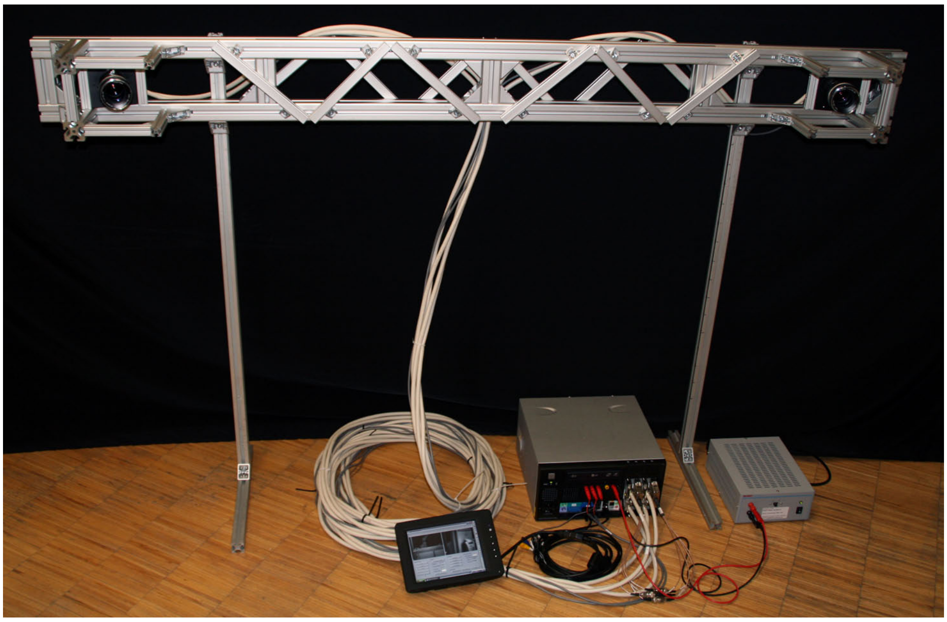


Figure 6.1: The image acquisition setup consisting of two cameras mounted to the camera stand, the industrial PC, the touch screen and an external power supply.

6.1 Image Acquisition Setup

The image acquisition setup (see Figure 6.1) consists of two CMOS cameras with a resolution of $2352 \times 1728 px$ that are mounted to a camera stand. An industrial PC equipped with two frame grabber cards manages the acquisition of the images via camera link. To achieve synchronization the two frame grabber cards are connected as a master slave system. The second card is externally triggered by the first one that generates a trigger by itself.

To meet the requirements of being mobile and flexible the PC is powered by a rechargeable battery using a DC-DC power supply. To achieve the flexibility of mounting the cameras the above mentioned camera stand has been designed of a total length of $1900mm$. At this camera stand the cameras can be mounted at an adjustable baseline between 800 and $1600mm$ and with the opportunity to choose any enclosed angle α. Figure 6.2 shows the 3D CAD model of the camera stand.

Using this setup it is possible to capture image sequences of $60fps$ of full resolution per camera. The recording time is constrained by the amount of memory the PC is equipped with. In this setup there are $8GB$ of RAM and so it can capture 900 frames per camera. This results in a total capturing time of 15 seconds. The reason for capturing to RAM is the high data rate of $487.7MB/s$ and the comparable low writing performance of a hard disk drive.

To evaluate the hardware to stand the requirements mentioned in Section 3.1 the synchronization of the two cameras has been measured. Further, the power consumption has been measured and leads to an operation time. Finally the acquisition buffer has been analyzed if

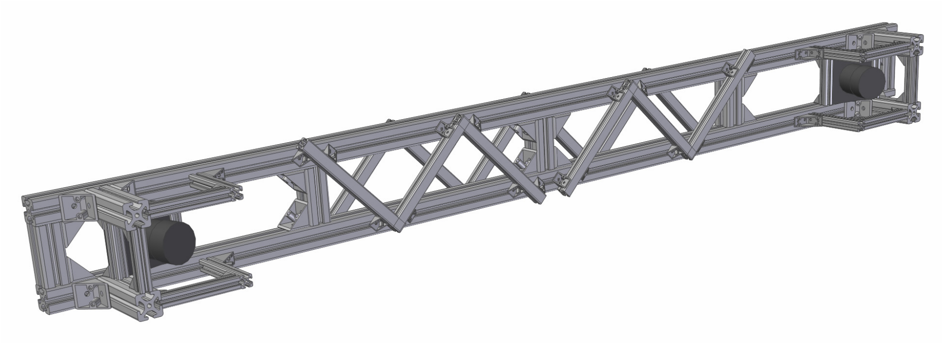


Figure 6.2: CAD model of the camera stand with the cameras mounted at a baseline of $1600mm$ and an enclosed angle α of 30°.

the usage of an other buffer management will improve the acquisition time. The next sections contain the results of the evaluations that have been done.

6.1.1 Synchronization

The synchronization has been evaluated by capturing a flashing light with the two cameras at a frame rate of $592fps$. Then mean of a mask containing the light source will be calculated. This results in an intensity value for each frame. To get rid of different shutter settings and the slightly different gradation graphs the intensity values are normalized per camera to [0..1] and compared in Figure 6.3. There it can be seen that the switching from dark to light and vice versa are at the same time. So the cameras operate synchronously.

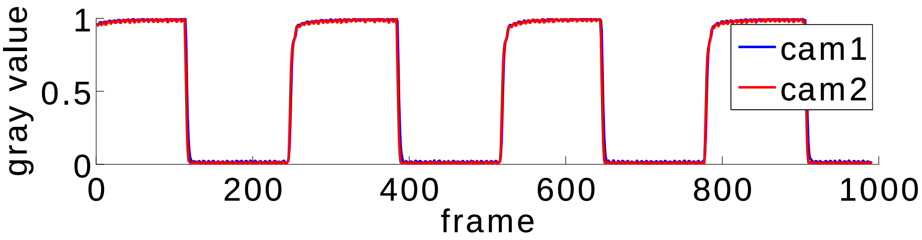


Figure 6.3: Evaluation of the synchronization.

6.1.2 Power Consumption

To figure out the power usage of the total setup the power consumption has been measured with a Wattmeter. At the maximal load a power consumption of $125W$ has been measured for the PC. In addition two times $10W$ for the cameras and $10W$ for the touch screen is required. The outcome of this is a total power consumption of $155W$. Using a rechargeable battery of $100Ah$ enables a runtime of more than seven hours which was calculated by:

$$\frac{Q * U}{P} = \frac{100Ah * 12V}{155W} = 7.74h \tag{6.1}$$

where Q represents the electric charge of the battery, U the voltage of the battery and P the total power consumption of the sensor.

6.1.3 Acquisition Buffer

The acquisition buffer has to deal with the data stream resulting from the two cameras. It has to process $487.7MB/s$ if the image sequence is captured with $4Mpx$ and $60fps$ per camera. Writing this stream directly to a hard disk drive is not possible for the industrial PC used. This leads to use a preallocated buffer in the PC's RAM. The current software solution stores the data to a preallocated memory and after the capturing the data are written to the disk. Using a ring buffer and writing to the disk in parallel would extent the recording time from 15 to 18 seconds which is a minor improvement in quality. Further, writing to the hard disc drive requires CPU time that is used for grabbing the images. Because of this a static buffer is used. The extent of the recording time has been calculated by measuring the writing time for one second blocks. One block needs $5.18s$ to write. So having a ring buffer of 15 blocks three blocks would be written in parallel to the hard disk before a buffer overrun would happen.

$$x * timePerBlock < x + numBlocks \tag{6.2}$$

$$x * 5.18 < x + 15 \tag{6.3}$$

$$x < \frac{15}{4.18} = 3.589 \tag{6.4}$$

6.1.4 Application

The image acquisition of the defense waves of the giant honeybees took place in the south of Nepal. There two bee colonies were selected to perform three experimental series. The facts of the experimental series are summarized in Table 6.1. For the positioning of the cameras a stand has to be constructed on-site. For the first experimental place it result in a 7 meter high bamboo construction as illustrated in Figure 6.4. At the second place the cameras could be mounted at the rear side of the roof as shown in Figure 6.5.

6.2 Evaluation of Segmentation Methods

The goal of the segmentation is to find as many bees as possible in the acquired image sequences. This can be done with different segmentation methods. Suitable methods are summarized in Chapter 4. The Evaluation of these methods is done by an evaluation of the recognition rate, the evaluation of the reconstruction rate and the evaluation of the execution time. The evaluation is done by three experiments that are calculated for each method.

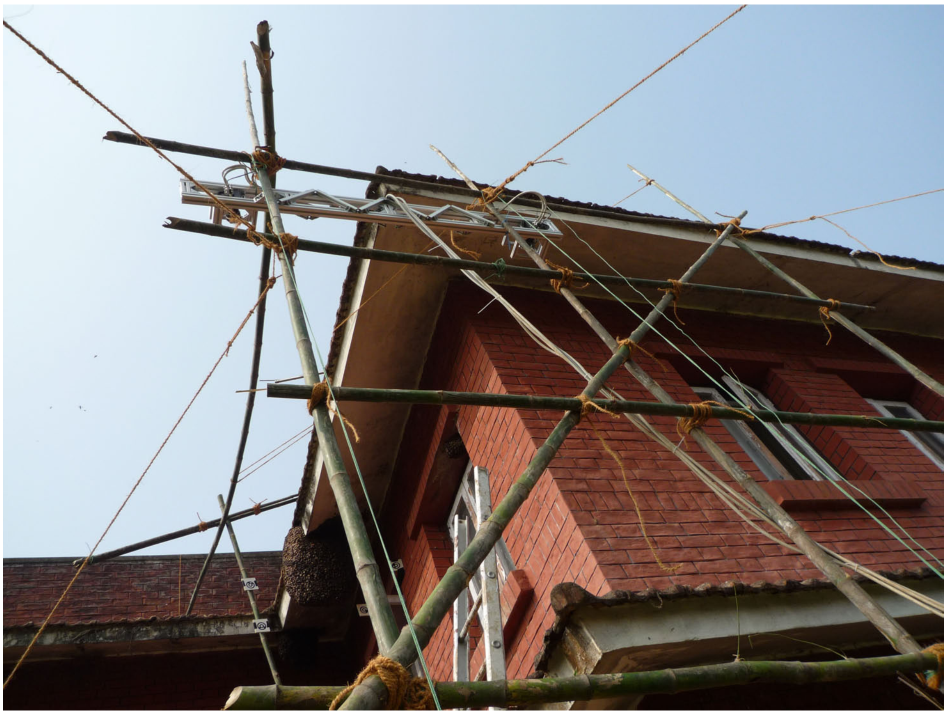


Figure 6.4: First experimental place with a bamboo construction of $7m$ to place the cameras.

Fact	**Experimental Series**		
	1	**2**	**3**
Location	Rampur	Sauraha	Sauraha
Distance to bee curtain [m]	3	3	2
Number of Sequences	154	263	104
Number of frames per sequence	900	900	900
Frame rate [fps]	60	60	60
Required storage [TB]	1.06	1.8	0.7
Experimental duration [days]	8	6	3

Table 6.1: Facts of the experimental series done to capture defense waves.

The following section describes these three experiments to evaluate the segmentation methods followed by a definition of the ground truth data set and the features of a single bee in Section 6.2.1. Next in Section 6.2.2 the parameters used for the single methods are defined and in Section 6.2.3 the results for the three segmentation methods described in Chapter 4 are presented. These methods are:

- Normalized Cross Correlation (NCC)
- Maximally Stable Extremal Regions (MSER)
- Shape Prior Segmentation (SPSeg)

The evaluation of the segmentation is concluded with a detailed evaluation of NCC because

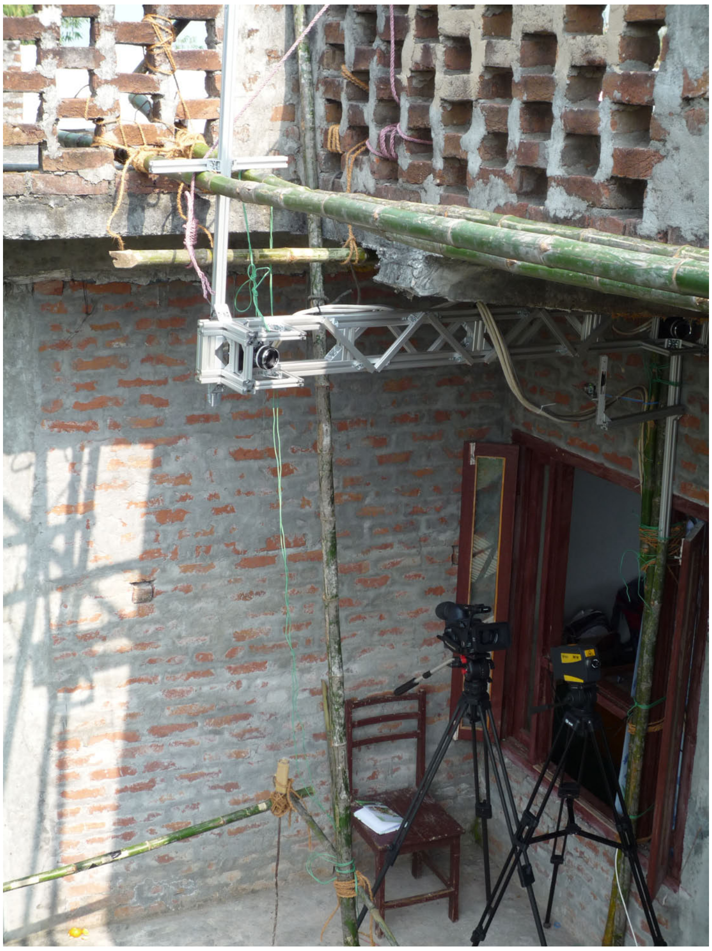


Figure 6.5: Second experimental place with the camera stand mounted at the rear side of the roof.

it outperforms the other methods and so an analysis of its parameters is presented in Section 6.3.

The experiments used to evaluate the performance of the segmentation methods are parted in three main experiments. These are the comparison of the recognition rate, the comparison of the reconstruction rate and the measurement of the required time.

The amount of detected bees represents the recognition performance of the segmentation method compared to a ground truth (GT). This recognition performance can be quantified by the number of true positives, false positives and false negatives. Positives are all bees that are segmented by the segmentation method. These positives can be parted in true and false ones, so true positives are bees that are labeled in the ground truth and are additionally segmented using the respective segmentation method. False positives represent the bees that are segmented by the method but were not labeled in the GT. Bees that are labeled in the ground truth data set but not segmented are called false negatives. True negatives are bees that are not segmented and not labeled in the GT. They can not be evaluated because the GT does not contain negative labels. The more true positives and less false positives the merrier the evaluated method represents the ground truth data.

Corresponding bees are bees whose distance of the center of gravity in the ground truth and in the segmentation is minimal and does not extent a threshold.

To be able to reconstruct a bee as a rigid body some features to describe the bee are required. These features are the position, length and the orientation of the bees and an overlay of the features onto the 'test_image_large' will be shown in the comparison of the results section. To have a first feeling of the reconstruction performance of the method the distribution of the length and the orientation of the single methods are compared. For a more detailed conclusion for the geometric accuracy the features are compared with the respective feature of the corresponding ground truth bee.

Geometric accuracy describes how well the single bees are segmented according to their position, length and orientation. This can be measured by calculating for each bee the differences of the features with the respective feature of the corresponding ground truth bee. These differences set up a distribution for each feature and so the mean can be calculated. A segmentation method with small mean values of these distributions result in a high geometric accuracy. The features evaluating the geometric accuracy are:

- the difference of corresponding positions,
- the difference of corresponding lengths,
- and the difference of corresponding orientations.

Because of the numerous test data the execution time of the segmentation method has to be considered too. It can be parted in three main steps the calculation time, the pre-processing time and the post-processing time. For MSER there is no pre-processing and the post-processing describes the filtering of overlapping regions. NCC also only requires a post-processing that contains the merging process of multiple templates used for segmentation. SPSeg is a method to refine the geometric accuracy and so it requires MSER for initialization.

6.2.1 The Ground Truth (GT)

To be able to compare the segmentation results of the single methods a ground truth data set is required. Here the ground truth is represented by manually annotated bees. The annotation is done by selecting three points of each bee, two points left and right of the torso and the third one at the end of the abdomen. A schematic view of the significant points is shown in Figure 6.6. Figure 6.7 shows a subset of the test image containing the annotation of the bees as yellow lines. The ground truth data set consists of two images representing the bees at different mean scales. The two sets are called 'small' and 'big', where the first one contains 454 and the second one 405 annotated bees. The sets are taken form a different series. For set ’small’ the cameras had a distance of 3 meters to the bee curtain. Set ’big’ is taken from an experimental series where the acquisition setup had a distance of 2 meters to the bee curtain. The set ’big’ got its name by the fact that the bees are represented in a higher resolution in the image.

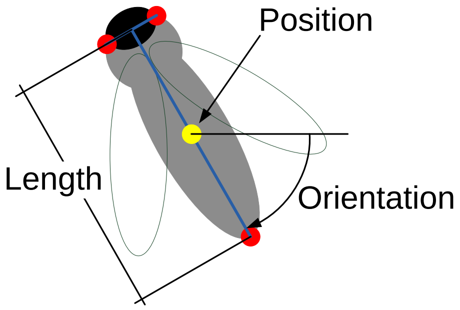


Figure 6.6: Schematic illustration of the significant points (red) of a bee used to define the ground truth data and the features describing a bee. These features are the position (center of gravity), the length and the orientation.

To evaluate the segmentation results of the different methods a common description for a bee has to be defined. The so called features for a bee consist of the following values:

- the position of the bee that represents the center of gravity,
- the length of the abdomen
- and the orientation according to the horizon

and are stored in a data structure. These features are illustrated in Figure 6.6 and in Figure 6.8 for test set ’big’.

6.2.2 Parameterization

As described in Section 6.1.4 the image sequences were captured at two different distances to the bee curtain. This leads the bees to be depicted at two different resolutions. For each resolution a parameter set is required to segment the bees. These two sets of parameters can be used for all captured sequences.

As described in Section 4.1 the MSER method is suitable to detect stable regions and can be parameterized by a stability threshold and three parameters to constrain the size of the stable region. For the two test sets the parameters are summarized in Table 6.2.

To initialize the shape prior segmentation the centers and orientations from the MSER experiment are used. SPSeg is parameterized by a search region in x- and y-axis that is

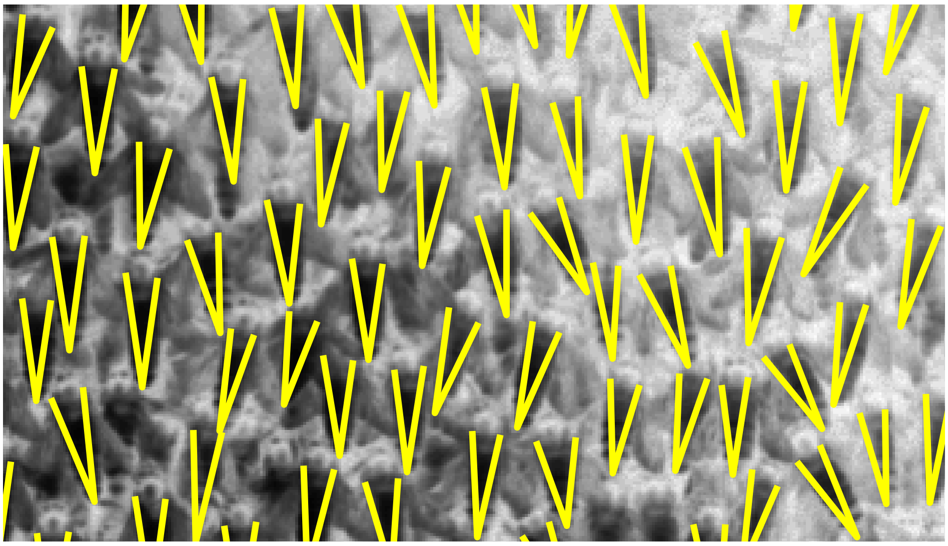


Figure 6.7: Subset of the manually annotated bees that are used as ground truth for the evaluation of the segmentation methods

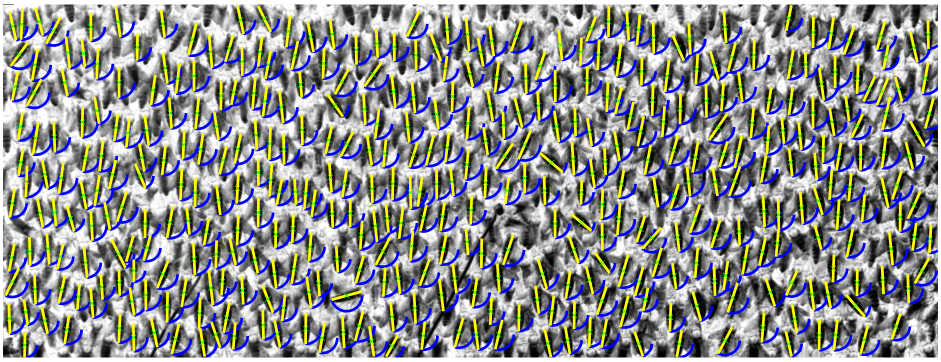


Figure 6.8: The features describing the 405 bees of the ground truth of set 'big'. The positions are marked with green 'x', the lengths of the abdomen are illustrated as yellow lines and the orientation are represented as blue arcs.

Parameter	**Test set**	
	'small'	**'big'**
Stability Threshold	20	20
Maximum Length [px]	80	100
Minimum Length [px]	30	40
Maximum Width [px]	30	40

Table 6.2: Parameters used for MSER segmentation of the two test sets.

scanned in x- and y-steps. Further, maximal allowed rotation and scale are set that are scanned in steps. For SPSeg a prior of a single bee has to be generated. The two priors use for the test sets are illustrated with the parameters in Table 6.3.

Parameter	**Test set**	
	’small’	**’big’**
Region X [px]	3	3
Region Y [px]	3	3
Region Rotation [°]	17.19	17.19
Region Scale	0.1	0.1
Step X [px]	1	1
Step Y [px]	1	1
Step Rotation [°]	2.86	2.86
Step Scale	0.05	0.05
Shape prior		

Table 6.3: Parameters used for SPSeg of the two test sets.

The NCC segmentation method can be parameterized by the maxDist that is responsible to find double segmented bees and filter them, startAngle and angleExtent that define the orientation variation range of one template, minScore that represents the threshold of the correlation score and maxOverlay that defines the maximal overlay of the regions using one template. The used parameters for the segmentation of the two test images are summarized in Table 6.4.

Parameter	**Test set**	
	’small’	**’big’**
Maximal Distance [px]	15	20
Start Angle [°]	-17.19	-17.19
Angle Extent [°]	34.38	34.38
Minimum Score	0.75	0.75
Maximum Overlay	0.3	0.3

Table 6.4: NCC parameters used for segmentation of the two test sets.

6.2.3 Segmentation Results

This section summarizes the results for the three experiments defined in 6.2 using the ground truth. The outputs of the experiments for the separate test sets do not vary that much and lead to combining of the absolute results to one. The illustrations of the results are demonstrated on test set ’big’ for the rest of this section.

Method	**Δ Position [px]**	**Δ Orientation [°]**	**Δ Length [px]**
NCC	2.757	2.120	3.330
MSER	3.303	2.750	8.504
SPSeg	3.791	2.464	6.347

Table 6.5: Geometric accuracy of the mean differences of the features. Δ Position describes the mean distance to the GT position in pixel. Δ Orientation describes the mean deviation in orientation in degree and Δ Length the mean length difference in pixel.

6.2.3.1 Recognition Results

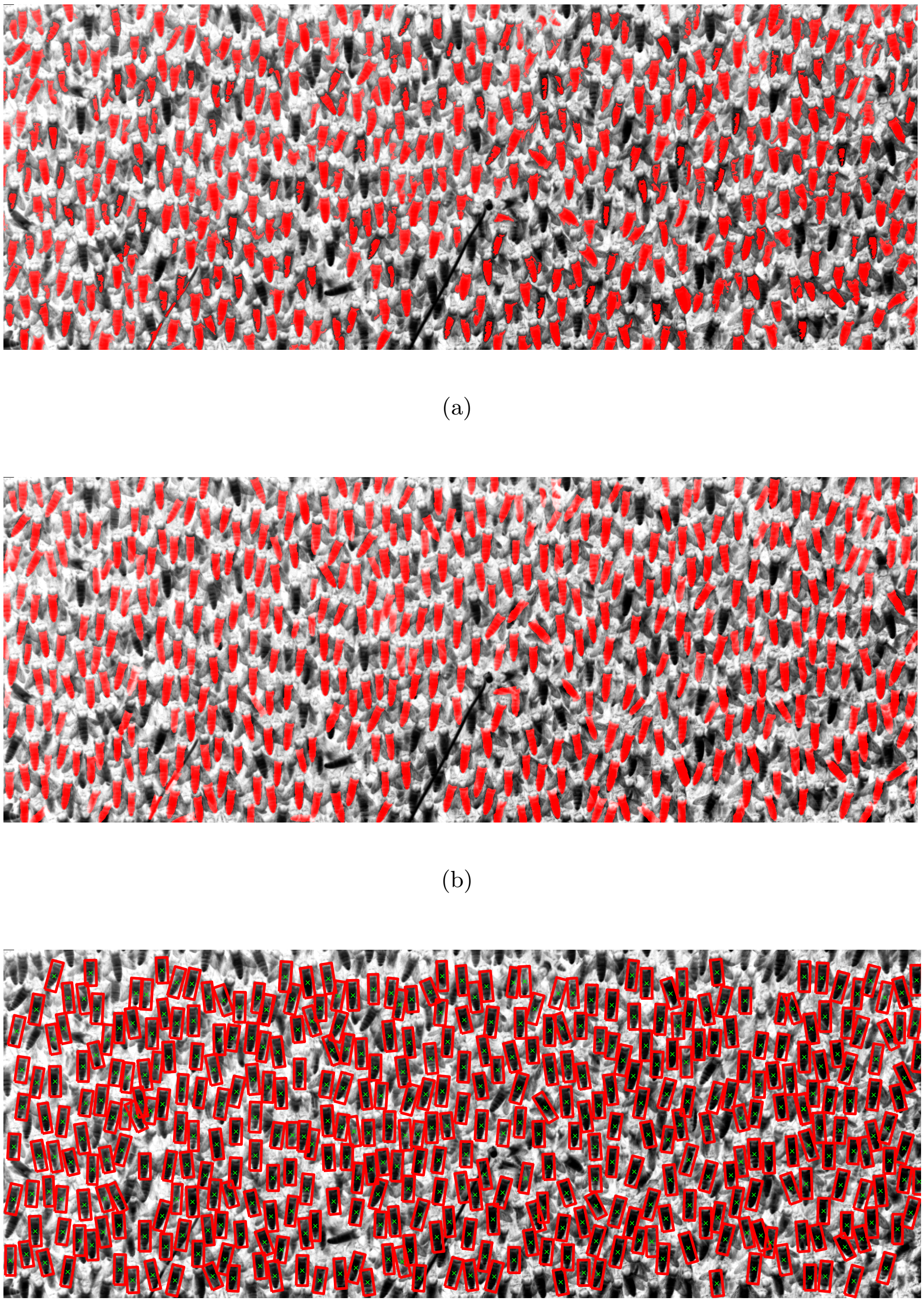
Figure 6.9 shows the results of the segmentation. On the one hand the regions are overlaid to the input images and on the other hand the boundary of the resulting area is plotted as an rectangle for NCC. The segmentation leads to each 895 regions for MSER and SPSeg and 897 for NCC.

The results for the recognition rate as described in Section 6.2 are summarized in Figure 6.10. There it can be seen that MSER and SPSeg perform nearly the same because SPSeg uses the centers extracted with MSER as initialization and does not add or remove any regions. The small differences relies on the calculation of the number of true positives as described in Section 6.2 and the movement of the centers caused by SPSeg. The just mentioned two methods are outperformed by NCC with 794 true positives.

6.2.3.2 Reconstruction Capability

To be able to compare the geometric accuracy the features for the segmented regions have to be calculated and are shown in Figure 6.11. The first results that give an indication of the capability of reconstruction are the comparison of the length and the orientation distributions of the bees with the GT as described in Section 6.2. Figure 6.2.3.2 shows that the distribution of lengths of NCC fit the GT distribution best. Although the fact that the length distribution of NCC has only few peaks is based on the templates exhibit distinct lengths and NCC does not scale the templates. Figure 6.2.3.2 shows the distribution of the angles where MSER seems to outperform the others.

A more detailed evaluation is done by calculating distributions for the feature differences as described in Section 6.2. Figure 6.14, 6.15 and 6.16 show the difference histograms for the evaluation of the geometric accuracy features. There having a distribution close to zero means that the method is suitable to represent the geometric feature. The mean for the single geometric features for the different segmentation methods are summarized in Table 6.5. It shows that NCC has the smallest mean for all three features. It can also be seen that using SPSeg results in a better representation than MSER although the mean position offset is smaller for MSER. The experiment shows that the NCC dominates the other two methods by having smaller mean values for each geometric feature.



(a)

(b)

(c)

Figure 6.9: Resulting segmentations for test set ’big’. In (a) the 439 segmented regions of MSER are overlaid to the original image in red. (b) shows the 439 regions resulting from SPSeg and (c) the 403 regions segmented using NCC that are drawn in the original image as rectangles.

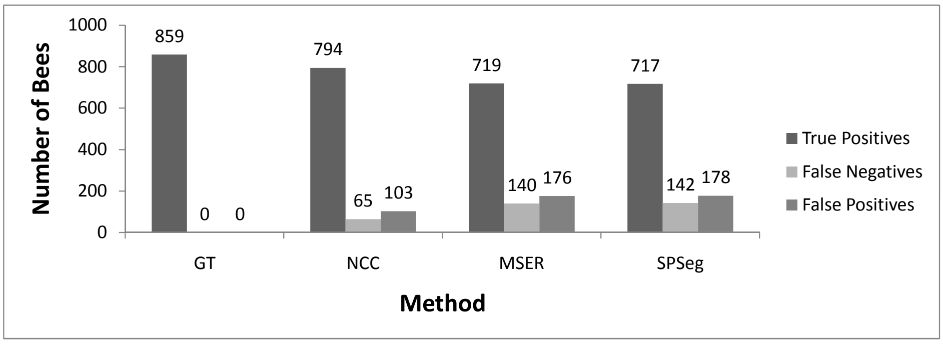


Figure 6.10: Statistical evaluation of the recognition performance of the three segmentation methods.

	NCC [s]	**MSER** [s]	**SPSeg** [s]
Algorithm	2.75	3.41	1650.00
Pre-processing	0.00	0.00	13.90
Post-processing	0.21	10.50	0.00
Total	2.96	13.91	1663.90

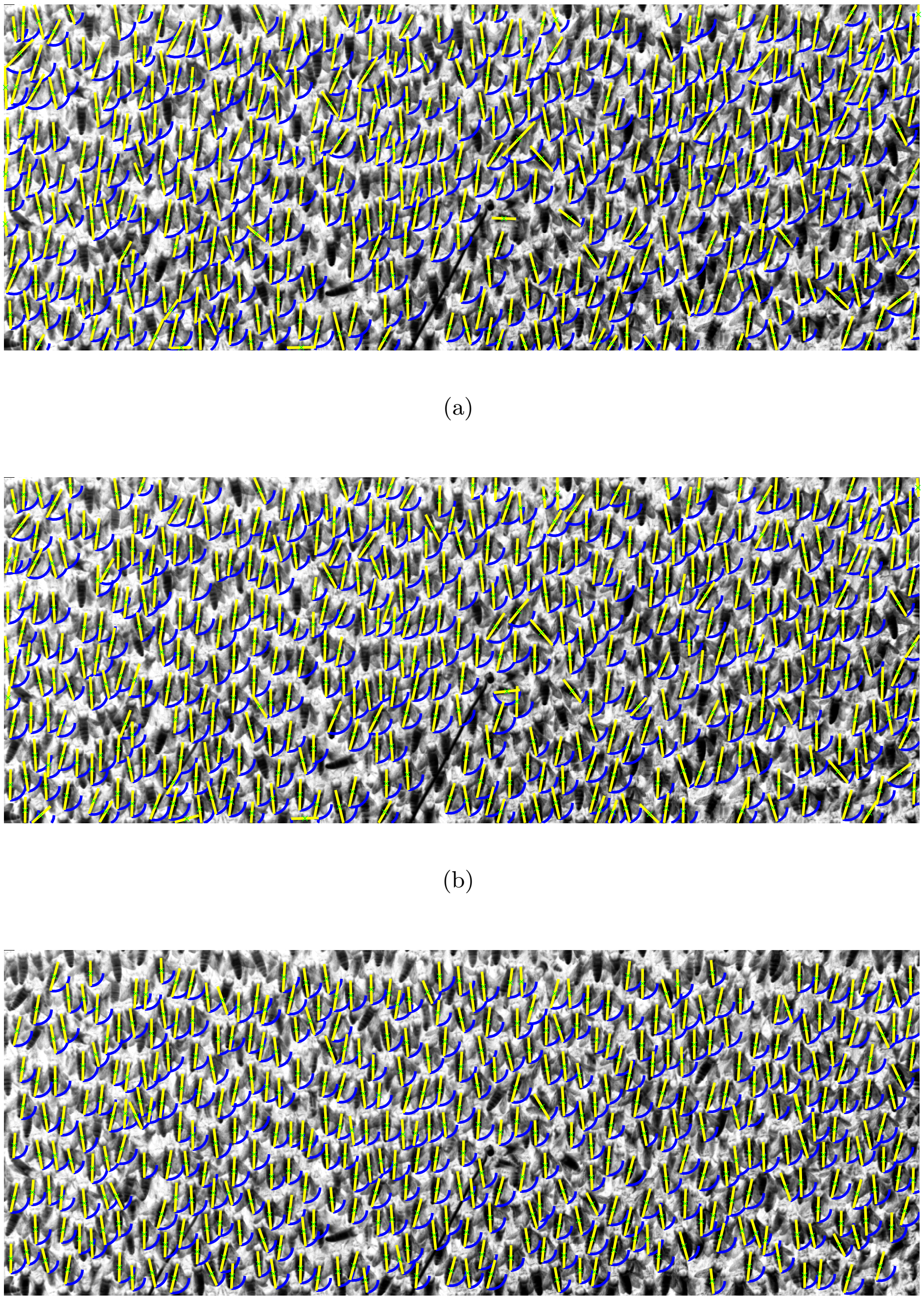
Table 6.6: Execution times. NCC does not need any pre-processing and the post-processing represents the merging of multiple templates. MSER also only requires a post-processing step to filter overlays. The pre-processing of the SPSeg denotes the initialization using MSER. For SPSeg no post-processing is required.

6.2.3.3 Execution Time

The evaluation of the runtime of the segmentation methods results NCC being the fastest with 2.96*s*. Next is the MSER with an execution time of about 4.7 times the time required for NCC and although the SPSeg is calculated on a GPU the execution time is bad compared to NCC. A detailed evaluation of the times can be found in Table 6.6 that shows the mean times for the different segmentation methods. The evaluation of the time has to be considered tentatively because the segmentation methods differ in implementation. For NCC a highly optimized library that utilizes multi core CPUs has been taken. MSER was implemented in C++ and SPSeg was a experimental version calculated on the GPU.

6.3 Detailed Evaluation of NCC Method

As described in Section 6.2.3 NCC is the appropriate method to segment bees so it is useful to make a detailed evaluation of this method to be able to get the best results using this method. In the following subsections the determination of the threshold, the influence of the template length, the amount of useful patches and the rotational constraint have been evaluated for test set 'small'.

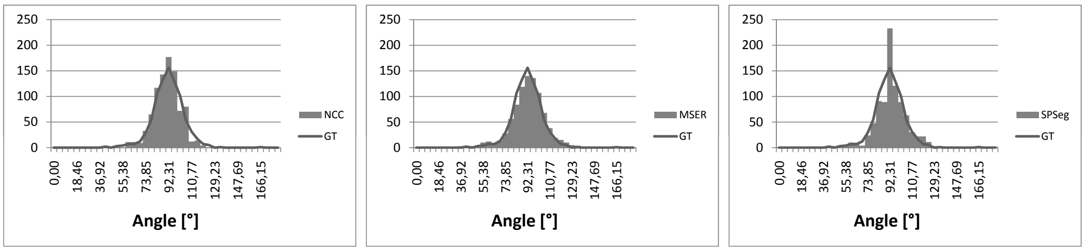


(a)

(b)

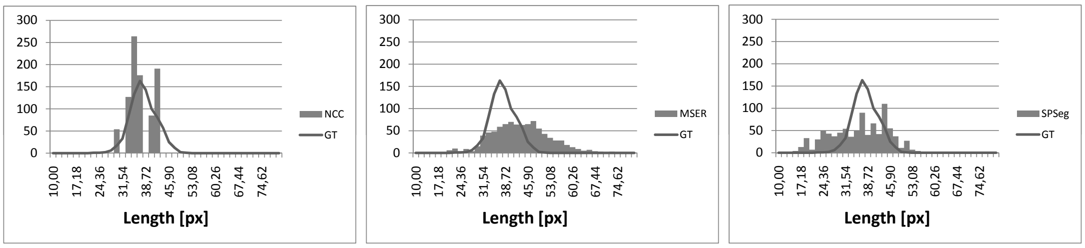
(c)

Figure 6.11: Resulting features for test set 'big'. (a) shows the features of MSER, (b) of SPSeg and (c) of NCC. The positions are marked with green 'x', the lengths of the abdomen are illustrated as yellow lines and the orientation are represented as blue arcs.



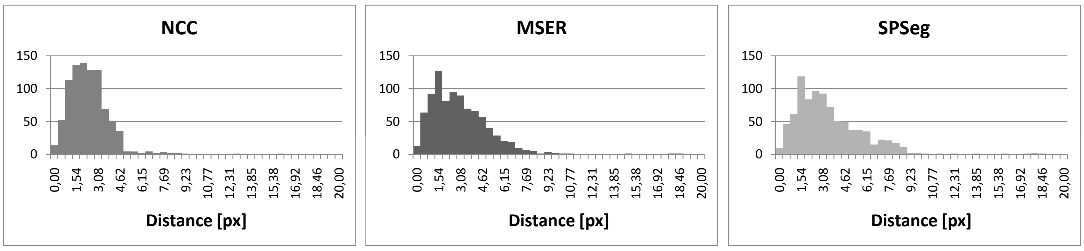
(a) NCC (b) MSER (c) SPSeg

Figure 6.12: Distribution of the occurring angles in degree of the bees. The dark gray line denotes the distribution of the GT.



(a) NCC (b) MSER (c) SPSeg

Figure 6.13: Distribution for the occurring lengths in pixel of the bees. The dark gray line denotes the distribution of the GT.

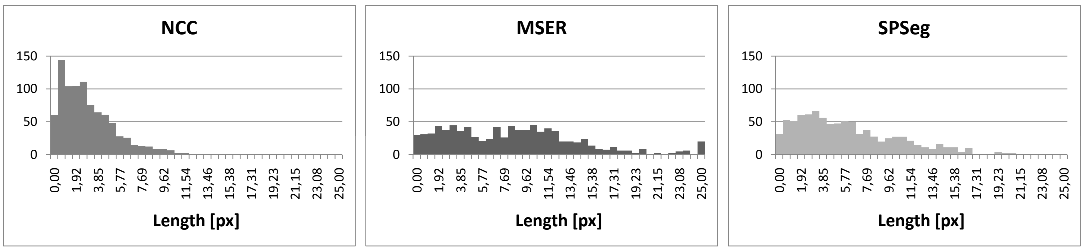


(a) NCC (b) MSER (c) SPSeg

Figure 6.14: Distribution of position differences of corresponding bees

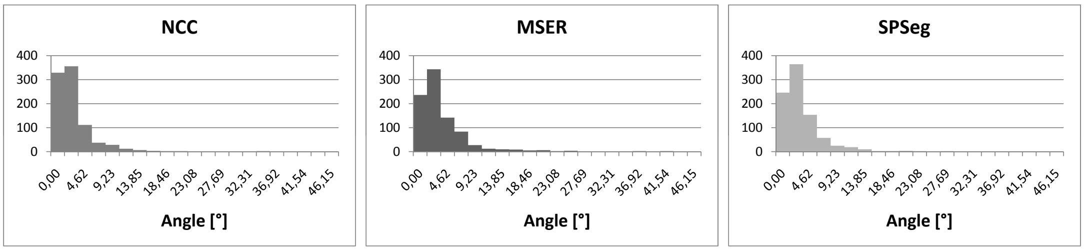
6.3.1 Determination of the Similarity Threshold

The similarity threshold of the NCC describes the minimum value of correlation. This means that the part of the image I and the template t must have a correlation greater than the similarity threshold. A good threshold is characterized by a high rate of true positives and a low rate of false positives.



(a) NCC (b) MSER (c) SPSeg

Figure 6.15: Distribution of length differences of corresponding bees



(a) NCC (b) MSER (c) SPSeg

Figure 6.16: Distribution of orientation differences of corresponding bees

The similarity threshold for the NCC has been evaluated by calculating the true and false positives of thresholds ranging from zero to one in steps of 0.01. Similarity values below zero represent inverted bees that are not possible for the input images and are not considered in the evaluation. For the calculation of NCC seven different templates of different lengths have been used. These results are illustrated in Figure 6.17. By examining the diagram it can be seen that a similarity threshold of 0.75 is a good choice to find about 95% of true positives. The choise has been validated by visual inspection for images of the same experimental series. The false positives are tolerated because the segmentation method segments more bees than are labeled in the GT and false matches are filtered by the preselection constraint using epipolar geometry later.

6.3.2 Influence of the Template Size

In this experiment different templates of a bee has been taken to calculate the NCC. The difference was characterized by the size of the abdomen. For the evaluation of the influence of the size a small, a medium and a large bee were taken as template.

In Figure 6.18 the evaluation of the segmentation results are plotted for a similarity threshold variation from zero to one. Taking a similarity threshold of 0.75 it can be seen

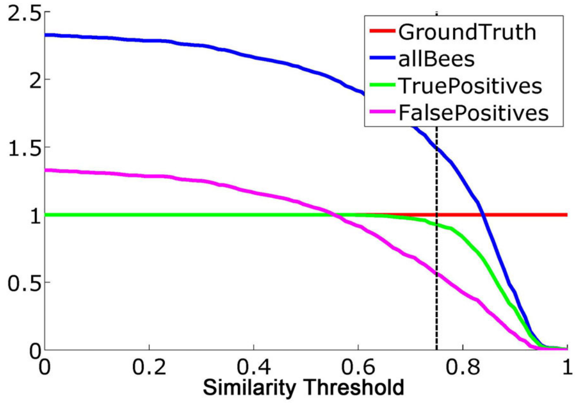
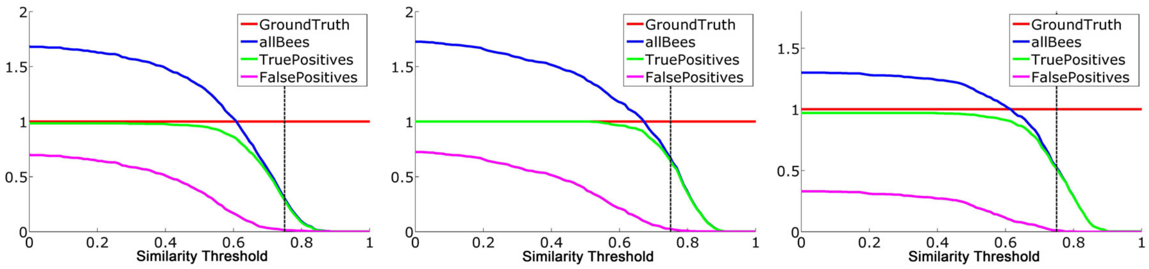


Figure 6.17: Evaluation of the NCC threshold normalized to the ground truth. The ground truth is represented in red, the true positives in green, the false positives in magenta and the sum of segmented bees in blue.

	Template size		
	Small	**Medium**	**Large**
True Positives	0.300	0.656	0.524

Table 6.7: Summary of the true positives for single templates with different sizes at a similarity threshold of 0.75.

that the medium sized bee template has the best true positives and an acceptable amount of false positives and represents the ground truth data best. The true positives at a similarity threshold of 0.75 are summarized in Table 6.7.



(a) Small

(b) Medium

(c) Large

Figure 6.18: Evaluation of the NCC threshold normalized to the ground truth for three different length of templates. The ground truth is represented in red, the true positives in green, the false positives in magenta and the sum of segmented bees in blue.

6.3.3 Number of Templates

To evaluate the relevance of the number of different sized templates the true and false positives have been calculated for one to eight templates. This has been done by consecutive adding a new template to the previous ones. The other parameters were fixed at a similarity threshold of 0.75, an angle step of 5.02° and an angle extent of 34.378° symmetrically around the template orientation. The single parameters are described in Section 4.3.

As the results in Figure 6.19 show for increasing number of templates the true positives and false positives increase but at different rates. So it can be seen that taking four templates leads to a good true positive value and a not that high false positive value so that the ground truth data is represented by a set of about four templates best.

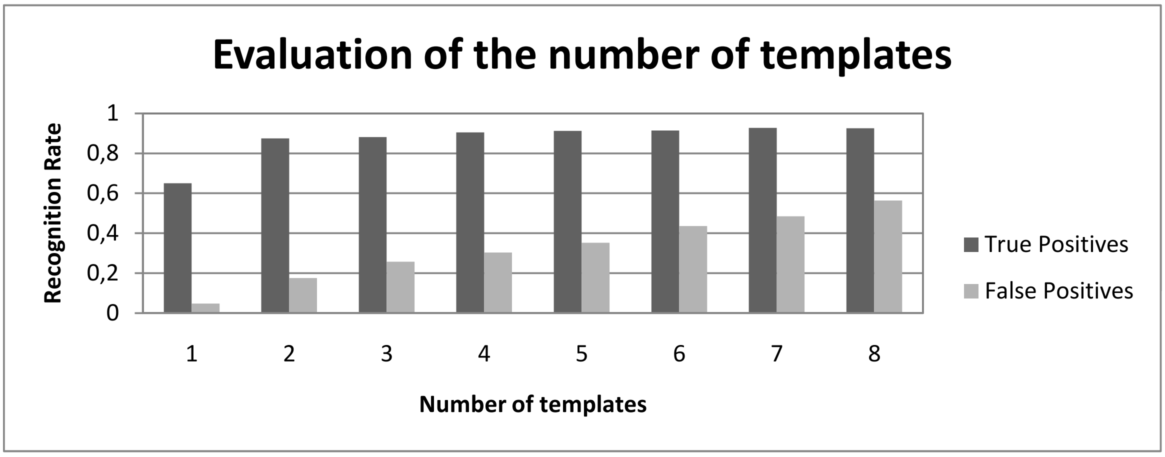


Figure 6.19: Evaluation of the true and false positives for increasing number of templates. The used parameters are a similarity threshold of 0.75 and an angle extent of 34.378°

6.3.4 Rotational Constraint

The last parameter used for the calculation of the NCC to evaluate is the angle extent as defined in Section 4.3. It describes how much the template is rotated symmetrically around the original pose. In this experiment the same set of four templates at a similarity threshold of 0.75 and an angle step of 2.8648° have been compared to each other according to the angle extent from 0 to 68.7° in seven steps.

The results illustrated in Figure 6.20 show that the higher the extent of the angle the more true and false positives are segmented. The choice of the angle extent is a trade of a high detection rate of true positives and a low detection rate of false positives. So an angle extent of 34.4° fulfills these requirements and the number of false positives are tolerated for the same reasons as in Section 6.3.1.

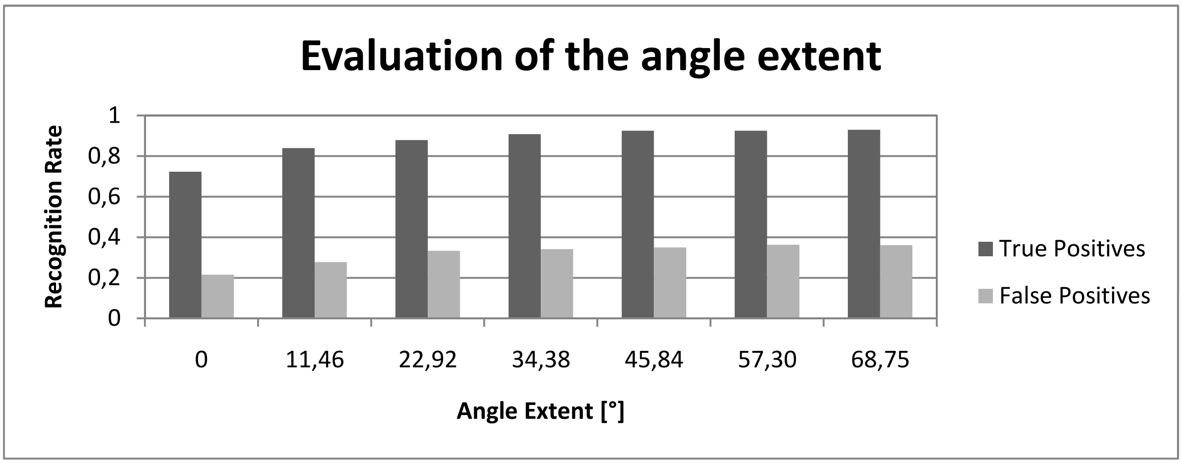


Figure 6.20: Evaluation of the true and false positives for increasing angle extents from 0 to 68.7°.

6.4 Evaluation of Matching Performance

This section deals with the evaluation of the stereo matching algorithm using reduced graph cuts as described in Chapter 5. First the functionality of the graph is demonstrated by varying the smoothness factor. Then the appropriate smoothness factor for stereo matching is evaluated. Followed by the evaluation of the horizontal search range and its effects on execution time.

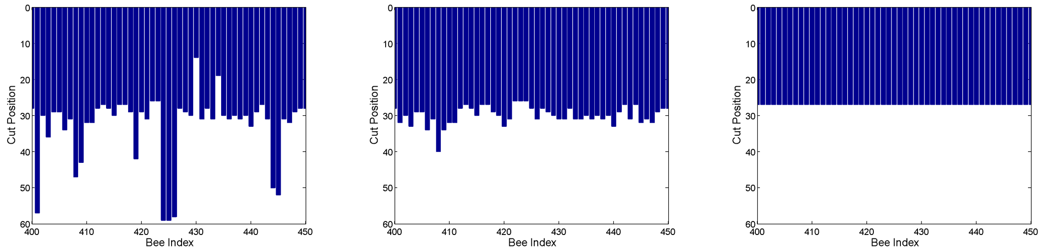
6.4.1 Functionality of Graph Cut

This experiment aims to show the functionality of the graph cut used for finding correspondences. To be more precise the influence of the smoothing term should be demonstrated. Given are two rectified stereo images, the segmentation of the bees by the centers of gravity and descriptions of the single bees by image patches. The tree is generated as described in Section 5.1.

Evaluating the positions of the minimum-cut in the tree for three different weights of the smoothing term should lead to:

- a horizontal cut through the graph for a high smoothing weight,
- a strong varying of the cut positions for no smoothing and
- a moderate varying of the cut positions for a value in between.

For the comparison the smoothing factor α was set to 0, 0.1 and 10. The resulting cuts through the tree are represented in Figure 6.21. A smoothing of 0 represents a graph that does not contain any neighborhood constraints and so the similarity maxima are selected for the cut as shown in Figure 6.21(a). It leads to a strong variation in positions. A smoothing factor of 10 forces the graph cut not to cut any neighboring links and so cuts the tree horizontally as illustrated in Figure 6.21(c). Using a factor in between results in a moderate variation of the positions of the tree as shown in Figure 6.21(b). This shows that the variation of α influences the smoothness of the cut as expected and controls the influence of the neighborhood.



(a) $\alpha = 0$ (b) $\alpha = 0.1$ (c) $\alpha = 10$

Figure 6.21: Illustration of the cut through the graph by different smoothing factors α.

6.4.2 Evaluation of Smoothness Factor (α)

Having α to control the smoothness of the cut, it has to be evaluated which smoothness factor is suitable to get good correspondences. Therefore a test set has been established. It consists of segmented bees in a stereo image pair. Further, to evaluate the correctness of the matches, matches were manually labeled and provide the ground truth (GT) as shown in Figure 6.22. The GT consists of 237 pairs. For the evaluation of true and false matches only bees of the first image that were labeled in the GT are matched with all the segmented bees in the second image.

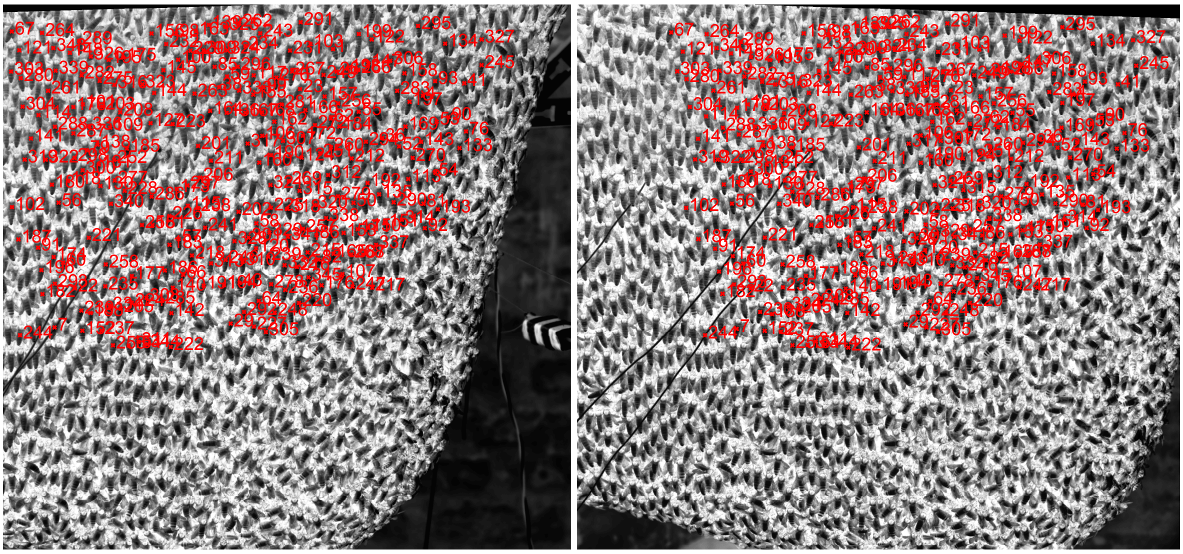


Figure 6.22: Ground truth set to evaluate the matching results.

α	**Ground Truth**	**True Matches**	**False Matches**
0	237	209	28
0.01	237	235	2
0.05	237	236	1
0.1	237	235	2
0.2	237	228	9
1	237	211	26
10	237	78	159

Table 6.8: Summary of the true and false matches in absolute values for varying smoothness factors α.

For the evaluation the stereo matching as described in Section 5.1 was calculated for smoothness factor α varying from 0 to 10. All other parameters were set constant ($precision = 100$; $span = 300$; $knn = 5$). The resulting values for true and false positives can be seen in Table 6.8 and shows that a smoothness factor of 0.05 leads to the best solution for the correspondence problem. Additionally it can be seen that associating bees with maximal similarity as corresponding ones (equals $\alpha = 0$) leads to worse results and so verifies the need of graph cuts.

6.4.3 Evaluation of Search-Region

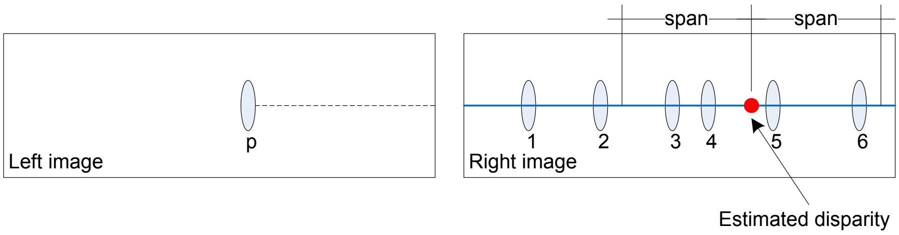


Figure 6.23: Schematic illustration of the horizontal search range parameterized by *span*. The bees 3 to 6 are possible candidates.

This experiment should evaluate the horizontal search range, *span*. For varying span the required computing time will be measured and the matching results will be compared to the same ground truth (GT) as in the previous experiment (Section 6.4.2) to evaluate the true and false matches. *span* constraints the number of possible correspondences by limiting the horizontal pixel range around the estimated disparity as illustrated in Figure 6.23. All other parameters were set constant ($precision = 100$; $\alpha = 0.1$; $knn = 5$). So choosing *span* too small leads to having no possible candidates if the estimated disparity is not that exact. The search region also influences the graph dimension and so the computation time. By increasing *span* the number of false matches should decreases and the computation time should increase.

Having a look at the results summarized in Table 6.9 it can be seen that the expected properties are satisfied. The number of false matches decreases and at a span of 300 only

Span	Ground Truth	True Matches	False Matches	Time [s]
20	237	107	130	27.7
30	237	152	85	29.9
40	237	199	38	35.0
50	237	216	21	37.7
100	237	221	16	51.1
200	237	233	4	90.2
300	237	235	2	127.3
400	237	235	2	179.2
500	237	236	1	256.2
600	237	236	1	372.5

Table 6.9: Summary of the true and false matches in absolute values for varying span and the required computing time @Intel Core2 Duo 2.54GHz.

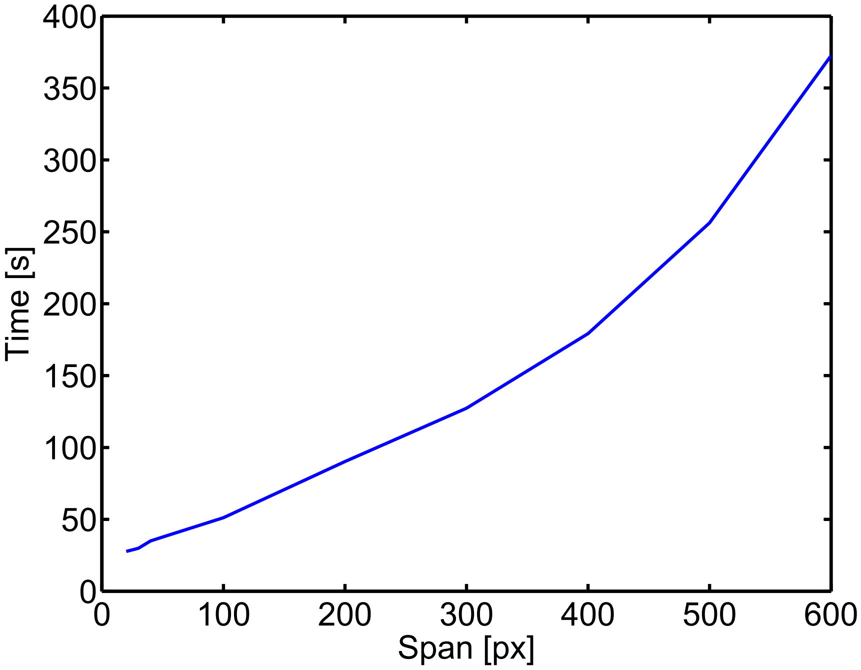


Figure 6.24: The calculation time at a Intel Core2 Duo 2,54GHz plotted over the *span*.

2 false matches of 237 is a good result. The computation time at the test computer (Intel Core2 Duo 2,54GHz) also shows the increasing time. Plotting the time over the *span* shows a growing of the computing time as depicted in Figure 6.24.

6.5 Summary and Conclusion

This chapter dealt of the evaluation of the image acquisition setup, the segmentation methods and the stereo matching algorithm. The image acquisition setup as described in Chapter 3 has been analyzed according the synchronization of the two cameras to have proper correspondences in time of the acquired image sequences. The synchronization was inspected by capturing a flashing light and leads to synchronous ramps for the illumination changes. To figure out if the setup can operate in nature for a working day with a rechargeable battery the total power consumption at maximum load was measured and resulted in an operation time of more than seven hours using a $100Ah$ battery. This has been proved when the acquisition setup has been in use in Nepal. A possible improvement of the recording time may be achieved by a accurate acquisition buffer method. The possible improvements of 3 seconds were not worth a redesign and with a recording time of 15 seconds the required capturing of one wave, which lasts about one second, is satisfied easily.

The segmentation methods that were evaluated are template matching using NCC, maximally stable extremal regions and shape prior segmentation as described in Chapter 4. The evaluation was parted in three experiments. First the recognition accuracy was evaluated and led template matching using NCC being the appropriate method to segment the bees because $92,4\%$ of the ground truth bees were segmented. Also for the reconstruction ability NCC outperformed the other methods and gained a mean deviation in position of $2.757px$, a mean deviation in orientation of $2.120°$ and a mean deviation in size of $3.330px$. Finally the execution time required by NCC is also shorter than the other methods. This has to do with the highly optimized implementation used for NCC calculation.

Because template matching using NCC outperformed the other methods a detailed evaluation was done for its parameters. This led to an appropriate set by selecting a similarity threshold of 0.75, an angle extent of 34.4° and choosing four templates of medium bee size.

Having a look at the length distribution of NCC (Figure 6.2.3.2) it does not approximate the GT data well. This results form NCC being calculated only for a fix template size. Extending the template by using a pyramid of scales the length distribution should approximate the GT better. The fact that the bees are represented in 3D only by one point permits to forgo without a scale space to save computation time.

The evaluation of the stereo matching algorithm using graph cuts as presented in Chapter 5 results in an optimization of the correspondences. It has been shown that the smoothing factor α controls the smoothness of the cut and so the influence of the neighborhood. A detailed evaluation of the smoothness factor led α to be selected between 0.1 and 0.01 to get qualitative stereo matches for the test image set. The selection of the horizontal search range results in a trade-off according the execution time and the accuracy. A very small search range will lead to a very fast computation, but it is likely that the corresponding bee is not contained in the set of possible candidates because of the rough disparity estimation. Selecting a high value for the search range the computation time will increase as shown in Figure 6.24. A horizontal search range of $300px$ makes a compromise between the accuracy and the required time, although focus lies on accurate correspondences.

Compared to a maximum similarity search the stereo matching algorithm using graph cuts fixes 96.4% of the false matches using optimized parameters. Using the parameter set ($precision = 100$, $\alpha = 0.05$, $span = 300$ and $knn = 5$) the best results were achieved and only 1 of 237 matches were false. However the maximum similarity search results in 28 false matches.

By performing the experiments parameter settings for the experimental series have been figured out. These parameters have not been validated. The usability of the parameters has only been tested exemplary using visual inspection for some images of the experimental series. So the validation of the parameter sets has to be done in future.

Chapter 7

Discussion and Outlook

A hardware setup to acquire image sequences in outdoor environments and a framework to reconstruct the bee curtain in 3D was presented. The hardware setup was designed to satisfy the requirements of capturing the defense wave of giant honeybees. These requirements are a field of view of about $700 \times 700 \times 100\text{mm}^3$ (height × width × depth), a resolution of 15px in width for a single bee, a frame rate of 60fps and the ability to operate outdoor.

During the expedition to Nepal the acquisition setup has been tested in the field. It emerged that the setup captured high quality images. The stereo images are suitable to reconstruct the single bees presented by a single point. To be able to reconstruct the abdomen flip of the bees a higher resolution in space and time would be required. The reason is the fast abdomen flips during a defense wave. These flips are only blurred represented in the images at a frame rate of 60fps and an exposure time of 5ms. Also the angle the bees flip their abdomen was underestimated. By analyzing the images it can be seen that the undersides of the bees are visible to the cameras that are mounted at the same level as the top of the nest. Because of this the cameras should be mounted higher and this would lead to constrain the field of view because of the limited depth of focus. To represent the same field of view multiple stereo setups would be required. But the positioning of such a setup would be difficult because the bee colonies are located at exposed places and largely under a jut. Selecting the two test colonies caused difficulties because of the lack of an ability to mount the acquisition setup with the constrained resources. In most cases a lifting platform that reaches 10 or more meters would be required.

In Chapter 4 three methods to segment the bees were proposed: Maximally stable extremal regions (MSER), shape prior segmentation (SPSeg) and template matching using NCC. The evaluation of the methods was presented in Section 6.2. By comparing the recognition rate and the reconstruction capability of the methods template matching performed best, although the length distribution of the segmented bees vary a lot (see Figure 6.13(a)). The few different lengths result from one fixed length for each selected template. Only by selecting multiple templates of different size the method was made slightly invariant to scale.

Inspired by the result of the compared segmentation methods a detailed evaluation of the parameters used for template matching was performed in Section 6.3. A set of parameters (similarity threshold = 0.75, angle extent = 34.37°, angle step = 2.86° and number of selected templates = 4) has been extracted to be suitable. Using this values a quite high rate of false positives compared to ground truth data are present (see Figure 6.17). The reason for tolerating the false positives are the fact that the false positives contain bees that were not labeled in the ground truth and by filtering according to the epipolar constraint false matches are eliminated.

The correspondence problem is solved by calculating the minimum cut of a reduced graph as described in Chapter 5. The graph is made up of chains for each bee. These chains represent possible position slots around an estimated disparity that are filled with possible matching candidates if present. Further, the chains are interconnected at each level by neighboring links that enable to control the smoothness of the cut. In Section 6.4 the functionality of the graph cut had been shown. Finally, the resulting correspondences were compared with manually labeled ground truth data. For an optimized parameter set that is suitable for one experimental series a performance of one false match of 237 matches has been gained. Compared to a simple maximum similarity search, 96.4% of the false matches were fixed. In absolute numbers a correction from 28 to 1 false matche was gained. The parameter sets do not have to be adjusted for each experiment. By visual inspection it has been shown that one set of parameters can be used for one experimental series. So three sets of parameters are required for all present data. A detailed validation of the parameter sets remains for future work.

Having a look at the tracking methods of insects presented by Veeraraghavan et al. [VCS08] and Maitra et al. [MSS09] they are tracking single individuals in 2D. Using the presented framework the bees can be reconstructed in 3D and gives the ability to track the reconstructed individuals in 3D. Having the 3D position of the single bees over the time the defense waves can be analyzed by the z-movement of the individuals, the speed, the spreading direction and the participation of second layer bees. A prototype stereo tracker has already been designed and the results satisfy the requirements of the zoologists to analyze the movement of the single bees during the shimmering behavior. An example of the movement towards the cameras of some bees can be seen in Figure 7.1. Further, the presented framework enables a semi-automatic segmentation of the individuals. Only by selecting some example bees the bee curtain can be segmented into individuals.

The reconstructed data has already been used by the zoologist for a publication at the Meeting of the Austrian Neuroscience Association (ANA) in Salzburg [WHM$^+$09]. There the z-movement of the bees has been overlaid color coded to the frames to illustrate the movements while shimmering behavior.

There are further expeditions planed to study the defense waves of the giant honeybees, but up to now it is not sure if the experiment with the stereo acquisition setup will be

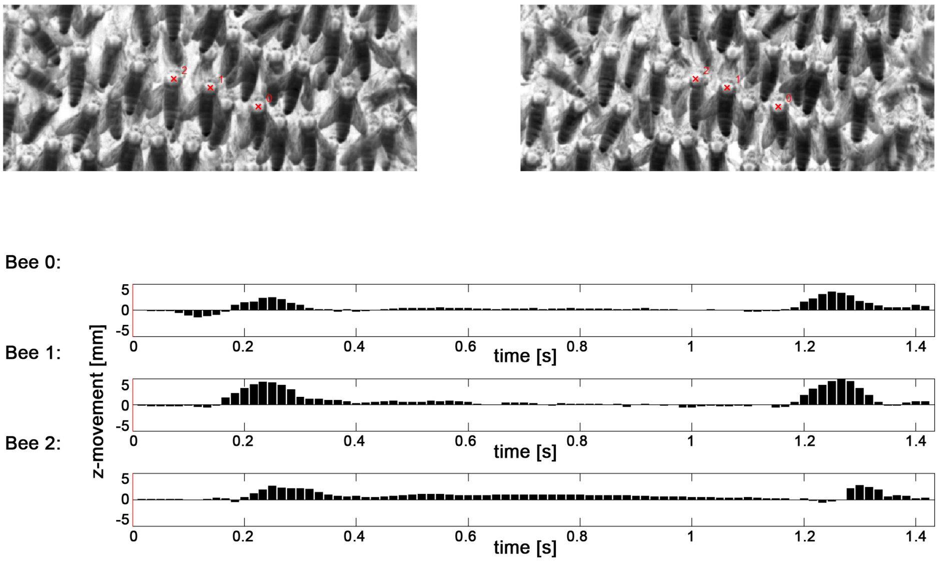


Figure 7.1: Z-movement of three example bees while performing defense waves from right to left.

repeated. Primarily the acquired data has to be processed using the presented framework and an adequate tracking method. With the prototype stereo tracker the reconstruction of a image sequence of 900 frames would last about 3 hours. So an improvement of the tracking is required.

If the experiments are repeated the stereo setup should be improved. The rig has to be made more stiff by using four instead of two main bars where the cameras are mounted on. Further, the location to mount the acquisition setup should be prepared more carefully and so prevent a deformation of the setup.

Future work considering the segmentation methods would be to improve the template based matching method by either using an universal template that has been learned and made scale and orientation invariant by resizing and rotating or to use a database of standard bees. This database should contain single bees of different sizes that were made orientation invariant.

An extension to the 3D reconstruction would be to represent the bees by a rigid model that is fitted to some supporting points in 3D. The rigid model could be an ellipsoid for the whole bee or to have a more detailed approximation, to use an ellipsoid for the head and the abdomen and a sphere for the thorax. The positions of the three elements should be constrained to represent only possible deformations of a bee.

The tracking of the bees could be realized by template matching in the stereo image sequences. There the variation of the representation has to be considered. By analyzing the image sequence it can be seen that the thorax changes its representation caused by the flipping wings. Further, the abdomen is visible from the underside and so undergoes a variation in representation. To combine the tracking results of corresponding image points the constraint of the epipolar geometry has to be verified and corrected for each frame. The verification of the tracking result in single points may be gained by comparing to measurements using a laser vibrometer. There the z-movements of the tracked point should be the same as the results of the laser vibrometer aimed at the same position. The direct comparison of the results is only possible if the orientation of the laser vibrometer is known in the coordinate system of the 3D reconstruction.

Appendix A

Abbreviations and Acronyms

A.1 Abbreviations

1D	one dimension(al)
2D	two dimension(al)
3D	three dimension(al)
s	second
ms	millisecond
μ**s**	microsecond
ns	nanoseconds
m	meter
mm	millimeter
kg	kilogram
px	pixel
Mpx	mega pixel
fps	frame per second
V	volt
A	ampere
W	watt
MB	megabyte
GB	gigabyte
TB	terabyte
Mbit/s	megabits per second (1.000.000 bits per second)
Gbit/s	gigabits per second (1.000.000.000 bits per second)

A.2 Acronyms

ATA	advanced technology attachment
CAD	computer-aided design

CCD	charge coupled device
CMOS	complementary metal oxide semiconductor
DOF	degree of freedom
GCS	growing correspondence seeds
GT	ground truth
HDD	hard disk drive
ISA	industry standard architecture
MSER	maximally stable extremal regions
NCC	normalized cross correlation
PnP	plug & play
RAM	random access memory
RMS	root mean square
ROI	region of interest
SAD	sum of absolute differences
SPSeg	shape prior segmentation
SSD	sum of squared differences
SSSD	sum of SSD
UPS	uninterrupted power supply unit
USB	universal serial bus

Bibliography

[BEV$^+$05] X. Bresson, S. Esedoglu, P. Vandergheynst, J. Thiran, and S. Osher. Global Minimizers of The Active Contour/Snake Model. In *Free Boundary Problems: Theory and Applications*, ISCAS. IEEE, 2005.

[BEV$^+$07] Xavier Bresson, Selim Esedoglu, Pierre Vandergheynst, Jean-Philippe Thiran, and Stanley Osher. Fast global minimization of the active contour/snake model. *J. Math. Imaging Vis.*, 28(2):151–167, 2007.

[BVZ99] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23, 1999.

[DB06] M. Donoser and H. Bischof. Efficient maximally stable extremal region (mser) tracking. In *CVPR*, pages I: 553–560, 2006.

[DZC07] H. Du, D.P. Zou, and Y.Q. Chen. Relative epipolar motion of tracked features for correspondence in binocular stereo. In *ICCV*, pages 1–8, 2007.

[GI89] Dan Gusfield and Robert W. Irving. *The Stable Marriage Problem: Structure and Algorithms*. The MIT Press, August 1989.

[GT88] Andrew V. Goldberg and Robert E. Tarjan. A new approach to the maximum flow problem. *Journal of the ACM*, 35:921–940, 1988.

[GW02] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*, chapter Image Segmentation, pages 567–642. Prentice-Hall, Inc. Upper Saddle River, New Jersey 07458, 2nd edition, 2002.

[HC04] Li Hong and George Chen. Segment-based stereo matching using graph cuts. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:74–81, 2004.

[HZ03] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.

[IG] Weiss Imaging and Solutions GmbH. Optimas - wissenschaftliche und industrielle bildanalysesoftware.

http://www.weiss-imaging.de/Webseiten/Software/Optimas/Optimas.htm. last visited: 2010.01.30.

[Kas] Gerald Kastberger. Communication and defense behavior of the asian giant honeybee apis dorsata. http://www.apis-dorsata.info. last visited: 2010.01.30.

[KK08] S. Kamiya and Y.S. Kanazawa. Accurate image matching in scenes including repetitive patterns. In *International Workshop on Robot Vision*, pages 165–176, 2008.

[KKZ03] Junhwan Kim, Vladimir Kolmogorov, and Ramin Zabih. Visual correspondence using energy minimization and mutual information. In *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision*, page 1033, Washington, DC, USA, 2003. IEEE Computer Society.

[Kon05] Thomas P Koninckx. Adaptive structured light. 2005.

[KZ01a] Vladimir Kolmogorov and Ramin Zabih. Computing visual correspondence with occlusions using graph cuts. In *In International Conference on Computer Vision*, pages 508–515, 2001.

[KZ01b] Vladimir Kolmogorov and Ramin Zabih. Computing visual correspondence with occlusions via graph cuts. Technical report, Ithaca, NY, USA, 2001.

[Lew95] J. P. Lewis. Fast normalized cross-correlation. In *Vision Interface*, pages 120–123. Canadian Image Processing and Pattern Recognition Society, 1995.

[LO05] Shingyu Leung and Stanley Osher. Global minimization of the active contour model with tv-inpainting and two-phase denoising. In *VLSM*, pages 149–160, 2005.

[MCUP02] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *In British Machine Vision Conference*, volume 1, pages 384–393, 2002.

[Med] MediaCybernetics. Image-pro plus - image processing, enhancement and analysis software. http://www.mediacy.com/index.aspx?page=IPP. last visited: 2010.01.30.

[MSS09] Protik Maitra, Stan Schneider, and Min C. Shin. Robust bee tracking with adaptive appearance template and geometry-constrainted resampling. *IEEE Workshop on Applications of Computer Vision*, 2009.

[oH] National Institutes of Health. Imagej - image processing and analysis in java. http://rsbweb.nih.gov/ij/index.html. last visited: 2010.01.30.

[RW00] Gerhard X. Ritter and Joseph N. Wilson. *Handbook of Computer Vision Algorithms in Image Algebra*. CRC Press, Inc., Boca Raton, FL, USA, 2000.

[SHB99a] Milan Sonka, Vaclav Hlavac, and Roger Boyle. *Image Processing: Analysis and Machine Vision*, chapter Segmentation, pages 123–227. International Thomson Publishing GmbH, 1999.

[SHB99b] Milan Sonka, Vaclav Hlavac, and Roger Boyle. *Image Processing: Analysis and Machine Vision*, chapter Matching, pages 190–194. International Thomson Publishing GmbH, 1999.

[SS01] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47:7–42, 2001.

[VCS08] Ashok Veeraraghavan, Rama Chellappa, and Mandyam Srinivasan. Shape-and-behavior encoded tracking of bee dances. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(3):463–476, 2008.

[Ved07] Andrea Vedaldi. An implementation of multi-dimensional maximally stable extremal regions. Technical report, University of California, LA, 2007.

[WHM$^+$09] Frank Weihmann, Thomas Höltzl, Michael Maurer, Madhusudan Man Singh, and Gerald Kastberger. 3-d patterning of social waves in the giant honneybee apis dorsata, 2009.

[Wik] Wikipedia. Apis dorsata. http://en.wikipedia.org/wiki/Giant_honey_bee. last visited: 2010.01.30.

[WPUB09] Manuel Werlberger, Thomas Pock, Markus Unger, and Horst Bischof. A variational model for interactive shape prior segmentation and real-time tracking. In *International Conference on Scale Space and Variational Methods in Computer Vision (SSVM)*, Voss, Norway, June 2009. to appear.

[ZCS03] L. Zhang, B. Curless, and S.M. Seitz. Spacetime stereo: shape recovery for dynamic scenes. In *CVPR*, pages II: 367–374, 2003.

[ZDC07] Ayman Zureiki, Michel Devy, and Raja Chatila. Stereo matching using reduced-graph cuts. In *ICIP07*, pages I: 237–240, 2007.

[Zha99] *Flexible camera calibration by viewing a plane from unknown orientations*, volume 1, 1999.