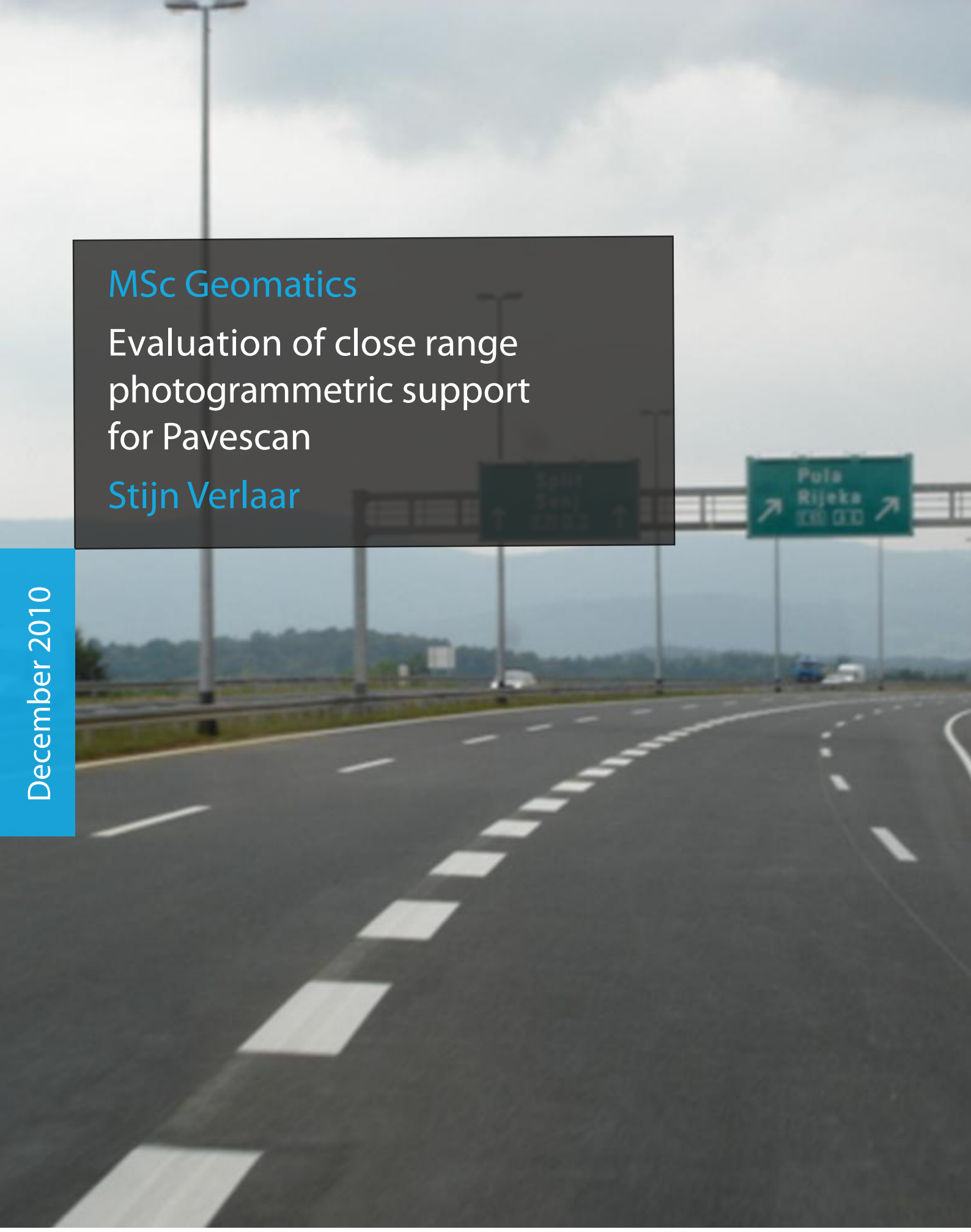


MSc Geomatics

Evaluation of close range photogrammetric support for Pavescan

Stijn Verlaar

December 2010

TU Delft Delft University of Technology

Department of Optical and Laser Remote Sensing

Department of Optical and Laser Remote Sensing
Faculty of Aerospace Engineering
Delft University of Technology

December 8, 2010

Evaluation of close range photogrammetric support for Pavescan

Stijn Verlaar

A thesis written for the Delft University of Technology for obtaining the degree
Master of Science in Geomatics

Graduation professor: Prof. dr. M. Menenti

Supervisors: Dr. R.C. Lindenbergh (TU Delft)
Dr. B.G.H. Gorte (TU Delft)
Ing. P.J. van Dueren den Hollander (Breijn)

Co-reader: Ir. M.F.C. van de Ven

Abstract

Pavescan is a low cost mobile system for road modelling survey. Because of the absence of navigation sensors it has several practical drawbacks compared to most of the other mobile mapping systems, but those sensors are very expensive and do not fulfil most of the accuracy requirements. Pavescan will be more attractive if some of the practical drawbacks are reduced. This thesis shows how close range photogrammetry can support Pavescan. The approach is able to achieve sub-centimetre absolute accuracy, but unfortunately not in combination with reducing practical drawbacks.

Pavescan measures road profiles across the road by laser scanning at a series of positions. The separate scans are linked via control points, which have to be measured in an additional survey. This is the main practical drawback of Pavescan and therefore it is desirable to reduce the required number of control points. The accuracy in height of the scans is most important (e.g. for volume calculations) and has to be in the range of millimetres. Integrating close range photogrammetry was chosen as the approach to fulfil these objectives for being low cost and its high accuracy potential. The approach will only be interesting for the client of this thesis (Breijn) if a high degree of automation is achieved.

A test survey was conducted with a downward looking camera, which was mounted close to the scanner at a height of around 3.6 metres above the road's surface. By automatic detection of corresponding points in overlapping images in combination with control point measurements, the position and orientation of the camera at each exposure is retrieved by bundle adjustment. These can be used to position and orient the laser scan. For every corresponding point, the 3D coordinate is calculated as well. Control points that are not used in the bundle adjustment are used as check point, which allows for comparing their calculated 3D coordinates with their (by tachymetry) surveyed coordinates. The difference is a measure of the accuracy of the object points and the bundle adjustment.

The measurement setup was compared with standard photogrammetric projects in order to estimate the potential quality. Quality is described as accuracy

in combination with reliability. The number of tie points (>1000 per image pair), their configuration and their accuracy (0.1 pixel $\approx$ 1.5 μm) is considered to be better with the proposed approach compared to standard photogrammetric projects. The number of control points is, however, lower and their configuration weaker. Thus, some aspects are better, while others are worse.

It is standard to use additional software for the bundle adjustment due to its complexity. Bundler is open source computer vision software for fully automatic 3D object reconstruction based on solely images. The result obtained by Bundler was not successful because only a few images could be processed simultaneously and residuals were too large. BINGO is specialized (close range) photogrammetric software that offers more control on the bundle adjustment. The consequence is that more manual steps are introduced. For instance, point correspondences should be obtained beforehand. The advantage is that all outliers could be removed by RANSAC and one can decide for the strictness of the matching. The latter is decisive for the number of point correspondences. It was found that a strict matching (i.e. fewer point correspondences but more distinctive) yields smallest residuals and thus highest accuracy.

Using accurate lens parameters appears to be of high importance. Estimating lens parameters in the bundle adjustment yields much better results than via a fast lens calibration with Photomodeler. However, control points are of highest concern. The accuracy of the bundle adjustment depended strongly on the configuration of the control points. Due to the narrow shape of a strip of images, this configuration will always be weak. For a sequence of 28 images ($\approx$43 metres), around five control points are needed to achieve subcentimetre accuracy of the object points. If more than 36 images or fewer than four control points were used, the bundle adjustment diverged (i.e. no solution could be obtained). A reduction of control points is therefore not possible. The practical feasibility for integrating close range photogrammetry into Pavescan is low, since too many control points are needed that should spatially be well distributed and measured with tachymetry (or with similar accuracy). Therefore it is not recommended to integrate photogrammetry in Pavescan.

Samenvatting

Pavescan is een low-cost mobiel systeem dat gebruikt wordt voor het modelleren van wegen. Door het ontbreken van navigatie apparatuur heeft het systeem echter enkele praktische nadelen ten opzichte van de meeste andere *mobile mapping systems*, maar deze zijn over het algemeen minder nauwkeurig dan Pavescan en bovendien duur. Pavescan wordt interessanter als enkele praktische nadelen verholpen zouden kunnen worden. Dit scriptie laat zien hoe close range fotogrammetrie Pavescan kan ondersteunen. Het is gebleken dat de methode in staat is om sub-centimeter absolute nauwkeurigheid te halen, maar helaas niet in combinatie met het verminderen van de praktische nadelen.

Pavescan meet dwarsprofielen van een wegoppervlak met een laserscanner. De scans worden uit stilstand gemeten en zijn op zichzelf staande metingen. De losse scans worden aan elkaar gerekend met behulp van paspunten die apart van de scans worden ingemeten. Dit is het grootste nadeel van Pavescan en het is daarom gewenst om het aantal vereiste paspunten te verminderen. Het belangrijkst van de scans is de nauwkeurigheid in hoogte (b.v. voor volumeberekeningen) en daarom is een nauwkeurigheid in de orde van millimeters vereist. Om deze doelen te verwezenlijken is de integratie van close range fotogrammetrie gekozen, omdat het goedkoop is en potentie heeft om millimeter nauwkeurigheid te behalen. Daarnaast zal de methode voor de opdrachtgever (Breijn) alleen interessant zijn als een hoge automatiseerbaarheid behaald kan worden.

Een testmeting was uitgevoerd met een naar beneden gerichte camera die naast de laserscanner is bevestigd op een hoogte van ongeveer 3,6 meter boven het wegoppervlak. Met behulp van paspunten en automatische detectie van corresponderende punten in overlappende foto's, kan de positie en stand van de camera voor elke foto worden achterhaald met een blokvereffening. Deze kunnen gebruikt worden om de laserscan te positioneren en oriënteren. Van elke puntcorrespondentie worden ook zijn 3D coördinaten berekend. Paspunten die niet in de blokvereffening gebruikt worden, worden als checkpunten gebruikt. Dit houdt in dat hun berekende coördinaten worden vergeleken met de (door tachymetrie) gemeten coördinaten. De ver-

schillen zijn een maat voor de nauwkeurigheid van de objectpunten en de blokvereffening.

De meetopzet is vergeleken met standaard fotogrammetrische projecten om een schatting te kunnen maken van de potentiële kwaliteit. Met kwaliteit wordt nauwkeurigheid in combinatie met betrouwbaarheid bedoeld. Het aantal verbindingspunten (>1000 per foto paar), hun configuratie en hun nauwkeurigheid (0,1 pixel $\approx$ 1,5 μm) zijn beter dan bij standaard fotogrammetrische projecten. Er zijn echter minder paspunten en de configuratie van deze punten is zwakker. Dus, waar sommige factoren beter zijn, zijn andere juist slechter.

Door de complexiteit van de blokvereffening, wordt hiervoor over het algemeen gebruik gemaakt van software. Bundler is open source computer vision software voor volledig automatische 3D reconstructie van objecten die gefotografeerd zijn vanuit verschillende standpunten. Het resultaat van Bundler was geen succes omdat slechts enkele foto's behandeld konden worden en bovendien met een lage nauwkeurigheid. BINGO is software die gespecialiseerd is in (close range) fotogrammetrie. Het biedt meer controle op de blokvereffening waardoor meer handmatige stappen worden geïntroduceerd. Zo moeten bijvoorbeeld vooraf de puntcorrespondenties worden verkregen. Het voordeel is dat grote fouten (outliers) al verwijderd kunnen worden en er bepaald kan worden hoe streng de matching zal zijn. Laatstgenoemde is bepalend voor het aantal puntcorrespondenties dat gevonden wordt. Een strenge matching (d.w.z. minder puntcorrespondenties maar meer onderscheidend) resulteerde in de hoogste nauwkeurigheid van de objectpunten.

Het gebruik van nauwkeurige lensparameters blijkt erg belangrijk te zijn. Door de lensparameters te schatten in de blokvereffening werd een veel beter resultaat behaald dan via een eenvoudige calibratie in Photomodeler. De paspunten zijn echter het meest van belang. De nauwkeurigheid van de blokvereffening hangt sterk af van de configuratie van de paspunten. Door de smalle vorm van een serie foto's zal de configuratie altijd zwak zijn. Een reeks van 28 foto's ($\approx$ 43 meter), heeft ongeveer vijf paspunten nodig om sub-centimeter nauwkeurigheid te behalen op de objectpunten. Gebruik van minder dan vier paspunten leidde tot divergentie van de blokvereffening. Ook als meer dan 36 foto's werden gebruikt resulteerde dat in divergentie van de blokvereffening. Het verlagen van het aantal paspunten is daarom niet mogelijk. De praktische haalbaarheid voor het integreren van close range fotogrammetrie in Pavescan is laag, omdat er te veel paspunten nodig zijn die bovendien een goede configuratie moeten hebben en gemeten dienen te worden met vergelijkbare nauwkeurigheid als tachymetrie. Daarom wordt niet aangeraden om close range fotogrammetrie te integreren in Pavescan.

Preface

About one year ago I started a project at engineer firm Breijn, which has lead to this thesis. Their request was to improve their laser scanning system, Pavescan, for road construction purposes. This thesis evaluates the possibilities to support Pavescan by integrating close range photogrammetry.
Basically, the thesis is written from a practical point of view. This implies that for every step it was considered if it would practically be feasible for Pavescan and Breijn. So, conclusions at the end of this thesis that are not beneficial for this project, could still be interesting for similar other projects with different requirements.

Without the help of many others, the quality of this thesis would not have been as it is now. Special thanks goes to my supervisors at TU Delft, Roderik Lindenbergh and Ben Gorte, who offered many ideas and help during the entire period. A lot of reading by Roderik made it possible to constantly improve the thesis. The discussions with Kourosh Khosh Elham were very important in order to draw the right conclusions regarding photogrammetric issues. The feedback from Massimo Menenti had lead to much more coherence between all chapters.

Pieter van Dueren den Hollander gave a lot of input throughout the year and always made time for me, even when he was extremely busy. Both Pieter as Edwin van Osch, showed their appreciation to me for which I am thankful. The test survey was never as successful without the help of Willy van Aarle who made a construction to mount the camera on Pavescan.
With Dinesh Kalpoe I had interesting discussions and made working in the Geolab nice. My parents may not be forgotten in my word of thanks, since they fully supported me throughout my entire study and stimulated me to start with the Master Geomatics. Last but not least my thanks goes to my girlfriend Kim for her continuous support and patience.

Stijn Verlaar

Delft, 2010

Contents

List of Figures

List of Tables

Glossary

accuracy	A measure of how well a measurement fits the true value. In this thesis, the main objective is to improve the accuracy of the laser scan point coordinates in height, where the height is the absolute height, i.e. with respect to NAP.
AOV	Angle of View. The maximum angle where light arrives at the CCD. The angle of view can be expressed in a horizontal, vertical or diagonal angle of view.
aperture	The opening in front of the lens that determines together with the shutter speed how much light reaches the image plane. The shutter speed and aperture should be well balanced for the right exposure of an image.
baseline	The baseline is the distance between two consecutive images. It is calculated by taking the difference of the position of the camera at the time of image i and image $i+1$.
collinearity	The principal that the objective center, an image point and its ground point fall on one straight line.
control point	A point that is known in both image coordinates as ground coordinates. Control points are used to link the images to a known coordinate system.

crossfall	The slope of a road in the direction perpendicular to the driving direction.
distance ratio	A number between 0 and 1 that defines how strict the SIFT matching performs. The lower the distance ratio, the stricter the matching and thus the fewer matches will be identified.
DTM	Digital Terrain Model. An elevation model of the terrain with all objects filtered out.
exterior orientation	The position and orientation of a camera. Every image has six exterior orientation parameters, three for its position and three for its orientation.
focal length	The length from the principal point to the perspective center. In this thesis, for focal length often the principal distance is meant.
ground resolution	The surface in the terrain covered by one pixel in the CCD chip.
INS	Inertial Navigation System. A system that calculates the position, acceleration and orientation of a moving object by measuring accelerations and rotations in all directions.
keypoint	A potential tie point, localized by SIFT. Keypoints become tie points if they can be matched with a keypoint in another image.
milling	The removal of asphalt by a cutter device.
object point	A tie point projected in object space.
perspective center	The position where all rays of light come together. The perspective center is also referred to as exposure station.

point correspondence	A match of two keypoints from overlapping images. These points are also referred to as tie points. In this thesis used when specifically the match is meant.
PPS	Principal point of Symmetry. The origin of the lens distortion.
precision	A measure for the reproducibility of a measurement. So, a very precise scan gives a high relative accuracy. Precise laser scan measurements are required if relative values are needed such as for slope determination or for modelling road damage like road rutting.
principal distance	The distance between the image plane and the point where all light rays intersect. Within aerial photogrammetry, the principal distance and focal length are often used interchangeably because they are practically the same.
principal point	The intersection of the image plane and the line perpendicular to the image plane, through the perspective center.
RD-NAP	The combination of the Dutch horizontal coordinate system (Rijksdriehoek) and the Dutch height system (Normaal Amsterdams Peil).
resolution	The number of pixels per unit distance or area. Thus, the actual resolution of a camera can only be calculated if besides the number of pixels also the chip size is known.
third degree tie point	A tie point that is measured in three images. These points are required for a successful bundle adjustment.
tie point	An image feature that is measured in at least two images. Tie points are used to calculate the relative orientation between images.

TIN	Triangular Irregular Network. A network of triangles through a given set of points. The triangles are created such that a circle through its three vertices have the smallest possible radius.

Chapter 1

Introduction

1.1 Motivation

Road modelling is important in the road construction industry for either making a new road design or maintenance of the existing road. New roads are often made on top of an old road. For the construction of such a road, it is of major importance how much and where there need to be milled from the old road. Milling is the removal of material (in this case asphalt) with a rotating cutter device. Measurements on the road's surface are needed to obtain a model of the road and to decide for the optimal milling amounts. Especially the accuracy in height is important, since this directly has an effect on how much and where there need to be milled. One can imaging that milling more than necessary, leads to unnecessary costs because more new asphalt is needed. In theory, relative measurements are sufficient. However, it is beneficial to obtain the measurements directly in the standard coordinate system for geodetic surveys and construction designs, which is RD-NAP in The Netherlands (the Dutch horizontal coordinate system in combination with the Dutch height system). Once the new road has been built, road models can check whether the road fulfils the requirements. An important factor here is the crossfall. Crossfall is the slope of a road in the direction perpendicular to the driving direction measured in degrees. This slope should avoid water to stay on the road's surface. Regarding maintenance of a road, deformations of the road's surface need to be modelled. Heavy traffic deforms the asphalt to a certain extent. The tracks that are created are called road rutting. This, or other kinds of damage, should be detected in road models. For both measuring the crossfall and road damage, relative measurements are sufficient.

Methods to perform this modelling task are abundant. However, each has its strengths and limitations. Conventional methods like levelling and tachymetry reach a high level of accuracy and precision at the price of a

low point density, much manpower and thus high costs. The main practical drawback of these methods is that roads need to be closed temporarily, [Mennink, 2008], which causes major inconveniences and very high costs. More advanced techniques like mobile mapping systems avoid the road to be blocked and even overcome the point density problem as well as the manpower, but are less accurate and precise and still expensive due to advanced equipment. Pavescan, a relative new system on the market, is a good trade-off by making cross profiles of a road with a laser scanner, each containing around 200 measurements. Pavescan drives on the emergency lane, so that the road does not have to be closed. Compared to the aforementioned conventional methods, the accuracy level is a bit worse while having a much better point density. Using not the most advanced equipment makes it also a strong financial competitor. So far, several projects with Pavescan have been executed by different companies. Among others the widening of the A12 between Zoetermeer and Gouda by Ballast Nedam, the provincial road N34 in Friesland and two parts of the A2 by Breijn B.V.. Breijn had the system operational since 2009 and at the moment of writing Pavescan is used for road modelling tasks for the widening of the A50 between the junctions Valburg and Grijsoord.

Each scan by Pavescan is a static measurement, which implies that for every scan the vehicle has to stop. Pavescan does not have a GPS/INS on board, so control points are needed to link all the scans to RD-NAP coordinates. These control points are measured in a separate survey with GPS or via tachymetry, depending on the desired accuracy. GPS measurements are fastest but are less accurate than tachymetry. The laser scan itself has millimetre precision, but both the absolute accuracy of the scan as the relative accuracy between multiple scans is only in the order of centimetres when GPS is used to measure the control points. An improvement of the accuracy in height of the laser scan points to millimetres or sub-centimetre is desired, so that milling jobs can be made more reliable, which is cost effective. Sometimes, additional levelling measurements on the control points are used to improve the accuracy in height. It is however most attractive to achieve a higher accuracy and reduce at the same time the number of control points. Excluding or reducing these measurements will lead to a less expensive survey and, also important, a much safer survey. A GPS/INS system will exclude the control point measurements, but this is very expensive and thus an alternative low cost solution is preferred. Besides, it is doubtful if such a system can reach to the desired level of accuracy. The proposed approach is to integrate close range photogrammetry with the laser scans. Close range images are very low cost, easy to acquire and may achieve millimetre accuracy, so it is an interesting method in combination with the traditional setup of Pavescan. Besides, images can cover the space in between two scans and may therefore be useful as well. With close range photogrammetry the

position and orientation of the camera and 3D coordinates of objects are provided. It is known that with close range photogrammetry also control points are required, but it is unknown how the number of control points will relate to the accuracy of the position and orientation of the camera and 3D coordinates of objects and thereafter the accuracy of the laser scan measurements.

1.2 Research objective and research questions

This section provides the research objective and research questions. The main research question will be subdivided into several sub questions in order to specify in more detail the workflow of answering the main research question.

The research objective of this thesis is to improve the accuracy in height of the laser scans from Pavescan and to connect all the separate scans with a reduced number of control points with low cost adjustments. As low cost solution, it was decided to evaluate these possibilities with close range photogrammetry. The solution should also be highly automated to avoid additional manpower or manual actions that require expert knowledge. From these objectives the main research question is stated as:

Is it possible within Pavescan to achieve millimetre accuracy in height in measuring road profiles and to reduce the control point measurements by integrating close range images?

This main research question is subdivided into three sub questions:

- *How can close range photogrammetry be integrated into Pavescan and what is the best setup?*
  How can the information obtained from close range photogrammetry be integrated into the laser scan? And what is the best setup of the adapted acquisition? It should be decided what type of camera and lens to be used and also the positioning of the camera is important.
- *What degree of automation can be achieved in the processing?*
  The degree of processing is dependent on several factors. The available software may have limitations and previous manual actions may be automated. The degree of automation is of high importance for the practical feasibility of the integration of close range images into Pavescan.
- *How does the number of control points relates to the accuracy of the road profile?*

If it is succeeded to define a solution where road profiles can be measured with millimetre accuracy, the question remains how many control points are needed. Or the other way around: if the number of control points is reduced, what will be the obtained accuracy. Especially the accuracy in height is important.

1.3 Thesis outline

In this thesis all relevant theory, methodology and results are presented. This section gives the outline of this thesis.

- Chapter 2 provides all relevant background information. It includes a system description of Pavescan and the alternative mobile mapping systems. Further, the chapter provides a description of the client, the concepts of accuracy and precision and a mathematical background on non linear least squares. Also for those less familiar with photogrammetry or feature matching the chapter is useful to read.
- Chapter 3 starts with how close range photogrammetry can be integrated into Pavescan. Then the focus is on which camera setup is most suitable and how the measurement campaign should look like. The available software and cameras are also part of this chapter.
- Chapter 4 is a quality analysis of the proposed approach. The error sources are defined and compared to standard aerial photogrammetric projects in order to find the potential quality of the proposed approach. The transfer to the accuracy of the laser scan is also included.
- In Chapter 5 a sequence of test images are obtained and processed. The processing is done with two different software programs. In between the processing with these programs, an existing algorithm to find point correspondences in images is optimized.
- In Chapter 6 the main results are presented. Also a validation of the approach is provided. The chapter concludes with the accuracy assessment.
- Chapter 7 summarizes the main results and an explanation of why these results are obtained is provided.
- In Chapter 8 the main conclusions are summarized. Also the recommendations to Breijn are presented.

Chapter 2

Background

This chapter provides all relevant theory for a better understanding of the scope of this thesis and for theoretical support on technical issues that are dealt with in this thesis. First, a description of the client is given in Section 2.1, followed by the concepts of accuracy and precision in Section 2.2 and the measurement requirements in Section 2.3. Then, the mobile mapping method as alternative of Pavescan is described in Section 2.4 and a detailed description of Pavescan itself in Section 2.5. The mathematical background in Section 2.6 is useful for those less familiar with nonlinear least-squares. In Section 2.7 the most important aspects of photogrammetry is provided before Section 2.8 introduces SIFT as approach for automatic tie point extraction. Section 2.9 describes how relative orientation works with epipolar geometry. The chapter concludes with Section 2.10 that is about different types of lenses.

2.1 Client

This thesis is performed in close co-operation with engineer firm Breijn B.V. that is part of Heijmans. Heijmans originally started as road construction company in the early twenties. Nowadays, it is one of the largest construction companies in The Netherlands with around 11,000 employees and a turnover of around 3.6 billion euros, [Website Heijmans]. Heijmans is divided in 5 major divisions: Real Estate, Utility, House-Building, Technique and Infra. Breijn belongs to the division Infra and has around 300 employees working at eight office locations.

2.2 Concepts of accuracy and precision

The terms of accuracy and precision are often used within quality analysis. There is, however, an essential difference between the two concepts. It is important to use these terms with proper understanding. Therefore, these

Figure 2.1: Head office Heijmans in Rosmalen

are explained in this section.

Accuracy is a measure of how well a measurement fits its 'true' value. The true value is marked here with quotes since the exact true value is never known. In practice, the true value is often considered to be a value obtained with measurements of higher quality. The closer the observations are to the accepted true value, the higher its accuracy.

Precision is the degree of which similar or repeated measurements show the same result. In other words, the closeness of the measurements to their mean. Often, precision is represented by the standard deviation σ where 1σ is the 68% probability threshold. For example, if the mean height of a 3D point is measured to be 100.00 mm with a precision 1σ=0.02 mm, then one measurement lies with 68% probability between 99.98 and 100.02. However, this does not tell how the measurement is related to its real value. Figure 2.2 illustrates the difference between accuracy and precision where the centre of the target represents the true value. It is shown that a high precision is required to obtain a high quality, but that highly precise measurements are not necessarily accurate.

In this thesis, the terms relative accuracy and absolute accuracy are also used. Two measurements with high precision have a high relative accuracy. Absolute accuracy is used when referring to local or world coordinate systems. How well the obtained coordinates fit into the coordinate system is the measure for absolute accuracy.

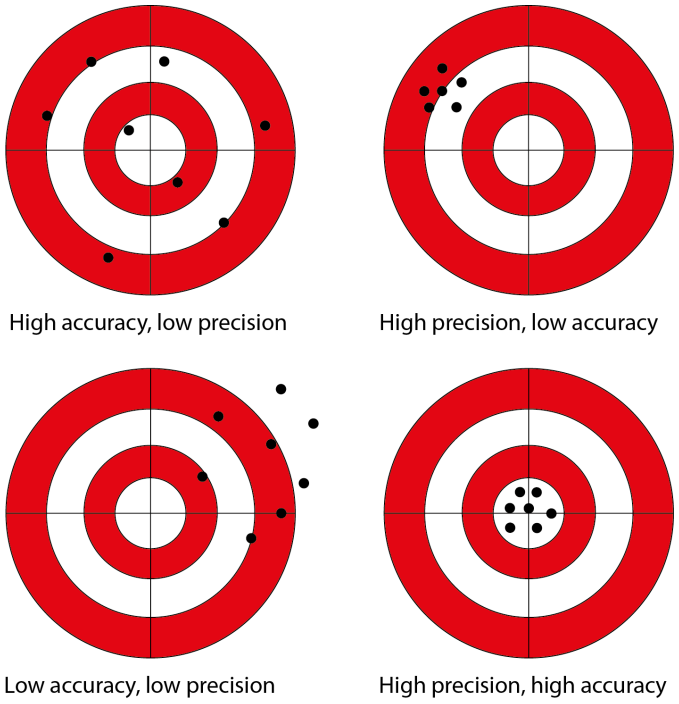


Figure 2.2: Concepts of accuracy and precision illustrated by target analogy.

2.3 Measurement requirements

As stated in Section 1.1, several measurements are required during the construction of a road or for road maintenance. For road maintenance it is of main importance to monitor the crossfall and local deformations such as road rutting. One can determine the crossfall and detect local deformations by measuring cross profiles of the road. The better the precision of these measurements, the higher the accuracy of the crossfall determination and the smaller the local deformations that can be detected. The absolute accuracy is of minor importance in this case.

For the construction of a new road, only the situation where the new road will be built on top of an old road will be considered. Partly, the same measurements are of interest as with the aforementioned road maintenance tasks where relative measurements were sufficient. However, the information from these measurements is used differently. All measurements have the purpose to find the optimal amount of old asphalt that need to be milled, and thus the optimal amount of new asphalt that should be used to construct the new road. Too much milling leads to unnecessary costs because more new asphalt is needed while too little milling may cause local deformations or

deep cracks not to be removed so that the quality of the new road becomes worse. So actually, one is looking for the optimum between costs and the quality of the new road. For example, the right hand lane of a highway is in general more deformed than the other lanes due to the greater intensity of heavy traffic. Therefore, from this lane, more asphalt need to be milled than from the other lanes to reach this optimum. The accuracy of the height is therefore most important to be able to determine the amount of milling accurately. The requirements for road construction measurements are besides a high relative accuracy also a high absolute accuracy. In theory, only relative measurements are required, but in practice it is much easier to have all measurements in RD-NAP, since everything in the field is defined in RD-NAP as well.

Regarding mobile mapping systems, the relative accuracy of all scan points will be dependent on the precision of the scanner that is used and the driving speed and accelerations. The absolute accuracy is dependent on the accuracy of the GPS in combination with INS. With Pavescan, the relative accuracy of scan points within one scan is dependent on the precision of the scanner while the relative accuracy of different scans and the absolute accuracy depends on the precision and accuracy of the control point measurements. The precision of the laser scanner is in the order of millimetres and therefore suitable for detecting local deformations and the crossfall accurately. The precision and accuracy of the control points are not sufficient if GPS is used, especially the height component.

2.4 Alternative methods

2.4.1 Mobile mapping

Pavescan is not the only advanced system that can give a detailed and fast model of a road. Mobile mapping systems offer a continuous point cloud and are also very fast in acquisition. Advanced mobile mapping systems like StreetMapper (as can be seen in Figure 2.3) and DRIVE-MAP, [Website DRIVE-MAP], are carrying high tech equipment (GPS/INS) for position determination and several 2D laser scanners for the point acquisition. These systems have the advantage, compared to Pavescan, that they measure continuously while driving with relatively high speed (specifications are based on a speed of 36 km/h). This makes the acquisition very fast, but at the price of a decrease in the level of accuracy. The continuous data stream also leads to enormous amounts of data that causes the processing to be slow. The relative accuracy of StreetMapper typically is 25 mm while the absolute accuracy ranges from 50 mm in clear sky environments up to 0.5 m in built-up areas where the GPS quality drops, [Hunter et al., 2006]. These accuracies do not fulfil the requirements that Pavescan demands.

(a) StreetMapper equipment (b) Practical example

Figure 2.3: The StreetMapper system

StreetMapper was designed as an alternative for aerial laser scanning systems to overcome the high mobilization costs. However, compared to Pavescan, it is much more expensive due to the high tech equipment. From unofficial sources it has been found that the StreetMapper system will cost around 1.2 million euros and DRIVE-MAP from Fugro around 240 thousand euros including software packages, while Pavescan costs around 40 thousand euros.

2.5 Pavescan

2.5.1 System description

Pavescan is different from mobile mapping systems as it is not directly connected to the driving vehicle. It is an individual trailer with a 4 metre (from the ground) high pole where the equipment is mounted on. A schematic illustration of the Pavescan system is shown in Figure 2.4.
For the generation of the cross profiles, the laser scanner is the most important device. On top of the pole, a line scanner (Sick LMS200-30106) is mounted. As with any other laser scanner, no reflectors or other special reflectivity properties are required and illumination of the target object is not necessary. The scanner operates by emitting a laser pulse and receiving its reflection that occurs if it meets an object. The time between transmission and reception, referred to as time of flight, is measured and is directly proportional to the distance from the scanner to the object. The pulsed laser beam is deflected by an internal rotating mirror so that a fan-shaped scan is made as can be seen in Figure 2.4. The mirror rotates after transmitting each pulse with 1° until an angle of view of 180° has been reached. This cycle will be repeated once more so that a resolution of 0.5° is obtained. From the 360 emitted pulses, around 200 pulses are received back, which will form the cross profile. Not all emitted pulses are received back because

Figure 2.4: The Pavescan system with a schematic laser bundle. (Illustration from [Mennink, 2008])

pulses with a near horizontal angle of emittance do not meet any objects.

Figure 2.5 shows the spot size and spot spacing over distance. From the linearly increasing spot size over the range it is calculated that the beam divergence is approximately 0.13°. Furthermore it can be seen that at emittance the spot size is approximately 1 cm. For Pavescan this means a spot size of around 0.03 m straight under the laser and less than 0.10 m at the far edge of the road. In Table 2.1 the most important specifications of the laser scanner are summarized.

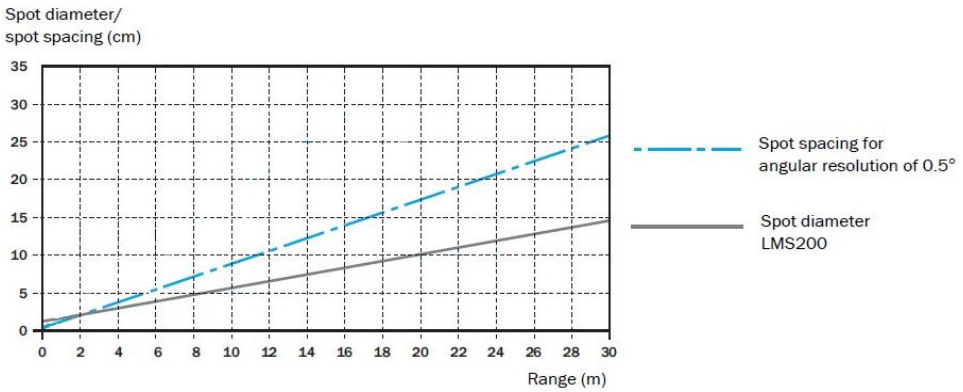


Figure 2.5: Spot size and spot spacing of the laser scanner as function from range. (Figure adapted from [Sick AG, 2003])

The wavelength of 905 nm is just outside the visible part of the electromagnetic spectrum (380-800 nm). The scanner scans with 75 pulses per second so it will take 4.8 seconds to complete one cross profile. The millimetre mode (the mode with the best performance) of the laser scanner has a systematic error of 15 mm and a standard deviation of 1σ=5 mm for ranges up to 8 metres. However, in practice a precision of better than 3 mm is achieved for distances up to 12 m. Also the accuracy was improved by implementing corrections for all ranges and angles based on control measurements. An additional check is performed by measuring a known distance to a small plate of roofing felt (having the same reflection property as asphalt).

In general, every ten metres a scan is made. Pavescan needs to stand still while measuring and is stabilized by two mechanical feet. These feet are also putting the trailer in an approximate horizontal position. An additional slope sensor is measuring the actual slope of the trailer at the time of acquiring the laser scan. In order to retrieve the absolute position of the scan points as well as the locations of the cross profiles relative to each other, control points are measured. A control point is a dot of paint that is placed on the ground by the operator by pushing a button just before a scan is made. Since it is not necessary to make scans at exact distance intervals, the operator looks at a distance meter to approximate the required distance. In this way, no time will be wasted on putting the vehicle on exact locations. After the scan is made, the control point will be measured by either tachymetry or GPS depending on the required accuracy. If needed, an additional levelling measurement is performed on the control points. As a result, each scan is positioned in the desired local coordinates. However, not all scan points will be known exactly, since the orientation of the scanner with respect to the vertical axis, i.e. the heading, is not measured. With the simple assumption that a scan profile is made perpendicular to the road,

Table 2.1: The most important properties of the laser scanner

Properties LMS200-30106	
Wavelength	905 nm
Angle of View	180°
Frequency	75 Hz
Angular resolution	1°
Max. range	80 m
Beam divergence	0.13°
Systematic error	15 mm
Standard deviation	5 mm

the scan points are approximately localized. So, within one scan, all scan points yield a high relative accuracy because of the millimetre precision of the laser scanner. The absolute accuracy of the scan depends on the accuracy of the control point measurement. Because the control point is located on the road's surface, the systematic range error of the scanner drops out. The relative accuracy of the location of two consecutive scans depend on the precision of the control point measurements.

Traffic may obscure a measurement since the laser beams may not reflect on the asphalt but on a vehicle instead. This traffic is removed from the data by automatic filtering. However, if the measurement is interrupted too much, the measurement needs to be repeated. In order to distinguish between vehicles and objects like trees or a crash barrier, the operator must enter the width of the road into the computer so that only objects within this width will be filtered out. The operator sees the result immediately on a screen and judges if it is of sufficient quality. An example of a cross profile is presented in Figure 2.6.

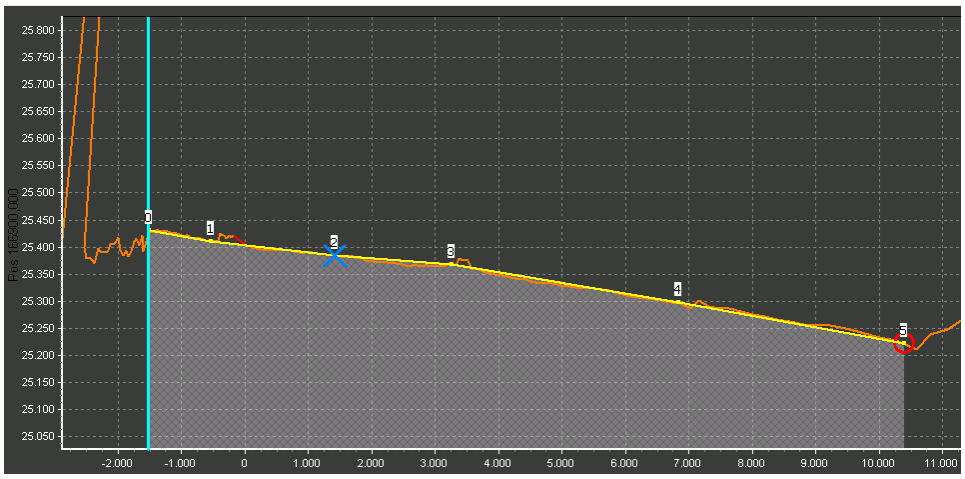


Figure 2.6: An example of a cross profile as presented in PaveCalc

The orange line above the gray marked area consists of measurements of the road's surface. It is practically not recommended to work with a cross profile containing this many points and having an unnatural shape due to measurement noise. Therefore, several of point measurements are selected manually to create a profile that follows the measurements best. Obviously these include the points of an abrupt change in gradient like on the axis of a road and in a lesser degree at lane edges. The line through those points, presented by the yellow line, represents the simplified cross profile. It is remarkable that road markings become 'visible' in the cross profile as

bumps (near the numbers 1, 3 and slightly at 4). This is due to the different reflection behaviour of road marking compared to asphalt. The laser scanner stops a measurement when a certain threshold of received energy has been reached. Since road marking reflects much better than asphalt, this threshold is reached faster and thus the scanner interprets this as a shorter distance. Road markings are very important for the final product because it's an easy indicator for creating length profiles. The cross profiles are an intermediate product to create a set of length profiles that will form the model of the road.

Besides the laser scanner, two video cameras (IQeye 703) are mounted on top of the pole, directed to the sides (as illustrated in Figure 2.7), and one video camera half way the pole (IQeye 705), directed to the back. The latter is used to give more clarity regarding the road marking and the near surrounding of the scan. Frames of the top two cameras give visual support of that part of the road where the scan was made. The software integrates the images with the laser points by indicating where a point on the image is located in the scan. For example, at the time the print screen in Figure 2.6 was made, the cursor was in the image on the exact edge of the road, represented by the vertical cyan line. The two cameras, are directed in different directions in order to obtain full coverage of the scan line. The IQeye 703 is a high-speed 3.1 megapixel (2048 H x 1536 V) network camera with a built-in web server and FTP server. The camera has a dichroic infrared mirror, which implies that it is not sensitive for infrared light. Thus, the laser pulses cannot be seen in the video recordings. The camera produces 12 frames per second at full resolution.

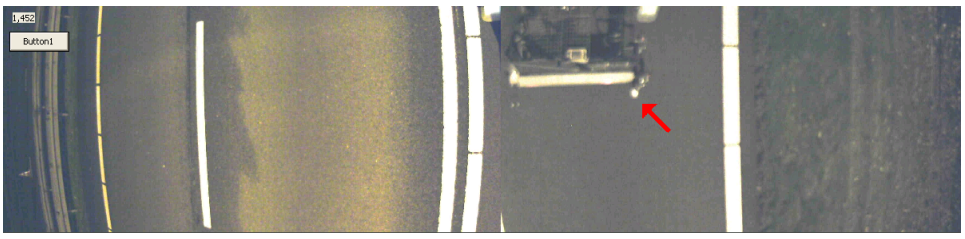


Figure 2.7: An example of a joint image. The control point is clearly visible.

An example of an image that is used in the processing of Pavescan data is presented in Figure 2.7. In this image the control point is clearly visible, but this does not hold for all images. The image consists of two joint frames of the two cameras in order to capture the entire width of the road. In order to obtain an image on the scan line, the operator presses a button while making the scan. The latest captured frames appear on a screen and if no vehicles are captured, the image is saved. It is clear that these images are

strongly distorted, but for the current application this is acceptable.

2.5.2 Drawbacks

Besides a lot of advantages, Pavescan still has some disadvantages. One of them is the stop and go character of the system. For every cross profile the vehicle needs to stop and scan before it can move on to the next stop. For detailed recordings of the road, this implies a stop every ten metres. A scan every ten (or more) metres also implies that there is no data in between two consecutive scans. Currently this space is simply being interpolated.
In the current state every control point is measured separately, which results in higher costs due to extra manpower. As also mentioned in Chapter 1, from the safety perspective it will be much better to have less people in the near vicinity of the (high speed) traffic. Research on possibilities to exclude or reduce these measurements is therefore useful.
Another drawback is that traffic may obscure the measurements. Although it may be assumed that one can find a time out of rush hours, where the traffic is calm enough, a sudden traffic jam will mess up the measurements. Besides the traffic status, also the weather may influence the success of a scan. Water on the road's surface influences the measurements to a certain extent and heavy rainfall obstructs the entire measurement. Thus, rainfall has a negative influence on the measurement campaign.

2.6 Mathematical background

In order to understand some future processes in this thesis, a mathematical background is provided.

2.6.1 Nonlinear least-squares

When one would like to determine the solution to a set of unknown parameters according to a set of measurements, least squares adjustment is the most common solution. A least squares adjustment, minimizes the sum of squared errors where the error is defined as the difference of the assigned value and the measured value. If the relation between observations and unknowns is linear, the system of equations can be written as:

$$y = A \cdot x + e \tag{2.1}$$

where y is a $m \times 1$ vector containing m (measurement) observations, x is a vector with n unknowns, A is the $m \times n$ model matrix and e is the $m \times 1$ error vector with the residuals of the observations. If there are just enough

observations to solve the model ($m = n$), then the error vector drops. Even when $m > n$, the error vector is often omitted and the model becomes:

$$y \approx A \cdot x \tag{2.2}$$

If all observations are expected to have the same weight, the unknowns can be estimated by:

$$\hat{x} = (A^T \cdot A)^{-1} \cdot A^T \cdot y \tag{2.3}$$

This solution is known as the least-squares solution; x is chosen such that $A \cdot x$ is as close as possible to the observation vector y:

$$\hat{x} = min \left\| y - A \cdot x \right\|^2 \tag{2.4}$$

This least-squares principle is not restricted to linear system of equations. It can also be used to solve

$$y \approx A(x) \tag{2.5}$$

where A is a nonlinear function of x. For such nonlinear minimization problems, in general no direct solution can be given. Therefore, one is designated to solve the problem iteratively. Starting with an initial approximation of the solution, one intends to obtain an improved approximation in each cycle of the iteration. Each successive approximation is usually based on the linearized version of the nonlinear model, [Teunissen et al., 2005]. The linearization of any function $A(x)$ is performed according to Taylor's theorem that defines a gradient vector in between x and x_0 where x_0 is an initial approximation of $\hat{x}$. This gradient vector is by definition linear and thus the model can be solved by normal linear least-squares. The approximation of $A(x)$ can now be written as:

$$A(x) \approx A(x_0) + J(x_0)(x - x_0) \tag{2.6}$$

where $J(x_0)$ is the $m \times n$ Jacobian matrix of partial derivatives:

$$J(x_0) = \begin{bmatrix} \frac{\partial y_1}{\partial x_1}(x_0) & \cdots & \frac{\partial y_1}{\partial x_n}(x_0) \\ \vdots & & \vdots \\ \frac{\partial y_m}{\partial x_1}(x_0) & \cdots & \frac{\partial y_m}{\partial x_n}(x_0) \end{bmatrix}$$

With this linear approximation of $A(x)$, the nonlinear model is approximated by:

$$\Delta y_0 \approx J(x_0) \cdot \Delta x_0 \tag{2.7}$$

where $\Delta y_0 = y - A(x_0)$ and $\Delta x_0 = x - x_0$. From this linear model, let $\Delta \hat{x}_0$ be the least squares solution:

$$\Delta \hat{x}_0 = \left[J(x_0)^T \cdot J(x_0) \right]^{-1} \cdot J(x_0)^T \cdot \Delta y_0 \tag{2.8}$$

Hopefully, $x_0 + \hat{x}_0$ is closer to the nonlinear least squares solution than the initial estimate x_0. The next approximation for $\hat{x}$ will therefore be:

$$x_1 = x_0 + \Delta\hat{x}_0. \tag{2.9}$$

In general, the iteration is described by:

$$\Delta\hat{x}_i = \left[J(x_i)^T \cdot J(x_i)\right]^{-1} \cdot J(x_i)^T \cdot \Delta y_i \tag{2.10}$$

and

$$x_{i+1} = x_i + \Delta\hat{x}_i. \tag{2.11}$$

Hopefully, this process converges to the least squares solution. The process will be terminated when the difference between two successive solutions is small enough. This iterative approach is called the *Gauss-Newton method*. The Levenberg-Marquardt (LM) approach is an alternative for the Gauss-Newton method relying on the same principle. Nowadays, LM is the standard for solving nonlinear least-squares problems, [Lourakis, 2005]. LM differs from Gauss-Newton mainly by using a damping strategy. Let H be the Hessian matrix that equals $J^T \cdot J$. In the LM approach, H is replaced by $(H + \lambda I)$, [Marquardt, 1963], where I is the identity matrix and λ the damping factor. By adding the damping factor to all diagonal elements of the Hessian matrix the solution to the linearized model will then look like:

$$\Delta\hat{x}_i = (H + \lambda I)^{-1} \cdot J(x_i)^T \cdot y \tag{2.12}$$

which is used to compute the updated parameters. If the error is smaller than the previous iteration (or smaller compared to the initial values in case of the first iteration), the correction is approved and the damping factor is decreased. The updated values are considered to be the new values. If the error is bigger than the previous iteration, the damping factor is increased until the error is smaller. This damping strategy is proven to guarantee a reduction of the error at each iteration. When the solution in an iteration is close to the correct solution, the damping factor approximates zero and thus the method will practically be equal to the the Gauss-Newton method, [Lourakis, 2005]. More theory about the damping factor or LM in general can be read among others in [Lourakis, 2005] and [Nielsen, 1999]. There is one inevitable risk when finding a solution to a nonlinear problem. That is when the solution converges to another minimum than the desired one. Therefore it is important for complex nonlinear models to have good initial approximates.

2.7 Close Range Photogrammetry

Photogrammetry can be described as "The science of deducing the physical three-dimensional measurements of objects from measurements on stereo

photographs that photograph an area from two different perspectives" [Maune et al., 2001]. Close range photogrammetry is basically the same as photogrammetry, except that close range photogrammetry is considered to be ground based and the distance to the object to be smaller. In this sense close range means that objects are maximally 300 meters away from the camera.

2.7.1 General theory

Photogrammetric applications involve the use of overlapping images. Two overlapping images are called a stereo pair. Each image has its own perspective center, the point where all rays intersect. Objects observed from different perspective centers show a displacement, which is referred to as parallax. This parallax is used to calculate the distance between the object and the camera. The nearer objects are, the more displacement they will have. In an image sequence, it is mostly attempted to keep the distance between two consecutive perspective centers approximately constant, so that an overlap of 60% is obtained. Theoretically, 50% overlap is sufficient to be able to create a full cover stereo model of the terrain. For safety, 60% overlap is used in practice. The distance from one perspective center to the next one is called the baseline. Figure 2.8 illustrates some important parameters used in photogrammetry. The ratio between the baseline and the height (B/H-ratio) determines how well the height of an object point can be estimated. The smaller the baseline, the smaller the parallax, which is needed for the height calculation. The greater the B/H-ratio, the more accurate the height estimation of an object point. The B/H-ratio should minimally be 0.5 to obtain a reliable height estimation.

Each image has a corresponding image scale associated with it. The image scale expresses the average ratio between a distance in the image and the same distance on the ground, [Leica Geosystems GIS and Mapping, 2003]. It is computed as the mean distance to the object(s) divided by the focal length. For example, with an average distance of 4 metres to the road's surface and a focal length of 20 millimetres the image scale will be 1:200. The main challenge within photogrammetry is to retrieve the position and orientation of the camera at the time of exposure and linking the images to a known coordinate system. To perform this task automatically is even a greater challenge. By using more than one image, the geometry associated with the camera, image and ground can be defined with greater accuracy and precision, [Leica Geosystems GIS and Mapping, 2003].

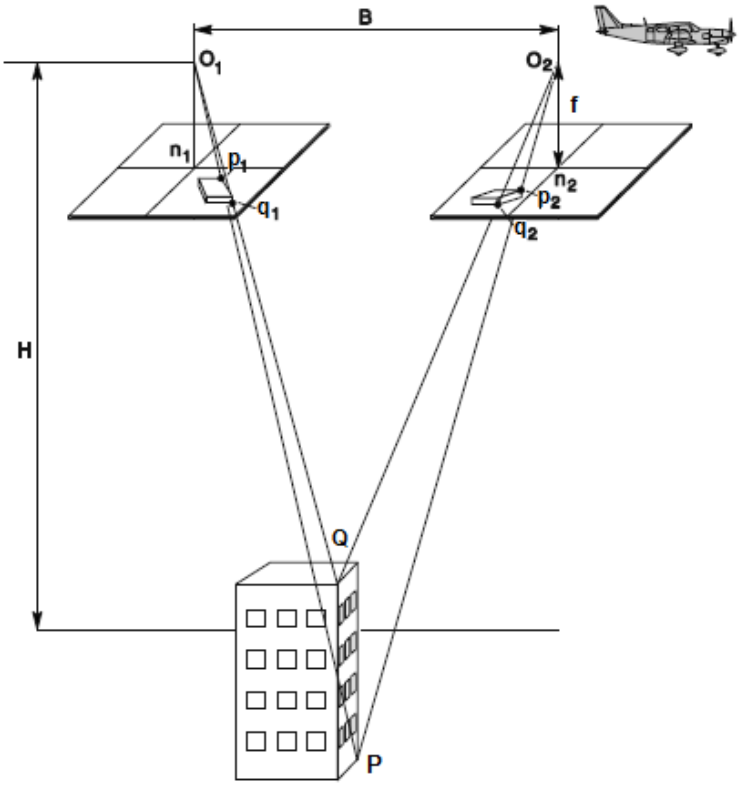


Figure 2.8: Principle of aerial photogrammetry with: P and Q = object points, p and q = image points, H = flying height, f = focal length, B = baseline, O = perspective center and n = principal point. Figure adapted from [Khosh Elham, 2010]

2.7.2 Coordinate systems

In general, photogrammetry involves establishing the relationship between the camera, the images and the ground. Each of these three must be defined with respect to a coordinate space and coordinate system.

Image coordinate system

The image coordinate system is related to the camera and its chip size. Its origin is at the center of the chip in the camera. This origin is also referred to as the principal point. Image coordinates are usually expressed in millimetres or microns. From now on image coordinates will be denoted as (x,y). If a height is added to the image coordinates, which equals the focal length, it becomes the image space coordinate system.

Pixel coordinate system

The coordinates of a digital image are usually defined in a pixel coordinate system. The pixel coordinate system is based on the actual image. The origin of this system is usually in the upper left corner of the image. The units are pixels, so the coordinates *(c,r)* should be seen as the column number and row number. The pixel coordinate system as well as the image coordinate system are represented in Figure 2.9.

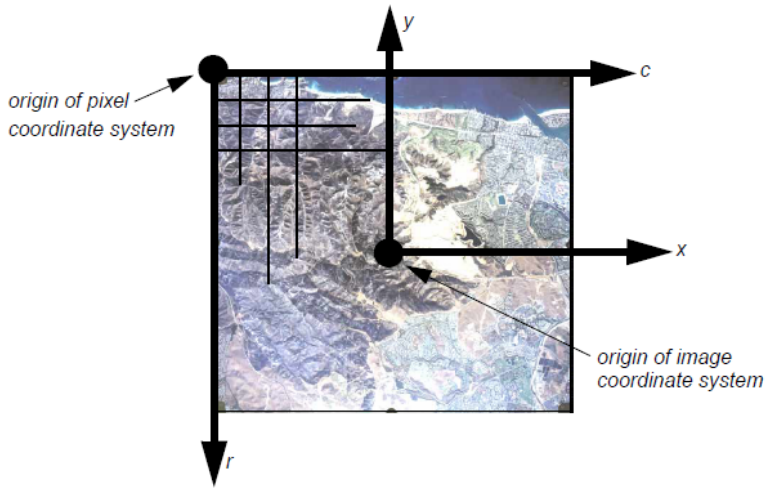


Figure 2.9: Pixel coordinate system and image coordinate system. (Illustration from [Leica Geosystems GIS and Mapping, 2003])

Object coordinate system

The object coordinate systems utilizes normally a known geographic map projection. In The Netherlands, RD-coordinates are often the standard. This system is two-dimensional and finds its origin 155 km west and 463 km south from the 'Onze Lieve Vrouwetoren' in Amersfoort. The height is in The Netherlands in general measured in NAP (Normaal Amsterdams Peil). From now on object coordinates will be denoted as (X, Y, Z).

2.7.3 Interior orientation

Interior orientation defines the internal geometry of a camera or sensor as it existed at the time of image capture. Interior orientation is primarily used to transform the pixel coordinate system to the image space coordinate system, [Leica Geosystems GIS and Mapping, 2003]. The internal geometry of a camera is defined by specifying the following variables:

- principal point
- focal length
- pixel size
- lens distortion

The principal point is defined as the intersection of the image plane and the line perpendicular to the image plane, through the perspective center. The length from the principal point to the perspective center is called the focal length, [Wang et al., 1990]. In traditional aerial photogrammetric cameras, so called fiducial marks are visible in each image, which are used for calculating the image position of the principal point. These fiducial marks are measured (in a pixel coordinate system) and compared with the coordinates that were calibrated in a laboratory. Using a 2D affine transformation, the relationship between the pixel coordinate system and the image space coordinate system is defined. An affine transformation consists of a linear transformation followed by a translation. The 2D-affine transformation to transform pixel coordinates to image coordinates can look like the following:

$$x = a_1 + a_2 \cdot c + a_3 \cdot r \tag{2.13}$$

and

$$y = b_1 + b_2 \cdot c + b_3 \cdot r \tag{2.14}$$

2.7.4 Exterior orientation

The position and orientation of the camera at the time of the image acquisition, is referred to as the exterior orientation. This exterior orientation is needed to link the image coordinate system to an object coordinate system. The location of the projection center consists of three parameters (X_0, Y_0 and Z_0) and another three parameters describe the orientation of this projection center (omega, phi and kappa). These six parameters are defined in the object coordinate system. The orientation is defined as a 3×3 rotation matrix, which is the product of three 3×3 rotation matrices representing sequential rotations around the X-axis, Y-axis and Z-axis of the object coordinate system. A rotation around the X-axis is represented by the parameter omega, a rotation around the Z-axis is represented by kappa and the rotation angle around the Y-axis is denoted as phi. All rotations can be joined in one 3×3 rotation matrix M as a product of three rotation matrices around the three axes:

$$M = M_\omega \cdot M_\varphi \cdot M_\kappa \tag{2.15}$$

where

$$M_\omega = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos(\omega) & sin(\omega) \\ 0 & -sin(\omega) & cos(\omega) \end{bmatrix}, M_\varphi = \begin{bmatrix} cos(\varphi) & 0 & -sin(\varphi) \\ 0 & 1 & 0 \\ sin(\varphi) & 0 & cos(\varphi) \end{bmatrix}$$

$$\texttt{and } M_\kappa = \begin{bmatrix} cos(\kappa) & sin(\kappa) & 0 \\ -sin(\kappa) & cos(\kappa) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The elements of the rotation matrix are used in the so called collinearity equations. These equations rely upon the principle of the central perspective projection, namely that the perspective center, an image point and its corresponding object point are on one line. This principle implies that:

$$\begin{bmatrix} x_p \\ y_p \\ z_p \end{bmatrix} = s \cdot M \cdot \begin{bmatrix} X_P - X_0 \\ Y_P - Y_0 \\ Z_P - Z_0 \end{bmatrix} \tag{2.16}$$

where the left hand side of the equation represents the image coordinates of a point P in object space, while X_P, Y_P and Z_P represent the object coordinates of P. The z component of the image coordinates equals the focal length f[1]. The object coordinate vector is subtracted by the position vector of the objective center and then multiplied by a scale factor s and the rotation matrix M. Straightforward mathematics lead to the two collinearity equations:

$$x_p = f \cdot \left[\frac{m_{11}(X_P - X_0) + m_{12}(Y_P - Y_0) + m_{13}(Z_P - Z_0)}{m_{31}(X_P - X_0) + m_{32}(Y_P - Y_0) + m_{33}(Z_P - Z_0)} \right] \tag{2.17}$$

and

$$y_p = f \cdot \left[\frac{m_{21}(X_P - X_0) + m_{22}(Y_P - Y_0) + m_{23}(Z_P - Z_0)}{m_{31}(X_P - X_0) + m_{32}(Y_P - Y_0) + m_{33}(Z_P - Z_0)} \right] \tag{2.18}$$

where m_{ij} are elements of the rotation matrix M. Points that are known in image coordinates as well as object coordinates are called control points. The collinearity equations can be used to calculate the exterior orientation parameters. The retrieval of the exterior orientation parameters of a single image with control points is called space resection. If in one image three control points are visible and measured, then it is possible to calculate the exterior orientation parameters of that image. The unknown parameters are the six exterior orientation parameters and there are also six equations from the control point measurements (in the image), so the system can be solved. Due to the rotation angles, there exists a nonlinear relationship between the

[1] Traditional photogrammetry (analogue) uses a negative focal length.

parameters and the observations and thus the solution is found iteratively. If for a sequence of images the exterior orientation parameters must be solved with space resection, then at least three control points should be visible in every single image. Another way of obtaining the exterior orientation parameters is via GPS/INS. The parameters will be directly available and can be used as initial estimates or as final result. However, as described in Section 2.4, the costs will be very high and therefore not interesting for this research.

Relative and absolute orientation

The most common way for obtaining the exterior orientation parameters is with the method of relative and absolute orientation. If a stereo pair of images is available, one image can be relatively oriented with respect to the other. For this relative orientation, image coordinates of tie points are required. Tie points are characteristic points, which appear in both images, and are also referred to as point correspondences. Relative orientation relies on the collinearity equations as well and requires image coordinates of at least five tie points (for technical clarification, see Section 2.9) in the overlapping area of the stereo pair. A 3D model of the terrain can be reconstructed, however in an arbitrary scale. At least three control points, each visible in at least two images and not lying on a straight line, are required to obtain the exterior orientation parameters for every image in the desired coordinate system. This last step is referred to as the absolute orientation.

2.7.5 Bundle Adjustment

Recovering the positions and orientations of the camera for each image and the calculation of 3D coordinates of all point correspondences is generally performed by bundle adjustment. The main objective of bundle adjustment is to obtain the 3D structure of the terrain and to estimate the camera position and orientation and possibly calibration estimates, [Triggs et al., 1999]. This is done by solving the collinearity conditions that were defined in Section 2.7.4. The solution is obtained by minimizing the reprojection error, i.e. the sum of distances between the observed and the predicted image points. Assume that n 3D points are observed in m images and let $\mathbf{x}_{ij}$ be the 2D projection of the ith point in image j. The minimization problem is then represented by:

$$min \sum_{i=1}^{n} \sum_{j=1}^{m} d(\mathbf{p}_{ij}, \mathbf{x}_{ij})^2 \tag{2.19}$$

where $\mathbf{p}_{ij}$ is the predicted projection of point i on image j and d(a,b) is the Euclidean distance between image points a and b.

This minimization problem can be formulated as a nonlinear least-squares problem due to the rotations that are involved in the model. The most successful algorithm for solving such nonlinear minimization problems is proven to be Levenberg-Marquardt due to its ease of implementation and its ability to converge quickly from a wide range of initial guesses, [Lourakis and Argyros, 2009]. In Section 2.6 a detailed explanation for solving a nonlinear least-squares problem can be read. In a bundle adjustment, the total number of minimization parameters is relatively large, even for short sequences. Therefore, bundle adjustment tends to be quite slow in finding its solution. However, many parameters do not interact with each other so that the Jacobian will have a regular sparse block structure (the valued elements are regularly spaced while the other elements are zero). Sparse Bundle Adjustment (SBA) takes advantage of this sparse structure and is therefore faster than the traditional bundle adjustment.

2.8 Scale Invariant Feature Transform[2]

The more tie points are found, the better the reliability of the relative orientation will be, since the model will be less sensitive for bad observations. Measuring tie points can be done manually, but nowadays photogrammetric software offers an automatic tool for obtaining the relative orientation parameters. Such a tool needs an algorithm to match corresponding points between images. Scale Invariant Feature Transform (SIFT) is a technique that detects distinctive features in an image. These features, from now on referred to as keypoints, will be used as candidate correspondences. A noteworthy property of SIFT is that the keypoints are invariant to image scale and rotation. There are four main stages [Lowe, 2004] for the extraction of the keypoints:

1. Scale-space extrema detection: Searches for candidate keypoints over the entire image and over all scales.

2. Keypoint localization: From all candidate keypoints, based on measures of their stability, the final keypoints are selected.

3. Orientation assignment: At each keypoint location an orientation is calculated based on the local image gradient directions.

4. Keypoint descriptor: The local image gradients are measured at the selected scale in the region around each keypoint. These are transformed into a descriptor for each keypoint, represented by a 128 items long vector.

[2]This section builds on [Lowe, 2004]

The scale-space extrema detection is done to find potential keypoints that are stable features across all possible scales. By using a continuous function of scale, locations that are invariant to these scale changes can be found. The filtering of these potential locations minimizes the computational cost by avoiding applying the more expensive operations at all locations. From all potential keypoints, some are being disregarded based on having low contrast or are poorly localized along an edge [Lowe, 2004]. Features with low contrast are considered to be sensitive to noise and therefore unstable. A threshold for minimum contrast is applied to filter out these unstable features. By assigning an orientation to each keypoint based on local image properties, future operations like setting up the keypoint descriptor can be represented relative to this orientation and therefore achieve invariance to image rotation [Lowe, 2004]. Peaks in the orientation histogram correspond to dominant directions of local gradients. All local peaks within 80% of the highest peak are used to create extra keypoints at those locations. Thus, for each location with multiple dominant directions, multiple keypoints are created with the same scale but different orientation. About 15% of the locations are assigned with more than one keypoint and contribute significantly to the stability of matching [Lowe, 2004].

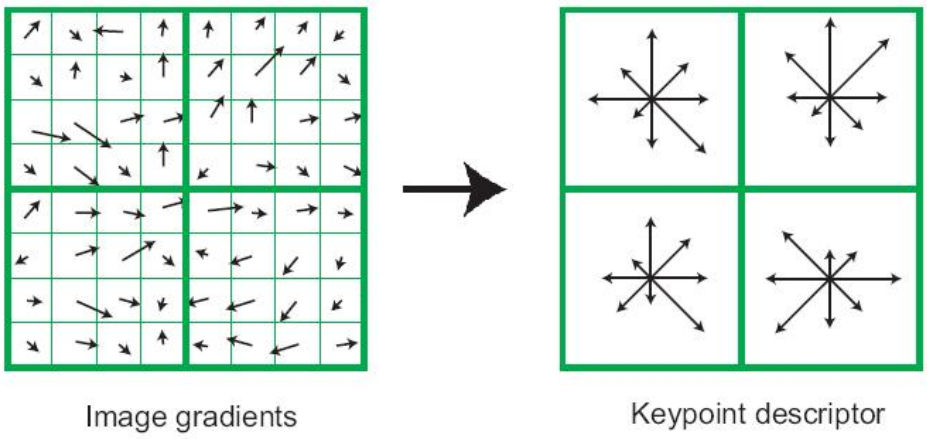


Figure 2.10: The creation of a keypoint descriptor. Illustration adapted from [Lowe, 2004]

The last step is to compute a highly distinctive descriptor for each keypoint based on its local image region. Figure 2.10 shows how the keypoint descriptor is made. First, the gradient magnitude and orientation for a set of pixel samples around the keypoint location is computed. These samples are then accumulated into eight directions summarizing the contents over 4×4 subregions. The length of each arrow corresponds to the sum of the gradient magnitudes near that direction within the region. Note that Figure

2.10 shows a 2×2 descriptor array computed from an 8×8 set of samples, whereas SIFT uses 4×4 descriptors computed from a 16×16 sample array. So, for each keypoint $4 \cdot 4 = 16$ descriptors are computed each containing gradient magnitudes of eight directions. Therefore, the descriptor vector contains $16 \cdot 8 = 128$ elements.

Matching SIFT keypoints

Furthermore, in [Lowe, 2004] a robust matching is provided across a substantial range of distortion, change in viewpoint and change in illumination. The features are claimed to be highly distinctive, so that a single feature can be matched correctly with high probability.

The best candidate match for each keypoint is found by identifying its nearest neighbour. The nearest neighbour is defined as the keypoint with minimum Euclidean distance for the invariant descriptor vector. The Euclidean distance between point p and q is defined as:

$$d = \sqrt{\sum_{n=1}^{k} (p_n - q_n)^2} \tag{2.20}$$

where the points are defined in the k^{th}-dimension. However, in the SIFT matching algorithm, the distance is expressed in a dot product. As explained in [Wan and Carmichael, 2005], the distance can also be calculated by taking the arc cosine of the dot product of the descriptor vectors:

$$d = cos^{-1}(\mathbf{p} \cdot \mathbf{q}) \tag{2.21}$$

For one keypoint, the distance to all keypoints in another image is calculated in order to find its nearest neighbour. For instance,

$$D = \begin{bmatrix} cos^{-1}(des_{11} \cdot des_{21}^T) \\ cos^{-1}(des_{11} \cdot des_{22}^T) \\ \vdots \\ cos^{-1}(des_{11} \cdot des_{2j}^T) \end{bmatrix} \tag{2.22}$$

where des_{ij} is the descriptor vector of keypoint j in image i, is the distance vector from keypoint 1 in image 1 to all keypoints in image 2. The smallest value in this distance vector corresponds to a keypoint j that is the nearest neighbour. Not every feature from an image should however be matched with a feature in the other image because they do not look enough like each other or arise from background clutter. Therefore, these features are discarded by comparing the Euclidean distance of the closest neighbour to that of the second-closest neighbour. The distance of the closest neighbour

divided by the distance of the second-closest neighbour is called the distance ratio. From Figure 2.11 it can be seen that matches for which the nearest neighbour was a correct match have a Probability Density Function (PDF) that is centered at a much lower ratio than that for incorrect matches. Matches above a certain distance ratio should be rejected to eliminate false matches. For example, a threshold of 0.8 indicates that 90% of the false matches are eliminated while less than 5% correct matches are disregarded.

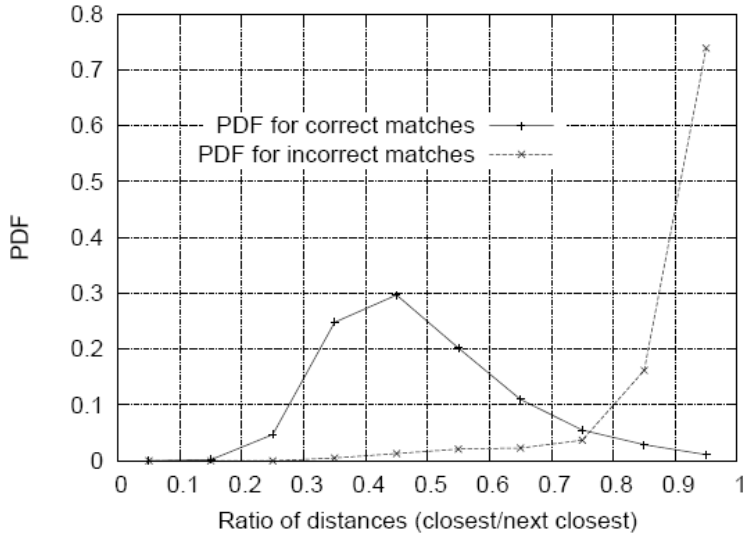


Figure 2.11: PDF of the distance ratio of correct matches against the PDF for incorrect matches. Illustration from [Lowe, 2004]

2.9 Relative orientation with epipolar geometry

Recovering the relative orientation between two images is an old issue. The original relative orientation consists of 3 rotation parameters and 2 translation parameters. One point correspondence gives image coordinates in both images, so four observations. Each observation results in three extra unknowns, i.e. the X-,Y- and Z-coordinate of the point in the object space. Five point correspondences results in a model with 20 observations and 20 unknowns and is therefore the absolute minimum to solve the relative orientation. Kruppa [Kruppa, 1913] was the first to describe a solution with only 5 point correspondences. Nowadays, the implementation of Nistér [Nistér, 2003] is common good. A more straightforward method is the one described by Hartley [Hartley, 1997] where the nonlinear 'five point' method is simplified to a linear 'eight point' method at the price that more observations

are needed. When more than eight point correspondences are available, a linear least-squares minimization problem can be formulated. In this thesis, it is expected that a lot more than eight point correspondences can be found automatically so the method of Hartley is described in more detail now.

The method relies on epipolar geometry. The principle of epipolar geometry is illustrated in Figure 2.12. The intersections with the baseline and image planes are called the epipoles and are represented in the figure by e and e'. Note that these points can also lie outside the physical image. Any plane π containing the baseline (and thus both epipoles), is an epipolar plane. The intersection of the epipolar plane with the image planes defines two epipolar lines l and l'. According to the principle of the central perspective projection (as explained in section 2.7.4), the objective center, an object point (U) and its corresponding image point (u) fall on a straight line. Each epipolar plane defined by an arbitrary object point contains both image points u and u', the baseline and the epipoles. Any point u' in the second image matching the point u must lie on the epipolar line l', [Hartley and Zisserman, 2003].

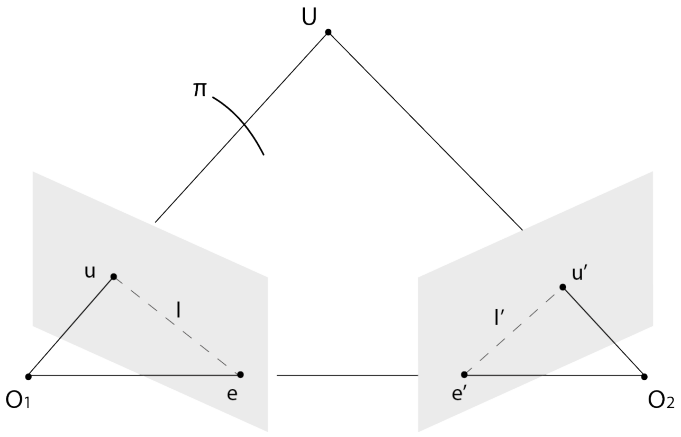


Figure 2.12: Epipolar geometry

The fundamental matrix F is the algebraic representation of epipolar geometry, [Hartley and Zisserman, 2003]. It is a 3×3 matrix defined by the following corresponding condition also known as the epipolar constraint:

$$\mathbf{u}'^T \cdot F \cdot \mathbf{u} = 0 \tag{2.23}$$

for every pair of corresponding points $\mathbf{u}'$ and $\mathbf{u}$. Given at least eight corresponding points, the fundamental matrix can be estimated. Writing $\mathbf{u} = (u, v, 1)^T$ and $\mathbf{u}' = (u', v', 1)^T$ leads to:

$$\begin{bmatrix} u_1u_1' & u_1v_1' & u_1 & v_1u_1' & v_1v_1' & v_1 & u_1' & v_1' & 1 \\ u_2u_2' & u_2v_2' & u_2 & v_2u_2' & v_2v_2' & v_2 & u_2' & v_2' & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_nu_n' & u_nv_n' & u_n & v_nu_n' & v_nv_n' & v_n & u_n' & v_n' & 1 \end{bmatrix} \cdot \begin{bmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{bmatrix} = 0 \quad (2.24)$$

or in simplified form

$$A \cdot \mathbf{f} = 0 \quad (2.25)$$

The vector $\mathbf{f}$ should be found such that it minimizes $A \cdot \mathbf{f}$ but is non-zero. One way to let $\mathbf{f}$ not be a zero vector is to make it fulfil the constraint that $\mathbf{f}^T \cdot \mathbf{f} = 1$. Hartley states that $\mathbf{f}$ equals the smallest eigenvector of $A^T \cdot A$ that can be found by singular value decomposition (SVD). By expressing $A^T \cdot A$ as a product UDU^T with SVD, the smallest eigenvector is the last column of U. The input (image coordinates of the point correspondences) is transformed before solving the linear least-squares minimization problem. The transformation consists of a translation and scaling and leads to an enormous improvement of the stability of the result [Hartley, 1997]. The translation changes the origin from the top left corner of the image to the center of the image. The points are then scaled (in each image separately), so that the average distance to the origin equals $\sqrt{2}$. Finally, the translation and rotation can be extracted from this fundamental matrix.

2.10 Lenses

2.10.1 Lens distortion

Images are always distorted to a certain extent due to imperfections of the lens. Even the best lenses need to be corrected for this lens distortion in order to apply (close range) photogrammetry with it. Lens distortion occurs when light rays through the lens are bent and as a consequence are intersecting the image plane at positions different from the norm, [Leica Geosystems GIS and Mapping, 2003]. There exists two types of lens distortion: radial and tangential lens distortion. Radial lens distortion, also referred to as symmetrical lens distortion, distorts image points along radial lines from the principle point. Thus, points with equal distance from the principle point have the same distortion. This distortion is determined for photogrammetric cameras during calibration in a laboratory. However, the cameras that will be used in this thesis will be off-the-shelf cameras and

thus a determination of the lens distortion is necessary. Photogrammetric software like PhotoModeler have options to perform this calibration with any type of camera. Photomodeler approximates the lens distortion by a sixth order polynomial:

$$\Delta r = k_0 r^2 + k_1 r^4 + k_2 r^6 \tag{2.26}$$

where Δr represents the radial lens distortion along the radial distance r from the principle point. The three coefficients, k_0, k_1 and k_2 need to be computed with statistical techniques. Tangential lens distortion, also referred to as decentering distortion, is the deviation of centering of the lens along the optical axis. Tangential distortion is usually smaller than radial distortion by a factor of ten or more, [Remondino and Fraser, 2006], and therefore often considered negligible. However, for high accuracy measurements it has to be considered. The formulas used by PhotoModeler for tangential distortion are: [Wiggenhagen, 2002]

$$\Delta p_x = P_1 \cdot (r^2 + 2 \cdot x^2) + 2 \cdot P_2 \cdot x \cdot y \tag{2.27}$$

$$\Delta p_y = P_2 \cdot (r^2 + 2 \cdot y^2) + 2 \cdot P_1 \cdot x \cdot y \tag{2.28}$$

where Δp_x and Δp_y are the distortion values and P_1 and P_2 are the two coefficients for the tangential distortion that need to be determined.

2.10.2 Zoom lenses

Cameras with a zoom lens are able to change the focal length. A longer focal length corresponds to a narrower angle of view while a shorter focal length corresponds to a wider angle of view. The principal distance is the distance between the image plane and the point where all light rays intersect. The principal distance is equal to the focal length when the lens is focused at infinity. However, if focused on objects that are closer to the camera the principal distance increases, [Website Photomodeler]. The most extreme example of this is when one focuses on a very close by and small object. The relation between focal length, principal distance and the distance to the object is given by:

$$\frac{1}{f} = \frac{1}{d} + \frac{1}{p} \tag{2.29}$$

where f is the focal length, d the distance to the object of focus and p the principal distance. For instance, a lens that has a 50 mm focal length has a 100 mm principal distance when focused at a 10 cm. Within aerial photogrammetry, the principal distance and focal length are often used interchangeably because they are practically the same due to the large distance to the object. Although for close range photogrammetry these are not the same, 'focal length' will be used for the remainder of this thesis instead of 'principal distance'. The variable character of the focal length

(principal distance) due to focusing at objects at different distances makes it recommendable to calibrate a lens for the distance to be worked with. Furthermore zoom lenses have larger lens distortions and therefore they need to be calibrated with a higher order polynomial.

2.10.3 Fisheye lenses

Fisheye lenses allow imaging a large area of the surrounding space by a single image. This property make fisheye images suitable for a diversity of applications. Although this large coverage seems helpful for photogrammetric applications, fisheye lenses produces extreme barrel distortion and also exaggerates the relative size of objects conform to its format. Barrel distortion gives the effect that an image is projected around a sphere as illustrated in Figure 2.13 and is in fact a strong radial lens distortion. Objects near the center appear larger than those at the edges and they show less barrel distortion.

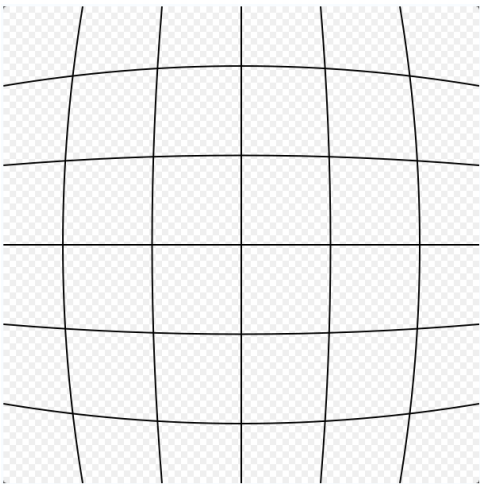
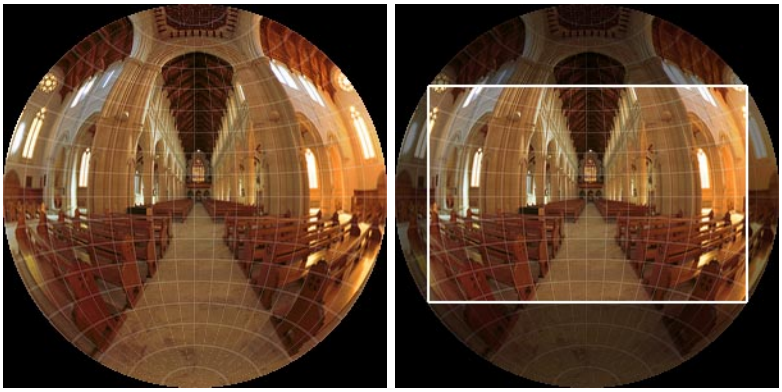


Figure 2.13: Barrel distortion

An example of a fisheye image is shown in Figure 2.14. Figure 2.14(a) is an original fisheye image with a horizontal AOV of around 180 degrees, which can be concluded from its circular shape. Rectangular images are, however, easier to work with, so often a rectangular cut is taken as can be seen in Figure 2.14(b). Making circular fisheye images rectangular is at the price of information loss. For fisheye lenses, the relation between diagonal AOV, horizontal AOV and vertical AOV is described by: $DAOV^2 = HAOV^2 + VAOV^2$. The diagonal AOV remains around 180° and respecting the standard image ratio (image width = 2/3 image length) gives a horizontal AOV of 150° and a vertical AOV of 100°.



(a) Original fisheye image (b) Rectangular cut

Figure 2.14: Fisheye images

Besides the barrel distortion, the geometry of images taken with fisheye lenses does not coincide with the central perspective projection, which was explained in Section 2.7. Therefore the collinearity equations cannot be used to mathematically describe the imaging process, [Schwalbe, 2006]. In order to realise a wider opening angle for a lens, the focal length needs to be decreased. However, in the central perspective projection this can only be done to a certain extent. This can be explained by a ray with an incidence angle of 90 degrees, which would be projected at an infinite distance from the principal point, independent on the focal length. A different projection model becomes necessary in order to project objects from all incidence angles onto the image plane. The fisheye projection is based on the principle that the distance between an image point and the principal point is linearly dependent on the angle of incidence. Therefore incoming rays are refracted in the direction of the optical axis as can be seen in Figure 2.15. For a fisheye lens, having a horizontal AOV of 180 degrees, an image point projected onto the outer border of the circular image corresponds to a ray with an incidence angle of 90 degrees. The fisheye projection is clarified further in Figure 2.16, which graphically shows the geometric relation between the object point and image point. The relation between the distance to the principal point and the incidence angle α is constant for the entire image. This relation results in the basic equation for the fisheye projection:

$$\frac{\alpha}{r} = \frac{90}{R} \tag{2.30}$$

where R is the radius of the image and r is the distance from the principal point to the image point:

$$r = \sqrt{x^2 + y^2} \tag{2.31}$$

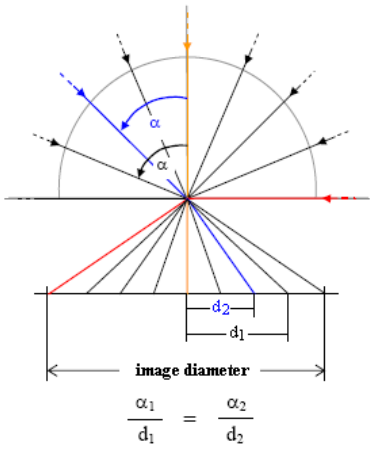


Figure 2.15: Fisheye projection model, [Schwalbe, 2006].

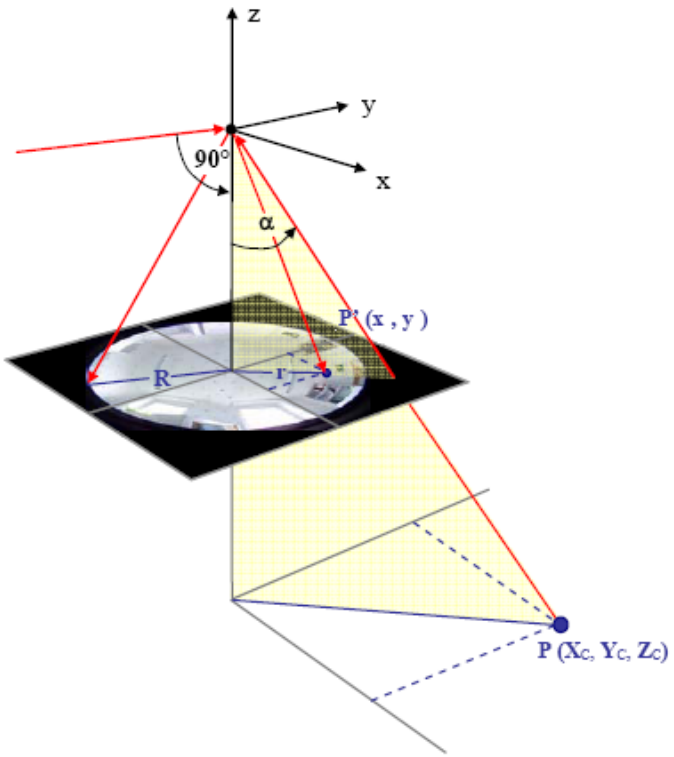


Figure 2.16: Geometric relation between object point and image point within a fisheye projection, [Schwalbe, 2006].

Fisheye lens calibration

Due to the highly distorted images produced by fisheye lenses, it is of paramount interest to perform a very precise lens calibration. This calibration has to estimate the interior orientation parameters as well as the lens distortion parameters. For fisheye lenses the interior orientation parameters consists of the image radius R and the principal point coordinates, [Schwalbe, 2006]. The parameter set for modelling lens distortion is equivalent to the parameters used for central perspective cameras extended by two parameters of an affine transformation (affinity and shear), [Schwalbe, 2006]. A method proposed by [Kedzierski and Fryskowska, 2008] determines arc curvature of line segments that are in reality straight lines. Then, from the curved lines, straight lines are obtained with no distortion. The used fisheye lens, having a 10.5 mm focal length, produces an image that is not hemispherical. The test field contained 230 points with known local coordinates having an error of ± 0.7 mm. The image was taken on 0.5 m to 2 m distance to the calibration points. Another calibration method of a fisheye lens, proposed by [Schwalbe, 2006], is done by spatial resection from a single hemispherical image. In the calibration room, 140 known object points having a standard deviation between 0.02 mm and 0.23 mm, were distributed in such a way that they form concentric circles on the image plane. Distances between neighbouring points on the respective circles are equal. Points projected into the outer part of the image have a larger distortion and thus a sufficient number of targets in the outer part of the image is necessary in order to obtain representative results in the calculation of the camera parameters.

Chapter 3

Possible setups and Methodology

The aim of this chapter is to find the best way to integrate close range images into Pavescan. Section 3.1 provides a general description of this potential integration. In Section 3.2 the available cameras and their properties are described. Section 3.3 is a test for how many tie points can be found with SIFT in images covering only asphalt. Three options for a new setup are proposed in Section 3.4. One of the options is chosen for this thesis and its measurement setup is described in Section 3.5. Software that can do the processing of close range images and that is available is described in Section 3.6. The chapter concludes with a camera calibration in Section 3.7 and the possibilities of DTM construction in Section 3.8.

3.1 Integration of close range images into Pavescan

The main goal of integrating close range images into Pavescan is to improve the absolute accuracy of the road profile measured by the laser scanner. This implies that the laser scanner stays the main measurement device with the images as support to strengthen the positioning of the scan. As can be read in Section 2.7, via the approach of (close range) photogrammetry the position and orientation of the camera can be retrieved for every image. For each point measured in two or more images, its 3D coordinates can be calculated in the same coordinate system as the control points. In this way, a DTM can be constructed of the area covered by the images. Regarding the control point reduction of the traditional survey, there are basically two approaches:

1. Measure the current control points in the images and calculate their 3D coordinates

2. Use the camera positions as control points

The first approach replaces the traditional measurements of the control points by image measurements. This even may be automated if the control points could be detected automatically in the images. Some control points still need to be measured with a traditional survey, because photogrammetry requires also control points to link the model to local coordinates (see Section 2.7.4). The second approach uses the camera positions as a reference for the scan. If a stable construction is realized, the vector between the camera and laser scanner remains constant. A drawback of this approach is that the range accuracy of the scanner should be taken into account as was provided in Section 2.5. An advantage of this approach is that the orientation of the camera will also be known, and consequently the orientation of the scanner too. Thus, the position of all scan points can be calculated. Therefore the second approach is preferred.

An important question is whether the number of control points needed for photogrammetry could be less than the number of control points used for Pavescan. If so, one likes to know how the number of control points relates to the accuracy of the road profile to be obtained.

3.2 Available cameras and lenses

In order to be provided with good images, the choice for the camera and lens should be taken with care. However, for this thesis the choice for a camera and lens is limited to the ones available at TU Delft and Breijn. Via Breijn, the video camera that is currently mounted on Pavescan is available. TU Delft has two more cameras available with different lenses:

Canon EOS 350D
The Canon EOS 350D is a 8 megapixels (maximum resolution is 3456×2304 pixels) camera. The image sensor size is 22.2×14.8 mm, which implies a rectangular pixel size of 6.4 microns. There are two lenses available for this camera; one lens with a focal length of 20 mm and one zoom lens with a focal length of 18-55 mm. The camera is limited in taking images on a high rate. It can produce maximally a sequence of 14 images in a high rate.

Nikon D300
The second camera available at TU Delft is the Nikon D300. The camera has 12.3 effective megapixels (4288×2848 pixels). This image sensor dimensions are 23.6×15.8 mm, which implies a rectangular pixel size of 5.5 microns. A zoom lens and a fisheye lens are available for this camera. The zoom lens has a focal length that varies between 18-70 mm. The corresponding horizontal angle of view (HAOV) is between 22.8° and 76°. The fisheye lens

Table 3.1: The main properties of the available cameras

		Focal length	Resolution	Horizontal AOV	Average ground res.
IQeye 703		4-10 mm	2048 × 1536	94.8°	4.3 mm
Canon 350D	Zoom lens	18-55 mm	3456 × 2304	61°51′	1.4 mm
	Static lens	20 mm		49°5′	1.1 mm
Nikon D300	Zoom lens	18-70 mm	4288 × 2848	76°	1.4 mm
	Fisheye lens	10.5 mm		149°46′	6.9 mm

has a focal length of 10.5 mm, which corresponds to a diagonal angle of view (DAOV) of 180°. The camera has an option to convert the fisheye images to regular wide angle lens projection images. The converted images may be of use, though it is unknown how this conversion is performed. The HAOV and VAOV of these images are around 150° and 100° respectively. The main properties of all available cameras are summarized in Table 3.1.

3.3 Test for automatic tie point extraction

An image capturing only asphalt is very monotonous for the human eye. It is therefore doubtful if point correspondences can be extracted automatically. Hence, several test measurements are set up with the objective to gain knowledge about automatically extracting point correspondences from such monotonous images. This is a critical test since if the results are bad, the feasibility of implementing close range images into Pavescan drops greatly. These first test measurements are also a good way to become more familiar with the cameras and software. For the most realistic result it is desired to approach the standard situation of Pavescan. This means photographing a road made of asphalt and at a height of around 3.6 metres. Three different locations for the tests where chosen in Delft. Each test measurement consists of two or three overlapping images capturing asphalt from a distance between 1.5 and 6 metres. It was attempted to approximate a 60% overlap between the two images. David Lowe's SIFT keypoint detector and its matching algorithm were used to find point correspondences. It is not only important to know how many matches could be found, but also how these matches are distributed over the images and if the matches are correct. The latter is especially important if the relative orientation needs to be automated.

The first image pair was captured vertically down from a viaduct from a

height of around 6 metres. Besides asphalt, also a strip of brick pavement was captured. Running SIFT on the images appears to require more memory of the computer (1.83 GHz processor). Hence, the resolution of the images is decreased from 3456×2304 pixels to 1500×1000 pixels. This time SIFT was successful and found a lot of matches. However, almost all of the matches are on the brick pavement as can be seen in Figure 3.1. The figure consists of two images above each other. A line represents a match by connecting two corresponding points. If two consecutive perspective centers have the same height and attitude, all 'match lines' are perfectly parallel. Two crossing lines indicate false matching if it is assumed that the height and attitude are approximately constant. In the asphalt part of the image, at least three non-parallel lines can be detected. The low number of matches on the asphalt is quite disappointing.

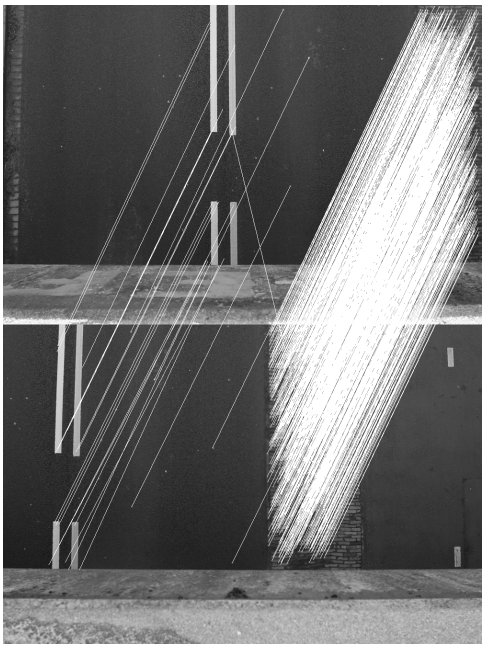


Figure 3.1: Matches from the two overlapping images of (pre-)test 1. Matches are represented by a line between two corresponding points. Crossing lines indicate false matching.

The second test is also captured from a viaduct from a height of approximately 5 metres. Three images are captured; one vertically down and two low oblique, all having overlap with each other. Even in the image pair with least overlap many matches were found as can be seen in Figure 3.2. Three false matches could be detected by eye when taking a closer look at the result. However, the result is promising, since this time all matches are on the asphalt and one image is captured low oblique.

The last test has the objective to test the effect of height. An image pair is captured vertically down, 1.5 metres above the asphalt pavement of a bicycle road. SIFT found around 150 matches and surprisingly no false matches could be detected by eye. There is no clear relation discovered between the height and number of matches. However, it seems that a smaller height is more successful for finding point correspondences. It is concluded that it is possible to find a lot of point correspondences on asphalt pavement with SIFT. Probably, a high resolution (i.e. small height) is beneficial for the number of point correspondences.

Figure 3.2: Matches from the stereopair of (pre-)test 2.

3.4 Proposed setups

Three possible setups are opted and examined:

1. Take one image at every stop (fisheye lens)
2. Take images while driving to the next stop
3. Make a video and use a selection of frames

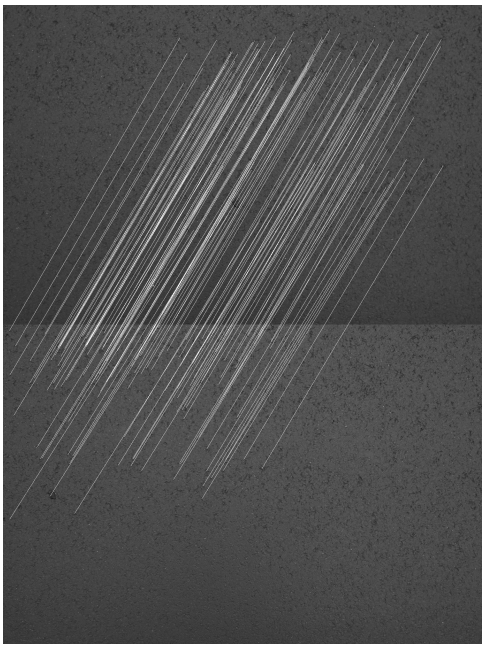


Figure 3.3: Matches from the stereopair of (pre-)test 3.

On the current system setup of Pavescan, the camera can be mounted on a height of maximally 3.6 metres. For each of the three options it will be determined whether it is practically possible to include it into this thesis. Independent from which method(s) is/are being researched, the challenges remain the same: perform an automatic relative orientation on the images/frames and retrieve the position and orientation of the camera for each image/frame. Thereby, all tie points measured in the images become known in local coordinates.

One practical drawback for all options is that if the camera is pointed vertically downwards, the images/frames will cover besides the road also the trailer. There are three ways to deal with this: cut off the part of the image that covers the trailer, rotate the camera to the back or position the camera behind the pole as is illustrated in Figure 3.4. Practically, the first two solutions are the most straightforward.

3.4.1 Option 1: Fisheye lens

The first option is to take one image per stop, which is illustrated in Figure 3.5. The minimum distance between two stops that is used now is 10 m, which is not desired to be decreased. When the images are captured vertically down, these should have a footprint of $\frac{10m}{1-0.6} = 25$ metres to meet the

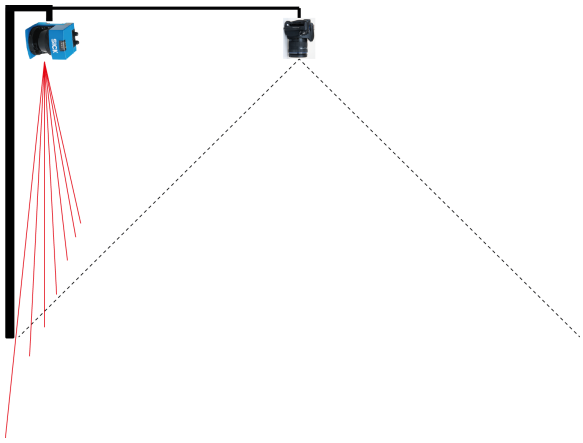


Figure 3.4: The positioning of the camera for an optimal footprint.

desired 60% of overlap between two consequtive images. If a camera with a normal (i.e. central projective) lens is mounted at a height of 3.6 metres, two consecutive images will not overlap at all. A larger angle of view or a higher position of the camera will increase the size of the footprint. However, the camera is limited in being mounted higher due to viaducts and tunnels to drive through. Thus, the most straighforward solution would be to use a lens with a larger angle of view. While still assuming the camera to be pointed vertically down, the angle of view should be $2 \cdot \arctan \frac{25/2}{3.6} \approx 148°$ to meet the requirements of 60% overlap. From Table 3.1 it is concluded that a fisheye lens, with its standard horizontal angle of view of 150°, is the only option.

Taking images only at stops brings many advantages. In this way the vehicle is stabilized and thus images will not be affected by vibrations or sudden rotations. Another advantage is that because the camera is not moving with respect to the road's surface, the images are protected from image blur. Image blur is a loss in sharpness caused by the movement of the camera during exposure. The degree of image blur is calculated by multiplying the movement velocity by the exposure time. For example if the camera moves with 1 m/s and the exposure time is 1/100 second, then the image blur is 1 cm. A third advantage is that fewer images need to be processed. This is not only advantageous for the data storage and computation time, but also for the error propagation. The more images that need to be processed, the more relative orientations must be done and thus the error is likely to propagate faster. The main drawback of taking images only at stops appears to be that one is forced to use a fisheye lens. The use of a fisheye lens brings some serious issues. In Section 2.10.3 it can be read that fisheye

lenses have a different projection and therefore differ from the standard photogrammetric mathematics. Its projection causes pixels at the edges of an image, when projected to object space, to have a much larger 'footprint' compared to pixels in the center of an image. The image measurements of control points can not achieve the accuracy that can be obtained in normal images. The quality of the absolute orientation will therefore be weaker than the absolute orientation that can be achieved with normal images. Although photogrammetric applications are not well known with fisheye lenses, calibration precisions in the order of millimetres have been achieved, [Heuvel et al., 2006]. In order to let this option have a chance for success, one has to deal with the different projection of fisheye lenses. Besides, a very extensive lens calibration is needed as was described in Section 2.10.3. This is not desired by Breijn and beyond the scope of this thesis. Therefore, this proposed option will not be examined further in this research.

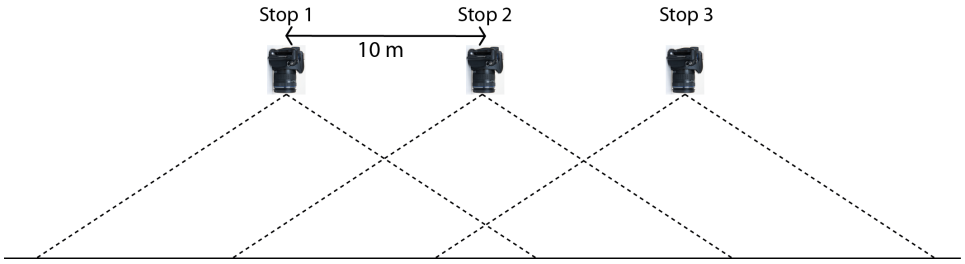


Figure 3.5: Schematic overview of option 1.

3.4.2 Option 2: Take images while driving

The second option is to take images while driving. Advantage of this option is that a normal, non-fisheye lens can be used so that known photogrammetric mathematical models can be applied and the fisheye lens calibration is avoided. The baseline (i.e. the distance between two consecutive exposure stations) is desired to be at least half the height of the camera (see Section 2.7.1). The longer the baseline, the better the height estimation of the object points and the less images are produced. Less images imply a faster relative orientation and consequently its error propagation is likely to be slower. With the camera at a height of 3.6 metres, the baseline is desired to be at least 1.8 metres. If the camera is pointed vertically down, an horizontal AOV of minimal $2 \cdot \arctan \frac{0.5 \cdot \frac{1.8}{1-0.6}}{3.6} \approx 64$ degrees is required to have 60% overlap. From Table 3.1 one can see that only the Nikon zoom lens is appropriate. If the camera is pointed slightly to the back, the footprint of an image will increase while less or nothing of the trailer will be covered. One can choose to keep the extra overlap that comes or increase

the baseline. These low oblique images have, however, different image scales within one image because the distance from pixels at the edge to the objective center is larger than the distance from pixels straight under the camera to the objective center. These scale differences may cause problems in the photogrammetric processing of the images, although no problems with SIFT are expected since it claims to be scale invariant.

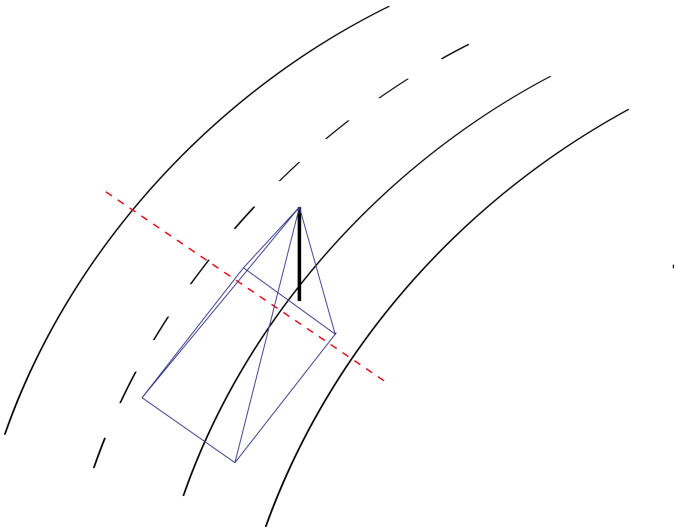


Figure 3.6: The footprint as covered by low oblique imagery. The blue rectangle represents schematically the footprint of an image. The red dashed line is the scan line.

3.4.3 Option 3: Videogrammetry

The third and last proposed option is to make use of video frames. Automatic relative orientation is generally not a straightforward task. According to the SIFT-tracking method as described in [Wortel, 2009], the relative orientation can be solved automatically for a set of largely overlapping image frames. However, the quality of the frames will never be as good as single shot images. This also reflects on the quality of the relative orientation. Another disadvantage is that a lot more data has to be processed. For each image frame, corresponding points should be found (e.g. by SIFT), which is a very time consuming task. Therefore it is not likely that videogrammetry will be the most attractive option.

3.4.4 Summary

Each of the proposed options has its advantages and disadvantages. The use of a fisheye lens limits the number of images, but does not rely on standard

photogrammetric mathematics. Also the difficult lens calibration makes the use of a fisheye lens practically not feasible for this thesis. Making images while driving results in more images and possible scale differences, but as long as the scale differences are not very large this is expected not to be a problem. The quality of video frames is much lower than the high resolution images. This reflects directly on the quality of the relative orientation. Besides, there will be so many frames that it will be extremely computational intensive since for every frame corresponding points should be found. It is concluded that the option to take the images while driving is expected to have the highest chance for success and therefore in this thesis the focus will be on this option. The video option will only be considered if the other option fails and there is more faith in a faster processing of the video frames.

3.5 Measurement setup

In Section 3.4 it was decided to focus this thesis on taking images while driving. An analysis of a sequence of these images should make it possible to answer the research question. The images will be taken with the camera rotated such that just the road is captured (and not the trailer).

A constant baseline is needed to regulate the total number of images. Often, photogrammetric images are captured based on an equal time interval. However, Pavescan is moving with a variable velocity, so images need to be captured based on an equal distance instead in order to maintain a constant baseline. A constant baseline can be realized by the principle of a cyclometer. One or more magnets should be mounted on the wheel of the vehicle and are registered by a sensor every time it passes the sensor. By knowing the perimeter of the wheel, the traveled distance can be computed with an accuracy depending on the amount of magnets on the wheel. Once the desired distance is traveled, the camera should be triggered. For this thesis it will be too complicated to realize such a system. Therefore, each image will be triggered manually. This implies that a remote trigger should be used since the camera will be at a height of 3.6 metres. A remote trigger is only available for the Canon camera. There exists remote triggers for the Nikon as well, but these are very expensive. In order to obtain the largest baseline, the lens with the highest possible horizontal AOV should be chosen. Also a static lens is recommended because it is of importance to capture all images at the same focal settings. Using a zoom lens cannot guarantee this requirement while a static lens can, as long as the Autofocus function is set off. Due to the availability of a remote trigger, the Canon camera must be chosen. Using a static lens is considered as more important than a high horizontal AOV. The choice for the camera and lens to be used is therefore the Canon EOS 350D with 20mm lens. Unfortunately, the horizontal AOV of this lens is just over 49 degrees while a minimum of 64 degrees is desired

as was described in Section 3.4. As a result, the baseline will be shorter and thus more images will be taken with a lower base-height ratio.

Furthermore, it is known that a large change in orientation between two consecutive exposures causes problems in the relative orientation. A curved road causes a rotation in the heading (rotation along the z-axis). Although the curves on a highway itself are not very sharp, an exit or junction can have a much shorter radius of curvature. If for example an exit has a radius of curvature of 100 m, the rotation of two consecutive images captured 2 m away is $\frac{2}{2\pi \cdot 100} \cdot 360 \approx 1°$. This should not give any problems. Rotations around the other axes (roll and pitch) are not likely to be much larger.

3.6 Software overview

The most common way to recover the position and orientation of each image and make a DTM of the covered area is via bundle adjustment. Due to the complexity of bundle adjustment, it is recommended to use additional software. Most of the appropriate software is not open source. Therefore, the possible software to be used in this thesis will be limited to those available at TU Delft and the few open source programs. Preferably for this research, the software is capable of running through this process fully automated. The first described software is Bundler, which is highly automated though not designed for (close range) photogrammetry. The second is BINGO, that is specialized photogrammetric software, however less automated. The third and last is Leica Photogrammetry Suite (LPS), that is also specialized photogrammetric software, and differs from Bundler and BINGO that it interacts more with the user.

3.6.1 Bundler

Bundler is a software package, written by Noah Snavely, for producing from a collection of images the scene geometry and the exterior orientation parameters of the camera(s) on the time of acquisition. The program is based on the Photo Tourism approach described in [Snavely et al., 2008], which was designed to view touristic hot spots three dimensionally from large collections of images available on the internet. Their main challenge was to deal with the difference of the image collections compared to the images in a traditional photogrammetry project. The images are unorganized, uncalibrated and having a wide variety in illumination, resolution and image quality. The approach consist of a *structure from motion* (SfM) procedure followed by *image rendering*. SfM aims at recovering the camera parameters, camera positions and orientations and reconstructing the unknown 3D scene from a set of point correspondences. Basically, SfM is similar to bundle adjustment but finds its origin in computer vision. The image rendering is not of interest in this research since it only has to do with the visualization

of the result and will therefore not be described.

The detection of point correspondences is performed by Lowe's SIFT keypoint detector followed by a matching algorithm (different than the one provided by SIFT). SIFT was chosen for its good performance over a reasonable range of viewpoint variation, [Snavely et al., 2008]. As described in Section 2.8, SIFT provides for each keypoint a descriptor. Bundler matches the descriptors using an approximate nearest neighbour kd-tree approach. It creates a kd-tree (a data structure in a k-dimensional space that divides the space at every point in two subspaces) from all descriptors in one image, i.e. a 128 dimensional space, and finds the nearest neighbour. For efficiency the approximate nearest neighbour search algorithm is used, limiting the number of visits of bins in the tree. The same ratio test as provided by the SIFT matching algorithm (see also Section 2.8) is used to assign matches. If more than one feature in the one image matches the same feature in the other image, all matches are removed since some of them must be incorrect. After the matching process, a fundamental matrix for the image pair is constructed according to Hartley's 8-points algorithm, [Hartley, 1997]. With RANSAC (Random Sample Consensus), the matches that do not fulfil the epipolar geometry constraint (see formula 2.23) to a certain threshold are removed.

Recovering the camera parameters and camera positions and orientations is executed with an incremental approach instead of recovering all parameters at once as in a standard bundle adjustment. Recovering the parameters can be formulated as an optimization problem. The process starts with choosing an initial pair of images with a high number of matches and a known focal length. For these two images the unknowns are estimated. Next, another image is added to the optimization. In the following bundle adjustment step, only the parameters of the new camera are allowed to change so the rest of the model is held fixed. Then, a sparse bundle adjustment (sba) is run to refine the entire model. This procedure is repeated (each time adding one image to the model) until no images are left for a reliable reconstruction of 3D points. Therefore, in general, not all images will be reconstructed. For the purpose of Bundler this is no problem since there are plenty of images that are highly overlapping. For Pavescan it would, however, be disastrous to skip some images since then there would be not enough overlap anymore and consequently, for the rest of the images nothing can be calculated. However, the images in the setup of Pavescan are expected to participate in the bundle adjustment in its full extent since all images are of comparable quality, i.e. same camera, approximate same illumination etc.).
The result is a relative 3D representation of the camera(s) and image points. For an absolute representation, the model should be aligned with a georeferenced image or map. The estimated camera locations are, in theory,

related to the absolute locations by a similarity transform (translation, rotation, and uniform scale). To determine the correct transformation the user interactively rotates, translates, and scales the model until it is in agreement with a provided image or map. Thus, the use of control points is not cooperated within Bundler. This implies that the result from Bundler should be transformed with a similarity transformation to be valuable for further processing in this research.

Bundler usage

Bundler does not have a graphical interface but consists of a script to be executed on a set of images. Bundler needs assistance of several other programs. The Bundler executable is actually the last in a sequence of steps that need to be run. First a list of JPEG images is created and with the program 'jhead' the focal length and CCD-width are extracted if available. Then, David Lowe's SIFT feature key point detector generates the SIFT features for each image (if you are planning to run Bundler yourself, it is recommended to read Appendix A). In the next step, these SIFT features are matched. These matches are the tie points in the photogrammetric process. The final step is to run the Bundler script with predefined options. The default options will result in calculations of lens distortion parameters as well as focal length for each image. In practice this implies that for each image a new camera will be detected since these parameters are never exactly the same. The output of Bundler is a file containing camera information and tie point information. The camera information consists of position, orientation and, depending on the options, focal length and lens distortion parameters. The tie point information lists the cameras that captured the point together with SIFT feature indicator and the position of the point. Another file lists all coordinates of tie points and projective centers. The latter can be used to import into (for instance) Matlab for visualization or modification of the points.

3.6.2 BINGO

BINGO is a program system that provides bundle adjustment for close range photogrammetry and aerial triangulation. Different from Bundler, this photogrammetric specialized software is not open source and thus a license is required. TU Delft has one license for BINGO and can therefore be used for this research.

Theory

BINGO uses bundle (block) adjustment for image orientation, which is the standard method in photogrammetry. The bundle adjustment optimizes the observations by the theory that all rays (from the projection center through

the measured image points) intersect in one point, i.e. the projective center. If a camera calibration is included in the bundle adjustment, the systematic errors in the images can be reduced considerably, [Kruck, 2003]. The search for data errors plays an important role in the bundle adjustment. Data snooping, an advanced search for data errors, first only detects errors in image measurements. A further process allows the automatic elimination of the detected errors in image measurements. However, outliers in the observations can fail the entire bundle adjustment so need to be removed carefully beforehand. Concerning all other types of observations like control points, GPS-data etc., the operator must correct or remove faulty observations himself. If 15 to 25 points are measured in each image, a stable block geometry is achieved. Wrong observations can then easily be detected and eliminated. BINGO also requires for every image i at least one corresponding point in image $i+1$ and $i+2$ (if these images exist) in the bundle adjustment. These special tie points are from now on referred to as third degree tie points. Each image pair is considered as one model, and the third degree tie points are needed to retrieve the scale between all models. Control points should also be measured in at least three images. Otherwise, no check of measurements is possible.

The equation system will be solved by a least-squares adjustment. The solution to this nonlinear model must be linearized and solved iteratively as described in Section 2.6. Initial approximates of the unknowns are required. These approximates are estimated in the BINGO component RELAX as far as these can be computed from the image measurements. When the bundle adjustment has converged to its solution, the search for errors starts. Together with the estimations for the camera positions and orientations, for each tie point its 3D object coordinates are estimated. These object points are reprojected to the images and its residuals are evaluated. Remember that minimizing this reprojection error is the main goal of bundle adjustment as was described in Section 2.7.5. If $e > 4 \cdot \sigma_0$, where σ_0 is the standard deviation of an image measurement, the observation will be eliminated from the model. The BINGO component SKIP lists these observations after each iteration. In the next adjustment many residuals will then be smaller. This is repeated until all the gross errors are eliminated. Finally, BINGO produces one output file with the exterior orientation parameters of each image and 3D coordinates of all object points. Another output file describes the quality of the bundle adjustment in terms of redundancy, variances, standard deviations etc.

3.6.3 Leica Photogrammetric Suite

Leica Photogrammetric Suite (LPS) is software for digital photogrammetric workstations, offering tools for a range of geospatial imaging applications.

LPS is a modular product where the core module is suitable for many standard applications. Modules that provide more advanced functionality can be added. Most important capabilities of LPS are ([Leica Geosystems Geospatial Imaging, 2008]):

- To define the internal geometry of imagery using measured fiducial marks
- Definitions for calibrated cameras and lenses can be defined and saved
- Automatic tie point extraction
- Triangulation; define the relationship between the imagery, the sensor model and the ground
- Removal of the scale variation caused by sensor orientation, topographic relief displacement and systematic errors associated with imagery, resulting in orthorectified images representing ground objects in their true positions, i.e. no relief displacement.
- Fully automatic DTM generation from two or more images

Although these capabilities sound very promising it is experienced that the software demands a very high level of interaction with the user. This automatically demands a high level of knowledge of photogrammetry from the user. The abundant settings and preferences are considered more advantageous for large photogrammetric projects than for Pavescan where photogrammetry will not be the main product. Therefore, LPS will not be considered as the best option to use for Pavescan so will not be used in this thesis.

3.7 Camera calibration

As stated in Section 2.10.1 it is necessary to determine the lens distortion. A camera calibration may offer besides lens distortion parameters also estimates for focal length, principal point offsets and chip size. One can choose to do an extensive and very accurate calibration in a special calibration room or a fast and straightforward calibration as offered in several photogrammetric software, which is less accurate. The latter is preferred, since such a calibration room is not available. A tool within PhotoModeler is used for the estimation of all these parameters. The calibration tool requires images covering a pre-defined pattern of dots, printed on a piece of paper, which positions are known in (local) coordinates. The images need to be taken perpendicular to each side of the grid under an angle close to 45 degrees, capturing all four control points located in the corners of the grid. Additionally, the grid should cover the image as much as possible. For the

best result, from each side an image must be taken in landscape (long image side in horizontal direction) as well as the two different portraits (long image side in vertical direction). One calibration image is presented in Figure 3.7.

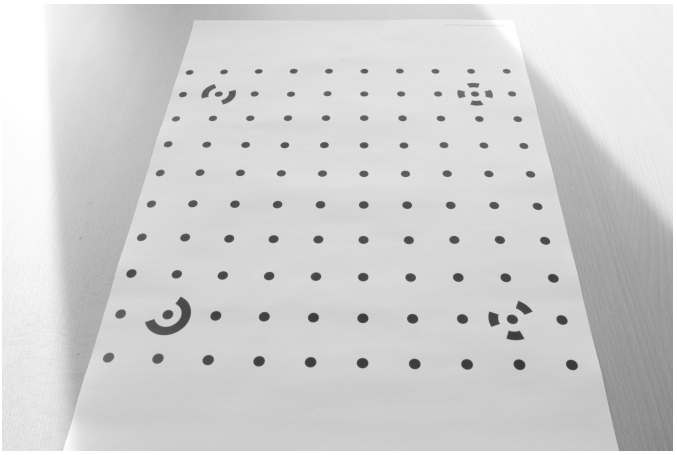


Figure 3.7: One of the calibration images (landscape). In the corners the control points are located.

Ideally the calibration is performed at the same focus setting (i.e. the same distance) as is used in practice. This is highly impractical since a camera on Pavescan will be mounted on a height of 3.6 metres and thus the calibration grid must also be several metres of width. For a first approximation this ideal setting is neglected and the calibration grid is printed on A3 paper, i.e. the images are taken at less than one metre. Following the steps from [Website Photomodeler], specific settings of the camera must be adjusted for an optimal result:

- Set the camera to black and white mode to prevent possible colour fringes around the targets.
- Disable the auto sharpening setting or set at lowest sharpness.
- Turn off the auto focus setting for a constant focal length.
- Disable the flash.
- Find an optimal shutterspeed / aperture proportion.

The focus has been set manually once for the used distance in advance of the series of calibration images. According to a rule of thumb, the shutter speed was set at one over the focal length, i.e. 1/20 second. The aperture value will be set subsequently in proportion to the shutter speed. An aperture

Table 3.2: Camera and lens parameters estimated by calibration in PhotoModeler.

Calibrated parameters of Canon EOS 350D, 20mm lens		
Focal length		20.9475 mm
Format size		22.1742 × 14.7828 mm
Principal point	X	10.9969 mm
	Y	7.2793 mm
Radial lens distortion	k_0	1.952 e^{-4}
	k_1	-3.326 e^{-7}
	k_2	0.0
Tangential distortion	p_1	2.926 e^{-5}
	p_2	6.298 e^{-6}

value of 10.0 seems good after trial and error. The twelve calibration images provide a highly redundant adjustment model resulting in the estimated calibration parameters presented in Table 3.2. From the parameters of radial lens distortion, an absolute distortion curve can be plotted according to Formula 2.26. This is illustrated in Figure 3.8.

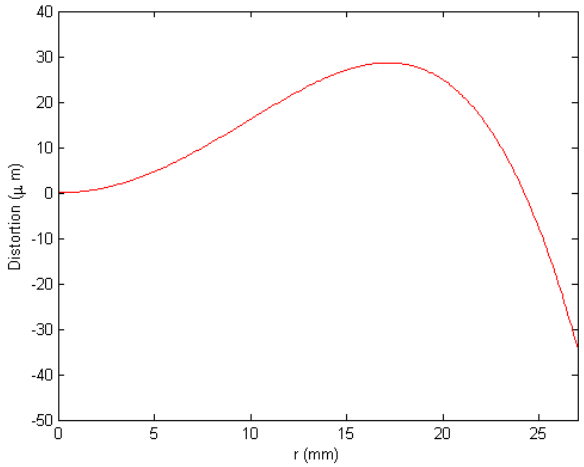


Figure 3.8: The absolute lens distortion based on the distance from the principal point.

The effect of making images at larger distances than the calibrated distance is as follows. When the distance to an object increases, the focal length decreases (actually, it is the principal distance that decreases), as was described in Section 2.10.2. A change in focal length causes a change

in direction of rays of light. Therefore, the amount of lens distortion is also slightly different. Besides, the accuracy of the focal length has a direct influence on the accuracy of the height estimation of object points. The calibrated values for focal length and lens distortion should therefore be used with care.

3.8 DTM construction

If it is assumed that the position and orientation of the camera can be retrieved, a DTM can be constructed for that part that is covered by at least two images. A DTM of the entire road would make the laser scan superfluous, provided that the accuracy is at least as good as the laser scan. This is however far from reality.

The footprint of a non-fishheye lens is maximally $2 \cdot 3.6 \cdot tan(\frac{76}{2}) = 5.6$ metres horizontally when captured vertically down or maximally $3.6 \cdot tan(\frac{76}{2} - 20) + 3.6 \cdot tan(\frac{76}{2} + 20) = 6.9$ metres horizontally when captured low oblique under 20 degrees. A two lane highway, the most common situation in The Netherlands, already covers a width of more than 10 metres. This implies that one camera is not sufficient for generating a DTM of the entire road's surface. A fisheye lens could be capable of covering the entire width of the road, but it is expected that the potential DTM will have a much lower accuracy at the edges since there the ground resolution of the images will be much higher.

From a previous pair of test images (see Figure 3.3) a DTM is constructed. Bundler will be used for calculating the 3D coordinates of the object points. The DTM is generated based on a TIN (Triangular Irregular Network) interpolation. The TIN interpolation is used since it is very fast and easy to implement. The result is, however, arbitrary scaled. The height assigned by Bundler is compared with the estimated height in the real situation. This ratio is used for an approximate scaling of the Bundler output. The DTM is presented in Figure 3.9 where the actual height is estimated at 1.5 metres. The DTM approximates a plane and has a clear trend. The trend indicates a gradually increasing elevation from the lowest elevated corner to the highest elevated corner. This height difference is 5-10 centimetres over one metre, which implies a slope of 5-10% . The bicycle road is expected to be nearly flat, so this is not in agreement. The use of better camera parameters (focal lenght and lens distortion parameters) may solve this problem.

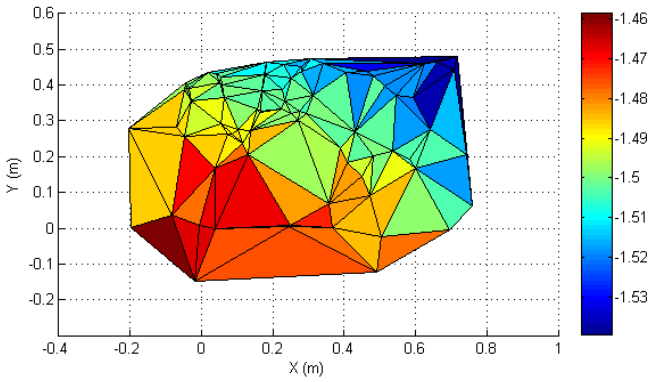


Figure 3.9: Top view of DTM of (pre-)test site 3, based on a TIN interpolation on the matches.

Chapter 4

Quality analysis

The aim of this chapter is to estimate the expected quality to be achieved with the approach as proposed in Section 3.4.4. Quality can be described as the combination of accuracy with reliability, [Triggs et al., 1999]. High reliability is achieved by minimizing the chance for outliers and creating high redundancy, i.e. using much more observations than required to estimate the unknowns. The chapter starts with the error sources and other factors that have an influence on the quality of the approach in Section 4.1. For each of the error sources or other factors it is attempted to quantify its accuracy or influence. Then, the quality potential of the bundle adjustment is discussed in Section 4.2 and the chapter concludes with the accuracy potential of the laser scan in Section 4.3.

4.1 Error sources

As was stated in Section 3.1, it is preferred to use the camera position as control point, since then also the orientation of the camera can be translated to the orientation of the laser scanner. The other option is to measure an actual control point in the image and calculate its object coordinates. During the process to the end result, several errors and other factors that influence the quality are introduced. These errors and factors are subdivided into a group that arise before bundle adjustment and a group that arise in the bundle adjustment. Before the bundle adjustment, one deals with image measurements and several control points while in the bundle adjustment one has to deal with more difficult to quantify factors such as the configuration of tie points and control points and possible lens distortion. The possible error sources and other influential factors are tabulated in Table 4.1.

In general, one can say that the observations contain errors and are introduced before bundle adjustment. Within the bundle adjustment other factors than the quality of observations have an influence on the result. These influential factors only operates within the bundle adjustment and

Table 4.1: The possible error sources and other influential factors that affect the end result.

Before bundle adjustment	**In bundle adjustment**
Decrease resolution of images	Configuration control points
Accuracy of SIFT-matching	Configuration tie points
Accuracy image coordinates control points	Number control points
Accuracy local coordinates control points	Number tie points
Lens calibration (optional)	Lens calibration (optional)
	Bundle adjustment itself

not before. One may notice that 'lens calibration' appears in both categories. This is because one can decide to calibrate the focal length and lens distortion beforehand and correct all image coordinates for this lens distortion or to estimate the lens distortion and focal length within the bundle adjustment. The bundle adjustment itself also introduces some error due to the fact that is has to converge to a solution. The solution is therefore, different than for linear least squares, not the exact least squares solution. It will be assumed that this error is very small and can be neglected. Figure 4.1 is a visualization of the process from observations to the accuracy of the end result. The decrease in resolution of the images and the accuracy of SIFT matching are considered together as one error source.

Next, according to Figure 4.1, the accuracy of the observations are obtained and it is attempted to estimate the effect of the other factors on the accuracy. This implies that also the control points measurements in object space and in image space are discussed together just as all other influential factors.

4.1.1 Accuracy of SIFT-matching

As stated in Section 3.3, running SIFT on the full resolution images demand too much memory of a computer with a 1.83 GHz processor. A decrease in resolution is the most simple solution. The resolution and the accuracy of SIFT are closely allied, since the accuracy of the SIFT matching is expressed in pixels and the pixel size is directly related to the resolution. According to Lowe [2004], SIFT is able to localize keypoints with sub-pixel accuracy. Via an empirical test the accuracy of the SIFT matching will be obtained. The test will be executed by matching keypoints of two images covering asphalt pavement with a resolution of 1500×1000 pixels. The images were taken vertically down to avoid large scale differences between pixels. In order to eliminate the effect of lens distortion, the images were captured in almost the

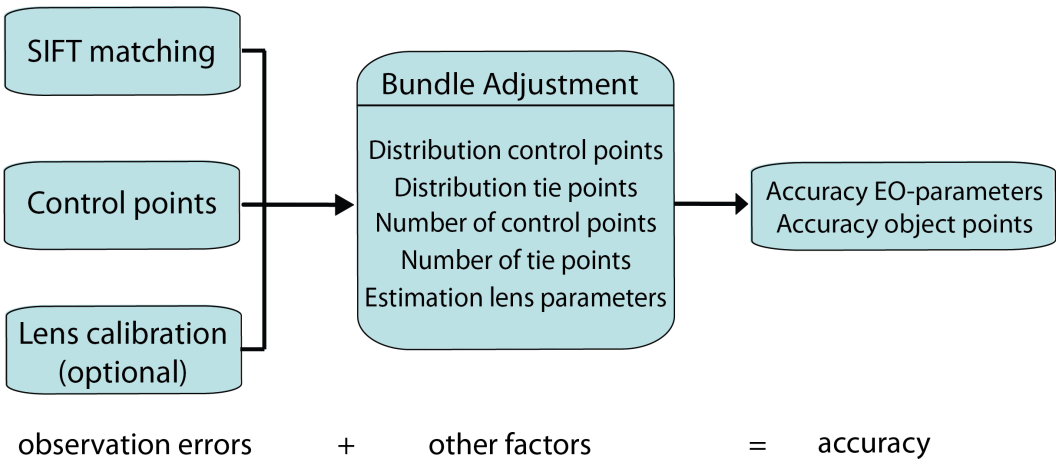


Figure 4.1: The factors that influence the accuracy of the object points and exterior orientation parameters.

same position and orientation. In this way, the coordinates of the matches will get an approximate constant and small shift. Locally, the variation of the shift is even smaller, since the effect of lens distortion will be similar. A surprisingly high number of more than 18,000 matches was found. For this pair of images, the shift was about 8 pixels in x-direction (column) and 1 pixel in y-direction (row). The approach to obtain the accuracy of the SIFT matching is as follows:

1. Obtain point correspondences for two images
2. Select an arbitrary match with its 20 nearest neighbours
3. Calculate the standard deviation of the 20 local shifts
4. Repeat step 2 and 3 for n times (where n is preferably a high number to create a homogeneous spreading of the selected matches)
5. The accuracy of the matching is considered to be the average of the standard deviations of the local shifts

The number of 20 nearest neighbours in step 2 is determined for these specific images. The high number of matches allows for selecting nearest neighbours that are actually at a very short distance from the selected match. The largest distance from an arbitrary match to a nearest neighbour is less than 50 pixels, for which the lens distortion is considered to be similar. The result of repeating step 2 and 3 for 500 times is illustrated in Figure 4.2. There seems to be no relation between the standard deviation and the location of

the match. Repeating step 2 and 3 more than 500 times seems therefore not beneficial. The average standard deviation of all local shifts is 0.10 pixel, which implies that for 68% of the point correspondences, the error is smaller than 0.10 pixel. Due to the decrease in resolution, the pixel size drops from the original 6.4μm (see Section 3.2) to approximately 14.7μm. The accuracy of the SIFT matching becomes therefore around 1.5μm when using the resized images.

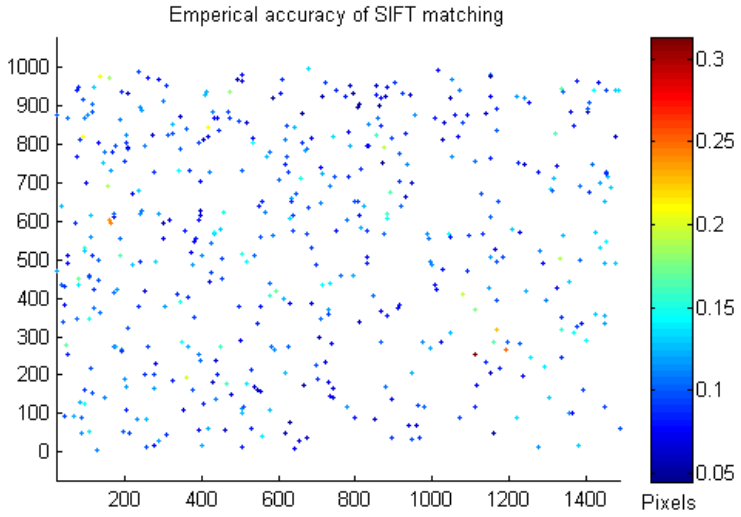


Figure 4.2: The empirical accuracy of SIFT matching.

4.1.2 Accuracy of control points

The control points are measured both in image space as in object space. The control points in the images should be measured manually, since these can not be detected automatically (yet). It will be assumed that an accuracy of 1-2 pixels can be obtained. This accuracy is a factor 10-20 worse than the accuracy of point correspondences, so it will be beneficial to improve these measurements. By implementing automatic feature detection, sub pixel accuracy may be achieved. The decision for this implementation should be made if the results from this thesis offer sufficient perspective for further development of the method.

The accuracy of the control point measurements in object space is considered to be the relative accuracy between all control points. The control points are measured with a total station (Leica TPS1201+). According to its specifications, [Leica Geosystems AG, 2009], its distance measurement

has a standard deviation of 1σ=1 mm and a bias of 1.5 mm per 1000 metres. The angle measurement has a standard deviation of 1σ=0.3 mgon. Overall, within a range of more than 100 metres, the accuracy will be in the millimetre range. However, one needs to position a prism manually right above a control point, which will have the dominant share in the accuracy. It will be assumed that this can be done with an accuracy of around 1 cm. The accuracies for the control point measurements are once more given in Table 4.2.

Table 4.2: The accuracy of control point measurements in image space and object space.

	image space	**object space**
accuracy	1-2 pixels	$\sim$ 1 cm

4.1.3 Accuracy of lens calibration

As stated in the beginning of this section, the lens calibration can be performed before or in the bundle adjustment. The aim of a lens calibration is to obtain estimates for the focal length, principal point offsets and radial lens distortion. In Section 3.7, the 20 mm lens of the Canon EOS 350D was calibrated. However, the calibration grid that was captured was at a much shorter distance from the lens than the distance from the road to the camera as it would be in practice. Both the focal length estimate as the radial lens distortion estimate are affected due to this difference in distance. So, the accuracy of these lens parameters is expected not to be very good. There are two solutions for this issue. The first is to perform the calibration with the calibration grid at the same distance as the road would be in practice. This requires a very large grid, since it should cover the entire image. The second solution is to obtain calibration estimates within the bundle adjustment. The latter seems practically better. If there is sufficient redundancy between observations and unknowns, one can add extra unknowns for the lens parameters to the model. Since it is known that a lot of point correspondences can be found, a highly redundant model is expected. One advantage of the latter solution is that it is based on the actual situation. A disadvantage is that it is not exactly known how the parameters are obtained or with what accuracy. However, bundle adjustment is capable of giving a precision for the focal length estimate and the principal point offsets but this precision depends also on other influential factors. Also within an actual calibration, precision values are obtained. High precisions do however not guarantee a high accuracy. In short, the accuracy of the lens parameters can not be retrieved with high reliability.

4.1.4 Effect of other influential factors

Other factors that influence the accuracy of the end result, i.e. object points and exterior orientation parameters, are the number of tie points and control points and the configuration of these points. Their influence is however complicated to quantify. The complexity of the relationship between the number of tie/control points and the accuracy of the objects points and the position and orientation of the camera is illustrated by the formula that was also presented in Section 2.7.4:

$$\begin{bmatrix} x_p \\ y_p \\ z_p \end{bmatrix} = s \cdot M \cdot \begin{bmatrix} X_P - X_0 \\ Y_P - Y_0 \\ Z_P - Z_0 \end{bmatrix} \tag{4.1}$$

where (x_p, y_p) is a tie point measurement in one image and (X_P, Y_P, Z_P) can be an unknown object point or a (known) control point. The unknowns for the camera position and orientation are respectively represented by (X_0, Y_0, Z_0) and rotation matrix M. If the focal length is also unknown, then the scale s and z_p are extra unknowns. Many unknowns are interlinked with each other and moreover by nonlinear relations due to the rotation matrix. Therefore it can not be said how the number of observations will affect the precision and accuracy of the unknowns. However, the reliability of the bundle adjustment does improve when using more tie/control point observations, provided that they are outlier free.

The configuration of the tie/control points is not incorporated at all, which makes it impossible to estimate mathematically its influence on the quality of the bundle adjustment. The only reference one can take are empirical results from other projects and reflect these on the situation of Pavescan. This will be done in the next section.

4.1.5 Summary

The quality of the object points and exterior orientation parameters is dependent on the quality of the observations that participate in the bundle adjustment and on the number and configuration of tie points and control points. For the observations (object coordinates of control points and image coordinates of tie points and control points), an accuracy is estimated. For the other factors that have an influence on the accuracy of the object points and exterior orientation parameters it was not succeeded to quantify their influence on the result. Also the accuracy of the lens parameters (focal length, principal point offsets and radial lens distortion) can not be retrieved with high reliability. The best prediction of the accuracy of the result can be made by comparing the situation of Pavescan with other projects (traditional) and reflect those results on the situation of Pavescan.

4.2 Quality potential of the bundle adjustment

The quality of the exterior orientation parameters and the object coordinates is dependent on several factors as was tabulated in Table 4.1. It is expected that the accuracy of the exterior orientation is strongly correlated with the accuracy of the object points that are calculated in the bundle adjustment. So first the accuracy of the object points is evaluated. A very optimistic and rough estimation for the accuracy of the object points is illustrated in Figure 4.3.

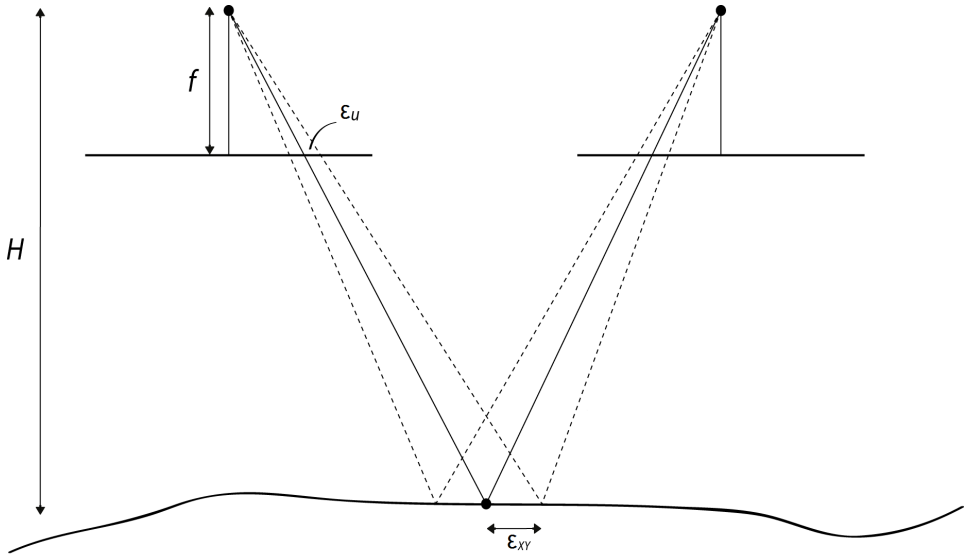


Figure 4.3: The accuracy of object points based on the accuracy of image measurements.

From image 4.3 the following formula is obtained:

$$\epsilon_{XY} = \epsilon_u \cdot s \tag{4.2}$$

where ϵ_{XY} is the accuracy of the object points in the horizontal plane, ϵ_u the accuracy of the image measurements and s the image scale, which equals $\frac{H}{f}$. The accuracy of the height of the object points, ϵ_Z, is ϵ_{XY} divided by the B/H ratio, [Vosselman, 1999]:

$$\epsilon_Z = \frac{\epsilon_{XY}}{B/H} \tag{4.3}$$

Although these equations are an idealization of the accuracy to be obtained (it only takes the accuracy of the image measurements into account), it seems useful to calculate these approximations in order to get insight on the

highest potential order of quality. If the accuracy of the image measurements is set to 1.5 μm (as was calculated in Section 4.1.1), and the image scale to 180 ($s = \frac{H}{f} \approx \frac{3.6}{0.02} \approx 180$), ϵ_{XY} is calculated to be 0.3 mm. The baseline is expected to be around half the height, which results in a ϵ_Z of 0.6 mm. It may be clear that these values can not be obtained in practice due to the diversity of other error sources that exist besides the error of the image measurements. A prediction of the accuracy of the orientation of the camera is also not available in formula form. However, if object coordinates would have millimetre accuracy, the camera orientation is expected to have sub degree accuracy. Predictions are best made with respect to standard photogrammetric projects, which is described next.

For standard aerial photogrammetric projects, some general requirements are defined for obtaining a stable model where demanded accuracies can be achieved. These general requirements are taken as a guideline, but one should bear in mind that these are extracted from some older references where digital photogrammetry was not yet the standard. According to Vosselman [1999], the accuracy of image measurements is in general between 10 and 20 μm. Automatic matching yields an accuracy between 0.3 and 0.5 pixels, [Habib and Kelley, 2001]. Regarding control points it is stated that these should be distributed along the edges of the block and some in the middle of the block to achieve a homogeneous accuracy for height measurements. The control points should be placed with a spacing of two model widths, i.e. a control point in every second image. For a homogeneous height accuracy, the control points should not lie on a line, [Vosselman, 1999]. The accuracy of the control points measured in object space is given by a rule of thumb provided by U.S. Army Corps of Engineers [1993]. The horizontal accuracy of the control points should be twice as good as the required horizontal accuracy for object points. The vertical accuracy of the control points should be similar to the required vertical accuracy for object points. Regarding tie points, Vosselman [1999] states that for an accurate result, in each of the six so called *von Gruber* locations two tie points should be measured.

The influence of the factors that are expected to be different from a standard photogrammetric project are presented in Table 4.3. As can be seen, the factors that involve tie points are expected to have a positive influence on the quality while the factors involving control points are expected to degrade the quality. The fact that there are much more tie point observations available compared to standard photogrammetric projects, the reliability of the bundle adjustment is likely to improve. The main question is to what extent the small number of control points and their poor configuration will influence the quality of the bundle adjustment.

Table 4.3: The influence of factors that are different from the norm on the quality.

Factor	Expected difference	Expected influence
Accuracy of image measurements	+	+
Number of tie points	++	+
Configuration of tie points	+	+
Number of control points	−−	−
Configuration of control points	−−	−

4.3 Accuracy potential of the laser scan

As was described in Section 2.5, in the traditional setup the individual scan points can not be localized very accurately because the driving direction is not measured and can not be estimated from one single control point. The camera positions were proposed to be used as control points instead, so that the orientation of the camera could be used for orienting the laser scan and thus also the scan points can be localized. For this, the camera needs to be linked to the laser scanner by calibrating the vector between the two devices. It will be assumed that this can be done very accurately and therefore this measurement error will be neglected. The accuracy of the position of the laser scanner will then be as good as the accuracy of the position of the camera. The potential accuracy for the laser scan points will then partly consist of the accuracy on the position and partly on the accuracy of the orientation of the camera. Additionally, the range accuracy of the laser scanner should be added. The way the accuracy of the position and orientation of the camera reflects on the accuracy of the laser points can be calculated via basic trigonometry. The object coordinate system is chosen such that the driving direction represents the x-axis. In this way the orientation angles are with respect to the vehicle, which is easier to understand and to visualize. Three basic formulas that describe the accuracy of the laser points are formulated:

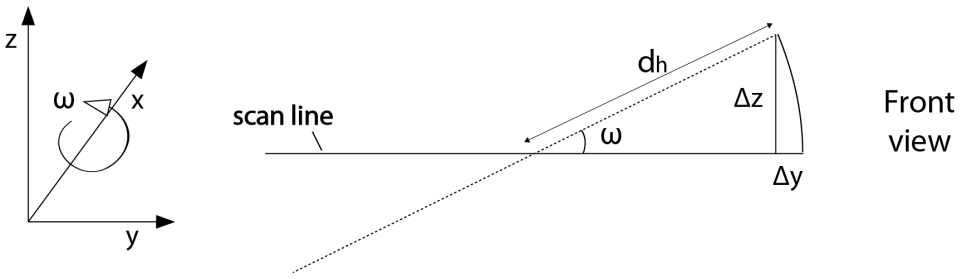
$$\epsilon_{\mathsf{X}} = \epsilon_{\mathsf{Xc}} + \epsilon_{\mathsf{X}}(\epsilon_{\varphi}) + \epsilon_{\mathsf{X}}(\epsilon_{\kappa}) \tag{4.4}$$

$$\epsilon_{Y} = \epsilon_{Yc} + \epsilon_{Y}(\epsilon_{\omega}) + \epsilon_{Y}(\epsilon_{\kappa}) \tag{4.5}$$

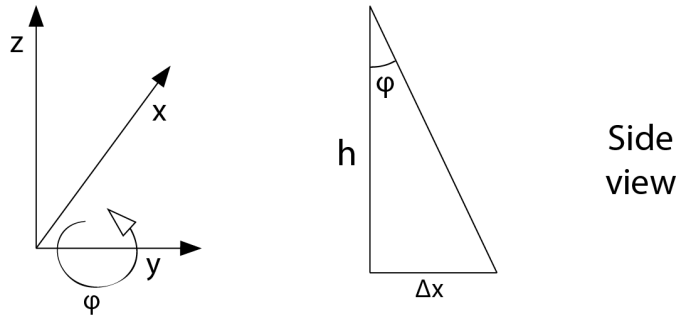
$$\epsilon_{Z} = \epsilon_{Zc} + \epsilon_{Z}(\epsilon_{\omega}) + \epsilon_{s} \tag{4.6}$$

where ϵ_{Xc}, ϵ_{Yc} and ϵ_{Zc} represent the accuracy of the position of the camera, and $\epsilon_X(\epsilon_{angle})$, $\epsilon_Y(\epsilon_{angle})$ and $\epsilon_Z(\epsilon_{angle})$ are the contributions from the accuracy of the orientation angles. The range accuracy of the laser scanner

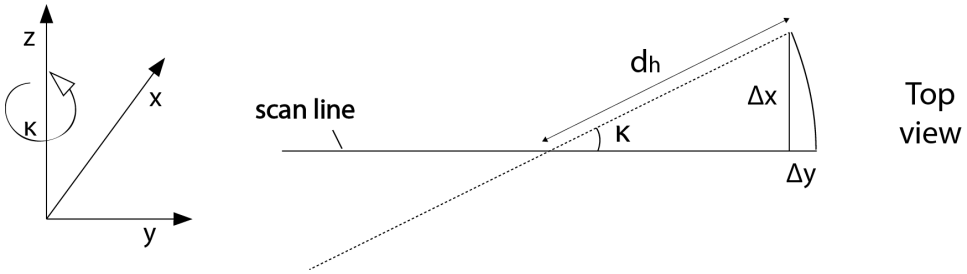
is represented by ϵ_{Zc}. Only those angles that affect the accuracy are incorporated, as can also be seen in Figure 4.4.



(a) Effect of change in omega



(b) Effect of change in phi



(c) Effect of change in kappa

Figure 4.4: The effects of changes around the three rotation axes.

The effect on the accuracy of the scan points depends on the horizontal distance between the scanner and the scan point or the height of the scanner. From Figure 4.4, the following relations are extracted:

$$\epsilon_X(\epsilon_\varphi) = h \cdot tan(\epsilon_\varphi) \tag{4.7}$$

$$\epsilon_X(\epsilon_\kappa) = d_h \cdot sin(\epsilon_\kappa) \tag{4.8}$$

$$\epsilon_Y(\epsilon_\omega) = d_h - d_h \cdot cos(\epsilon_\omega) \tag{4.9}$$

$$\epsilon_Y(\epsilon_\kappa) = d_h - d_h \cdot cos(\epsilon_\kappa) \tag{4.10}$$

$$\epsilon_Z(\epsilon_\omega) = d_h \cdot sin(\epsilon_\omega) \tag{4.11}$$

where h is the height of the laser scanner (≈ 3.6 metres) and d_h the horizontal distance from a scan point to the laser scanner. It can be seen that scan points are most affected in height and x-direction. However, relative height measurements are available from the laser scan, so that $\epsilon_Z(\epsilon_\omega)$ can be omitted. The propagation from the accuracy of the orientation of the scanner to the accuracy of the scan points is calculated for different accuracies as function of horizontal distance. For three different accuracies of the orientation angles, the position accuracy is calculated up to a horizontal distance of 12 metres. Figure 4.5 shows the position accuracy of a scan point in x-direction. It should be noted that the accuracy of the position of the camera is left out. These (yet unknown) accuracy values should however be added to the values observed in Figure 4.5.

Even if the accuracy of the orientation of the camera is very good, the position accuracy of the scan points is in the order of centimetres. This may seem not so good, however, these are absolute position accuracies, while in the traditional setup of Pavescan these are much less accurate. For example, a scan differing 2° from perpendicular to the road, already results in an error of 42 centimetres in x-direction at a horizontal distance of 12 metres. Obtaining the accuracy of the orientation of the camera requires testing with for instance an INS, in order to have reference values.

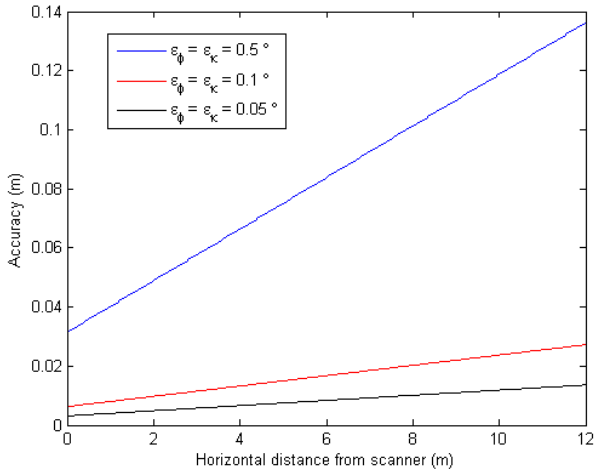


Figure 4.5: The accuracy of a scan point in x-direction for different orientation accuracies.

Chapter 5

Image acquisition and processing

In this chapter a sequence of images is acquired and processed with two of the available software programs. Section 5.1 describes the image acquisition according to the proposed setup in Section 3.4. These images are first processed with Bundler in Section 5.2. Then, the existing algorithm of SIFT to find point correspondences in images is optimized in Section 5.3. Finally, the optimized algorithm is used within the processing of BINGO in Section 5.4.

5.1 Image acquisition

In Section 3.3 it was stated that sufficient tie points can be found on asphalt. It is therefore useful to acquire a sequence of images as proposed in Section 3.4. For this acquisition, the Canon EOS 350D with the 20 mm lens is mounted on Pavescan at a height of approximately 3.6 metres.

The choice for the Canon camera and the lens was explained in Section 3.5. The camera is rotated manually to the back to capture just the road's surface. This *pitch* angle is estimated by eye at 20-30 degrees. The test site is the parking lot 'Transferium' in Den Bosch, which is in the vicinity of the office of Breijn in Rosmalen.

The images are taken in full colour and with manual focus. An example of an image is presented in Figure 5.1. As can be seen, a small strip still covers the trailer of Pavescan. Therefore the images need to be cropped slightly. For 1500×1000 pixel images, the crop will be 150 pixels leaving images of 1350×1000 pixels. The weather conditions were very good for taking images at the time of acquisition. The temperature was 23-26 °C, the wind speed was 1.6 km/h with maximum wind gusts of 6.4 km/h (weather statistics are from [Website Website Weather Underground]).

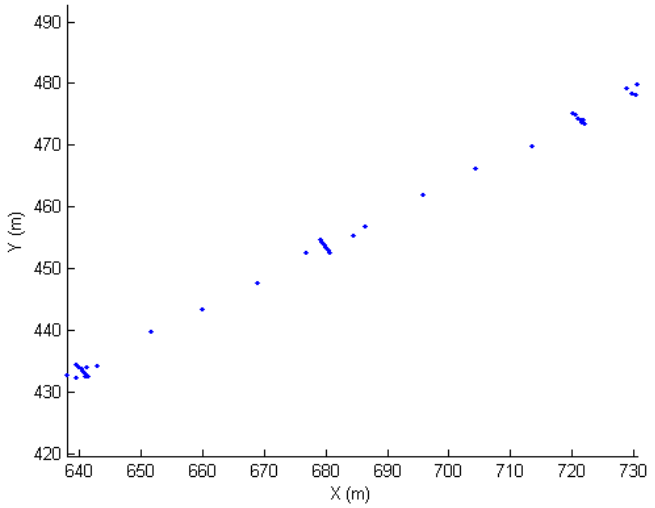
In total 66 images are taken with an aimed baseline of 1.6 metres, i.e.

Figure 5.1: One of the images captured at the test site.

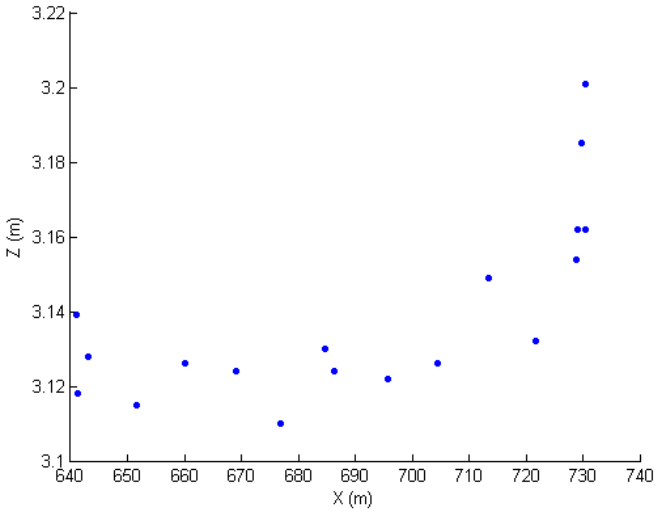
one long strip of around 100 metres. At the beginning, halfway and at the end of the strip three nails were hit into the asphalt and measured in a local grid (RD) with a total station. These points, provided that they are visible in the images, can be used as control points to link the image frame to the local frame. An extra control point is made every ten metres and also a scan is made by Pavescan for additional laser data. Additionally, three sets of check points are measured on the road's surface, which are not marked by paint or a nail, i.e. these are not visible in the images. Each set consists of six points and are approximately located on a line perpendicular to the driving direction. These points can be used for an independent check on a future result. The control points and check points are visualized in Figure 5.2(a)[1].

Because it is known which part of the road is captured, one is able to make a prediction of what the DTM of the road should look like along the trajectory. Figure 5.3 illustrates the part of the road that has been captured. Together with the height information of the control points provided in Figure 5.2(b), the expected shape of this part of the road is described as an approximate plane having a subtle constant crossfall and is gradually decreasing in height along the drive line for the first ten metres and then stabilizes.

[1]The RD coordinates are subtracted with a constant for a clearer visualization. X=X_{RD} - 150,000 and Y=Y_{RD} - 412,000. This is also done for coming figures with RD coordinates.



(a) The control points and check points projected in RD.



(b) Height of the control points

Figure 5.2: The control points

5.2 Bundler

The first software to process the images is Bundler. Bundler is described in more detail in Section 3.6.1. Bundler was chosen for its highly automated character. Basically, the Bundler script is run on a folder containing images. The major output file contains the position and orientation of the camera for each image. The remainder of the file lists image coordinates and 3D

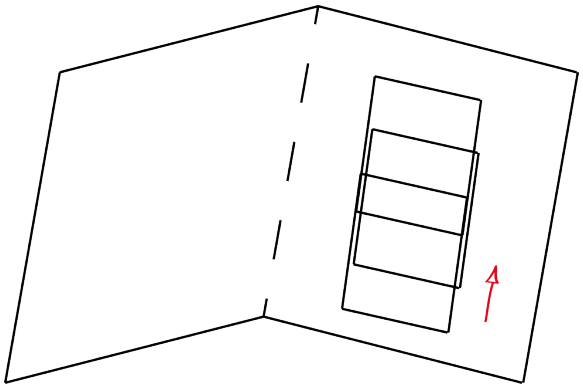


Figure 5.3: The part of the road that has been captured.

coordinates of the tie points. A very important detail is that the output is at an arbitrary scale because control points can not be incorporated in the Bundler processing.

5.2.1 Workflow

The main objective of this chapter is to retrieve the positions and orientations of the camera for each image and assess its accuracy. Another goal is to generate a DTM of the terrain captured on the images and its accuracy. The workflow to achieve these goals with Bundler is illustrated in Figure 5.4. Before the images are used in Bundler, they can be corrected for lens distortion. This procedure is explained in Section 5.2.2. Another preparation may be to crop and/or resize the images for better performance. For the transformation to a local grid, at least three control points must be known in image coordinates as well as in object coordinates.

5.2.2 Lens correction

The parameters for lens distortion can be estimated by Bundler in the bundle block adjustment. Another way to deal with the lens distortion is to correct for it before the bundle adjustment. The advantage of the latter is that no extra unknowns for lens distortion are introduced in the bundle adjustment, which results in a higher redundant model. The lens calibration as performed in Section 3.7 can be used to correct the images. This calibration is however not error free, due to the different focal setting between the calibration images and the actual images. Therefore, in the first processing the lens distortion will be estimated in the bundle adjustment. If good results are obtained, the processing will be repeated with corrected images, which may lead to a slight improvement of the result.

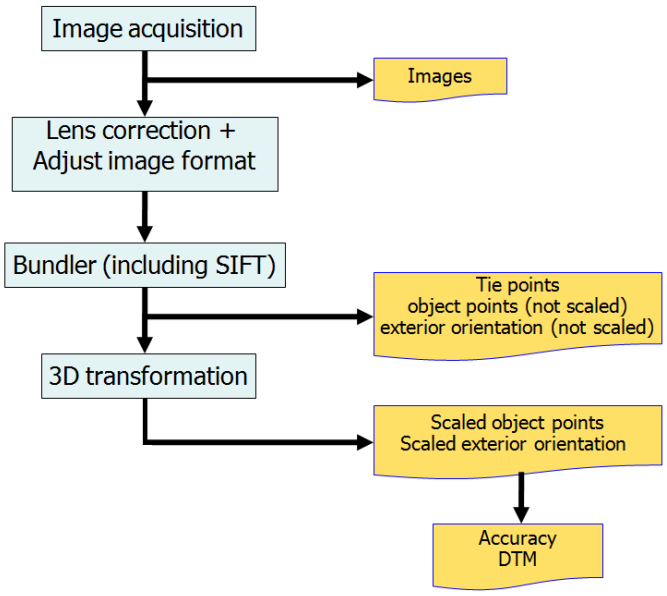


Figure 5.4: Workflow diagram of Bundler processing.

Correcting images for lens distortion is a pixel by pixel operation that is based on the formulas presented in Section 2.10.1. The correction for radial lens distortion is the most straightforward. First for every pixel, the distance to the principal point is determined. This distance is the only variable in calculating the magnitude of the correction. The direction of the correction is in line with the vector from the principal point to the pixel. With standard trigonometrics a new position for every pixel will be calculated. The result is a set of non uniformly spaced points. Resampling is required to recover the rectangular and regular grid of an image.

5.2.3 Bundler processing

First only five images are used as input for Bundler, since this is much faster while the result will still be interpretable. As stated in the previous section, these images are not (yet) corrected for lens distortion. Since Bundler makes use of SIFT, the full resolution images (3456×2304 pixels) are resized to 1500×1000 pixels to avoid computation problems. Besides, 10% was cut off from the left in each image, so that a crop of 1350×1000 pixels remains. Bundler is set up in a way such that the user does not have much influence on the individual steps, input and result. There is one 'options file' where you can set some preferences. Here you can choose for instance to disregard lens distortion or to set the focal length tight for all images. Since all images are captured with the same camera and equal focal settings, it is

highly recommended to set the focal length tight. Bundler estimates the focal length in the bundle block adjustment but it is not possible to set it manually. So, using the calibrated focal length that was found in Section 3.7 is not possible. The bundle adjustment converged to a solution, but the result cannot be judged properly since the output is not scaled right. In the next section, the transformation will be described and performed.

5.2.4 Transformations

Due to the fact that Bundler can not incorporate control points, the output is arbitrary scaled. Therefore a 3D transformation of the output is required. This transformation requires control points that should be known in both image coordinates as object coordinates. Control points are generally not detected by SIFT as tie point and thus these should be measured otherwise. For this thesis, these are measured manually, but an automatic measurement by *target recognition* is preferred. Figure 5.5 presents a diagram with all steps to transform the arbitrary scaled Bundler output (from now on referred to as Bundler coordinates) to local coordinates. To obtain the Bundler coordinates of the control points, first its 2D coordinates are estimated via a 2D transformation from its image coordinates. The height should be interpolated with its surrounding since image coordinates are two dimensional. Once the estimates for the control points in Bundler coordinates are determined, the parameters for the transformation to local coordinates are estimated via a 3D transformation. With these parameters all object points and the position and orientation of all images can also be transformed to local coordinates so that also a DTM in local coordinates can be produced. Next, the transformation of the control points from image coordinates to Bundler coordinates is described in more detail followed by the transformation from Bundler coordinates to local coordinates.

Image coordinates to Bundler coordinates

The first step is to estimate the Bundler coordinates of the control points according to their image coordinates. When imaging an entirely flat surface, the relation between the image coordinates of tie points and their coordinates in an object space will be a 2D similarity transformation. Since the height differences are relatively small, it will be assumed that locally the relation between the image coordinates and Bundler coordinates approximates a 2D similarity transformation. Since image coordinates have no z-value, only the xy-coordinates of the control points can be estimated in the Bundler coordinates with a 2D similarity transformation. The height will be interpolated according to their surrounding tie points. Inverse distance interpolation (with p=2) is chosen to estimate the height for its straightforward and fast implementation. There are four unknown parameters in

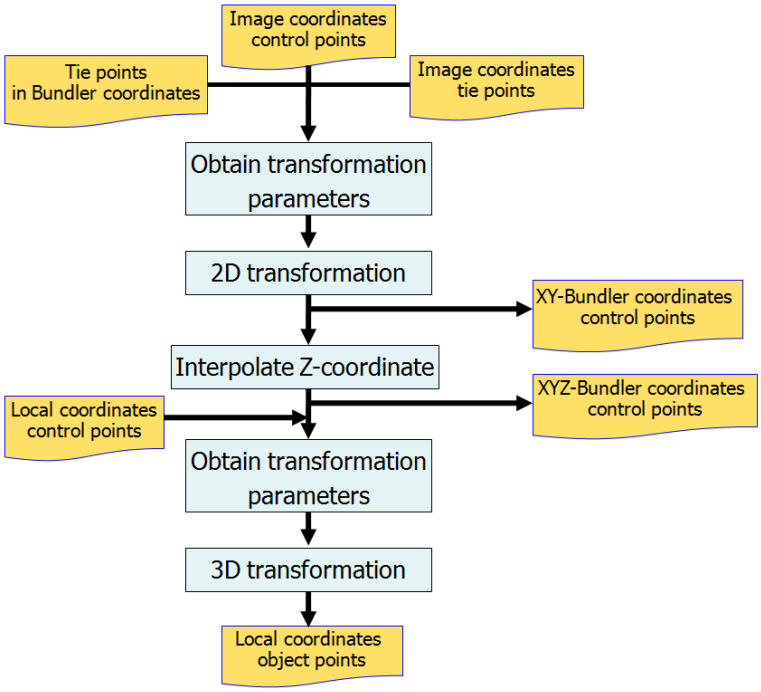


Figure 5.5: The process to link the Bundler output to a local frame

a 2D similarity transformation; one rotation parameter ω, two translation parameters (X_0,Y_0) and one scale factor λ. At least two tie points need to be known in image coordinates (x_P, y_P) and Bundler coordinates $(x_{P'}, y_{P'})$ to solve the model. To avoid a nonlinear model (due to rotation ω), the model observations are written in a linear notation as proposed in [Kodde, 2004]:

$$x_{P'} = a \cdot x_P - b \cdot y_P + X_0 \tag{5.1}$$

$$y_{P'} = a \cdot y_P + b \cdot x_P + Y_0 \tag{5.2}$$

where

$$a = \lambda \cdot cos(\omega)$$

$$b = \lambda \cdot sin(\omega)$$

and P is the point to be transformed and P' the transformed point. Since it is assumed that a similarity transformation holds only on a local scale, separate transformation parameters are estimated for each control point. The observations in this least-squares adjustment are the coordinates of the nearest surrounding tie points. Normally it holds true that the more observations the better the estimation. Since the transformation holds only

on a local scale, it implies that the observations should not be too far away from the control point. In order to find the optimum number of tie points to be used, the error for each scenario is calculated. The error is calculated by letting tie points act as control points. The difference between the calculated coordinates and the Bundler coordinates is the error of the transformation at the location of the tie point. The RMS error of 500 random locations (i.e. tie points) for different number of used surrounding tie points is calculated and presented in Figure 5.6. The trend is that the more tie points are used, the lower the accuracy of the transformation. Below 13 points, the smallest RMS error is obtained and is also approximately constant. From 13 points there seems to be trend that the RMS error is linearly increasing. Therefore, for estimating the transformation parameters it is recommended to use less than 13 tie points. The expected error of the transformation is around $1e^{-3}$ Bundler units. The quality of this transformation cannot yet be judged well since the scaling between Bundler coordinates and object coordinates is not yet known.

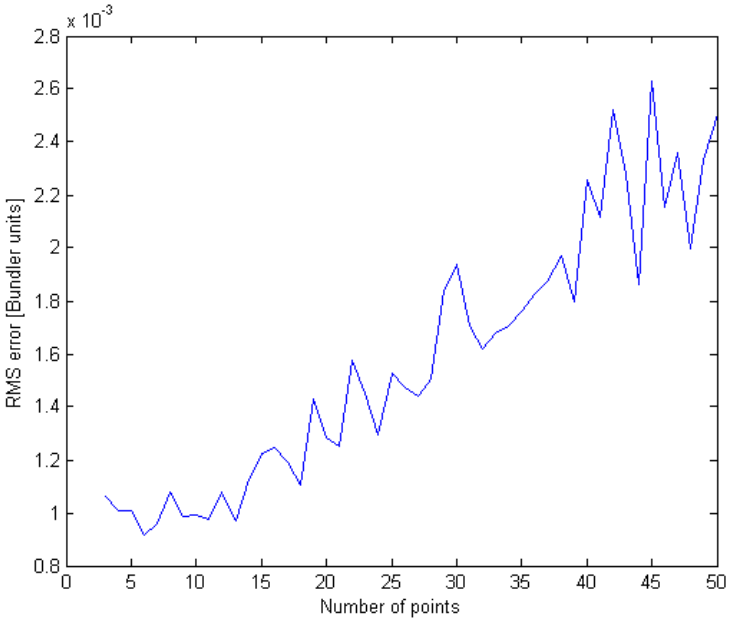


Figure 5.6: The number of tie points versus the RMS error of the local transformation parameters. A low number of tie points should be used for estimating the transformation parameters.

Bundler coordinates to local coordinates

The transformation from Bundler coordinates to local coordinates is a 3D similarity transformation which consists of seven unknowns; three rotation parameters $(\varphi, \omega, \kappa)$, three translation parameters (X_0, Y_0, Z_0) and one scale factor λ. This implies that at least three control points need to be known in both the Bundler frame and the local frame. The model observations are written best in matrix form:

$$C' = s \cdot M \cdot C + T \tag{5.3}$$

where

$$C' = \begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix}, C = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}, T = \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix}$$

and M is a 3×3 rotation matrix which is the product of three rotation matrices around the three axes as described in Section 2.7.4.

The model cannot be written in linear equations due to the extensive form of the equations with many rotations. Therefore a Levenberg-Marquardt (LM) approach is used. This method is described in Section 2.6. The Jacobian matrix in the linearized model looks like:

$$J = \begin{bmatrix} \frac{\partial y_1}{\partial \varphi} & \frac{\partial y_1}{\partial \omega} & \frac{\partial y_1}{\partial \kappa} & \frac{\partial y_1}{\partial X_0} & \frac{\partial y_1}{\partial Y_0} & \frac{\partial y_1}{\partial Z_0} & \frac{\partial y_1}{\partial s} \\ \frac{\partial y_2}{\partial \varphi} & \frac{\partial y_2}{\partial \omega} & \frac{\partial y_2}{\partial \kappa} & \frac{\partial y_2}{\partial X_0} & \frac{\partial y_2}{\partial Y_0} & \frac{\partial y_2}{\partial Z_0} & \frac{\partial y_2}{\partial s} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial y_n}{\partial \varphi} & \frac{\partial y_n}{\partial \omega} & \frac{\partial y_n}{\partial \kappa} & \frac{\partial y_n}{\partial X_0} & \frac{\partial y_n}{\partial Y_0} & \frac{\partial y_n}{\partial Z_0} & \frac{\partial y_n}{\partial s} \end{bmatrix} \tag{5.4}$$

where

$$y = \begin{bmatrix} X_1 & Y_1 & Z_1 & X_2 & Y_2 & Z_2 & \dots & X_n & Y_n & Z_n \end{bmatrix}^T$$

containing the coordinates of the control points.

The basis for the algorithm used to find the transformation parameters can be found in [Mittrapiyanuruk, 2008]. The convergence to the control points is fast in the beginning and slow in the end. The iteration process was stopped when the corrections of the parameters dropped below $1e^{-5}$ degree or metre. An example of the convergence to three control points is illustrated in Figure 5.7. Around 700 iterations were needed to get the definite values. For better visualization, each dot represents 50 iterations. The quality of the transformation as well as the image measurements of the control points is judged by the differences of the transformed coordinates with the actual coordinates of the control points. In horizontal direction the differences are up to almost 3 cm while the height differs less than 1 cm.

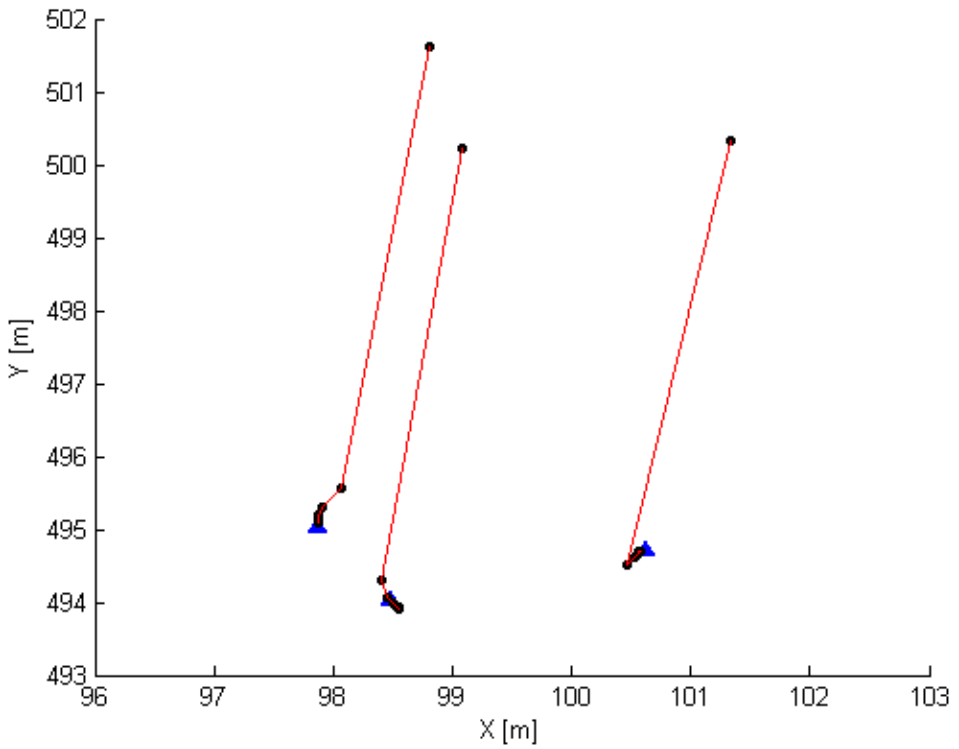


Figure 5.7: The convergence of the 3D transformation. The convergence is fast in the beginning, but very slow in the end. Around 700 iterations were needed. Each dot represents 50 iterations.

5.2.5 Results

A nice way of checking the result is to visualize the transformed point cloud of the object points. This point cloud should represent the shape of the road as was presumed, i.e. an approximate plane with a constant crossfall. The result of the transformation of the Bundler output (of five images) is presented in Figure 5.8. Each dot represents a tie point that is in reality on the road's surface. The shape of the road appears far from flat and it seems that the quality degrades exponentially when the distance to the control points increases. Within four metres the terrain rises over 10 cm, while the measured control points indicate no significant height difference (see Figure 5.2(b)). A visual check in the terrain confirms this presumption and therefore the result from Bundler is definitely wrong. The positions and orientations of the camera are therefore also not reliable.

It was also attempted to process more than five images, but Bundler does not succeed in calculating all exterior orientation parameters of the camera. As can be read in Section 3.6.1, the images are added one by one to

the bundle adjustment until no images are left for a reliable reconstruction of 3D points. For several images, the position and orientation of the camera are not calculated while the overlap is in theory sufficient and the number of corresponding points is more than sufficient. If one image is rejected by Bundler, the following images in the sequence are rejected as well since there is no sufficient overlap anymore. For a standard Bundler project, it is not disastrous when an image is rejected (it is even very common), since there are generally more than enough images. However, for an application in Pavescan every single image is necessary to participate in the bundle adjustment in order to obtain a result. Apparently, Bundler considers adding more images as not reliable. This makes sense when looking at Figure 5.8 and imaging that the observed trend would proceed. However, the reason for the observed trend is not known. It could be either wrong observations or false assignments of lens parameters. This can unfortunately not be discovered since Bundler is very limited for modification.

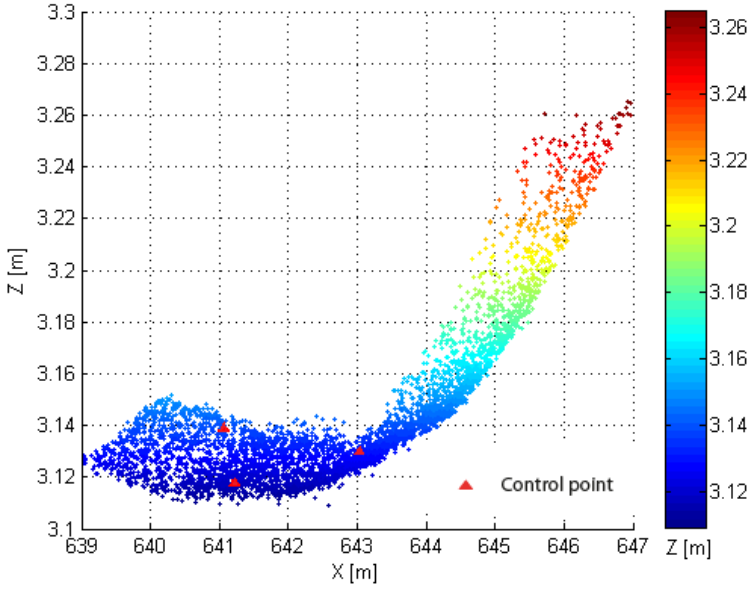


Figure 5.8: The transformed point cloud of Bundler. The shape cannot be certified and is definitely wrong.

5.2.6 Conclusion

Bundler was chosen for its highly automated performance. However, several issues with the usage of Bundler were experienced. The obtained result from a strip of five images was incorrect and due to the low modification character of Bundler it could not be detected why. Within the application of Pavescan,

there is known information that cannot be incorporated in Bundler. The main shortcomings and limitations of Bundler for this research are:

- Bundler considers each image as captured with a different camera
- Bundler has its own matching algorithm which cannot be modified
- The output of Bundler is arbitrary scaled

The first is a shortcoming that directly influences the accuracy of the result. Every image is treated differently by Bundler and consequently the corrections per image are different. A proper way to deal with the information that only one camera is used, is to calibrate the focal length and lens distortion parameters of the camera before or in the bundle adjustment. Within the bundle adjustment, these parameters should not be changed anymore. In this way every image is treated the same and approximates the real situation best. The second item is a limitation of Bundler, which is also a general one. In many steps, the exact method of Bundler remains unknown and is not adaptable. Bundler uses the SIFT algorithm for the detection of keypoints, but uses its own matching algorithm, that can not be modified. The image coordinates of the matching keypoints are the only observations in the model and therefore of paramount interest. Since the images are captured with a constant baseline, the position of a match in the next image can be estimated quite accurately. This knowledge could be used to detect false matches or to prevent false matches by searching for matches in a limited area. Limiting the search area is also much faster than comparing every keypoint in one image by every keypoint in the other image. The scaling issue of Bundler is discussed in this chapter and the proposed solution is to transform the Bundler output to known local coordinates. Although the transformation works, it is an extra step that interrupts the automation and introduces small errors. An alternative method that incorporates control points is therefore recommended.

All discussed shortcomings and limitations have to do with the fact that Bundler is no photogrammetric software. Therefore it has been decided to switch from Bundler to a photogrammetric software. Another issue is the quality of the image coordinates of the corresponding points. The next section deals with improving SIFT.

5.3 Optimization SIFT

5.3.1 Original SIFT

As stated in Section 5.2.6, the procedure to match keypoints has potential to be made faster and more reliable. A free (for research purposes) demo

program of SIFT is used. This demo provides a Matlab program that can find the keypoints and match them. During this research two issues regarding matching SIFT keypoints have shown up:

- A computer with a 1.83 GHz processor cannot match images at full resolution (3456×2304 pixels) while matching resized images (1500×1000 pixels) still takes a very long time.
- Road markings are sensitive for incorrect matches

An image at full resolution has almost 8 million pixels that all need to be examined for being a candidate keypoint. Matlab R2008b is not capable of running SIFT on these images successfully on a computer with a 1.83 GHz processor. Therefore, for this research the images are resized to 1500×1000 pixels. Even on these resized images the matching is very slow. Finding keypoints takes around 13 seconds per image while the matching takes another 80 seconds. So, processing a strip of 100 metres, i.e. around 60 images, will take over 90 minutes. For the practical feasibility of using SIFT, a faster computer is required or the algorithm should be made more efficient (or a combination of both). As stated in Section 2.8, each keypoint in one image is compared with every other keypoint in the other image. Per image around 10,000 keypoints are found, which resulted in around 1,000 matches per image pair. The SIFT matching algorithm also plots the matches between two images by drawing lines between the locations of the matching keypoints from one image to the other. A selection of matches, visualized by this program, is illustrated in Figure 5.9. The figure presents two images where the overlapping area (of image 1) is on the left hand side from the black dashed line. A match is represented by a cyan coloured line running from the image coordinates of a point to the image coordinates of its corresponding point. However, in the red rectangle matches are drawn starting from the right hand side of the black dashed line, so these are obvious incorrect matches. It is remarkable that all incorrect matches in this example are on or near road markings. Apparently repetitive patterns are sensitive for wrong matching. Wrong matches have a negative influence on the accuracy of the bundle adjustment and should therefore be eliminated.

5.3.2 Elimination of wrong matches by RANSAC

The detection of wrong matches is performed by RANSAC. As can be read in Section 3.6.1, Bundler also uses RANSAC for removing wrong matches. RANSAC was described first by Fischler and Bolles [1981]. The method estimates parameters of a model out of a dataset containing outliers. In the case of point correspondences, the model can be described by the epipolar constraint:

$$\mathbf{u}'^{T} \cdot F \cdot \mathbf{u} = 0 \tag{5.5}$$

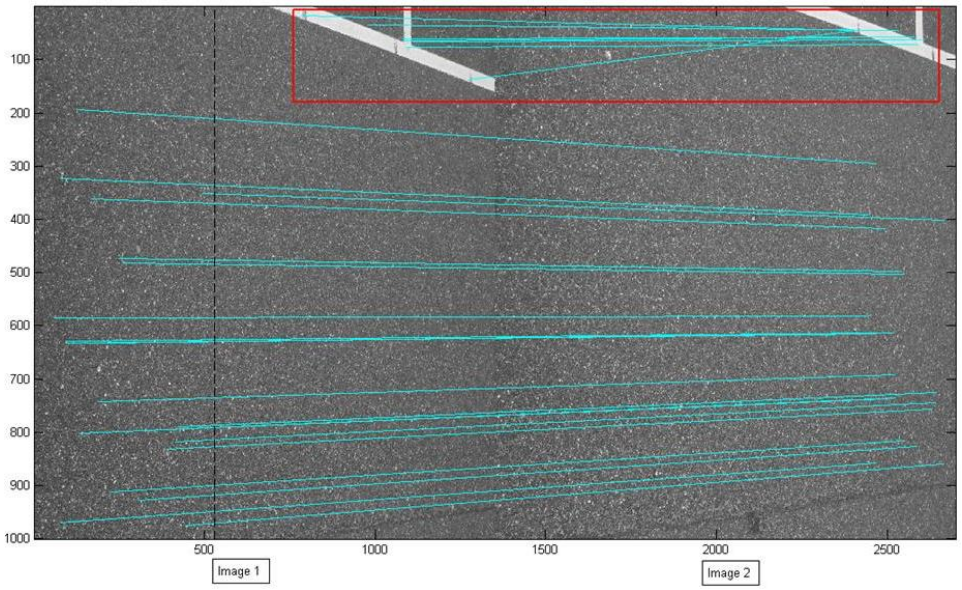


Figure 5.9: Incorrect SIFT matching. In the red box the incorrect matches are illustrated.

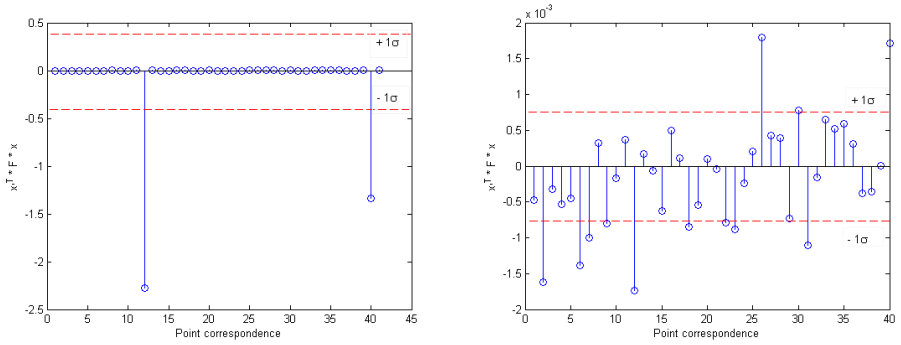
as explained in section 2.9. The relation between two corresponding points $\mathbf{u}'$ and $\mathbf{u}$ is described by fundamental matrix F and can be estimated if at least 8 point correspondences are known. The idea is now to generate a number of samples, containing randomly selected point correspondences. For each sample the fundamental matrix is estimated. The support for this fundamental matrix is measured by the number of point correspondences that fulfil the epipolar constraint within a threshold. All point correspondences within the threshold are the so-called inliers. The sample with highest support (i.e. highest number of inliers) will be considered as best fit to the model. The intuition here is that if a sample contains an outlier, it will not gain much support. It is not necessary to try every possible sample. The number of samples is chosen, so that the probability that at least one sample contains no outliers is very high. According to [Hartley and Zisserman, 2003] the minimum number of samples required is calculated by:

$$N = \frac{log(l-p)}{log(l-(1-e)^s)} \tag{5.6}$$

where p is the probability that at least one sample has no outliers for a given proportion outliers e and a sample size s. The sample size equals the number of point correspondences used to estimate F. In this case the sample size should be at least eight, however a safer sample size of twelve is used. Considering a safe proportion of outliers (e = 0.05) and a high

probability number (p = 0.99), the number of samples is calculated to be at least 6. However, the proportion of outliers when matching image i with image $i + 2$ is considerably higher. This is because the amount of matches is much less than when matching image i with image $i + 1$ (because there is much less overlap) while the total amount of outliers stays approximately constant. For this situation, the number of samples should be increased to 144 (e = 0.25) to guarantee, with a probability of 99%, that at least one sample contains no outliers.

All inliers of the sample with the highest number of inliers are used to estimate the final and optimal fundamental matrix. All matches are now judged on how well they fulfil the epipolar constraint. Outliers should now easily be detected. However, there is also the possibility that there are no outliers. This situation is not desirable since RANSAC assumes at least one outlier. The difficulty is illustrated in Figure 5.10.



(a) Containing outliers

(b) No outliers

Figure 5.10: The complexity to set a general threshold for outlier removal.

In the case of at least one outlier, the threshold can be set to a low value, e.g. 1σ. If there are no outliers, a threshold of 1σ may falsely assign lots of point correspondences as outlier. It is experienced that giving the threshold an absolute value also does not work in every case, since outliers have sometimes a lower value than inliers. An explanation for this is that the images are still distorted by lens distortion. A consequence is that matches that are wrong, but not deviating more than the amount of lens distortion can not be detected. Since it is better to remove correct matches than to keep outliers, the image pairs with no outliers should sacrifice correct matches. Therefore the threshold is kept at 1σ and only if the detected number of outliers is far beyond the expected number of outliers, there are probably no outliers at all.

The number of outliers for the entire sequence of 66 test images is illus-

trated in Figure 5.11. The average number of outliers is just less than 4, but never exceeds 13 outliers.

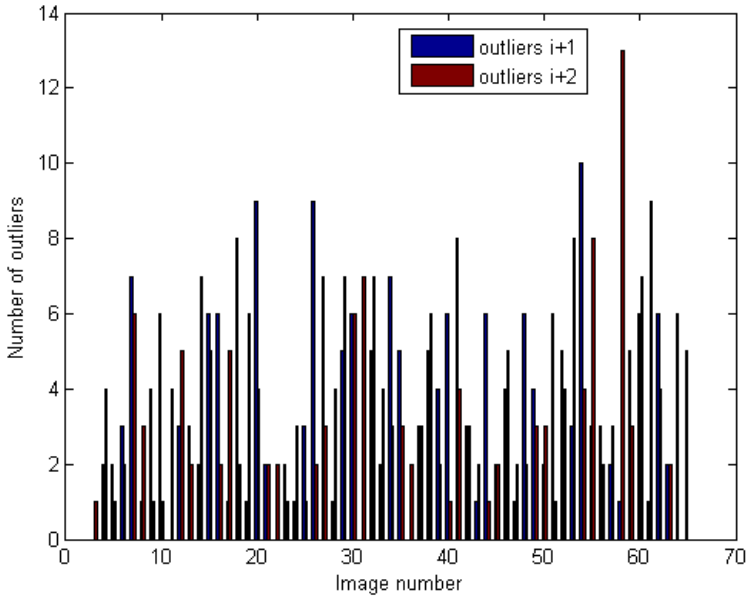


Figure 5.11: The number of outliers for every test image matched by the original algorithm.

The outliers are removed from the list with point correspondences to exclude them from any further processing. However, the original matching is still not very efficient. Therefore, the matching algorithm is adapted to run faster. This is described in the next section.

5.3.3 Optimization of the matching algorithm

The optimization of the matching algorithm is achieved by implementing the knowledge that the images are captured in an approximate straight line and with a baseline of around 1.6 metres. If only those keypoints are compared with each other that are located in the same area, the algorithm will be much more efficient. At the same time, it is expected that also most of the outliers are avoided, since matches will only be found at approximate right locations. For this approach the position of a possible match needs to be estimated according to the position of a keypoint in the previous image. The estimation of the location is decomposed in an x-shift and a y-shift. Let us assign the long side of the image as the x-axis, i.e. approximately the driving direction. In the ideal scenario, a point in one image appears in the next image with a constant x-shift, directly related to the length of the baseline and with no y-shift. However, in the test case of this research there

are several things to account for when estimating the position of a possible match in a subsequent image:

- the image scale is different for each pixel due to the tilt of the camera
- the heading (driving direction) changes due to manual steering of the vehicle
- for the test case, the length of baselines can differ ±10 cm, i.e. up to 50 pixels for resized images

Due to the tilt of the camera, the footprint of the images have the shape of a parallelogram in object space as can be seen in Figure 5.12. A tie point is therefore located in different image scales from image to image. This has an effect on the y-shift; image points are 'bended' towards the median y-value. This implies that, for an image of 1500×1000 pixels with a pixel coordinate system (origin at top left corner), pixels having a y-value lower than 500 are getting a positive y-shift and pixels having a y-value higher than 500 are getting a negative y-shift.

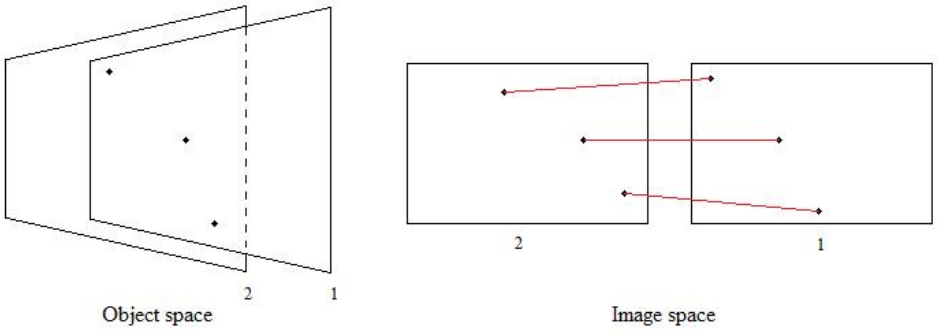


Figure 5.12: The effect of tilt. In the object space images are parallelograms, so that in the image space tie points will get different y-values.

If the tilt angle is known, then the x-shift can be corrected easily for the varying image scale. For the test acquisition, the tilt angle is estimated by eye at 30 degrees. In the test case, it is attempted to capture the images on a straight line. However, the steering of the vehicle is done manually, so the drive line is never perfectly straight. This has the largest effect on the y-shift as long as the angle difference in heading is small. Correcting for these heading changes is not possible since the heading is not known for the test survey. A difference of 2 degrees in heading can lead to a y-shift of more than 100 pixels. It is attempted to capture the images with a baseline of 1.6 metres. Differences in baseline lengths could be up to 10 cm. This corresponds to more than 50 pixels at locations with the smallest image

scale, i.e. vertically under the camera. When the images could be captured automatically based on an equal baseline, this uncertainty is brought down significantly.

It is concluded that the area for finding matches should be taken quite large, i.e. at least 150 by 150 pixels. The improved search for matches was implemented in Matlab. The total number of matches increased with an average of 3.3%. At first, one may expect a slight decrease of the matches because the false matches should be dissappeared. However, in the small search area, potential matches are sooner considered distinctive because there are much fewer keypoints to compare with. The distance ratio (as described in Section 2.8) could be decreased a bit, so that matches should look even more distinctive with respect to the rest of the keypoints in the surrounding area and less incorrect matches are selected. The number of keypoints and matches for both algorithms can be seen in Table 5.1.

The compution time for the matching decreased to 12 seconds per image pair. This is an improvement of 68 seconds per image pair compared to the original matching algorithm. Some of the keypoints can not only be seen in the next image but also in the second next image since the images have more than 50% overlap. Point correspondences in three images are better than in two images since it adds observations to the model without adding unknowns, so the redundancy increases. Besides, software like BINGO (as described in Section 3.6.2), requires for every image i at least one corresponding point in image $i+1$ and $i+2$ (if these images exist) for a successful bundle adjustment. Thus, it is recommended to match an image with its second next image as well. The original algorithm will almost take as long as matching two consecutive images while the optimized algorithm is much faster (2 seconds) because only a small part of the keypoints will be considered, i.e. those in the overlapping region. The total processing time for the optimized algorithm is less than one sixth of the original procesing

Table 5.1: Number of keypoints and matches found by the original algorithm and the optimized algorithm.

		keypoints	matching i+1	matching i+2
Original				
	resized	10,099	1376	37
	full	n.a.	n.a.	n.a.
Optimized				
	resized	9,994	1422	9
	full	$\pm 150,000$	$\pm 15,500$	± 300

Table 5.2: Computation times for finding and matching keypoints for the original algorithm and the optimized algorithm.

		keypoints	matching i+1	matching i+2	total
Original					
	resized	13s	80s	77s	173s
	full	n.a.	n.a.	n.a.	n.a.
Optimized					
	resized	13s	12s	2s	27s
	full	120s	$>$ 20m	$>$ 5m	$>$ 27m

time. The computation time for both images can be seen in Table 5.2.

Varying the distance ratio

As described in Section 2.8, the distance ratio is a measure for the strictness of the matching. The default distance ratio is 0.6, but in some cases one would like to change this value. Decreasing this ratio is the most obvious change, since it will reduce the number of outliers by making the matching more strict. This also results in a decrease in the number of correct matches. However, there are plenty of matches and it may be even beneficial to reduce the amount of matches, since it will also decrease the processing time of the following bundle adjustment. Table 5.3 provides the relation between the distance ratio and the number of matches. Important to notice is that when the distance ratio is lower than 0.4, no point correspondences can be found anymore in an image with its second next image. In order to guarantee a sufficient number of these third degree tie points with a distance ratio lower than 0.4, one can make the matching of just the third degree tie points less strict. This implies that the distance ratio is only decreased (to for instance 0.6) if and only if a keypoint in image i matches a keypoint in both image $i+1$ and $i+2$ to that decreased distance ratio. In this way, one can decrease the distance ratio to 0.3 without getting a lack of third degree tie points. The disadvantage is that the chance for outliers increases, which need to be filtered out.

Further optimization

Full resolution images still cannot be matched since the search for keypoints is already too heavy on a 'normal' computer. In order to solve this problem, the image is split into four equal regions and the SIFT algorithm is run on those regions separately. In the end, all information from the keypoints are merged into large matrices as would be produced directly from the full image.

Table 5.3: The relation between the distance ratio and the number of matches.

distance ratio	matches i+1	matches i+2
0.6	1422	18
0.5	1156	9
0.4	651	1
0.35	361	0
0.3	132	0
0.25	23	0

In this way the SIFT algorithm runs on a full resolution image in 120 seconds and finds between 110,000 and 150,000 keypoints. This number of keypoints is a factor of ten times more compared to the number of keypoints found on the resized images. This huge amount of keypoints makes the matching very slow and therefore practically not feasible to use. A solution could be to stop the matching process when a pre-selected number of matches are found, but one should be careful with this solution because it may not longer be guaranteed that the matches will have a good configuration. The approach of dividing an image into four equal regions is also applied on the resized images. The computation time is equal to the original algorithm since it only depends on the total number of pixels in the image. The number of found keypoints slightly reduces since near the edges of the split regions there are no keypoints. This is because keypoints are selected based on their surrounding, and near edges there are too few surrounding pixels. So, for resized images there is no need to use the adapted SIFT algorithm.

5.4 BINGO

BINGO is, just as Photomodeler and LPS, one of the photogrammetric software programs available at TU Delft. BINGO is chosen to do the improved processing with for its highest degree of automation and its suitability for close range photogrammetry. This specialized software should overcome the problems of Bundler. Different from Bundler, this photogrammetric specialized software is not open source and thus a license is required.

5.4.1 Workflow

The usage of BINGO has some advantages compared to Bundler. On the one hand it is designed for (close range) photogrammetric applications so it is better capable of setting preferences as is normal within photogrammetry (e.g. handle calibrated lens corrections and focal length). On the other

hand it allows for (and even demands) the use of control points. The 3D transformation to a local grid is therefore not needed anymore. Another difference is that the user has control on what observations contribute in the model. BINGO requires image coordinates of point correspondences and therefore the matching should be performed first. Besides the point correspondences, BINGO also demands the image coordinates and the local coordinates of the control points. The disadvantage is that the degree of automation is a bit lower, because the matching takes place in a separate step. BINGO should be able to do the relative orientation of all images, estimate the exterior orientation parameters for all images and obtain object point coordinates. A diagram of this workflow is illustrated in Figure 5.13.

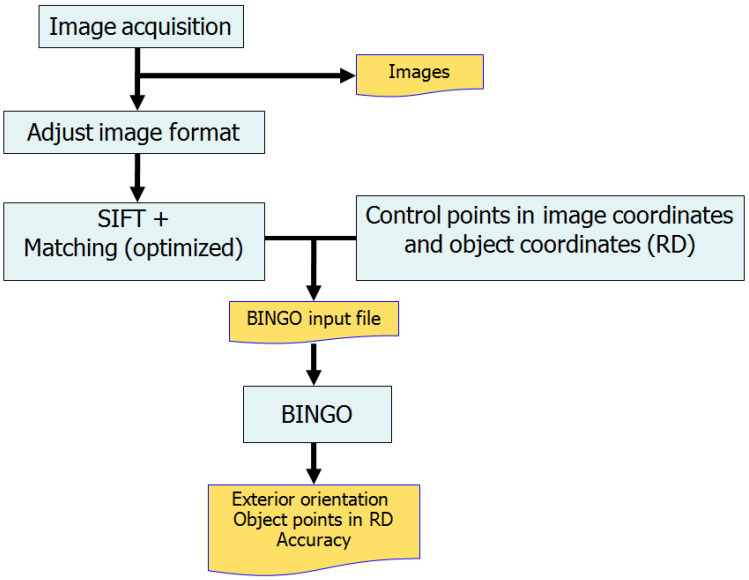


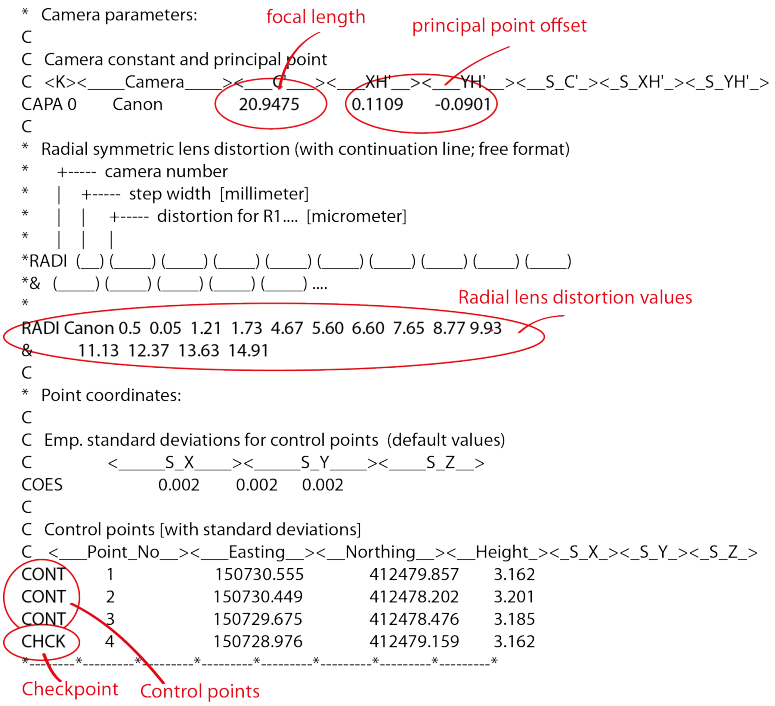
Figure 5.13: Workflow diagram of BINGO processing.

5.4.2 Preparation and processing

The preparation for the processing requires more effort than the processing itself. BINGO requires three input files before starting the bundle adjustment. The first file contains the image coordinates of point correspondences and control points. In Matlab, a program is written to generate automatically this input file from a set of images. Before this program is run, the control points should be measured manually in the images. For the image coordinates of the point correspondences, the optimized SIFT matching algorithm is used as described in Section 5.3 followed by a RANSAC outlier removal.

The second input file contains project settings. Once set, this file can be used for every similar project. In this file basic settings are defined, for instance not to reject control point measurements, a maximum number of iterations or a minimum number of corresponding points per image pair.

In the last input file, information about the camera and control points is provided. An example of this input file is illustrated in Figure 5.14. In this file, the focal length, principal point offset and radial lens distortion can be inserted if desired. If not, BINGO considers these parameters as unknowns in the model and estimates their values. The lens distortion is defined differently with respect to Photomodeler as described in 2.10.1. For a set of distances from the principal point, the absolute distortion is filled in. Furthermore, the local coordinates of all control points should be in this file. A control point can also be marked as check point. A minimum of three control points is required for the transformation to local coordinates, spread equally and preferably near the edges of the block. A check point is treated as any other tie point. Its calculated 3D coordinate will be compared with the measured 3D coordinate. The residuals provide a measure of accuracy of the object points.




```
*  Camera parameters:
C
C  Camera constant and principal point
C  <K><____Camera_____><____C____><___XH'__><___YH'__><__S_C'_><_S_XH'_><_S_YH'_>
CAPA 0       Canon          20.9475      0.1109     -0.0901
C
*  Radial symmetric lens distortion (with continuation line; free format)
*     +-----  camera number
*     |    +-----  step width  [millimeter]
*     |    |    +-----  distortion for R1....  [micrometer]
*     |    |    |
*RADI  (__) (____) (____) (____) (____) (____) (____) (____) (____)
*&   (____) (____) (____) (____) (____) ....
*
RADI Canon 0.5  0.05  1.21  1.73  4.67  5.60  6.60  7.65  8.77 9.93
&         11.13  12.37  13.63  14.91
C
*  Point coordinates:
C
C  Emp. standard deviations for control points  (default values)
C              <_____S_X____><_____S_Y____><____S_Z__>
COES                0.002       0.002     0.002
C
C  Control points [with standard deviations]
C  <___Point_No__><___Easting__><__Northing__><__Height_><_S_X_><_S_Y_><_S_Z_>
CONT        1              150730.555          412479.857       3.162
CONT        2              150730.449          412478.202       3.201
CONT        3              150729.675          412478.476       3.185
CHCK        4              150728.976          412479.159       3.162
*---------*---------*---------*---------*--------*--------*--------*--------*
```

Figure 5.14: Example BINGO input file called geoin.dat.

The processing starts with calculating initial approximations of the camera positions by the BINGO component 'RELAX' as was described in Section 3.6.2. Then, the bundle adjustment can start. In general, the bundle adjustment needs multiple cycles of iterations to detect gross errors or weak observations and converge to the final solution. During a cycle of iterations, a 'skip file' may be generated listing points that are considered to be a gross error or a weak observation. These points will then be skipped in the next cycle. Generally, the bundle adjustment performed in the next cycle is more precise and probably more accurate. The final result is obtained when during a cycle of iterations there are no points marked anymore to be skipped. Two output files are created. One with the actual result, listing all exterior orientation parameters and 3D coordinates of all object points. The other file describes the quality of the bundle adjustment in terms of redundancy, variances, standard deviations etc.

There are two measures for the quality of the result. The first measure is the internal precision of the bundle adjustment, i.e. how well the collinearity equations fit, and the second measure is the external accuracy, i.e. the actual obtained accuracy in object space. The internal precision is listed by BINGO in an output file. Root mean square (RMS) precision values are provided for the position and orientation of the camera and for object points. The external accuracy is, however, the most important quality measure since this is a comparison with the true values. The external accuracy is evaluated with the use of check points. Check points are known in object coordinates and preferably also in image coordinates. Check points can be control points that are not considered as control point during the bundle adjustment. These are measured manually in the images. Their object coordinates are calculated in the bundle adjustment. The difference between these calculated coordinates and their actual object coordinates is the measure for the obtained accuracy. Other check points can be points that are not measured in the image as were introduced in Section 5.1. Due to the lack of image coordinates, the latter check points can only be used as a check for the height.

Chapter 6

Results

In this chapter the results from the BINGO processing are provided. It start with the result of a small strip of five images in Section 6.1. The process to obtain new estimates for lens parameters is described in Section 6.2 and the method is validated in Section 6.3. The effect of the number of tie point observations on the obtained accuracy is provided in Section 6.4. The chapter concludes with the effect of the number of control points and their configuration in Section 6.5.

6.1 Sequence of five images

The first processing in BINGO is performed on the same five images used in the Bundler processing as was described in Section 5.2. The aim of this processing is on one hand to get familiar with BINGO and on the other hand to check if the unexpected shape that was found by Bundler (see also Figure 5.8) has disappeared.

The optimized algorithm of SIFT (as was described in Section 5.3.3) was used to find point correspondences. The distance ratio (i.e. the strictness of the matching procedure as described in Section 2.8 and 5.3.3) was set at the default 0.6. An average of 1675 point correspondences were found for two consecutive images. All of them were used in the bundle adjustment. The calibrated values for principal point offset, focal length and radial lens distortion were used. A plot of the calculated object points is shown in Figure 6.1[1]. The result is different from the result of Bundler (see Figure 5.2.5), but the point cloud still does not approximate a plane as was expected in Section 5.1. The maximum height difference, of more than 35 centimetres, is even larger than with the Bundler approach. But it still holds that if the

[1]The RD coordinates are subtracted with a constant for a clearer visualization. X=X_{RD} - 150,000 and Y=Y_{RD} - 412,000. This is also done for coming figures with RD coordinates.

distance to the control points increases, the height difference increases, i.e. a degradation of the quality occurs.

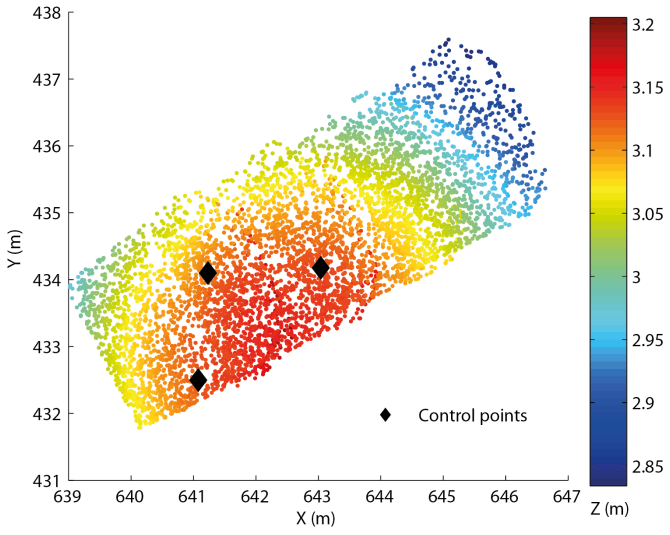


Figure 6.1: Object coordinates of last 5 images of the sequence in RD.

It is clear that there is still a gross error in the model. There are two possibilities that are most likely to be the cause of this problem. The first is an undetected outlier in the set of point correspondences, which could distort the model. The second possibility is that the calibrated lens distortions and focal length are different from the actual values.

In order to exclude the possibility that undetected outliers are the reason, a selection of 20 tie points per consecutive image pair are used for the bundle adjustment. In the strip of five images, all tie points ($4 \cdot 20$) can be checked by eye on outliers quite fast. No outliers could be found while the result has not significantly changed. Thus, an outlier free set of observations will not solve the problem. In the next paragraph, new values for radial lens distortion and focal length are estimated by BINGO.

6.2 Focal length and lens distortion estimation

As was described in Section 3.7, the calibrated focal length and radial lens distortion should be used with care since the actual image conditions are different from the conditions during calibration. BINGO can consider these parameters as unknowns in the bundle adjustment and estimate their val-

ues itself. It is recommended to obtain the point correspondences with a distance ratio lower than 0.6 in order to decrease the chance for outliers (the lower the distance ratio, the stricter the matching). From Table 5.3 a distance ratio of 0.30 is considered to be good, since it minimizes the chance for outliers and still keeps a lot of point correspondences. In order to avoid a lack of third degree tie points, a distance ratio of 0.6 is used for matching keypoints in image i with keypoints in both image $i+1$ and image $i+2$. The price of considering these third degree tie points, is that an additional check on outliers is required since its matching was less strict.

First, the calibrated radial lens distortion is set to zero and marked as unknown, so that its value will be estimated in the bundle adjustment. Then, the estimated radial lens distortion is used as new input for the next, hopefully improved, bundle adjustment. Within this second bundle adjustment, the focal length is also marked as unknown in order to find a new estimate. Both the focal length as the radial lens distortion have changed considerably. The focal length decreases from the calibrated 20.9475 mm to 20.1508 mm with a precision of $1\sigma = 0.073$ mm. The difference in radial lens distortion is illustrated in Figure 6.2.

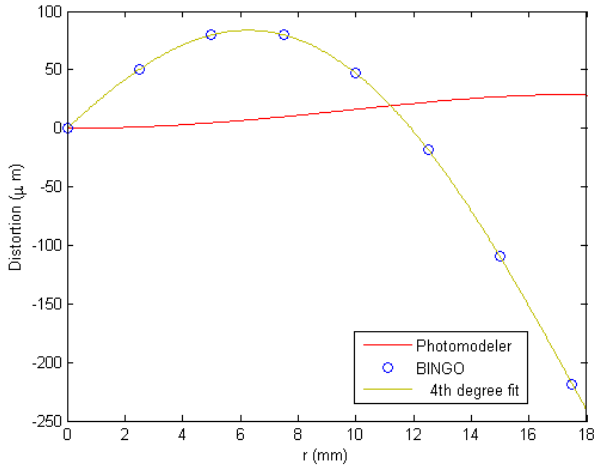


Figure 6.2: Difference in radial lens distortion between the calibration of Photomodeler and the estimation of BINGO. The horizontal axis is the distance from a pixel to the principal point in millimetres.

It can be seen that the difference between the calibration by Photomodeler and the BINGO estimation is large. The peaks of both curves are largely separated and the magnitude of the peak of the BINGO curve at around 6 mm, is more than two times higher than the peak observed in the curve

from Photomodeler at around 18 mm. However, both curves have a similar shape as can be seen best when looking at Figure 3.8 where the calibrated curve is scaled better.

The effect of the new parameters on the result is large. The best way to show that is again via a plot of the calculated object points that should follow the shape of the road. This result is presented in Figure 6.3. When looking at the height of the points, a clear trend can be seen perpendicular to the driving direction. The trend seems to be linear and thus the shape of the point cloud is in consensus with the expected shape as formulated in Section 5.1. It is concluded that the new radial lens distortion values are better than the calibrated parameters. The best explanation for this, is that BINGO uses the actual images, with many tie points to estimate and check the lens distortion. As expected, the focal length is estimated a bit lower than the calibrated value, since the distance to the 'object' is larger and thus the focal length decreases, see also Section 2.10.2. So, the lens calibration by Photomodeler is considered to be not good enough, since estimating the lens parameters during the bundle adjustment yields much better results. However, a lens calibration executed in a laboratory or a special calibration room will probably yield even better results, but this requires a lot of time, expertise and money.

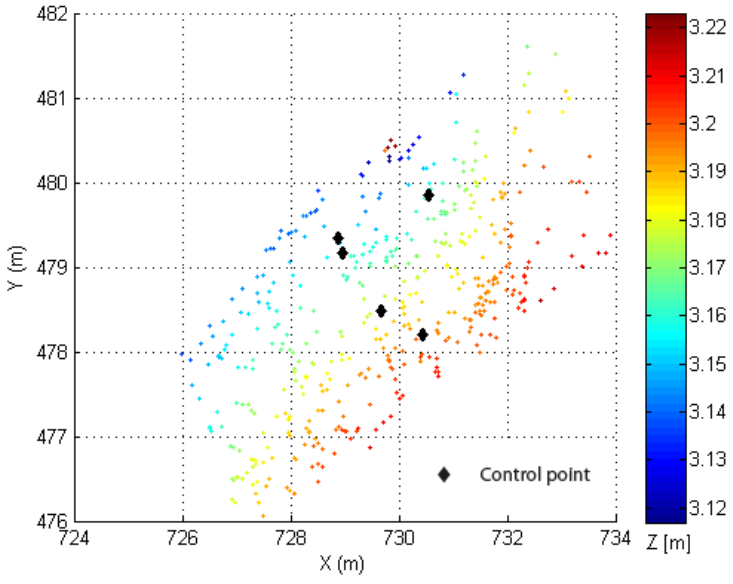


Figure 6.3: Object coordinates of the first 5 images of the sequence with new estimates of lens distortion and focal length

6.3 Validation of the method

The obtained result so far is not of great value without some validation of the accuracy. Since there are no true values available for the position and orientation of the camera, the validation has to come from the obtained 3D point cloud, i.e. the 3D coordinates of the tie points. In order to be able to verify that the obtained object points are accurate, they are compared with a set of check points. These check points are measurements on the road's surface and are not visible in the images. Hence, these points can not participate in the bundle adjustment, which makes them fully independent and therefore ideal check points. As described in Section 5.1, a set of check points consists of six points that are approximately on a line perpendicular to the driving direction. The set of five images used in the previous section are extended with another five images so that the covered surface contains also a set of six check points. No settings have changed with respect to the previous section, i.e. the same matching, focal length and radial lens distortion parameters.

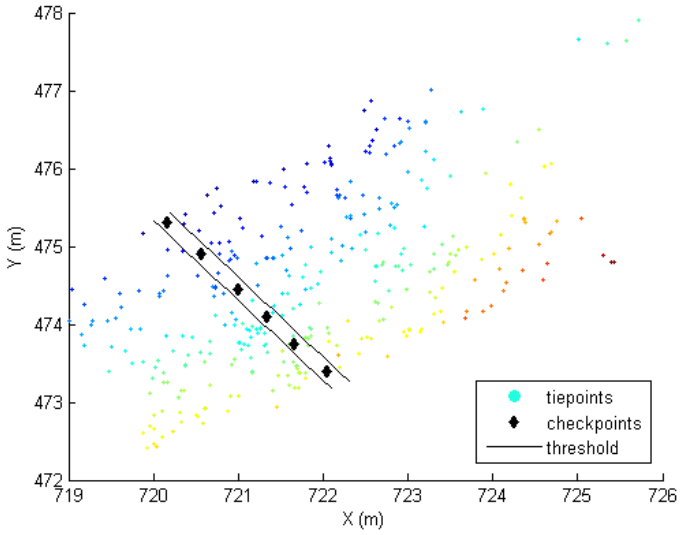


Figure 6.4: The selection of tie points within a small distance from the check points.

First, the height component for both check points and object points is neglected. Then, a least squares line is fit to the check points and all object points within a small distance threshold from this line are selected for comparison. This is also illustrated in Figure 6.4. The height of the selected object points is compared with the height of the check points in order to

estimate the accuracy of the height of the object points. In Figure 6.5 the y-coordinates of the selected object points and the check points are plotted against their height. In this way, the slope of the road (of the part that was covered by the images) becomes visible. The least squares line fit through the check points shows very little difference with the least squares line fit through the object points. Absolute differences are 1-2 mm while the difference in slope is 0.022 degree.

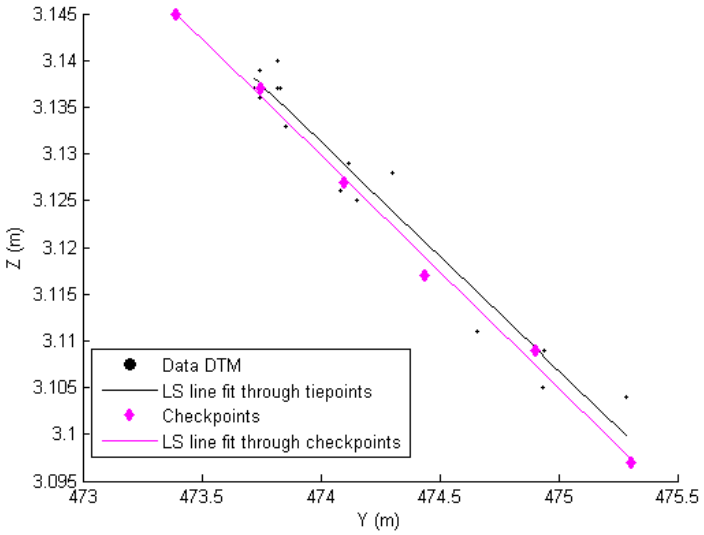


Figure 6.5: Accuracy check of the object points by comparing them to independent check points.

As stated before in this section, there are no true values for camera positions or orientations. Nevertheless, the approximate known baseline does give an indication of the distance between two consecutive camera positions. The aimed baseline is 1.6 metres as stated in Section 5.1. With a measure tape the locations of the stops were determined and marked. It will be assumed that the markings are placed with a precision of 1σ=0.01 m and the vehicle can be put on the right location with an accuracy of 0.05 m. These assumptions lead to an expected deviation of less than $0.01 * \sqrt{i-1} + 0.05$ m from the aimed distance from the first image, where i is the image number starting from $i = 2$. Figure 6.6 shows the difference in distance from the first image between the calculated distance by BINGO and the aimed distance (i.e. $1.6 * (i-1)$ metres). These differences are much smaller than expected. For instance, for image 10 the deviation of the aimed distance is expected to be less than $0.01 * \sqrt{10-1} + 0.05 = 0.08$ m, while the observed difference

is less than 0.03 m. The difference never even exceeds 0.04 m. Officially, this check does not validate the correctness of the method, because it says as much about the driver's skills to stop the vehicle at the right location. However, its success is a requirement for a high accuracy.

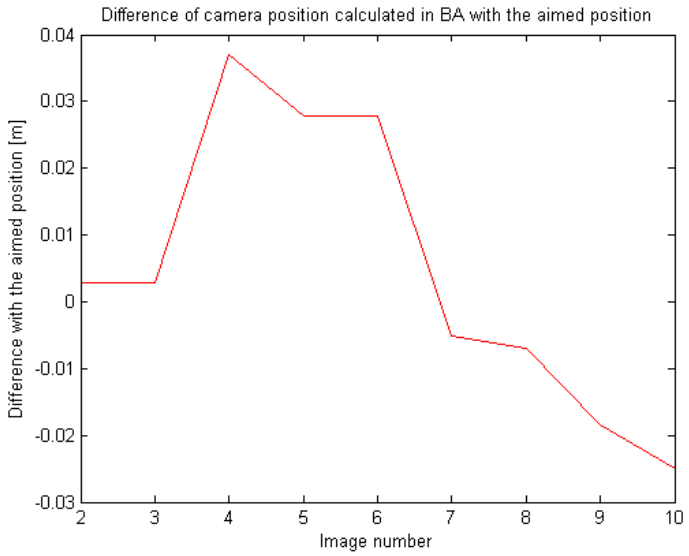


Figure 6.6: Difference between the actual camera position as calculated by BINGO with the aimed position.

6.4 Varying the number of observations

It is expected that if more tie point observations are used in the bundle adjustment, the accuracy of the result of the bundle adjustment will improve, as can be seen in Table 4.3. Another logical consequence of using more observations is that the processing time will increase. It is wise to use the optimal number of observations in relation to processing time and obtained accuracy, so that unnecessary processing time can be avoided. Of course the accuracy will be more important than the processing time, but the processing time should stay within practical bounds. In order to extract this relation, multiple bundle adjustment were executed where only the number of tie points was varied. In fact, it is the distance ratio that was changed. Four different distance ratios were used, which resulted in an average number of matches per image pair of respectively 28, 87, 223 and 1155. The processing time per image pair is shown in Figure 6.7. A linear relationship is observed. Apparently, if the matrices within the bundle adjustment double in size, the number of calculations also doubles.

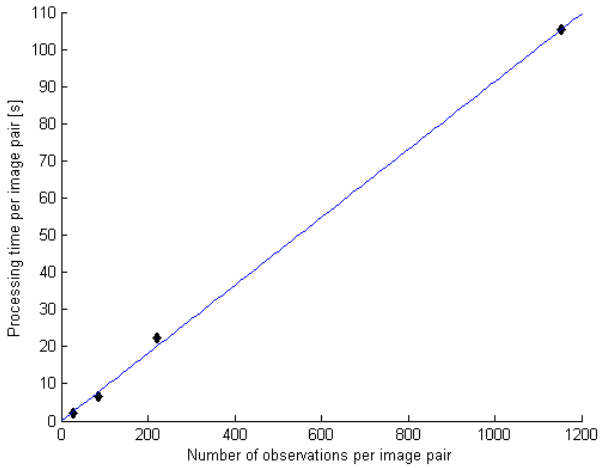


Figure 6.7: The number of observations per image pair versus the processing time per image pair. A linear relationship is observed.

The effect of increasing the number of tie point observations to the quality of the bundle adjustment is evaluated next. A sequence of 28 images was processed with almost all available control points. The only variable is the number of tie point observations, obtained with variable distance ratios. If the distance ratio is decreased, not only the number of point correspondences will drop, but the point correspondences will also be more distinctive. As described in Section 5.4.2, there are two measures of quality, internal precision and external accuracy. The effect of varying the number of tie point observations to the internal precision is shown in Table 6.1 and the effect on the external accuracy is shown in Table 6.2. It is observed that the more tie point observations are used, the better the internal precision of the bundle adjustment. The external accuracy does however not improve if more tie point observations are used. The contrary seems even true. Best accuracies are observed when using the least number of tie point observations. This is not in line with the expectations as were stated in Section 4.2. An explanation for this is that the less tie point observations, the lower the distance ratio and thus the more distinctive the point correspondences are. These point correspondences may be of higher quality than the ones obtained with a higher distance ratio. It is concluded that varying the number of observations has a different impact on the accuracy than on the internal precision. For the remainder of this chapter, only 28 tie point observations per image pair are used, since this yields best results and saves processing time.

Table 6.1: The RMS precision of the bundle adjustment for camera position and orientation for different number of tie point observations per image pair.

Average number of observations	(mm)			(°)		
	X	**Y**	**Z**	φ	ω	κ
28	7	12	5	0.52	0.09	0.57
83	4	7	3	0.29	0.04	0.31
223	2	4	2	0.15	0.02	0.16

Table 6.2: Check point residuals (mm) for different number of tie point observations.

Check point #	**28 obs.**			**83 obs.**			**223 obs.**		
	ΔX	**ΔY**	**ΔZ**	**ΔX**	**ΔY**	**ΔZ**	**ΔX**	**ΔY**	**ΔZ**
3	3	2	-1	4	2	-1	4	2	0
4	2	1	1	2	1	1	2	0	3
34			0			4			1
35			1			1			2
36			-1			1			5
37			-2			1			8
38			1			-4			7
39			-1			-3			10
RMS	3	2	1	3	2	2	3	1	6

6.5 The influence of control points

Although the observations are free from outliers, the bundle adjustment is not guaranteed to converge to a solution. For the sequence of available test images, it is experienced that the bundle adjustment converges only if not more than the first 28 images are used while using almost all available control points. The possible causes were tabulated in Table 4.1. It is not likely that the image coordinates of the point correspondences cause this divergence, since they are proven to be highly precise (see Section 4.1.1) and outlier free (see Section 5.3.2). The quality of the control point measurements is also considered to be high, because for its object coordinates measurements reliable tachymetry is used while the image measurements of the control points are still at pixel accuracy (see Section 4.1.2). The number of tie points and its configuration is considerably better than a standard photogrammetry project, so this is also excluded to be the disturbing factor. The lens distortion and focal length were determined in Section 6.2 and

yields high accuracy for the object points in Section 6.3. So, it is most likely that the number of control points and their configuration are the two most dominant factors for a successful bundle adjustment in this thesis.

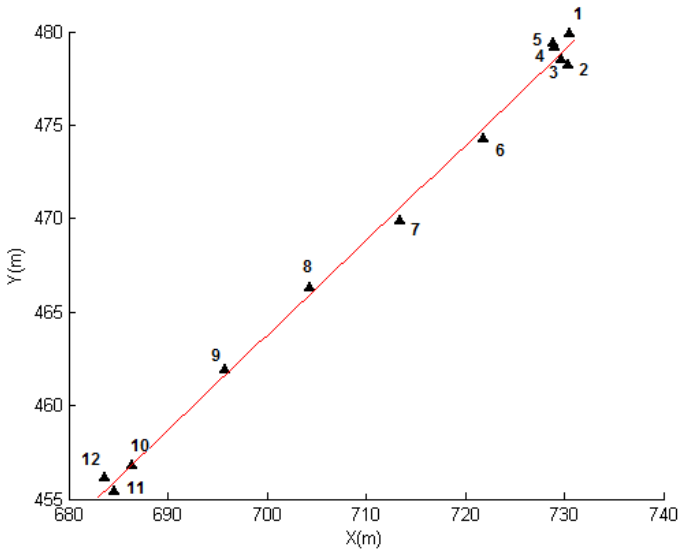


Figure 6.8: The control points that are visible in the first 36 images. A least squares line is fitted to emphasize that the control points are almost on a line.

The effect of varying the number of control points is evaluated next for the sequence consisting of the first 28 images. An average of 28 tie point observations per image pair was used in the bundle adjustment. The number of control points was decreased such that the distance between two consecutive control points is respectively 10, 20 and 30 metres, except for the beginning of the strip where three control points nearby each other were used. The remaining control points were used as check points. The result is shown in Table 6.3. Check points 34-39 were measured only in object space and are not visible in the images. These points are located on a line perpendicular to the driving direction and thus form a road profile. Their horizontal coordinates were used to find the closest object points (horizontally) from the bundle adjustment. A least squares plane was fitted through these object points to calculate the residuals in height of these check points. As expected, smallest residuals are observed in the case with every ten metres a control point. For the other interval distances, it is not entirely clear which one is best. Check point 3 and 4 are nearby control points 1, 2 and 5 as can be seen in Figure 6.8, and have therefore in all cases small residuals. So, it is more interesting to look at the other check point residuals. Check point 8

has the largest residual in height for the case with one control point every 20 metres, while the other check points residuals are a bit larger in height in the case of every 30 metres a control point. Using even fewer control points did not result in a successful bundle adjustment anymore, i.e. no convergence to the solution.

Table 6.3: Check point residuals (mm) for different control point interval distances.

Check	**10 m**			**20 m**			**30 m**		
point #	**ΔX**	**ΔY**	**ΔZ**	**ΔX**	**ΔY**	**ΔZ**	**ΔX**	**ΔY**	**ΔZ**
3	3	2	-1	3	2	-1	3	2	-1
4	2	1	1	2	1	0	2	1	1
6				1	3	-2	-10	-2	-3
7							-28	-13	6
8				16	11	31			
34			0			3			2
35			1			4			4
36			-1			3			3
37			-2			1			2
38			1			4			7
39			-1			2			6
RMS	3	2	1	8	6	10	15	7	4

Next, the effect of the configuration of the control points is evaluated. Due to the long and narrow shape of the strip, the configuration of the control points will always be weak. The control points that are visible in the first 36 images are visualized in Figure 6.8. A least squares line was fitted through the control points to emphasize that the control points are almost on a line. As stated in Section 4.2, this has especially a negative effect on the accuracy in height. Those control points with the largest distance to the line as visualized in Figure 6.8, i.e. the points that are least on one line, are considered as best distributed. At the beginning of the strip, control points 2 and 5 are considered as best and control points 1 and 3 are considered as weakest. Both combinations are executed in a bundle adjustment complemented with control points 7 and 9. The residuals on the check points are shown in Table 6.4. The strongest configuration of control points yields much smaller residuals in height. Besides, with the best distributed control points, a longer strip of images could be processed. Instead of the first 28 images, the first 36 images could be processed, which is just over half the sequence of available images. So, a slight improvement of the configuration of the control points yields a significant better accuracy in height and makes it possible to process more images. From Table 6.4 it is also observed that

the residuals are in general in the order of millimetres, except for check point 8 which has residuals in the order of centimetres.
It is concluded that a good configuration of control points is needed to obtain high accuracy in height and for a more robust bundle adjustment. For a short strip, the configuration of the control points has less impact as can be concluded from Section 6.3. With the best configuration, 28 images ($\approx$ 43 metres) could be processed with sub-centimetre accuracy and five control points. The configuration of the control points is too weak to process more than 36 images. Thus, the weak configuration decreases the robustness of the processing approach. By cutting the bundle adjustment into pieces of let's say 28 images, sub-centimetre accuracies could be achieved, provided that per bundle adjustment at least five control points are used with a good configuration and these are measured by tachymetry (or with similar precision and accuracy).

Table 6.4: Check point residuals (mm) for for the weakest and strongest control point configuration.

Check	**Weakest configuration**			**Strongest configuration**		
point #	**Δ X**	**Δ Y**	**Δ Z**	**Δ X**	**Δ Y**	**Δ Z**
1				2	0	1
2	-5	0	-10			
3				3	2	-1
4	1	-2	14	2	1	0
5	-1	-4	16			
6	9	8	4	0	4	-5
8	14	10	39	18	12	32
34			8			6
35			3			7
36			-5			5
37			-12			3
38			-17			6
39			-26			3
RMS	8	6	17	8	6	11

6.6 Summary

It was found that the calibrated lens parameters, as were obtained in Section 3.7, caused a major error in the bundle adjustment. Focal length and radial lens distortion parameters are therefore estimated during bundle adjustment. The method is validated with a bundle adjustment on ten images.

The check point residuals were in the order of millimetres and the calculated slope of the road differed 0.022 degree with the slope measured by tachymetry. The effect of using more tie point observations by decreasing the distance ratio on the bundle adjustment was not as expected. The case with least tie point observations lead to the best accuracy. This is probably because the higher the distance ratio, the fewer tie points, but the more distinctive these are. The configuration of control points is proven to be important. Due to the narrow and long shape of the image strip, a weak configuration of the control points is inevitable. Choosing those control points that deviate most from the least squares line through the control points, yields the best accuracy in height. However, the configuration is still too weak to guarantee a robust bundle adjustment. By cutting the bundle adjustment into small pieces, sub-centimetre accuracies could be achieved, provided that per bundle adjustment at least five control points are used with a good configuration and these are measured by tachymetry (or with similar precision and accuracy).

Chapter 7

Discussion

This chapter summarizes the most important results from this thesis. It also provides an interpretation of the results and the consequences for Breijn.

A maximum sequence of 36 images could be handled successfully with BINGO, independent on the number of available control points or tie points (see also Sections 6.4 and 6.5). The bundle adjustment could not converge to a solution when using more than 36 images. However, a sequence of ten overlapping images with four control points can be processed with BINGO with millimetre absolute accuracy. The requirements for a successful bundle adjustment include having *accurate*, *sufficient* and *well distributed* tie points and control points. Section 4.1.1 demonstrates that the image coordinates of the tie points are more precise than in a standard photogrammetric project (0.1 pixels versus 0.3-0.5 pixels). Also the number of tie points (>1000) and their configuration is considered to be better. Regarding control points it was concluded that the image measurements are of sufficient accuracy (1-2 pixels) while their object coordinates have millimetre accuracy (see Section 4.1.2). Consequently, the cause for the failure has to be found in the number of control points and their configuration.

It was found that the accuracy of the bundle adjustment depends significantly on the number of control points used. If fewer control points are used than one per 10 metres, the accuracy of calculated object points degrades to the order of centimetres (as can be seen in Table 6.3). In Section 6.5, it was shown that the configuration of the control points is very important. A slight improvement of the configuration resulted in that more images could be processed and better accuracies were obtained.

It was expected that a large number of tie points would have a positive effect on the accuracy, so that fewer control points were needed. As can be seen in Table 6.2, best accuracies are observed when using only 28 tie point observations per image pair. An explanation for this is that the fewer tie

point observations, the lower the distance ratio was set and thus the more distinctive the point correspondences are. These point correspondences may be of higher quality than the ones obtained with a higher distance ratio.

The original goal was to improve the accuracy of Pavescan and to reduce the number of control points. Even if more than 36 images could have been processed, the accuracy is not expected to get better unless more and better distributed control points are used. The configuration of the control points can not be improved much more due to the narrow shape of the strip. Besides, it may demand photogrammetric knowledge from the operator. Cutting the bundle adjustment in small pieces may work. The configuration of the control points is much better in this way. Using strips of 28 images ($\approx$ 43 metres) would however still require at least five control points that should be measured by tachymetry (or with similar precision and accuracy). It is concluded that close range photogrammetry does not improve the accuracy of Pavescan measurements in combination with decreasing the number of control points. An improvement of accuracy is only possible if at least the same amount of control points are used that are highly accurate and spatially well distributed. The practical feasibility differs therefore from the theoretical feasibility.

The use of other software is not expected to give a better result, since BINGO is designed for (close range) photogrammetry and all settings were optimized. There are, however, other possibilities to overcome the problems with processing a strip of images. Although it is known that the scan provides a precise profile of the road's surface, such a measurement can not be incorporated in photogrammetric software. A solution could therefore be to code a new bundle adjustment program with additional constraints that the object points should follow the laser scan measurements.

A recent development within Pavescan provides the system with GPS measurements. These GPS measurements are related to the scan and therefore can be considered as control points. The horizontal accuracy is expected to be in the order of centimetres and the vertical accuracy a bit weaker. As was stated in Section 1.1, the accuracy of the height is most important, so for an accurate height of the scan additional measurements are still required. Usage of these points as control points in the bundle adjustment seems not very advantageous since the accuracy of these control points is rather weak. However, with the GPS measurements on board one can limit the additional survey to just a levelling survey on the control points on the road's surface provided that the GPS reception will be good enough.

Chapter 8

Conclusions and recommendations

This chapter contains the main conclusions and recommendations of this thesis.

8.1 Conclusions

8.1.1 Main conclusion

With integrating close range photogrammetry into Pavescan, it is not possible to achieve millimetre accuracy of the laser scans with using fewer control points. This answers the main research question as stated in Section 1.2. Reducing the number of control points is not possible since the bundle adjustment will not converge to a solution anymore. Cutting the bundle adjustment in small pieces may be a solution, but still the desired level of accuracy could only be achieved if at least the same amount of control points are used that are highly accurate and well distributed. The configuration of the control points is very important. The fact that the images construct one long and narrow strip, makes it in practice difficult to guarantee a good configuration of the control points. Besides, it may demand photogrammetric knowledge of the operator. The theoretical feasibility differs therefore from the practical feasibility.

8.1.2 Other conclusions

- The best setup for integrating close range images into Pavescan is to make the images while driving and pointed such that the trailer of Pavescan is not covered. Images are preferred above video frames because images are of better quality and thus contain more detail. Besides, SIFT is even after the optimization computational expensive. Since there would be much more video frames than images, the use of

video frames is not recommended. The use of a fisheye lens is also not recommended, since it is expected that the accuracy of the absolute orientation will be worse than the accuracy of the absolute orientation obtained with normal images. Besides, fisheye images bring serious practical drawbacks. The projection is different from the central perspective projection as with normal images and an extensive calibration of the fisheye lens is needed.

- The (for the human eye) very monotonous image of asphalt pavement is highly suitable for finding automatically point correspondences. The number of point correspondences is dependent on the height of the camera above the asphalt pavement. SIFT was able to detect around 1,400 point correspondences per image pair for the test images. It is however not desired to use all the point correspondences in the bundle adjustment since the processing time is linearly increasing with the number of tie points. A good solution is to make the matching more strict. This not only reduces the number of matches, but also results in the most distinctive matches.

- By using estimates of the baseline and orientation, SIFT matching was made a factor six faster. For all keypoints in one image, the position can be estimated in the other image. By just comparing those keypoints in that local neighbourhood, the original matching is optimized. With this optimized algorithm, matching an image i with image $i + 1$ and image $i + 2$ takes 27 seconds instead of 173 seconds with the original algorithm.

- The point correspondences found by SIFT have an accuracy of 0.1 pixel as was tested on images from the Canon EOS 350D with 20 mm lens. This accuracy is significantly better than the 0.3-0.5 pixel accuracy that is standard for automatic image measurements. The images were resized from 3456×2304 to 1500×1000 pixels, leaving a square pixel size of around 14.7 μm. The resulting accuracy of 1.5 μm is around ten times better than the accuracy of 10-20 μm that is used for manual image measurements in aerial photogrammetry.

- The program Bundler is not the solution for Pavescan. It offers a great advantage compared to other programs in its automation capability, but one is not able to keep sufficient control of the bundle adjustment. Besides, control points can not be implemented and the result from Bundler therefore requires an additional 3D transformation.

- Close range photogrammetry integrated in Pavescan could achieve a high degree of automation. However, control points should always be measured in the terrain and in this thesis the control points are measured manually in the images as well. The latter is expected to be

able to automate. Point correspondences of a sequence of images can be found automatically and when using BINGO only minor manual actions are required. The results of the bundle adjustment can directly be linked with the laser scan.

- Increasing the number of tie point observations by making the matching less strict does not improve the accuracy of the bundle adjustment. The contrary was observed in Section 6.4 where the least number of tie point observations yielded best accuracies. The usage of more tie point observations does improve the internal precision of the bundle adjustment, i.e. how well the collinearity equations fit.

- If fewer control points are used than one per 10 metres, the accuracy of calculated object point coordinates degrades to the order of centimetres. A sequence of ten overlapping images (covering around 15 metres) with four control points was processed with BINGO with millimetre accuracy. The accuracy of the height of the laser scan will have a similar accuracy level as the accuracy of the height of the object points, while the horizontal accuracy of the scan points depends on the accuracy of the orientation of the camera. Obtaining the accuracy of the orientation of the camera requires testing with for instance an INS, in order to have reference values.

- Estimating lens parameters in the bundle adjustment yields much better results than obtaining these via a simple lens calibration with Photomodeler. The calibration by Photomodeler was not comparable with the actual situation of the survey as was also described in Section 3.7. Calibration in a laboratory or a special calibration room will probably yield best results.

8.2 Recommendations

8.2.1 Main recommendation

The main recommendation to Breijn based on the results obtained in this thesis is not to use close range images to improve the accuracy of Pavescan. Although it is theoretically possible to improve the accuracy of the scan points, in practice it gives many difficulties. It would require at least as many control points, that should have millimetre accuracy. Also the configuration of the control points should be improved, which is complex with one narrow strip of images. This thesis focused on a low cost solution. Now close range photogrammetry appears to be not feasible in practice, it is recommended to Breijn to look at the possibilities with mobile mapping systems. Although these systems will be more expensive, it is likely that prices will drop. More and more companies offer mobile mapping systems, which leads

to competition and thus reduced prices. In Section 2.4 it was stated that mobile mapping systems do not fulfil the accuracy requirements. However, these systems are in full development and thus improvements are expected. More research is recommended on whether the quality of such systems can fulfil the requirements Breijn demands.

8.2.2 Other recommendations

In case one decides to do more research on improving the accuracy of the scan points with close range photogrammetry or on obtaining a more accurate and robust bundle adjustment in general, the following is recommended:

- It should be considered to code a new bundle adjustment program instead of using an existing program. The main advantage will be that additional information could be taken into account. In case of Pavescan, it is known what the shape of the road looks like. If this can be implemented as a constraint in the bundle adjustment, it would be beneficial.
- Regarding the accuracy and robustness of the bundle adjustment it is recommended to capture the images with a constant distance interval to keep control on the overlap. Therefore, the camera needs to be triggered for instance by wheel rotations. The camera itself should have a larger angle of view, such that a better baseline-height ratio is obtained. Furthermore, it is recommended to place the control points with even more care, such that an optimal configuration is achieved. In order to obtain a higher automation, it is recommended to replace the nails as control points by automatically detectable targets. Manual measurements of the control points in the images are then avoided.

Bibliography

A. Fischler and R. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 1981. http://www.ai.sri.com/pubs/files/836.pdf.

A. Habib and D. Kelley. Automatic relative orientation of large scale imagery over urban areas using Modified Iterated Hough Transform. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2001.

R. Hartley. In defense of the eight-point algorithm. *IEEE Transactions on pattern analysis and machine intelligence*, 1997. http://www.cs.ucf.edu/courses/cap6938-02/refs/hartley.pdf.

R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003. http://books.google.nl/books?id=si3R3Pfa98QC&printsec=frontcover&dq=multiple+view+geometry&hl=nl&ei=dZ9OTK_kE9O7jAeHr7TIBw&sa=X&oi=book_result&ct=result&resnum=1&ved=0CCwQ6AEwAA#v=onepage&q&f=false.

F. van den Heuvel, R. Verwaal, and B Beers. Calibration of fisheye camera systems and the reduction of chromatic aberration. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVI, 2006. http://www.isprs.org/commission5/proceedings06/paper/1267_Dresden06.pdf.

G. Hunter, C. Cox, and J Kremer. Development of a commercial laser scanning mobile mapping system - StreetMapper. *Second International Workshop The Future Of Remote Sensing*, 2006. http://www.isprs.org/proceedings/XXXVI/1-W44/www.pegasus4europe.com/pegasus/workshop/documents/contributions/Hunter_full.pdf.

M. Kedzierski and A. Fryskowska. Precise methods of fisheye lens calibration. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVII:765–768, 2008. http://www.isprs.org/proceedings/XXXVII/congress/5_pdf/134.pdf.

K. Khosh Elham. *Lecture notes Photogrammetry.* TU Delft, 2010.

M. Kodde. *Dictaat Meetkundige Geodesie.* Hogeschool van Utrecht, 2004. http://www.martinkodde.nl/downloads/dictaat_mgd.pdf.

E. Kruck. *BINGO 5.0 Users Manual.* Aalen University, 2003.

E. Kruppa. Zur Ermittlung eines Objectes aus zwei Perspektiven mit innere Orientierung. *Math. Naturw., Vienna*, 1913.

Leica Geosystems AG. Leica TPS1200+ Series. Brochure, 2009. http://www.leica-geosystems.com/downloads123/zz/tps/tps1200/brochures/Leica_TPS1200+_brochure_en.pdf.

Leica Geosystems Geospatial Imaging. *Leica Photogrammetry Suite Product Description*, 2008. http://www.sccon.com.br/pdfs/erdas_lps_92_description.pdf.

Leica Geosystems GIS and Mapping. *Leica Photogrammetry Suite OrthoBASE & OrthoBASE Pro User's Guide.* Leica Geosystems GIS & Mapping, 2003.

M. Lourakis. A Brief Description of the Levenberg-Marquardt Algorithm Implemened by levmar. Technical report, Institute of Computer Science - Foundation of research and Technology, Hellas, 2005. http://www.ics.forth.gr/~lourakis/levmar/levmar.pdf.

M. Lourakis and A. Argyros. SBA: A software package for generic sparse bundle adjustment. *ACM Transactions on Mathematical Software*, 2009. http://www.ics.forth.gr/~lourakis/sba/sba-toms.pdf.

D. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 2004. http://people.cs.ubc.ca/~lowe/papers/ijcv04.pdf.

D. Marquardt. An algorithm for least squares estimation of non-linear parameters. *SIAM journal of applied sciences*, 1963. http://edu.chem.tue.nl/6KM06/Backgroundreadingmaterial/DMarquardt1963.pdf.

D.F. Maune, J.B. Maitra, and E.J. McKay. *Digital Elevation Model Technologies and Applications: The DEM Users Manual.* American Society for Photogrammetry and Remote Sensing, 2001.

J. Mennink. Vastleggen dwarsprofiel in het verkeer. *Asfalt*, 2008. http://www.vbwasfalt.org/cms/Media/BladAsfalt2008-2/2-2008Mennink.pdf.

P. Mittrapiyanuruk. A Memo on How to Use the Levenberg-Marquardt Algorithm for Refining Camera Calibration Parameters, 2008. `http://cobweb.ecn.purdue.edu/~kak/courses-i-teach/ECE661.08/homework/HW5_LM_handout.pdf`.

H.B. Nielsen. Damping parameter in Marquardt's method. Technical report, Technical University of Denmark, 1999. `http://www2.imm.dtu.dk/documents/ftp/tr99/tr05_99.pdf`.

D. Nistér. An efficient solution to the five-point relative pose problem. *IEEE conference on Computer Vision and Pattern Recognition*, 2003. `http://cmp.felk.cvut.cz/cmp/courses/EP33VKR/2005/Nister-PAMI2004.pdf`.

F. Remondino and C. Fraser. Digital camera calibration methods: considerations and comparisons. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVI, 2006. `http://www.photogrammetry.ethz.ch/general/persons/fabio/Remondino_Fraser_ISPRSV_2006.pdf`.

E. Schwalbe. Geometric modelling and calibration of fisheye lens camera systems. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVI, 2006. `http://www.isprs.org/proceedings/XXXVI/5-W8/Paper/PanoWS_Berlin2005_Schwalbe.pdf`.

Sick AG. *LMS 200 / LMS 211 / LMS 220 / LMS 221 / LMS 291 Laser Measurement Systems, Technical description*, 2003.

N. Snavely, S. M. Seitz, and R. Szeliski. Modeling the world from internet photo collections. *International Journal of Computer Vision*, 80(2):189–210, 2008. `http://phototour.cs.washington.edu/`.

P. Teunissen, D. Simons, and C. Tiberius. *Probability and observation theory*. Department of Earth Observation and Space Systems, TU Delft, 2005.

B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle adjustment - a modern synthesis. In *Workshop on Vision Algorithms*, pages 298–372, 1999. `http://lear.inrialpes.fr/pubs/2000/TMHF00/Triggs-va99.pdf`.

U.S. Army Corps of Engineers. Photogrammetric mapping. Technical report, 1993. `http://gisceu.net/U5.HTM`.

M.G. Vosselman. *Fotogrammetrie II*. TU Delft, 1999.

V. Wan and J. Carmichael. Polynomial Dynamic Time Warping Kernel Support Vector Machines for Dysarthric Speech Recognition with Sparse Training Data. *Interspeech 2005*, 2005. `http://www.dcs.shef.ac.uk/~natty/PolynomialDynamicTimeWarping.pdf`.

J. Wang, B. Zhang, and Y. Song. *Principles of photogrammetry (with remote sensing)*. Wuhan Technical University of Surveying and Mapping, 1990.

Website DRIVE-MAP. `http://www.drive-map.nl/dl/DRIVE_MAP_A4_NL_LR2.pdf`.

Website Heijmans. `http://www.heijmans.nl`.

Website Photomodeler. `http://www.photomodeler.com`.

Website Weather Underground. `http://www.wunderground.com`.

M. Wiggenhagen. Calibration of digital consumer cameras for photogrammetric applications. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIV, 2002. `http://www.isprs.org/proceedings/XXXIV/part3/papers/paper009.pdf`.

T. Wortel. Automating quantity surveying in road construction using UAV videogrammetry. Master's thesis, TU Delft, 2009.

Appendix A

Bundler usage in windows

If you are planning to use Bundler under Windows, it is recommended by the creators to use Cygwin, assuming that bash and perl are being installed. Cygwin provides a Linux look and feel under the Windows environment. Files with .sh and .pl extensions need to be converted from DOS to Unix format in order to run properly in Cygwin. This is due to usage of different line endings. Every time a change is made in such a file, the conversion needs to be applied. This should be achieved with the command:

d2u -U dir-path/file.sh

Before running the bundler script, the base path needs to be defined. This is the directory where all the Bundler files are installed. The C-drive is accessible in Cygwin via

/cygdrive/c/

Running the Bundler script requires a lot of computation power. If the processing takes too long or an error pops up telling the computer ran out of memory, it is recommended to resize the images. Be aware that resizing the images leads inevitably to information loss. Having ImageMagick installed, resizing goes like this:

convert -resize 1000 image1.jpg image2.jpg

where the number 1000 indicates that the new image (image2.jpg) has 1000 pixels in length. If some parts of the image contain unwanted objects e.g. the trailer, it is recommended to make a crop of the image. This goes similar like resizing:

convert -crop "1100x1000+400+0" image1.jpg image2.jpg

where the first two numbers within the quotes represent the crop size and the last two numbers represent the upper left corner of the crop both in pixels in x and y direction. Original size in this example is 1100 + 400 = 1500 pixels width and 1000 + 0 = 1000 pixels height.
The output of Bundler has a .ply extension which cannot be read without additional software. Conversion to .wrl is an easy solution with the srcipt xyz2wrl. To do so, copy the xyz2wrl.txt file into the directory of the .ply and convert with a command like:

gawk -f xyz2wrl.txt < file.ply > file.wrl

This wrl-file contains all the coordinates of the tie points and the coordinates of the projective center which can easily be viewed via vrml-view or Matlab.