Photogrammetric Surveying of Wood Piles on Handheld Devices

Dissertation

zur Erlangung des akademischen Grades

Doktoringenieur (Dr.-Ing.)

angenommen durch die Fakultät für Informatik
der Otto-von-Guericke-Universität Magdeburg

von

Christopher Herbon, M.Eng.
geboren am 24.10.1987 in Göttingen

Magdeburg, den 20.10.2015

Gutachter:

Prof. Dr.-Ing. Klaus-Dietz Tönnies

Prof. Dr.-Ing. Bernd Stock

Prof. Dr.-Ing. Dietrich Paulus

Danksagung

Zusammenfassung

Das Interesse an der automatisierten Vermessung von Holzstapeln, sogenannten Holzpoltern, hat im letzten Jahrzehnt bedeutend zugenommen. Dies zeigt sich insbesondere im stark angestiegenen Bedarf der Holzwirtschaft an fotobasierten Baumstammzählungs- und Vermessungsmethoden. Infolge eines vergleichsweise geringen Grades an Automatisierung in diesem Industriezweig existiert ein hohes Potenzial für technologischen Fortschritt, was Forschung und neue Ideen in diesem Bereich besonders interessant macht. Obwohl bereits einige Ansätze präsentiert wurden, die die Erkennung und Segmentierung von Baumstämmen zum Thema haben, werden viele wichtige Fragestellungen nicht berücksichtigt. Die meisten Methoden funktionieren nur unter Laborbedingungen und es findet keine Untersuchung unter realen Bedingungen statt. Die eigentliche Vermessung wird selten thematisiert und es erfolgt keine ausreichende geometrische Beschreibung von Holzpoltern. Darüber hinaus werden oftmals lediglich Einzelbilder für die Auswertung herangezogen, was bei einer praktischen Anwendung nicht realisierbar ist, da Holzpolter normalerweise zu groß sind, um von einem einzigen Bild erfasst zu werden. Diese Schwächen werden in Ansätzen sichtbar, die darauf abzielen, Teilprobleme zu lösen, aber in der Praxis versagen und keine Aussage über die Holzpoltervermessung als Mittel zur Volumenbestimmung machen.

Die hier vorgelegte Dissertation vereint neue Lösungen für eine Reihe von Unterproblemen zu einem robusten photogrammetrischen Holzpoltervermessungsansatz. Um diese Zielstellung zu erfüllen, werden Holzpolter als geometrische Metaobjekte beschrieben. Relevante Ansätze in der Literatur werden aufgearbeitet und erweitert. Es wird gezeigt, auf welche Weise Baumstammerkennung als fundamentales Problem nicht nur effizient sondern auch mit hoher Genauigkeit und Wiederholbarkeit gelöst werden kann. Basierend auf einem neuen Algorithmus für planare Panoramabildverarbeitung wird ein Ansatz zur zweidimensionalen Vermessung vorgestellt. Alternativ zur objektbasierten Vermessung wird eine Evaluierung einer Methode zur Segmentierung der Poltervorderfläche durchgeführt. Das Herzstück dieser Dissertation ist eine photogrammetrische Vermessungsmethode, welche auf Multiperspektiven-Rekonstruktion und Quadrikenfilterung basiert. Im Vergleich zum zweidimensionalen Ansatz wird erkennbar, dass der Multiperspektivenansatz eindeutige Vorteile aufweist. Alle 354 Testdatensätze mit 7655 Bildern wurden öffentlich verfügbar gemacht und die Erstresultate für die vorgeschlagenen Benchmarks werden angegeben. Die präsentierten Methoden sind auf handelsüblichen Smartphones und Tablets lauffähig. Zusätzlich wird die Performanz der photogrammetrischen Methode für diese Art von Gerät quantitativ evaluiert.

Abstract

The interest in automatic wood pile surveying has significantly increased in the last decade. This manifests itself especially in the timber industry's strong need for image-based wood log counting and surveying methods. Due to a comparatively low degree of automation in this industrial area, a high potential for technological advances makes research and novel ideas in this area especially interesting. Although some approaches which address wood log recognition and segmentation have already been presented, many important issues have not been considered. Most methods only work under ideal laboratory conditions and do not analyze performance under real world conditions. When using such methods, actual surveying is rarely addressed and no sufficient geometric description of the wood pile is given. Additionally, only single images are used for evaluation, which is not feasible in practice, since wood piles are usually too large to be captured by a single image. These deficiencies become apparent in approaches which aim to solve individual problems, but fail in practice and do not address wood pile surveying as a means of dermining the wood volume.

This dissertation unifies novel solutions for a number of subproblems associated with a robust photogrammetric wood pile surveying approach. To achieve this objective, wood piles are described as geometric meta-objects. Relevant approaches in literature are reviewed and extended. It is shown how wood log recognition as a fundamental problem can not only be solved efficiently, but also with high accuracy and repeatability. Based on a novel planar panoramic imaging algorithm, a two-dimensional surveying method is proposed. As an alternative to object detection-based surveying, a method for the segmentation of the wood pile front surface as a quasi-planar area is evaluated. The centerpiece of this dissertation is a photogrammetric surveying approach, which builds on multiple view reconstruction and quadric filtering. When compared to the two-dimensional approach, it becomes clear that using a multiple viewpoint algorithm yields significant advantages. All 354 test datasets with 7655 images were made publicly available and initial results for the defined benchmarks are given. The proposed methods are all capable of running on commercially available consumer-level smartphones and tablets. Additionally, the photogrammetric method's performance is quantitatively assessed for these types of devices.

Ehrenerklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; verwendete fremde und eigene Quellen sind als solche kenntlich gemacht. Insbesondere habe ich nicht die Hilfe eines kommerziellen Promotionsberaters in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Ich habe insbesondere nicht wissentlich:

- Ergebnisse erfunden oder widersprüchliche Ergebnisse verschwiegen,
- statistische Verfahren absichtlich missbraucht, um Daten in ungerechtfertigter Weise zu interpretieren,
- fremde Ergebnisse oder Veröffentlichungen plagiiert,
- fremde Forschungsergebnisse verzerrt wiedergegeben.

Mir ist bekannt, dass Verstöße gegen das Urheberrecht Unterlassungs- und Schadensersatzansprüche des Urhebers sowie eine strafrechtliche Ahndung durch die Strafverfolgungsbehörden begründen kann. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form als Dissertation eingereicht und ist als Ganzes auch noch nicht veröffentlicht.

Magdeburg, den 20.10.2015

Christopher Herbon

Contents

Chapter 1

Introduction

This introduction provides an outline of the research presented in this thesis. The goals and objectives are formulated based on the academic motivation for this dissertation. Several prior publications serve as the foundation for the proposed methods and the experiments conducted. Connections between different partial problems are drawn, to show how these parts can be combined to establish sophisticated wood pile surveying methods.

1.1 Motivation

A wood pile is a special case of an object cluster. Many such clusters can be found in nature and in artificial environments. An object cluster typically consists of objects with similar properties which are subject to defined spatial constraints. Examples include cells in microscopic scenes, piles of fruit (e.g. apples or potatoes), mussels, fish, and many more [GUTZEIT and LUKAS, 2013]. While this thesis focuses on the detection, segmentation, and surveying of wood piles, applicability to other areas is always retained and it is shown that many of the novel methods described in this thesis can be applied to a variety of other problems.

Wood pile surveying has been a research topic for many years. A large number of patents has been filed in this area and many scientific contributions have been published in the last two decades. Due to high costs for the measurement of roundwood and recent advances in research, photo-optical surveying continually gains importance in the forestry industry. The logging of roundwood in the European Union (EU-27) increased from 392 million cubic meters (m^3) in 2009 [EUROPEAN COMMISSION and EUROSTAT, 2012] to an estimated 433 million m^3 in 2011 [EUROPEAN COMMISSION and EUROSTAT, 2013] (average 2004-2011: 423 million m^3). Germany alone accounted for 56 million m^3 in 2011 and 53 million m^3 in 2013 [EUROPEAN COMMISSION and EUROSTAT, 2013; STATISTISCHES BUNDESAMT, 2014]. The area covered by trees in Germany is approximately 32%, which corresponds to 112,000 km^2

Figure 1.1: Example of a wood pile under ideal conditions

[Bundesministerium für Ernährung und Landwirtschaft (BMEL), 2014].

In order to optimize the processing, dispatching, and distribution of roundwood in an economical fashion, it is important to accurately measure wood piles on site. Manual methods for wood pile surveying are time-consuming and thus expensive. In Germany, 38,900 AWU (annual work units) were used in forestry and logging [European Commission and Eurostat, 2013], which is the equivalent of 38,900 people working full-time. Even though forest administrations are mostly governmental organizations, they are required to operate in a profit-oriented manner. For this reason, the efficiency of all workflows is subject to continuous optimization.

Photo-optical surveying methods have long been known to exhibit significant potential for time savings in comparison to manual surveying. The time for measuring the solid wood volume of a wood pile is usually between 30 minutes and several hours, depending on the size of the wood pile. Automatic methods have the potential of reducing this time to a few minutes or even seconds. Apart from the measurement itself, automatic procedures also export digital results, which are beneficial for the traceability and transparency of wood log logistics. Current computer-aided methods for wood pile surveying are solely based on desktop computers or cloud servers. In many remote, rural areas there is no mobile internet connection and it is not always feasible to carry a desktop or laptop computer. The processing power of handheld devices, such as smartphones and tablets, has increased significantly over the last years. Although the computational power of desktop computers is still higher, many computer vision tasks can now be solved quickly on handheld devices. Computational efficiency and a

small size combined with good portability make such devices predestined for off-road surveying methods. For this reason, this thesis will focus on the applicability of handheld devices to wood pile surveying.

1.2   Objective and contributions

The motivation for conducting the research in this thesis is used to outline the requirements for the proposed methods and accompanying experiments. The objective of this dissertation is to answer the question:

Can photogrammetric surveying of wood piles be performed on a handheld device?

While it has already been established that the general principles for recognizing clustered objects can be applied to wood logs [GUTZEIT and LUKAS, 2013], no research has so far addressed the problem of on-site wood pile surveying with handheld devices. It is thus unknown if such methods are feasible and if a sufficient accuracy can be achieved. The goal of this dissertation is to investigate the applicability of mobile photogrammetric measurement methods for wood pile surveying. The contributions obtained as a result of such investigations can be summarized as follows:

Contributions

1. A new method for adaptive image stitching is proposed, with which a panoramic image of a wood pile front surface or any other planar surface can be generated.
2. A novel state of the art image-based wood log detection method is presented.
3. This dissertation is the first to contribute algorithms for image-based wood pile model generation.
4. The novel image stitching and wood log recognition methods are combined for a new two-dimensional wood pile surveying method.
5. A recognition-independent method for the segmentation of the wood pile front surface or any other quasi-planar surface is presented, which is based on image registration and block matching.
6. Three-dimensional photogrammetric surveying is achieved through a novel multiple view reconstruction and image-based wood log recognition approach.
7. Different scale estimation techniques for the surveying object are discussed and quantitatively assessed.
8. The *HAWKwood* database, consisting of 7655 real, synthetic, and ground truth images, is provided free of charge for academic use.

It can be seen from the contributions that the objective of this dissertation is to establish different wood pile surveying methods and to compare the results to manually- and automatically-obtained ground truth. The first method works in 2D space, by stitching multiple images and performing measurements based on object detection and segmentation. The second method operates in 3D space. It combines multiple view reconstruction with object detection, reconstruction, and surveying methods. Both approaches need to be optimized for handheld devices with regards to computational efficiency and memory usage. A special focus lies in the applicability of the methods to real world scenarios, a problem which is rarely addressed by state-of-the-art algorithms. The results of both methods will be compared and it will be proven that, indeed, both approaches work robustly and efficiently on handheld devices.

To achieve these objectives, state-of-the-art wood log detection and segmentation methods are extended, significantly improved, and integrated into the reconstruction approaches presented in this thesis. Many of the existing methods only work under laboratory conditions [FINK, 2004] or impose very strong constraints on the geometry of the wood pile and the conditions under which it is photographed [GUTZEIT et al., 2011; GUTZEIT and VOSKAMP, 2012]. The presented approaches are optimized to work on a handheld device, under real world conditions, and to be used with very few constraints on the image acquisition process. Smartphones and tablets often only provide limited memory resources and computational power (which is strongly linked to battery drainage). It is therefore crucial to optimize all algorithms to consider these constraints.

For the photogrammetric surveying methods to work robustly, multiple partial problems must be solved. The most substantial of these problems is the image-based detection and segmentation of wood log faces. Many of the proposed methods are built on the assumption that wood logs can be detected with sufficient accuracy in images, so that these detection results can be used as cues for further processing. It is therefore shown how wood log faces can be detected and which parametrization should be chosen for the detection process in order to optimize the true and false positive rate. A parametrizable model yields significant advantages with regards to different applications. For example, in the case of multiple view wood pile reconstruction it may be desirable to work with a higher true positive rate (and thus an inherently higher false positive rate), as outliers can be filtered through geometric constraints. On the other hand, the application to panoramic images requires a medium to low false positive rate.

Both the two-dimensional and the three-dimensional surveying method will be extensively evaluated. In order to establish a framework for reproducible testing, a large number of test cases are defined and the results are compared for the three most important measurement parameters: the number of wood logs, the solid wood volume, and the contour volume. In

most cases the wood pile width is used as the scale reference. Therefore, the accuracy of the wood pile width measurement is also subject to an evaluation and it is shown that it is the most accurately measurable scale reference.

1.3 Outline of this thesis

In this section, the structure of this dissertation is briefly outlined to give an overview of the presented methods, evaluation methodologies, and findings. Following this introductory chapter one, the case study *wood pile surveying* will be examined in chapter two. The motivation for choosing mobile photogrammetric wood pile surveying as a research field is discussed in depth. Scientific, economic, and environmental challenges, which have a significant impact on the scope of this thesis as well as the scope of the discussed methods, are considered and the objective is concretized.

To establish logically and mathematically adequate models, the wood pile as a measurement object must be well defined. It will be established that a wood pile is considered to be a meta-object, consisting of single wood logs, which can be described by a set of parameters used for surveying. In order to compare the obtained results to ground truth data and manual measurements, current manual and automatic wood pile surveying methods are discussed and compared. Each of these methods features advantages and disadvantages, which mostly consist of time constraints and measurement errors. To minimize errors, the environmental properties which can lead to measurement errors are further explored. For an extensive evaluation of the proposed methods, the large, publicly available *HAWKwood* database is introduced. Concluding chapter two, related applications, to which the presented algorithms are relevant, will be discussed.

The *Related work* chapter reviews relevant computer vision foundations, the current state-of-the-art wood pile surveying methods, and gives an overview of national and international patents. Computer vision foundations and state-of-the-art methods include image-based object recognition, segmentation, panoramic image stitching, multiple view reconstruction, and structure from motion. The proposed algorithms are compared to current methods and the scope of this thesis is placed in the context of related algorithms and methods.

Chapter four describes the basis for the 2D surveying method proposed in this thesis. It takes as input a set of overlapping images and produces a panoramic image through a novel image stitching technique. The images are weakly constrained to partly depict a planar surface. Subsequently, the aim is to detect and segment objects (wood log faces) in single and stitched images, which is discussed and evaluated in chapter five.

As an alternative step, a novel segmentation procedure for the front surface of a wood pile is presented in chapter six. This method has the advantage of being independent of the

detection of single wood logs, as it directly segments the wood pile front surface as a whole.

This thesis aims to provide a three-dimensional wood pile surveying method in chapter seven (*Three-dimensional surveying*). The proposed approach relies on the research results of the preceding chapters. Detected wood log faces are used as cues for the 3D reconstruction of the wood pile. Following the generation of the 3D model, different parameters are measured and quantitatively evaluated, such as the detected number of wood logs, the solid wood volume, and the contour volume. At the end of chapter seven the use of handheld devices is examined.

The *Conclusion* chapter reviews the contributions of this thesis in light of the obtained results, while future work in wood pile surveying and further applications of the proposed algorithms are outlined.

Chapter 2

Wood pile surveying as a case study

This chapter introduces the case study of this thesis. Wood pile surveying has been done for hundreds of years, with the first single log measurement techniques having been applied in the 18th century [KRAMER and AKÇA, 2008, p. 1; PRODAN, 1965, p. 3]. Kramer and Akça [KRAMER and AKÇA, 2008, pp. 1–2] portray how statistical methods and mechanical measurement instruments gained importance during the 18th century. The development of measurement techniques focused especially on measuring the diameter of individual wood logs, which is important when surveying single logs either individually or inside a wood pile. In Germany, the surveying of wood piles was first standardized by the *HOMA Holzmeßanweisung : Bestimmungen über die Ausformung, Messung und Sortenbildung des Holzes in den deutschen Forsten vom 1. April 1936* [HOMA, 1936] and later replaced by the *Verordnung über gesetzliche Handelsklassen für Rohholz vom 31. Juli 1969* [BUNDESMINISTERIUM FÜR ERNÄHRUNG, LANDWIRTSCHAFT UND FORSTEN, 1969], where surveying with several types of measurement procedures is described for different types of wood piles. In 2015, the new agreement for the roundwood industry called *Rahmenvereinbarung für den Rohholzhandel in Deutschland (RVR)* [DEUTSCHER FORSTWIRTSCHAFTSRAT E.V. UND DEUTSCHER HOLZWIRTSCHAFTSRAT E.V., 2014], which mostly conforms to the *HKS*, came into effect. The sorting of wood and classification of wood pile types [FROMMHOLD, 2013] will be discussed in depth in Section 2.5 (*Manual surveying methods*). Furthermore, this chapter includes a view of the environmental constraints, research challenges, geometric constraints of a wood pile, economic aspects of wood pile surveying, error estimations for different surveying techniques, and a detailed description of the data basis which is used in this thesis.

2.1 Environmental constraints

For the work presented in this dissertation, the applicability to real world scenarios under authentic conditions is emphasized. For this reason, there exists a set of constraints which

must be considered during the research and implementation of the methods discussed below.

> **Constraints**
>
> 1. All methods must be capable of operating on a consumer-level smartphone or tablet with a single monocular camera.
> 2. No additional hardware may be used, except for forestry rangers' standard equipment.
> 3. An internet connection must not be required.
> 4. German standards for the felling, piling, and selection of tree species can be presumed.

The first constraint limits the usable third-party software, as not all algorithms are implemented as multi-platform open source software. Some algorithms which can by applied to photogrammetric wood pile surveying (e.g. [FURUKAWA and PONCE, 2010; SNAVELY et al., 2006; SNAVELY et al., 2008]) can only be used on desktop PCs with a specific CPU platform (e.g. Intel) or are implemented for a single operating system (OS) only. While the portation to a mobile platform seems possible in general, it is not feasible in the scope of this thesis. Apart from computational time constraints, the results without such software exhibit a similar accuracy because the proposed methods are specifically optimized to work with compatible software frameworks.

For scientific purposes, the accuracy of wood pile surveying in a laboratory environment has been evaluated and discussed, e.g., in [FINK, 2004]. This thesis focuses on methods which can be applied in a rural, rough terrain environment. While it is surely possible to use large equipment such as laser scanners [NYLINDER et al., 2008] or car mounted stereo cameras [DRALLE A/S, 2014] for scientific purposes or in wood processing plants, such measurement instruments do not meet the portability requirements for real world use. Dietz [DIETZ, 1985] emphasizes that the rough conditions of forestry terrain must be considered in timber harvesting, as costs increase with terrain roughness. Thus tools aiding the surveying process are limited to instruments that forestry rangers can easily carry. This includes measurement tape, which will be used as a scale reference in some of the proposed methods.

Even in many technologically advanced countries, not all areas are covered by mobile internet and mobile broadband internet is often only available in urban areas [STIFTUNG WARENTEST, 2011]. The proposed methods must therefore be able to process data on-site. As with all algorithmic measurement techniques, a measurement might have to be repeated in some cases, e.g., when the image acquisition produces unusable images due to bad lighting conditions or motion blur. A high-performance cloud server could be used when processing data off-site, as proposed, e.g., by [JØRGENSEN and KRISTIANSEN, 2008], but a successfully-processed measurement cannot be guaranteed. An on-site measurement method that does not

require an internet connection thus yields significant advantages for real world applications.

Due to the large variety in standards for wood piling and the hundreds of existing tree species around the world, this thesis focuses on German standards and conventions. When the directions for wood piling are known, constraints which represent the known geometric properties of a wood pile can be incorporated into the algorithms. The constraints for automatic surveying correspond to the constraints that are stipulated for manual surveying [BUNDESMINISTERIUM FÜR ERNÄHRUNG, LANDWIRTSCHAFT UND FORSTEN, 1969; DEUTSCHER FORSTWIRTSCHAFTSRAT E.V. UND DEUTSCHER HOLZWIRTSCHAFTSRAT E.V., 2014]. As a consequence, manual and automatic methods can be compared for the same data.

2.2   Research challenges

At first glance, wood pile surveying seems like a simple problem. One may prematurely come to the conclusion that this problem could be solved by using simple thresholding and morphological operations. In some specific cases and under laboratory conditions, this is indeed possible. For real world applications, which take place outdoors in a rough terrain environment under varying light conditions, methods with higher robustness must be employed. During the research for this thesis, a number of challenges have arisen, which will be discussed in this section. Figure 2.1 shows a generic photo-optical surveying method, as used e.g. in [HERBON et al., 2015a], which will be used to demonstrate the specific difficulties of each step.

Figure 2.1: A generic surveying method with intermediate steps. Most wood pile surveying methods perform a similar process.
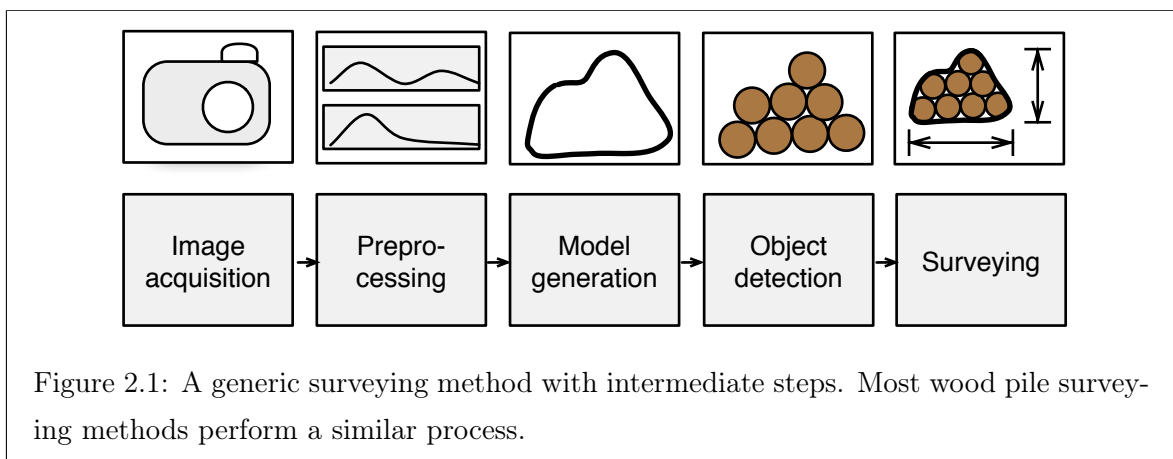
Image acquisition   During image acquisition, many environmental factors must be considered, the most problematic of which are changes in illumination. Especially in an environment with many trees occluding the sky, cast shadows pose a challenge. Direct sunlight can lead to blooming artifacts, which Langford et al. [LANGFORD et al., 2012] define as "[...] halos or streaks recorded around images of bright light sources or other intense highlights." Figure 2.2

Figure 2.2: Left and center: cast shadows on a wood pile front surface. Right: direct sunlight and blooming.

Figure 2.3: Illumination change in adjacent images due to automatic illumination adjustment by the digital camera. Well-exposed regions in the left image are underexposed in the right image, while the left image suffers from overexposed areas.

shows some examples of these phenomena. The combination of bright and dark regions makes it difficult to perform thresholding operations. A camera's autofocus must consider a larger focus area, otherwise over- or underexposed regions may occur. Direct sunlight is problematic in all image acquisition scenarios and should be avoided if possible.

When taking multiple overlapping pictures, illumination changes between images are problematic for feature matching. Many feature descriptors are invariant to illumination changes up to a certain point, e.g. *SIFT* [BROWN and LOWE, 2007; LOWE, 1999; LOWE, 2001; LOWE, 2004; MIKOLAJCZYK and SCHMID, 2005; MORENO et al., 2009], but when an image exhibits large over- or underexposed regions, even such invariant features fail. It can be seen in Figure 2.3 that shadows lead to exposure problems, which prevent feature detection and matching algorithms from working as intended.

Preprocessing   Following the acquisition process, the input images can be preprocessed for further usage. Such preprocessing methods include histogram-based operations and scaling. The scaling of the images is carried out for performance reasons. Mobile devices in particular provide limited memory, which quickly becomes insufficient when dealing with a large number of images. Since the actual wood log size is unknown, as is the distance from the camera to the wood pile, a sufficiently large image size must be chosen in order to retain detectability and segmentability of the wood log faces.

Model generation   The model generation step of the surveying pipeline can be implemented through different approaches. As discussed in the introduction, a model can be generated in 2D (e.g. through image stitching) or 3D (e.g. through structure from motion). Both approaches impose different constraints on the scene.

Panoramic images rely on the assumption of a planar scene, in other words a scene which only consists of a two-dimensional surface, such as a wall, a floor, or a table [CAPEL and ZISSERMAN, 1998; SZELISKI, 1996; SZELISKI and SHUM, 1997; ZOGHLAMI et al., 1997], is required. When performing rotational image stitching about the camera's optical center, the scene is a special case of a planar surface, as it can be described by a sphere [BROWN and LOWE, 2007; CHEN and KLETTE, 1999; SHUM and SZELISKI, 2000; SZELISKI, 2006; SZELISKI and SHUM, 1997]. In [HERBON et al., 2014b] we have presented a hybrid approach, which is able to automatically detect the panorama type, rotational or planar, from a set of images and perform adaptive stitching. When handling images of wood piles, the constraint of a planar surface can be approximately satisfied in many cases. In Section 2.1 (*Environmental constraints*) environmental constraint #4 was introduced, through which it can be assumed that German standards for roundwood piling are met, such as the ones presented in [BAYRISCHE STAATSFORSTEN, 2013]. This is necessary in order for a stitching mechanism to work robustly. Nevertheless some errors are introduced when the wood pile front surface is not strictly planar. The impact of these errors will be discussed in Section 2.6 (*Measurement errors and error propagation*).

In contrast to panoramic image stitching, the scene must not be planar when performing 3D reconstruction through structure from motion, as it leads to the well-known planarity degeneracy [CHUM et al., 2005; DECKER et al., 2008]. Hartley and Zisserman [HARTLEY and ZISSERMAN, 2004, pp. 295–296] show that not only is it a degenerate configuration if all world points lie in a plane (degenerate structure), but also a purely rotational motion of the camera is a degenerate case (degenerate motion). Both of these scenarios are eligible for panoramic image stitching but must be avoided for 3D reconstruction. When reconstructing wood piles and especially wood pile front surfaces, planar degeneracy always poses a threat and should be avoided if possible. This can be achieved by including a larger foreground and

background, so that not all points lie in a single dominant plane. If an image does indeed depict large non-wood areas, it is all the more important for wood log detection methods to work robustly in terms of a low false positive rate.

Object detection Object detection in the scope of this thesis refers to the detection of roundwood log faces. As will be shown in Section 2.3.2 (*Attributes of wood piles*), a large number of tree species must be considered. In addition to the different tree species, object detection must be robust under different environmental conditions. The most important of these are:

1. snow
2. mud
3. leaves on the ground
4. decay of the wood
5. spray paint
6. tree branches covering the front surface

Figure 2.4 shows examples of environmental challenges and underlines why wood log surveying methods that work under laboratory conditions are not sufficient for real world usage. In addition to these examples, the illumination changes shown in Figures 2.2 and 2.3 make it difficult to establish a histogram-based model for wood log detection, as the color and brightness distribution is non-uniform, even for wood logs that are theoretically exactly alike. Some of the complications shown in Figure 2.4, such as vegetation and snow, can be manually removed prior to image acquisition. Although this is not feasible in many cases, it could improve detection results. On the other hand, removing mud or spray paint seems exhaustive under real world conditions and decay of the wood is impossible to exclude. Methods for wood log detection must thus take these environmental conditions under consideration and if possible account for these obstacles while retaining a high detection rate.

Surveying Although wood piles are usually separated by a few meters, in some cases piles are very close to each other or a pile is visible in the background, when processing another pile. When rainfalls occur during harvesting of the stems or during piling, stems might sink into the muddy ground. This makes surveying difficult, as the pile is only partly visible. The same holds true for leaves covering the front faces. To enable automatic surveying procedures, some constraints for piling are employed, which will be discussed in Section 2.3.3 (*Geometric constraints for roundwood piling*).

Figure 2.4: Examples for challenges in wood log detection. (a)-(c) different tree species, (d) mud, (e) leaves (f) decay, (g)-(h) spray paint, (i)-(j) snow, (k)-(l) vegetation.

2.3   Wood pile as a geometric meta-object

Historically, wood measurement was restricted to single roundwood logs exclusively. Only in the last century have advances been made to classify a wood pile as an object cluster consisting of individual logs. In Germany, the regulations for measurement procedures are subject to constant improvement [PLATTFORM FORST & HOLZ, 2012; PLATTFORM FORST & HOLZ, 2014]. In this section, the geometric properties of wood logs and wood piles will be investigated.

2.3.1   Attributes of roundwood logs

Parameters   A roundwood log as a geometric object provides a number of properties which are relevant for determining its merchantable volume and thus its value. West [WEST, 2009, pp. 11–15] defines these properties, of which the following are most important for the methods in this thesis:

Figure 2.5: (a) wood log face, (b) stems with approximately equal length

1. diameter

2. length

3. cross-sectional shape (form factor)

4. bark thickness

The stems that are considered in this thesis are always of a known length, called the ordered length (because this is the length that a buyer orders), as shown in Figure 2.5(b). West [WEST, 2009, p. 14] emphasizes, that while a tree cross-section in general differs from an ideal circle (Figure 2.5(a)), its average diameter nevertheless provides important information about the stem. It closely correlates with the wood volume and its weight and therefore gives information about its value. According to [KRAMER and AKÇA, 2008, pp. 35–39] and [WEST, 2009, pp. 14–15] deviation from the ideal circular form is caused by branches from the stem, wind, fire damage, disease, or insects. In some very rare cases, some extraordinary irregularities can be induced through trees being deformed by another tree or a solid object, when growing on steep slopes, or through odd branching [MATTHECK et al., 1991; WEST, 2009, pp. 14–15].

Diameter   Kramer and Akça [KRAMER and AKÇA, 2008, pp. 38], West [WEST, 2009, pp. 14–15], and Van Laar and Akça [VAN LAAR and AKÇA, 2007, p. 66] agree that despite the well-known irregularities, the measurement of a roundwood log can be accurately performed by measuring its (average) diameter and approximating the wood log face area as a circle. Husch et al. [HUSCH et al., 2002, p. 85] define the goal of the circular approximation as determining a circle with the same area as the area of the cross-section. The bias from this approximation is negligible [WEST, 2009, pp. 14–15]. Usually the diameter is measured at a single point for stems with a diameter of less than 20cm and at two orthogonal points for

larger stems [BUNDESMINISTERIUM FÜR ERNÄHRUNG, LANDWIRTSCHAFT UND FORSTEN, 1969]. Prodan [PRODAN, 1965, p. 84] describes that the repeatability accuracy for measuring the diameter is higher when measuring the circumference of a stem and deriving the diameter implicitly, a method which is difficult to use when wood logs are stacked.

Stem volume   There exist several mathematical models for stem forms and volumes of entire trees. The most accurate measurement instrument is the xylometer, which is based on Archimedes' principle of filling a tank with water and measuring the water's displacement [VAN LAAR and AKÇA, 2007, p. 86]. As this procedure is expensive in practice, a less accurate but more time-efficient method can be applied, which consists of dividing a stem into sections and calculating the volume of each section based on the diameter at its midpoint. For relative sections (the number of sections is an integer), Equation 2.1 shows the computation of the stem volume V_{stem}, where l_{s} is the constant log section length and d_i is the mid point diameter [KRAMER and AKÇA, 2008, p. 36; VAN LAAR and AKÇA, 2007, p. 87].

$$V_{\text{stem}} = \frac{\pi}{4} l_{\text{s}} \sum_{i=1}^{k} d_i^2 = \frac{\pi}{4} l_{\text{s}} \left(d_1^2 + d_2^2 + d_2^2 + \cdots + d_k^2 \right) \tag{2.1}$$

When dividing a stem into sections of a fixed length (absolute sections), the following formula applies, where, in addition to the parameters of Equation 2.1, l_{t} is the length and d_{t} is the midpoint diameter of the top section [VAN LAAR and AKÇA, 2007, p. 87].

$$V_{\text{stem}} = \frac{\pi}{4} l_{\text{s}} \sum_{i=1}^{k-1} d_i^2 + \frac{\pi}{4} l_{\text{t}} d_{\text{t}}^2 = \frac{\pi}{4} l_{\text{s}} \left(d_1^2 + d_2^2 + d_2^2 + \cdots + d_{k-1}^2 \right) + \frac{\pi}{4} l_{\text{t}} d_{\text{t}}^2 \tag{2.2}$$

Additional formulae have been suggested to simplify the volume computation of roundwood logs without performing sectioning. West [WEST, 2009, p. 26] names the following, of which Huber's formula (Equation 2.3) is the simplest and the most commonly used [FINK, 2004]. g_{m} is the midpoint cross-section area and g_{u} and g_{t} are the cross-section areas at the lower and upper end. The Smalian formula (Equation 2.4) also takes cases into account, where the diameters at the end points of the wood log are significantly larger or smaller than the midpoint diameter.

$$V = g_{\text{m}} l \tag{2.3}$$

$$V = \frac{g_{\text{l}} + g_{\text{u}}}{2} l \tag{2.4}$$

Bark   In most cases, the merchantable volume does not include bark. The bark thickness is different for various types of wood. Depending on the wood type, either the diameter is reduced by a certain absolute amount or a specific percentage of the volume is subtracted

to account for the bark [PRODAN, 1965, pp. 105–106]. For the methods presented in this thesis, all measurements are performed by including bark. This way, a better comparability is ensured, while bark subtraction can be performed afterwards if needed.

2.3.2   Attributes of wood piles

A wood pile consists of wood logs of the same length, wood quality, and, most of the time, the same wood type. Since it is a cluster of similar objects, we consider it to be a meta-object [HERBON et al., 2015a]. By choosing this representation, additional information apart from the roundwood logs can be included, such as the space between wood logs, information about the ground surface and the background, as well as about adjacent wood piles. We showed in [HERBON et al., 2015a] that the most important parameters of a wood pile, which concern wood pile survey are the following:

1. defined wood log length l
2. tree species
3. wood quality
4. number of wood logs N
5. solid wood volume V_s
6. wood pile contour volume V_c

The first three properties can be easily determined by a forestry ranger through visual inspection. They do not require special focus during the surveying process. Nevertheless, these parameters are important in determining the sales price.

Defined wood log length   The length of the roundwood log is not measured, as the actual log is usually longer than the defined length [KRAMER and AKÇA, 2008, p. 38; PRODAN, 1965, p. 78; BUNDESMINISTERIUM FÜR ERNÄHRUNG, LANDWIRTSCHAFT UND FORSTEN, 1969]. The reason for this is that a certain minimum length is required for wood processing, which can only be guaranteed if the actual length is higher. For volume computation in manual and automatic methods, the defined wood log length, denoted as l, is used rather than the actual length.

Tree species   51 different tree species exist in Germany. The four most common tree species, which make up 73% of all trees, are the following in descending order: [BUNDESMINISTERIUM FÜR ERNÄHRUNG UND LANDWIRTSCHAFT (BMEL), 2014]

1. spruce
2. pine

Figure 2.6: Panoramic image of a wood pile. The logs have been counted by a forestry ranger by using red spray paint. The number of logs ($N = 214$) is indicated on the front face of a wood log.

3. beech

4. oak

Considering all possible German tree species implies a number of research challenges and additional constraints, which are discussed in Section 2.2 (*Research challenges*).

Number of wood logs N The number of wood logs is a well defined integer quantity. Figure 2.6 shows a panoramic image of a wood pile, where the number of wood logs has been manually determined by a forestry ranger. A common procedure is to mark all counted logs with spray paint; usually a dot is made, but sometimes small lines are used. The result of the counting procedure is then indicated on one of the wood log faces. The next section will show that the manual determination of the number of wood logs is not impeccable and that it can benefit from automatic procedures.

Solid wood volume V_{s} The solid wood volume V_{s} is defined as the sum of the individual wood log volumes. As demonstrated in Section 2.3.1 (*Attributes of roundwood logs*) the volume of a single log is approximated by its midpoint diameter and the Huber's formula, as defined in the HKS [Bundesministerium für Ernährung, Landwirtschaft und Forsten, 1969]. Figure 2.7 shows the digitally marked front side diameters (green circles), which can be used for the approximation of the solid wood volume.

Contour volume V_{c} The solid wood volume is sometimes not of interest because the wood quality is low, so-called industrial wood (see next paragraph), or the wood logs are very small, which would make the measurement disproportionately time-consuming [Ministerium für Umwelt, Forsten und Verbraucherschutz Rheinland-Pfalz, 2013]. In such cases, the contour volume V_{c} is calculated and the solid wood volume is derived by multiplication

Figure 2.7: Solid wood volume V_s measured by accumulating the diameters (green). The cumulated area of the log faces A_s is multiplied with the defined length l to obtain the solid wood volume V_s.

Figure 2.8: Contour volume V_c measured through a concave hull (green). The front area of the wood pile A_s is multiplied with the defined length l to obtain the contour volume V_c.

with a defined factor called the stacking coefficient [KNYAZ and MAKSIMOV, 2014; MINISTERIUM FÜR UMWELT, FORSTEN UND VERBRAUCHERSCHUTZ RHEINLAND-PFALZ, 2013; RUNDHOLZSORTIERUNGSVORSCHRIFT, 1988]. In contrast to the solid wood volume, V_c includes the solid wood as well as the space between the wood logs (see Figure 2.8) [MINISTERIUM FÜR UMWELT, FORSTEN UND VERBRAUCHERSCHUTZ RHEINLAND-PFALZ, 2013; RUNDHOLZSORTIERUNGSVORSCHRIFT, 1988].

Wood quality According to the HKS [BUNDESMINISTERIUM FÜR ERNÄHRUNG, LANDWIRTSCHAFT UND FORSTEN, 1969] and the RVR [DEUTSCHER FORSTWIRTSCHAFTSRAT E.V. UND DEUTSCHER HOLZWIRTSCHAFTSRAT E.V., 2014] the quality of wood is defined by the four classes *A, B, C,* and *D*. Sauter et al. [SAUTER et al., 2012] define that class *A* indicates the best possible quality with very few defects and *D* is the class with the lowest sales prices, but which still contains at least 40% commercially usable wood. For wood that belongs to one of these four classes, the three aforementioned parameters, number of wood

logs N, wood pile contour volume V_c, and solid wood volume V_s, are applicable.

As previously mentioned, in some cases it is not desirable to measure each wood log individually. In addition to the four categories A through D, there exists a type of wood called industrial wood, which is usually mechanically or chemically processed [BUNDESMINISTERIUM FÜR ERNÄHRUNG, LANDWIRTSCHAFT UND FORSTEN, 1969; DEUTSCHER FORSTWIRTSCHAFTS-RAT E.V. UND DEUTSCHER HOLZWIRTSCHAFTSRAT E.V., 2014], e.g., for the production of paper. For industrial wood, only the contour volume and sometimes the number of wood logs are relevant, while the solid wood volume is not considered for the sales price [FINK, 2004].

2.3.3   Geometric constraints for roundwood piling

Based on Sections 2.2, 2.3.1, and 2.3.2 the constraints for roundwood piling that make automatic surveying possible can be identified:

1. Wood logs must be piled in a way so that the wood log faces align as much as possible.
2. Large vegetation, such as tree branches and tall grass, must be removed.
3. Spray paint can be used, but should not cover very large areas.
4. Wood log faces should be more than 75% visible.

Under these conditions, automatic surveying is possible. Obviously some of the deficiencies of the log faces, as shown in Figure 2.4, can cause automatic detection methods to fail. Model generation is still possible in such cases and surveying can be amended by user interaction.

2.4   Necessity for digital wood pile surveying

The main motivation for researching and developing automatic wood surveying methods is to save time and to process wood more efficiently. If the volume and the number of wood logs are known, the logistics for wood distribution can be planned accordingly, but the price for a wood pile can also be determined based on the outcome of the on-site measurement. Fink [FINK, 2004] summarizes, based on Maler [MALER, 1997], that on-site wood measurements are usually used for the determination of the sales price between the negotiating parties.

As previously stated, the manual measurement of wood logs and wood piles is a time-consuming task. According to Dietz [DIETZ, 1985] the costs for harvesting and piling wood are on average 31.50 DM (16.11 €) per solid cubic meter without bark. Schöttle et al. [SCHÖTTLE et al., 1999] report costs of 48.50 DM (24.80 €) per solid cubic meter for harvesting and piling. The piled wood must then be manually surveyed, which costs between 15 DM and 20 DM (7.67 € to 10.23 €) per solid cubic meter [FINK, 2004; GUGLHÖR, 1994; WEGELAAR, 1997] and thus accounts for a significant part of the overall costs. This can potentially be reduced through automatic or half-automatic surveying methods [FINK, 2004].

Apart from being time-consuming, manual surveying methods have also been proven to be error prone. Bort et al. [BORT et al., 1989] report that in their measurement the number of wood logs in a pile was underestimated by 0.3% to 7.1% (average: 0.5%), which linearly reduces the value of the wood pile. In practice, not all wood logs of a wood pile are measured, but rather a subset, which is defined as a representative sample [BORT et al., 1989] (see Section 2.5 *Manual surveying methods*). According to Bort et al. [BORT et al., 1989] the difference between a full measurement and the measurement of a meaningful sample was between -1.4% and +3.7% with an average of +1.7%. For an accurate surveying method it therefore seems preferable to measure all wood logs instead of a subset, but manually measuring all individual wood logs is not feasible in many cases [MINISTERIUM FÜR UMWELT, FORSTEN UND VERBRAUCHERSCHUTZ RHEINLAND-PFALZ, 2013]. Automatic procedures can potentially solve this problem and the methods proposed in this thesis aim to provide sufficient accuracy for real world applications.

2.5 Manual surveying methods

Fink [FINK, 2004] gives an overview of manual surveying methods that are applied by German forestry rangers. These are the complete midpoint diameter inventory (CMDI), the line sampling method, the free sampling method, the contour sampling method, and the contour volume method. In this thesis, only the complete midpoint diameter inventory and the section volume method are considered, because methods which only use a subset of wood logs are not sufficiently accurate as a reference.

Complete midpoint diameter inventory (CMDI) The most time-consuming procedure is the complete inventory of all wood logs, where the diameter of all wood logs is measured before or after piling. In Figure 2.9(a) all diameters (in [cm]) were written on the wood log faces. The solid wood volume V_{s} is then determined by accumulating all wood log volumes. We have explicitly defined this formula (shown in Equation 2.5) in [HERBON et al., 2015a], where N is the number of wood logs, d_i is the diameter of the ith log, and l is the defined wood log length. A visual representation is depicted in Figure 2.7.

$$V_{\mathrm{s}} = \sum_{i=1}^{N} \left(\frac{d_i}{2} \right)^2 \pi l \tag{2.5}$$

The *HKS* [BUNDESMINISTERIUM FÜR ERNÄHRUNG, LANDWIRTSCHAFT UND FORSTEN, 1969] and its successor the *RVR* [DEUTSCHER FORSTWIRTSCHAFTSRAT E.V. UND DEUTSCHER HOLZWIRTSCHAFTSRAT E.V., 2014] require the diameter to be measured at midpoint. What seems problematic here is that the midpoint diameter can only be determined before piling, because most midpoints are occluded by stems lying on top. The midpoint diameter

Figure 2.9: Manual surveying methods. (a) Complete midpoint diameter inventory. All diameters in [cm] are written on the wood log faces. (b) Section volume method for industrial wood. Sections of equal width are used to determine the average height of the wood pile and the height in [m] is indicated with spray paint.

is necessary for the application of Huber's formula (Equation 2.3), while Smalian's formula (Equation 2.4) only requires the diameters at the upper and lower end. Although it has been proven by de Leóna and Uranga-Valenciaa [DE LEÓNA and URANGA-VALENCIAA, 2013] that the error from applying Huber's formula is in fact smaller than the error induced through Smalian's formula [DE LEÓNA and URANGA-VALENCIAA, 2013], a midpoint diameter cannot be practicably obtained for already piled wood logs. The solid wood volume can be defined by Smalian's formula as shown in Equation 2.6, where the midpoint diameter is calculated as the average of the upper $(d_{u,i})$ and the lower diameter $(d_{l,i})$ of the ith log.

$$V_{s} = \sum_{i=1}^{N} \left(\frac{d_{u,i} + d_{l,i}}{4} \right)^{2} \pi l \tag{2.6}$$

Calculating the midpoint diameter first and deriving the solid wood volume from it is equivalent to calculating the solid wood volume through the diameters on the front and on the back of the pile and then averaging the results. This procedure is often used in practice. The formula for this method is shown by Equation 2.7, where $d_{f,i}$ and $d_{b,i}$ are the diameters of the ith wood log on the front and the back of the pile.

$$V_{s} = \frac{1}{2} \left(V_{s,f} + V_{s,b} \right) l = \frac{1}{2} \left(\sum_{i=1}^{N} \left(\frac{d_{f,i}}{2} \right)^{2} + \sum_{i=1}^{N} \left(\frac{d_{b,i}}{2} \right)^{2} \right) \pi l \tag{2.7}$$

For wood logs with a diameter larger than or equal to 20cm the standards [Bundesmin-
isterium für Ernährung, Landwirtschaft und Forsten, 1969; Deutscher Forst-
wirtschaftsrat e.V. und Deutscher Holzwirtschaftsrat e.V., 2014] require two
measurements, orthogonal to each other, which are then averaged. This way, deviations from
the assumed circular model can be accounted for. Wood logs with a diameter smaller than
20cm only have to be measured once, since it is generally assumed that small wood logs do not
differ from the circular model as much as large wood logs do. Figure 2.10 shows how the di-
ameters on the front face should be measured in accordance with [Bundesministerium für
Ernährung, Landwirtschaft und Forsten, 1969; Deutscher Forstwirtschafts-
rat e.V. und Deutscher Holzwirtschaftsrat e.V., 2014].

Figure 2.10: Orthogonal diameter measurement. (a) Pure horizontal measurement of the
diameter for wood logs with $d <$20cm, (b) orthogonal measurements for wood logs with
$d \geq$ 20cm. Image source: *HAWKwood* [Herbon, 2014b]

Section volume method   The section volume method is mainly used for industrial wood
but sometimes it is used in addition to the complete midpoint diameter inventory [Fink, 2004].
The cost and time consumption for the section volume method is estimated by Fink [Fink,
2004] to be lowest of all methods. It is generally only capable of determining the contour
volume V_c, but it is possible to derive the solid wood volume based on the tree species and
wood pile specific factors [Van Laar and Akça, 2007, p. 88; Kramer and Akça, 2008,
p. 39; Fink, 2004]. If the number of wood logs in the pile is known, an average log volume
and log diameter can be determined in addition to the contour volume. We have defined the
contour volume based on the section volume method in [Herbon et al., 2015a]. Equation
2.8 shows this definition with w_p being the wood pile width (which can be expressed as the
product of the equidistant section width w_s and the number of sections n_s) and h_i being the

measured height in the center of the ith section.

$$V_{\mathrm{c}} = w_{\mathrm{p}}l\frac{1}{n_{\mathrm{s}}}\sum_{i=1}^{n_s} h_i = w_{\mathrm{s}}ln_{\mathrm{s}}\frac{1}{n_{\mathrm{s}}}\sum_{i=1}^{n_{\mathrm{s}}} h_i = w_{\mathrm{s}}l\sum_{i=1}^{n_{\mathrm{s}}} h_i \qquad (2.8)$$

Figure 2.9(b) shows how the sections are marked on a wood pile and the height of each section (in [m]) is indicated with spray paint. The width of a section is defined in [Ministerium für Umwelt, Forsten und Verbraucherschutz Rheinland-Pfalz, 2013], where it varies between $w_{\mathrm{s}} = 1$m for $w_{\mathrm{p}} < 10$m and $w_{\mathrm{s}} = 16$m for $w_{\mathrm{p}} > 200$m. The choice of a large w_{s} seems problematic for accurate surveying of ground truth data in this thesis. Therefore all contour volumes have been measured with a section width smaller than the one proposed by [Ministerium für Umwelt, Forsten und Verbraucherschutz Rheinland-Pfalz, 2013], usually $w_{\mathrm{s}} = 0.25$m. . . 0.5m.

The solid wood volume can be estimated from the contour volume via a correctional factor $f_{\mathrm{s,c}}$, called the stacking coefficient [Knyaz and Maksimov, 2014]. This coefficient is dependent on the tree species, wood quality, and log length, and is between 0.58 [Sachsse, 2003] and 0.8 [Kramer and Akça, 2008, p. 39]. Equation 2.9 introduces the solid wood volume calculated from the contour volume $V_{\mathrm{s,c}}$.

$$V_{\mathrm{s,c}} = f_{\mathrm{s,c}} \cdot V_{\mathrm{c}} = f_{\mathrm{s,c}} \cdot w_{\mathrm{s}}l\sum_{i=1}^{n_{\mathrm{s}}} h_i \qquad (2.9)$$

2.6 Measurement errors and error propagation

A large number of parameters influence the outcome of manual and automatic surveying techniques. In this section it will be clarified which measurement parameters affect the results and which magnitude of error should be expected for different methods. In literature the following error inducing parameters can be found:

- number of wood logs [Bort et al., 1989]
- wood log length [Kramer and Akça, 2008, pp. 38–39; Prodan, 1965, p. 78]
- midpoint diameter / cross-section area [Kramer and Akça, 2008, pp. 38–39; Prodan, 1965, pp. 76–78]
- conversion from S_{c} to $S_{\mathrm{s,c}}$ [Kramer and Akça, 2008, pp. 38–39]

2.6.1 Solid wood volume

This subsection aims to derive an error margin for manual measurements of the solid wood volume, as to gain insight on the flaws of manual surveying techniques. As shown by [Bort et al., 1989], the average error when manually determining the number of wood logs in a

pile is −0.5% with a maximum error of −7.1%. Sachße [SACHSSE, 2003] provides results of extensive testing. The number of wood logs was underestimated by 2.9% on average with a range of 1.5% to 5.0%. This must be taken into account when calculating the volume of a wood pile based on the number of wood logs. The length of a wood log is theoretically measured with an error of 0.001% [KRAMER and AKÇA, 2008, p. 38]. In practice, the length is not measured, because a defined length is used for calculations, which differs from the actual length. Therefore the length error does not need to be accounted for.

The measurement of midpoint diameter is influenced by a number of possible errors. [KRAMER and AKÇA, 2008, p. 38–39] and [PRODAN, 1965, pp. 76–78] point out that the following must be considered when anticipating the resulting error:

1. construction flaws of the measuring device
2. random errors in diameter measurement
3. irregular stem surfaces
4. non-representative diameter measurements
5. approximation of the cross-section with a circular model
6. shrinkage over time
7. rounding according to forestry standards

Points 1 and 6 do not apply in the cases considered in this thesis, as we use calibrated measurement devices and measurements are performed within one day, so the shrinkage is negligible. [KRAMER and AKÇA, 2008, pp. 38–39] and [PRODAN, 1965, p. 89] show that the average absolute value for random errors of the diameter is 0.5% and an irregular stem surface causes an error of approximately 1%. Non-representative diameter measurements can be accounted for by measuring the diameter twice, with the second measurement being orthogonal to the first. The remaining error is considered to be insignificant. Importantly for the research of the methods in this thesis, the approximation of the cross-section with a circular model causes only very small errors, which Kramer and Akça declare to be negligible. The largest error is induced be rounding according to forestry standards, where diameters are always rounded downward to whole centimeters [BUNDESMINISTERIUM FÜR ERNÄHRUNG, LANDWIRTSCHAFT UND FORSTEN, 1969; DEUTSCHER FORSTWIRTSCHAFTSRAT E.V. UND DEUTSCHER HOLZWIRTSCHAFTSRAT E.V., 2014]. This procedure causes an error between -3% and -6% depending on the diameter of the wood logs [KRAMER and AKÇA, 2008, p. 39]. The overall error for the diameter is said to be -2% to -8%. The detailed empirical study conducted by Staudenmaier [STAUDENMAIER, 2012] confirms that the average volume error of a single log is -8%, when comparing measurements as required by [BUNDESMINISTERIUM FÜR ERNÄHRUNG, LANDWIRTSCHAFT UND FORSTEN, 1969; DEUTSCHER FORSTWIRTSCHAFTS- RAT E.V. UND DEUTSCHER HOLZWIRTSCHAFTSRAT E.V., 2014] to ground truth.

The solid wood volume is defined in Equation 2.5 as the product of the sum of midpoint areas and the length of the wood log. This equation can also be rewritten to take the average diameter as a parameter, respectively the average cross-section area $\overline{g}$ calculated from the average diameter.

$$\overline{d} = \frac{1}{N} \sum_{i=1}^{N} d_i \tag{2.10}$$

$$V_\mathrm{s} = N \left(\frac{\overline{d}}{2} \right)^2 \pi l = N \overline{g} l \tag{2.11}$$

As mentioned above, in literature the error for $\overline{g}$ is said to be between -2% and -8% [KRAMER and AKÇA, 2008, p. 39] and the maximum error for the number of wood logs in empirical evaluations is -7.1% [BORT et al., 1989]. Using these values, the error margin can be estimated. [PAPULA, 2008, pp. 674–688] shows that the relative error for the solid wood volume $\Delta V_\mathrm{s}/V_\mathrm{s}$ can be defined as the partial derivative shown in Equation 2.12.

$$\frac{\Delta V_\mathrm{s}}{|V_\mathrm{s}|} = \left| \frac{\partial V_\mathrm{s}}{\partial N} \Delta N \right| + \left| \frac{\partial V_\mathrm{s}}{\partial \overline{g}} \Delta \overline{g} \right| + \left| \frac{\partial V_\mathrm{s}}{\partial l} \Delta l \right| \tag{2.12}$$

Since only relative errors are known, the following must be defined. (Note that $\Delta l = 0$ as it is a defined length and not a measurement.)

$$\frac{\Delta N}{N} = -0.071 \qquad \frac{\Delta \overline{g}}{\overline{g}} = -0.08 \tag{2.13}$$

When inserting Equations 2.13 and 2.11 into Equation 2.12, the relative error margin can be calculated. Since the solid wood volume must always be positive, it is safe to declare $|V_\mathrm{s}| = V_\mathrm{s}$.

$$\frac{\Delta V_\mathrm{s}}{V_\mathrm{s}} = \frac{0.071 \cdot N \overline{g} l + 0.08 \cdot N \overline{g} l}{N \overline{g} l} = 0.071 + 0.08 = 15.1\% \tag{2.14}$$

It should be noted that this is the largest possible error, which can be concluded from empiric observations. Real world measurements will, on average, possibly be much closer to the real value. Sachße [SACHSSE, 2003] confirms that the number of wood logs, as well as the diameter, are error-prone when measured in practice. In accordance with the error propagation in Equation 2.14, Sachße provides a maximum error from empirical observations of 13.7% for the solid wood volume.

2.6.2 Contour volume

The error propagation for the contour volume cannot be determined easily. This subsection will discuss what types of errors can occur during measurements and which errors are induced by the procedure itself.

It was shown in Equation 2.8 that the contour volume is the product of wood pile width, defined log length, and average section height. The wood pile width is influenced by some of the same parameters that influence the estimation of the wood log diameter. The most significant of these parameters are flawed measurement devices, random errors, and irregular stem surfaces. Sachße [Sachsse, 2003] suggests that vegetation around the pile can influence the measurement of the pile width and lead to overestimation. While errors for the width of the wood pile can be estimated this way and the defined log length is always constant, the average section height must be subject to deeper analysis. In Figure 2.9(b) it was shown how the manual sectioning of a wood pile is performed. By forestry standards, the section width is $w_s \geq 1$m. The section height is measured in the middle of a section, as a means to approximate the section's average height. Although this procedure is convenient in practice, a number of errors can occur:

1. The center section height is not necessarily representative for the whole section.
2. The chosen stacking coefficient $f_{s,c}$ is likely to differ from the true correctional factor, although it may be correct on average.
3. The section center can be influenced by an offset, due to incorrect measurement of the pile width. [Sachsse, 2003]
4. Wood logs on the bottom of the pile can sink into the ground, making a measurement difficult. [Sachsse, 2003]
5. Protruding wood logs can make measurements at the section center impossible. [Sachsse, 2003]
6. The measurement device usually does not provide a mechanism for exact vertical measuring, making the upper and lower point of the section height ambiguous.

Due to the complex nature of the error-prone parameters, a deterministic evaluation of the error propagation seems infeasible in the scope of this thesis. Instead, results from empirical observations will be discussed to provide insight into the nature of errors induced by the contour volume method.

Sachße [Sachsse, 2003] compares the results of the section volume method to the results for the complete midpoint diameter inventory. Interestingly, the number of wood logs was lower by 3.0% for the full inventory, although the average contour volume was higher by 0.35%. In conclusion, the volume error of the full inventory would be even higher if all wood logs had been measured. Sachße [Sachsse, 2003] also reports results for $V_{c,s}$ in comparison to the true value, as determined by a certified roundwood measuring device. On average, the contour volume method overestimated the volume by 4.5% with values ranging from −6.2% to +18.4%. It can be concluded that the volume measurement method, as it is currently used in practice, provides rough estimates at best. In the following sections it will be discussed

how this method can be applied and improved to become adequate for automatic surveying.

2.6.3 Digital surveying

In digital surveying, new possible errors are introduced which can be compared to their counterpart in analog surveying, as shown in Table 2.2.

digital error	analog counterpart
scale reference	width and height measurements
non-planarity of front surface	non-planarity of front surface
pixel-based quantization	reading accuracy

Table 2.2: Comparison of digital and analog errors

Scale reference Due to reconstruction ambiguities [HARTLEY and ZISSERMAN, 2004, pp. 264–265], the scale of a photo or a multiple view reconstruction is lost. For surveying purposes, an adequate scale reference must be introduced, through which metric measurements are enabled. The error of this scale produces a quadratic error in the result because the width, as well as the height, of the digital model are estimated based on this scale. The depth on the other hand is known via the defined length and is therefore unaffected by the scale reference. In advance of Section 7.3 (*Scale reference*), it can be said that the error for scale estimation is small if the correct end points of the pile are used.

Non-planarity of the front surface A non planarity of the front surface is only relevant if a planar surface model is used, e.g., for image stitching. From empirical observations it was found that small non-planarities only affect the volume error in a linear fashion because the images usually undergo planar projection in the image stitching process. Large non-planarities lead to failure of the stitching pipeline. 3D reconstruction approaches are not affected by the front surface geometry, unless heuristics which assume partial planarity are used.

Pixel-based quantization Although some image processing methods provide sub-pixel accuracy (e.g., [DA and ZHANG, 2010; GEHRIG and FRANKE, 2007]), most measurements are still subject to pixel-based quantization. This is comparable to the reading accuracy of analog instruments, such as a measuring tape. When measuring distances in an image, the error always influences the result linearly. 3D models from multiple view stereo can provide sub-pixel accuracy through sub-pixel feature detection and matching [TOMASI and KANADE, 1991; ZHENG and CHELLAPPA, 1995] and are thus more influenced by the reconstruction pipeline than the pixel quantization.

2.7 Data basis - *The HAWKwood Database*

The methods presented in this thesis aim to estimate the volume of different types of wood piles. The results need to be subject to quantitative evaluation. As demonstrated in Section 2.4 (*Necessity for digital wood pile surveying*), digital wood pile surveying is an active research topic. One of the contributions of this thesis is a large set of wood pile images, which is provided for academic and non-commercial use to researchers interested in this field of study. The provided collection of images and ground truth data is called *The HAWKwood Database* [HERBON, 2014b]. All images were taken by a standard digital or smartphone camera. One of the main purposes of this database is to establish a platform for algorithm comparison and benchmarking. The results of the methods proposed in the following chapters define the initial benchmark values in the hope that further research will pick up and build upon the advances provided by this thesis.

The database is divided into two different benchmarks, to account for different research strategies. The single image benchmarks provide individual images of wood pile front surfaces, combined with ground truth object location and segmentation information, on a pixel level. For the surveying of entire wood piles, the multi-image benchmark supplies image sets of wood piles where the images are taken from different viewpoints. Each set consists of between 5 and 60 images of the same pile, which can be used for panoramic image stitching and 3D reconstruction. Ground truth data are provided for different parameters of the pile, which may be used for evaluation purposes. The *HAWKwood* database includes a total of 7655 images, which are clustered into 354 data sets. A subset of these images was generated synthetically. This way, ground truth can be calculated highly accurately without relying on manual measurements. In the following sections, the benchmarks will be discussed in detail, based on the elaborations of [HERBON, 2014b], and extending implementation descriptions when necessary.

2.7.1 Categories

Each benchmark, single and multi-image, is divided into three categories, which address different problems of the respective research area with regards to wood log detection and wood pile surveying. The categories are ordered incrementally so that each category can theoretically, but not necessarily, build on the results obtained in the previous category.

Single image benchmark

    Category S.1: wood log detection

    Category S.2: wood log faces segmentation

    Category S.3: front surface segmentation

	S.1	S.2	S.3
images (real)	121	40	30
images (synthetic)	36	-	-
data sets (real)	121	20	10
data sets (synthetic)	36	-	-
ground truth (real)	man. marked	man. segmen.	man. segmen.
ground truth (synthetic)	man. marked	-	-

Table 2.4: Number of images, data sets, and ground truth type for the single-image benchmark

Multi-image benchmark

Category M.1: wood log detection

Category M.2: solid wood volume computation

Category M.3: contour volume computation

	M.1	M.2	M.3
data sets (real, low overlap)	72	34	117
data sets (real, high overlap)	147	71	206
data sets (synth., high overlap)	40	40	40
ground truth (real)	man. counted	CMDI[1]	Sec. Vol.[2]
ground truth (synthetic)	dig. counted	computed	computed

Table 2.6: Number of images, data sets, and ground truth type for the multi-image benchmark

2.7.2 Single-image benchmark

Category S.1: wood log detection Image-based wood log detection is one of the most fundamental tasks in wood pile surveying. The *S.1* category includes 121 real and 36 synthetic images of wood pile front faces for which ground truth is available (Figure 2.11). In Section 2.3.1 (*Attributes of roundwood logs*), it has been shown that in literature performing a circular approximation of the wood log front face is well-accepted. In the *S.1* category, wood logs are manually marked with a circle, which is stored as a bounding rectangle. A number of possible occlusions can make the wood log detection task complex, as shown in Section 2.2 (*Research challenges*). In order for algorithms to prove invariance to a reasonable amount of occlusion, all wood logs which are occluded by up to 25% and which do not intersect the image borders

---

[1]Complete midpoint diameter inventory, described in Section 2.5 (*Manual surveying methods*)

[2]Section volume method, described in Section 2.5 (*Manual surveying methods*)

Figure 2.11: Sample images and ground truth from category *S.1*. (a)-(c) Input images. (d)-(f) Manually marked ground truth (blue: valid wood logs, red: partially occluded wood logs). Image source: *HAWKwood* [HERBON, 2014b]

are considered. This definition is consistent with the geometric piling constraints introduced in Section 2.3.3. Wood logs which are largely occluded are also marked manually, but it is noted that these logs must not be considered. Should an algorithm detect one of these objects, it can be assured that it is not counted as a false positive.

Category S.2: wood log faces segmentation Category *S.2* provides manually segmented ground truth for a subset of the images from the *S.1* benchmark. All wood log front faces, including bark, have been manually segmented on a pixel-level. The goal for algorithms performing this benchmark should be to perform accurate binary segmentation of all wood log faces and evaluate the results through the pixel-based comparison of the provided ground truth segmentation masks. Figure 2.12 shows samples from category *S.2* including segmentation masks.

Figure 2.12: Sample images from category *S.2* with different tree species and wood qualities. (a)-(c) Input images. (d)-(f) Binary segmentation masks. Image source: *HAWKwood* [HERBON, 2014b]

Category S.3: front surface segmentation In contrast to category *S.2*, even partially occluded wood logs are considered in the *S.3* benchmark. This category is not concerned with the segmentation of individual wood logs but rather with the segmentation of the wood pile front surface as a whole. Wood logs that were previously excluded due to visibility constraints are now included in the front surface definition. Each set offers two images of the same wood pile surface taken from slightly different positions, similar to a stereo camera setup. Additionally, a segmentation reference mask is provided to the left image. Algorithms aiming to perform front surface segmentation may either use one or both images. When the second image is used also, stereo auto-calibration can be performed. The pixel-based ground truth is marked manually, as can be seen in Figure 2.13.

Figure 2.13: Sample images from category *S.3*. First column: left input image. Second column: right input image. Third column: Manually marked ground truth for the left image. Image source: *HAWKwood* [HERBON, 2014b]
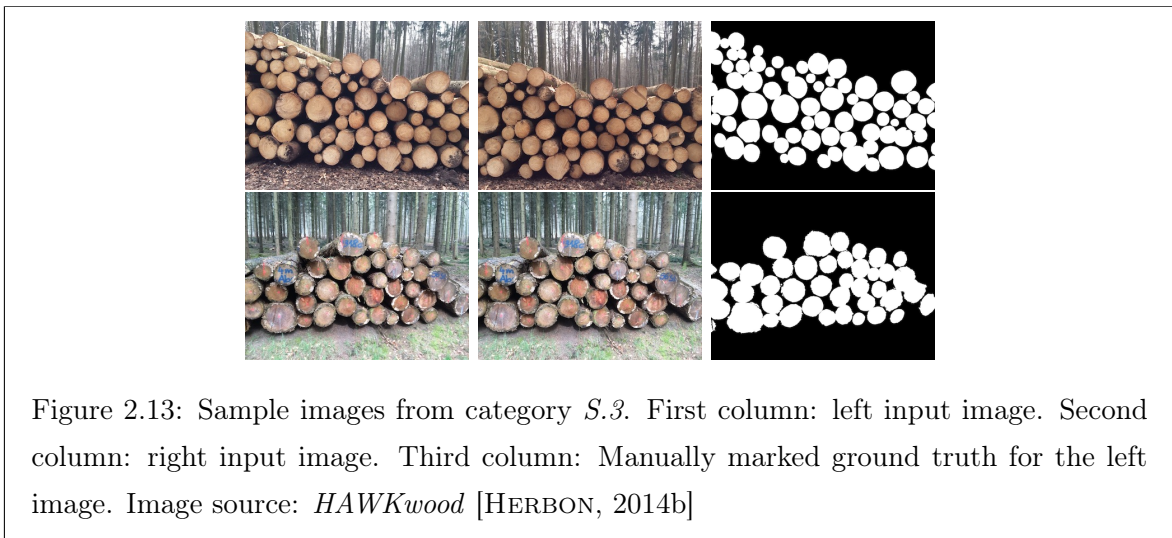
2.7.3 Multi-image benchmark

The multi-image benchmark supplies data sets consisting of overlapping images of the same wood pile. These wood piles are usually too large to be captured by a single image and therefore a model must be established which fuses these images into a reconstruction of the wood pile. Two possible fusion models are panoramic image stitching and multiple view reconstruction. For each of the models a distinct set of images is defined:

1. *Small overlap* datasets can be used for panoramic image stitching (Figure 2.14).
2. *Large overlap* datasets can be used for multiple view reconstruction, e.g. structure from motion (Figure 2.15).

Adjacent images with small overlap show approximately 50-70% of the same wood log pile front surface. Large overlap is considered to be about 90%. Since the images were taken by hand, these values cannot be guaranteed and are merely an approximation. Most of the data

Figure 2.14: Images with small overlap. Samples from category *M.1-M.3*. Top row: real images, bottom row: synthetic images. Image source: *HAWKwood* [HERBON, 2014b]

Figure 2.15: Images with large overlap from category *M.1-M.3*. Top row: real images, bottom row: synthetic images. Image source: *HAWKwood* [HERBON, 2014b]

sets only depict one side of the wood pile. In real world scenarios, both sides of the wood pile are measured individually and the results are averaged. In Equation 2.7 the validity of this approach has been proven. Therefore, the digital surveying of the front surface is sufficient to prove the validity of the surveying method.

Category M.1: wood log detection It has been shown in Section 2.4 (*Necessity for digital wood pile surveying*) that the counting of wood logs can be error prone. For the determination of the ground truth number of wood logs, the wood log faces were counted by seven persons, independently of each other. When discrepancies occurred, the counting procedures were repeated until consistency was reached, thus assuring the correct determination of the wood log count. Some wood piles are put on underlays, which are additional wood logs, whose purpose it is to keep the wood pile from sinking into the mud or snow. These underlays are not included in the total number of wood logs.
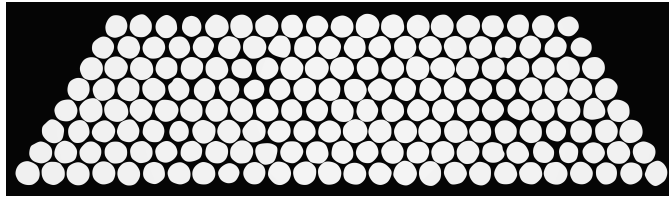
Figure 2.16: Orthographic projection of a wood log front surface and subsequent thresholding of the image

Category M.2: solid wood volume computation For the data sets of real wood piles, the solid wood volume V_s has been measured through the complete midpoint diameter inventory, where each wood log diameter is measured once when it is smaller than 20cm and twice when it is larger than or equal to 20cm [BUNDESMINISTERIUM FÜR ERNÄHRUNG, LANDWIRTSCHAFT UND FORSTEN, 1969; DEUTSCHER FORSTWIRTSCHAFTSRAT E.V. UND DEUTSCHER HOLZWIRTSCHAFTSRAT E.V., 2014]. In Section 2.6.1 it has been shown that the CMDI is not necessarily accurate in all cases. The ground truth data obtained this way should not be seen as absolutely accurate, but rather as an approximation with which the feasibility can be tested. This is sufficient by forestry standards and it is even more precise than what is usually measured in practice [FINK, 2004], but it is not sufficient for a scientific evaluation. For this reason, the *M.2* category also provides synthetic data, for which ground truth can be computed highly accurately. The solid wood volume is determined by performing an orthographic projection of the wood log front faces, as shown in Figure 2.16.

Since an image cannot directly provide the scale of its contents, scaling information must be provided additionally. In Section 7.3 (*Scale reference*) it will be shown that the most accurate scale reference for a wood pile is the largest measurable distance, i.e. its width. For the *M.2* and *M.3* categories, the width of the wood pile (the largest possible distance) is provided. In the case of real wood piles, the distance was measured by hand ten times and then averaged. For synthetic wood piles, the width is computed automatically. The solid wood volume is then calculated through the multiplication of the metric front surface with the defined wood log length.

Category M.3: contour volume computation The contour volume V_c of a wood pile must be determined in accordance with manual surveying methods. For the data sets which depict real wood piles, the section volume method is used. This manual method requires the average height to be calculated from a number of equidistant height measurements. For the manual ground truth in category *M.3*, a section width of $w_s = 0.5$m is used. The volume is then calculated by multiplying the measured width, the average height, and the defined wood log

length (depth of the pile). Similar to the *M.2* benchmark, the ground truth for real wood piles in the *M.3* category is not free of error and should therefore be seen as an approximation, which is defined by forestry standards [Bundesministerium für Ernährung, Landwirtschaft und Forsten, 1969; Deutscher Forstwirtschaftsrat e.V. und Deutscher Holzwirtschaftsrat e.V., 2014] and which is generally accepted as the consensus in the forestry industry. Precise ground truth for the contour volume of a wood pile can only be obtained by using synthetic data. The computation of the contour volume is done similarly to the solid wood volume, where the first step is to perform an orthographic projection of the wood log front faces. Equation 2.8 shows how the contour volume is measured by averaging the height. In real world scenarios it is not feasible to use a section width $w_{\mathrm{s}} < 0.25$m. For synthetic data however, the computation can be performed on a pixel basis by measuring the height at each x-location in the image (within the horizontal borders of the wood pile).

$$V_{\mathrm{c,gt}} = l \sum_{i=1}^{w_{\mathrm{px}}} h_{i,\mathrm{px}} \cdot f_{\mathrm{mppx}}^2 \tag{2.15}$$

$$f_{\mathrm{mppx}} = \frac{w_{\mathrm{p}}}{w_{\mathrm{px}}} \tag{2.16}$$

Equations 2.15 and 2.16 express the computation of the ground truth contour volume $V_{\mathrm{c,gt}}$. The volume is calculated by accumulating all pixels through the height at each x-location $h_{i,\mathrm{px}}$ and then by multiplying the result with the defined wood log length l and the correctional factor (meters per pixel) f_{mppx} squared. The correctional factor f_{mppx} is necessary to convert height and width measurements from pixels to meters. It is defined as the quotient of wood pile width in [m] w_{p} and the wood pile width in [px] w_{px}.

Chapter 3

Related work

The scope of this thesis includes a number of different topics which are combined for the purpose of creating photogrammetric wood pile surveying approaches. This related work chapter is divided into different sections, each of which addresses an individual subproblem. To retain focus on the scientific contributions of this thesis, only methods are discussed which are either directly relevant to the proposed novel methods or which are employed by current state of the art wood pile surveying approaches. At the end of this related work chapter, further areas of application are discussed to emphasize the importance of wood pile surveying research. Starting with Chapter 4 (*Adaptive image stitching*), the remaining chapters of this thesis are concerned with showing how the related work can be improved and incorporated into distinct surveying approaches.

3.1 Prerequisites: 2D object recognition and segmentation

In recent years, a number of approaches have been presented which are able to accurately recognize objects in digital images. Object recognition methods usually rely on local features or the windowed processing of subregions in an image. It has been proven that these approaches provide robust and accurate results when aiming to detect different object types or combinations of objects. Object recognition information can be used as priors for segmentation procedures. Due to the hundreds of algorithms proposed in the last decade, only the most prominent approaches with direct applications to this thesis will be discussed. The actual methods and the application of these methods are distinguished between for different scenarios. In accordance with their implementations in the next sections, object recognition methods will be discussed prior to segmentation methods.

3.1.1 Object recognition methods

Scale Invariant Feature Transform (SIFT) Object recognition via SIFT (Scale Invariant Feature Transform) [BROWN and LOWE, 2007; LOWE, 1999; LOWE, 2001; LOWE, 2004] is based on the idea of extracting key points from a template image and matching these key points with an image of unknown content. For each key point, a feature descriptor is computed which is rotation and scale invariant. According to Lowe [LOWE, 1999], SIFT features are, within limits, robust to distortions, such as partial occlusion, scaling, changes in illumination, and affine transformations. Apart from object recognition tasks, SIFT features are often used in panoramic image stitching (Section 3.3 (*Panoramic image stitching*)) or 3D reconstruction (Section 3.4 (*3D reconstruction / structure from motion*)) in order to perform image registration. An improvement of the original SIFT features is PCA-SIFT [KE and SUKTHANKAR, 2004], which computes features by including a principal component analysis (PCA). The results are shown to be both more accurate and faster than the original SIFT implementation.

Local Binary Patterns (LBP) One of the earliest classification / detection approaches, based on a texture operator, is LBP (Local Binary Patterns) [OJALA et al., 1994; OJALA et al., 1996]. LBP considers a neighborhood of pixels where binary thresholding is performed and a local histogram is calculated, which is then used as a texture descriptor. The extension of LBP with a contrast measure [OJALA and PIETIKÄINEN, 1999], called LBP/C, provides good results in unsupervised texture segmentation. Zhao and Pietikainen [ZHAO and PIETIKAINEN, 2007] propose to extend LBP to a third dimension with a volumetric approach (VLBP). This way, LBP can be applied to a spatiotemporal domain, e.g., for use in videos to detect temporal self-similarities.

LBP has received attention for its capability to detect human faces, as first proposed by Ahonen et al. [AHONEN et al., 2006]. In this approach, faces are divided into regions, for which features are computed and combined into a feature vector. The entire approach is based on the multi-resolution LBP implementation by Ojala et al. [OJALA et al., 2002]. The computed vector is then used as a face descriptor. In comparison to other texture descriptors, the LBP-based face detection framework performs considerably better than gray-level difference histograms, homogeneous texture descriptors, and texton histograms [AHONEN et al., 2004]. Based on these findings, face recognition has been improved by including motion pictures to derive information about the person's gender [HADID and PIETIKÄINEN, 2009]. A number of further applications has been proposed [FENG et al., 2005; TAN and TRIGGS, 2010; ZHAO et al., 2009], from which it can be seen that Local Binary Patterns are applicable to a variety of problems. Hence LBP is a strong candidate for object recognition in the case of wood log faces.

Histograms of Oriented Gradients (HOG) Histograms of Oriented Gradients (HOG) were first proposed by Dalal and Triggs [DALAL and TRIGGS, 2005], with specific application to the detection of humans. The method is based on support vector machines (SVMs, see e.g. [CRISTIANINI and SHAWE-TAYLOR, 2000; KECMAN, 2001]) and uses normalized local histograms of image gradient orientations. During detection, an input image is divided into cells and a histogram for gradient or edge orientations is computed. The obtained histograms are then used as a feature vector. The original paper [DALAL and TRIGGS, 2005] includes a study on two distinct data sets, containing 509 and 1805 training images respectively, which show pedestrians in an urban environment. HOG features are shown to perform well in comparison to PCA-SIFT [KE and SUKTHANKAR, 2004], shape contexts [BELONGIE et al., 2001], and generalized Haar wavelets [MOHAN et al., 2001]. Some improvements have been made to HOG features, such as a speed-up of the computation time [ZHU et al., 2006] or an adaptation to sketch-based image retrieval (SBIR) [EITZ et al., 2011], and the so-called Gradient Field Histograms of Oriented Gradients (GF-HOG) [HU et al., 2010; HU and COLLOMOSSE, 2013].

Haar-like features Haar-like features, as proposed by Viola and Jones [VIOLA and JONES, 2001], are named after wavelets with Haar basis functions [HAAR, 1910]. Based on the integral image representation, which is also introduced in the original paper, a number of features are computed from a training data set. Via AdaBoosting [SCHAPIRE and SINGER, 1999], a feature selection procedure is run which chooses distinct rectangle features from a set of 180,000 features in a 24 by 24 pixel window. Similar to HOG and LBP features, Viola and Jones train their detector on images of human faces and images which do not include faces. The training procedure can be applied to any other object class as well.

3.1.2 Object segmentation methods

Object segmentation methods can be roughly divided into histogram-based, region-based, edge-based, and model-based methods [JÄHNE, 2005, pp. 479–518]. Additionally, multi-scale approaches consider information of different scales, as a means to achieve scale invariance [TÖNNIES, 2005, pp. 206-208]. For histogram-based methods, the goal is to define one or more thresholds, which are optimal in some sense, as defined e.g., by the well-known Otsu method [OTSU, 1975]. Since the computed thresholds can be applied to each pixel individually, histogram-based methods are considered to be a global approach. Methods which segment an image based on homogeneity criteria, such as texture or gray-scale value distribution, are called region-based [TÖNNIES, 2005, pp. 213–218]. In contrast to homogeneity information, edge-based approaches use discontinuities (e.g. peaks in the image gradient) to detect edges. Edges are considered to be natural borders between image segments and can thus be used to

separate image regions while being less prone to oversegmentation caused by inhomogeneous region characteristics [Tönnies, 2005, p. 221].

Watershed transform The basic idea of the watershed transform as an edge-based method is to view a gray-scale image as a topographic surface [Beucher and Lantuéjoul, 1979]. Adjacent regions are separated by edges. Geographically, a watershed is the border between two basins which are continually flooded with water. Flooding usually starts at local minima. Oversegmentation poses a challenging problem for watershed algorithms. For this reason, marker-based watershed has gained popularity. Markers can either be manually inserted or computed using a priori information or heuristics. [Tönnies, 2005, pp. 225–227]

Some efforts have been made to optimize watersheds for computational efficiency, e.g., through parallelization [Roerdink and Meijster, 2000] or GPU implementation [Kauffmann and Piche, 2008]. Medical imaging is one of the areas where watersheds are often used [Grau et al., 2004; Hahn and Peitgen, 2000; Ng et al., 2006; Shojaii et al., 2005; Umesh Adiga and Chaudhuri, 2001]. Road segmentation / obstacle detection [Beucher et al., 1990; Beucher and Bilodeau, 1994; Yu et al., 1992] and cell segmentation [Jiang et al., 2003; Tek et al., 2005] often also achieve good results by applying the watershed transform.

Expectation maximization Expectation maximization (EM) is a histogram-based method, which was introduced by Dempster et al. [Dempster et al., 1977]. [Köhler, 2005, p. 77] sees it as one of the most important estimation algorithms in digital signal processing. It starts from an initial estimate and iterates until convergence is achieved. In image processing, it is realized through a mixture of multivariate normal distributions which represent, e.g., color or texture information [Carson et al., 2002; Permuter et al., 2006; Tai et al., 2005]. The goal is then to refine the parameters of the model to achieve a maximum likelihood estimation. Different authors [Caetano and Barone, 2001; Greenspan et al., 2001; Kakumanu et al., 2007; Phung et al., 2005; Yang and Ahuja, 1999; Zhu et al., 2004; Zhu et al., 2000] propose to segment human skin in digital images or videos by using a mixture of Gaussian models. The mixture model can also be combined with depth information, e.g., for segmenting people in RGB-D data [Harville et al., 2001].

Graph Cuts Graph cuts are edge-based energy minimization algorithms, first described by Greig et al. [Greig et al., 1989]. They can be applied to a wide variety of low-level computer vision tasks where the problem can be formulated as an energy minimization problem. Image segmentation by graph cuts is described in [Toennies, 2012, pp. 236–250], where the graph representation of an image defines pixels as nodes while neighboring pixels are connected by an edge. The segmentation can be achieved by assigning a cost to each edge and performing a cut by computing the minimum cut along the edges. Boykov and Kolmogorov [Boykov and

Kolmogorov, 2001; Boykov and Kolmogorov, 2004] propose an efficient implementation for the computation of the max-flow of a graph in computer vision. The applications in image segmentation are broad and include image restoration [Boykov et al., 2001], stereo vision [Bleyer and Gelautz, 2007; Boykov et al., 2001; Hornung and Kobbelt, 2006; Kolmogorov and Zabih, 2001; Kolmogorov and Zabih, 2002; Vogiatzis et al., 2007; Vogiatzis et al., 2005], motion computation [Boykov et al., 2001], medical image processing, and geodesic segmentation [Boykov and Kolmogorov, 2003; Price et al., 2010].

Rother et al. [Rother et al., 2004] propose to extend graph cuts with interactive user input, an approach they call Grab Cut. The user defines a rectangular region in an image, in which the object of interest is located. Regions outside the rectangle are marked as background. From the color information, a Gaussian mixture model (GMM) is extracted, which is then used to perform global optimization, as described by Boykov and Jolly [Boykov and Jolly, 2001]. According to Rother et al. [Rother et al., 2004], one of the main advantages of the Grab Cut approach is the non-necessity of foreground labeling while requiring only minimal user interaction.

Level sets   Level sets were originally described by Osher and Sethian [Osher and Sethian, 1988] as an edge- / contour-based method which aims to track the motion of a front (e.g. the boundary between two regions). The idea behind level sets is to describe an n-dimensional contour by an $n+1$-dimensional surface. A curve in an image would thus be defined by a three-dimensional surface and its intersection with the xy-plane, called the zero level set of a defined level set function. The change of the curvature is described by a speed function F, which may depend on different factors of the front, such as local properties (e.g. curvature or normal direction), global properties (shape and position of the front), and independent properties which are not affected by the front itself [Sethian, 1999, p. 4]. Early approaches are the geodesic active contours [Caselles et al., 1997] (Caselles et al.) and active contours without edges [Chan and Vese, 2001] (Chan and Vese). Rousson and Paragios [Rousson and Paragios, 2002] demonstrate how shape priors can be incorporated into level sets for 2D closed structures. In recent years, level sets have become popular in image segmentation because the result yields an appealing representation of regions and their boundaries [Brox and Weickert, 2004].

Based on the active contour model without edges, a multi-phase level set framework was introduced by Vese and Chan [Vese and Chan, 2002] with which multiple regions can be segmented without vacuum or overlap. Brox and Weickert [Brox and Weickert, 2004] follow a similar approach and extend level sets to work with a dynamic number of regions, as opposed to the classical foreground and background segmentation. Paragios and Deriche [Paragios and Deriche, 2002] use level sets for supervised texture analysis, inspired by the

geodesic active contour model.

So far state-of-the-art practices in wood pile surveying do not use level sets. Therefore related applications will only be described briefly. Level sets are especially popular in medical imaging. Nosrati and Hamarneh [NOSRATI and HAMARNEH, 2014] demonstrate an approach where prior knowledge is incorporated into a level set framework, with the goal of obtaining more accurate and more plausible results for the segmentation of cells. Similar to the problem of cast shadows on wood pile surfaces, Li et al. [LI et al., 2011] propose a level set based method for medical image segmentation, which is able to account for intensity inhomogeneities.

3.2 Recognition and segmentation of wood logs and clustered objects

The first photo-optical measurement method for determining solid wood volume was introduced by Gläser [GLÄSER, 1955; PRODAN, 1965, p. 95]. A regular grid is laid over an image of a wood pile front with points in equidistant locations. The number of points that are on the wood are placed in relation to the overall number of points. This way, a factor can be determined with which the solid wood volume can be calculated from the contour volume. This very basic method lays the foundation for all photo-optical surveying methods, although Gläser does not account for distortions of any kind, such as distortions from the camera or distortions induced by not taking photos orthogonally to the wood pile front surface.

Another early approach was proposed by Meyer [MEYER, 1995], in which analog pictures of wood logs are taken, digitized, and evaluated. The segmentation procedure relies on grayscale thresholding to separate foreground (wood logs) from background. Fink [FINK, 2004] points out that in this procedure it is not possible to measure each individual wood log and that only a cumulative computation of the solid wood volume is possible. A similar method, through which the load of timber trucks can be measured, was presented by Carvalho et al. [CARVALHO et al., 1993]. The goal of their implementation is the computation of the stacking coefficient $f_{s,c}$, with which the solid wood volume can be derived from the contour volume. Results for this factor vary from 0.55 to 0.8, depending on the tree species, the diameter of the wood logs, and the wood quality. Tian and Murphy [TIAN and MURPHY, 1997] implemented a computer vision system for detecting trimmed and occluded branches in tree stems. This system aims to provide information about the texture, especially existing knots in freshly harvested wood logs. Although this is usually done by laser scanning, Tian and Murphy show that a computer vision system is also feasible for this task.

Many modern wood log segmentation procedures aim at reducing the segmentation problem to a setup, in which a simple thresholding operation suffices. An important contribution is the detailed analysis of Fink [FINK, 2004], which elaborates the needs of the forestry industry,

argues for the development of image processing methods for wood log detection through economic observations, and proposes a variety of methods with which wood logs can be detected and segmented. Unfortunately, the proposed image processing pipelines are limited to the evaluation of a single image under laboratory conditions. Fink's work can be seen as the extension of the work of Meyer [MEYER, 1995] and Carvalho et al. [CARVALHO et al., 1993]. Different methods for the segmentation of individual wood logs are proposed, which include active contours (snakes) [KASS et al., 1988], the watershed transform [BEUCHER and LANTUÉJOUL, 1979], and even stereo-imaging methods.

Two important problems are not addressed by Fink, specifically the spatial distribution of wood logs (the large dimensions of wood piles and the self-similarity through geometric constraints) and the conditions under which wood piles are measured outside of a factory setting. The computer vision techniques mentioned are only applicable to single images and how to extend these methods to multiple images or how to fuse several images, e.g., through panoramic image stitching, is not discussed. All experiments were carried out in a laboratory setting, which is not comparable to real world conditions, as shown in Section 2.2 (*Research challenges*).

Dahl et al. [DAHL et al., 2006] extend the watershed segmentation approach of Fink. The use of a scale space watershed segmentation procedure is proposed, which, in the end, produces binary blobs (homogeneous regions in an image) that are merged and filtered. Similar to the previously discussed approaches, Dahl et al. assume that the problem can be reduced to an intensity based segmentation task. As shown at different points in this thesis, this constraint can only be satisfied under laboratory conditions.

A histogram-based method for the determination of wood logs in a wood pile is shown by Noonpan and Chaisricharoen [NOONPAN and CHAISRICHAROEN, 2013]. The use of a mixture of morphological operations combined with histogram backprojection (see [SWAIN and BALLARD, 1992]) is proposed. An estimation accuracy of 78% is reached. For real world applications this detection rate is not sufficient. It seems possible to use such an approach as a preprocessing method. Similarly, Rahman et al. [RAHMAN et al., 2011] propose simple image processing techniques for wood log detection on timber trucks. Building on this approach, Yella and Dougherty [YELLA and DOUGHERTY, 2013] extend the methodology by using a circular Hough transform to detect wood logs. The detection rate in this approach is low in comparison to state-of-the-art methods, which will be discussed below.

The definition of a wood log face as a circular feature, along with the proposition for a detector, is also used by Knyaz and Maksimov [KNYAZ and MAKSIMOV, 2014]. Based on prior work [KNYAZ and VIZILTER, 2001], a stereo camera setup is proposed, which aims to calculate the stacking coefficient. The shape of a wood log is assumed to be strictly circular and a circle detection method is used to identify wood logs. From the locations and sizes, the

stacking coefficient is derived. The proposed method requires a rather complex stereo camera set up. On the one hand, this is needed to remove distortions, which is an important task in photo-optical surveying, on the other hand, state-of-the-art methods for wood log detection are able to perform a much more precise description and authentication of wood logs which are not dependent on a strong circularity constraint. In Section 2.2 (*Research challenges*) it has been shown that the volume of a wood log may be accurately approximated with a circular model, but this does not necessarily hold true for object recognition. Since no actual detection rate results are given, it can only be assumed that the circular detection model can be significantly improved by state-of-the-art detection methods.

Some of the more advanced approaches are presented by Gutzeit et al. [GUTZEIT et al., 2010; GUTZEIT et al., 2011; GUTZEIT and VOSKAMP, 2012]. In [GUTZEIT et al., 2011], the authors describe a graph cut based segmentation procedure for wood log front faces. A sub-image in the center of the original image is extracted, which is used to estimate a foreground and background model. Based on this color information, the entire image is segmented via graph cuts. Different weight-setting metrics for the graph cut procedure are evaluated. A comparison between a histogram-based weight setting, k-means clustering [FORGY, 1965; LLOYD, 1982; MACQUEEN et al., 1967], Gaussian mixture models with expectation maximization, and a novel kd-tree-based [BENTLEY, 1975] nearest neighbor approach called *KD-NN* is performed. It is shown that the *KD-NN* segmentation generally performs best. Although the proposed method seems to perform well on the test image set, the acquisition constraints on the images are very restrictive. The wood pile must be located at the image center, because the initial sub-image is extracted from this region. No spatial information is considered, which would seem like an obvious constraint, given the geometric properties of a wood pile. Furthermore, it is assumed that the extracted sub-image is well-segmentable by the intensity values, which cannot be guaranteed during weather conditions like snow or rain.

The aforementioned shortcomings of [GUTZEIT et al., 2011] are addressed in [GUTZEIT and VOSKAMP, 2012], where Gutzeit and Voskamp propose the use of Haar-like features [VIOLA and JONES, 2001] as an initial detection step. This way, the location of the wood pile in the image is not constrained to the center anymore and the spatial correlation of the wood logs faces is taken into account when filtering the detected objects. The suggested method performs an initial detection first, refines the locations of the detected objects, and then uses the detected objects to derive information about the color distribution in the foreground and background. The color model is used in the next step to set the weights for a graph cut procedure, as described in [GUTZEIT et al., 2011]. After the first segmentation, "holes" in the binary image are used to find additional wood log faces. All remaining blobs are considered to be wood logs. The results are compared to a distance-based watershed procedure (watershed on the distance transform of the binary image) and to a binary image, consisting of the

detected objects as circles. It is found that the graph cut based segmentation performs best, with the distance-based watershed delivering comparable results. What seems problematic for the use of this method is the assumption that blobs in the binary image are wood log faces. From preliminary experiments for this thesis it became clear that many of these blobs are not part of the wood pile. This issue will be addressed in Chapter 5 (*Two-dimensional surveying*) based on our findings presented in [HERBON et al., 2014c].

In addition to academic literature, a number of patents have been filed which concern the process of wood log surveying. Not all of the following patents have been granted, but all relevant patent applications will also be discussed, as they contribute to state-of-the-art techniques in wood pile surveying. Davis [DAVIS, 1990] proposes a stereographic surveying technique, which measures the cross-section and length of each log on a truck and derives the cumulated volume from the individual logs. Since a stereo camera is needed, this patent is not applicable to surveying with monocular cameras, as exhibited in consumer-level smartphones.

A patent for measuring the external dimensions, the average diameter, and the length of wood logs of a wood pile is claimed by Kauppinen [KAUPPINEN, 1993]. This method requires a beam of light to be projected onto the wood pile, which is then photographed from different angles. At least three sources of light and at least three video cameras are required for the application of the patent, which disqualifies it for mobile, outdoor usage. Dralle and Tard-Johansen filed patents [DRALLE, 2005; DRALLE and MADS, 2004; DRALLE and TARD-JOHANSEN, 2010] and were granted a patent [TARP-JOHANSEN and DRALLE, 2012] for a vehicle-based stereo camera system, which measures the size, shape, surface, and location of the wood logs. The computation is performed on a computer inside the vehicle and the wood piles are referenced via GPS. A similar system, which measures wood on a gripper arm, has been filed as a patent application by Seto [SETO, 2011].

Willmann [WILLMANN, 2009] describes a process in which digital images of wood piles are taken and transferred to a PC, where they are processed to determine the contour volume, the number of wood logs, and the average wood log diameter of a wood pile. A similar patent was filed by Bombosch et al. [BOMBOSCH et al., 2011a; BOMBOSCH et al., 2011b], where overlapping images of wood pile front surfaces are taken to determine the aforementioned quantities, while the device is mounted on a forestry vehicle. [ILUMETS, 2013] is a patent application in which an apparatus consisting of cameras and lasers determines the volume, and a number of other parameters, of wood on a truck. The patent described in [FREISTAAT SACHSEN STAATSBETRIEB SACHSENFORST UND GIS-DIENST GMBH, 2012] builds on the composition of a panoramic image, from which the number of wood logs and the contour volume is derived, based on the known depth of the pile and an unspecified reference object. The evaluation is performed externally on a PC. A patent based on markers, which are mounted to the wood pile, was granted to Scheller et al. [SCHELLER et al., 2014]. The markers

display bar codes, through which the individual sections of the wood pile are referenced. During the image acquisition process, live feedback is given to the user in order so show if a sufficient number of markers is recognized. Only the image acquisition, not the image processing, is performed on site.

From the number of patents claimed, especially in the last five years, it becomes clear that wood pile surveying is not only a very active field of academic research but also an economically vital industrial area. This thesis aims at improving state-of-the-art methods, while also providing faster methods with higher usability.

3.3   Panoramic image stitching

Panoramic image stitching refers to the task of registering and fusing two or more images, which depict parts of the same scene [SZELISKI, 2006]. Most digital cameras and especially smartphone cameras have a very limited field of view (FOV). This is a significant limitation, because many natural or man-made scenes are too large to be captured in a single image. A common approach to solve this problem is panoramic image stitching. Many modern smartphones, tablets, and digital cameras offer powerful processing units, which can stitch images in a matter of seconds.

The registration and stitching of panoramic images is a well-studied and well-understood research area, with many important contributions over the last two decades. It has broad applications in computer vision, such as robot navigation [BOURQUE et al., 1998; HRABAR and SUKHATME, 2003; UYTTENDAELE et al., 2004; YUEN and MACDONALD, 2002; ZHENG, 2003], video stabilization [BATTIATO et al., 2007; HU et al., 2007; LEE et al., 2009; LITVIN et al., 2003; SHEN et al., 2009], or creating image mosaics [BROWN and LOWE, 2007; CHEN and KLETTE, 1999; SHUM and SZELISKI, 2000; SZELISKI, 2006; SZELISKI and SHUM, 1997]. Panoramic image stitching relies on image registration. The degrees of freedom (DOFs) that are necessary for image registration are dependent on the image acquisition geometry, as well as on a priori information about the intrinsic parameters of a camera. Direct and feature-based approaches are usually differentiated. Brown and Lowe [BROWN and LOWE, 2007] elaborate that direct approaches work on a pixel level and can provide very accurate registration results when a sufficiently close initialization is provided. Feature-based methods are independent of initialization values, but it is important to choose a feature type, which is invariant to the kind of transform that relates a set of images, e.g., a pure translation, as seen (in the ideal case) in some microscopic images, or a projective transform in the general case. In the context of image stitching, adjacent images are related by a 3x3 homography matrix, which is a matrix that relates two images of a plane by preserving straight lines [SZELISKI, 2006]. According to Brown and Lowe [BROWN and LOWE, 2007], SIFT features [LOWE, 2004] are

well-suited for image registration, as the homography transform can be approximated by an affine transform when linearization is performed. [Lowe, 2004] shows that SIFT features are invariant to affine transforms when only a small image part is considered.

A homography matrix is a projective transform with eight degrees of freedom [Szeliski, 2006]. Since the homography itself is homogeneous, it is only defined up to scale. Although a full rank 3x3 matrix generally offers nine degrees of freedom, one DOF must be subtracted to account for arbitrary scaling. When registering images, this enables the computation of a homography matrix through four point correspondences, as each point in $\mathbb{R}^2$ consists of two degrees of freedom itself [Hartley and Zisserman, 2004, p. 88]. [Hartley and Zisserman, 2004, pp. 34–36] point out that a homography is induced through a world plane, meaning that all world points captured by a camera lie in a common two-dimensional plane. A special case of the common plane constraint occurs when all points lie on a common sphere. Szeliski [Szeliski, 2006] shows that this case can be found when a camera undergoes a pure rotational motion.

Panoramas, which are composed based on a common world plane constraint, will be referred to as planar panoramas in this thesis. The stitching itself is defined as planar image stitching. In the case where the optical center of a camera remains approximately stationary, the term rotational panorama is used and the stitching process is called rotational image stitching.

In panoramic image stitching, many scenarios satisfy the common plane constraint. Agarwala et al. [Agarwala et al., 2006] propose a half-automatic (user-aided) system which is capable of stitching facades of houses in a street to a large panoramic image. The result is a multi-viewpoint panorama which seems natural to an observer. User interaction is mainly required to remove or refine off-plane objects, such as cars, bicycles and pedestrians. In microscopy, image stitching gains importance for the stitching of high resolution images, which can be used as virtual slides [Altinay and Bradley, 2011; Appleton et al., 2005; Steckhan et al., 2008; Steckhan and Paulus, 2010; Zhang et al., 2013]. The use of virtual microscopy is especially important in telepathology, where microscopes can be used through a network service [Weinstein et al., 2009]. Document mosaicing is another important area, in which planar fragments of a document are stitched together. Whichello and Yan [Whichello and Yan, 1998] propose to stitch images from a digital camera or a scanner by using a priori knowledge about the overlapping segments. The approach is pixel-based and relies on cross-correlation [Russ, 2011, pp. 385-391].

A feature-based approach, using PCA-SIFT, is presented by Liang et al. [Liang et al., 2006]. In the first step, perspective distortions are removed and the registration problem is reduced to a three-dimensional problem, leaving only two translational parameters and one for scaling. The scale and translations are then recovered through a histogram-based multi-scale approach. These parameters are further refined via a cross-correlation procedure. In the end,

a projective transform for each image is recovered and the images are composed seamlessly. Such a feature-based approach is also feasible for the use on mobile devices, as shown by Hannuksela et al. [HANNUKSELA et al., 2007]. During the process, online motion estimation is performed to guide the user when taking pictures. The high resolution image is processed offline on a desktop computer and the resolution is found to be sufficient for OCR (optical character recognition).

When dealing with rotational panoramas, the eight DOFs which a planar homography uses can be reduced to three DOFs, one for each rotation direction [BROWN and LOWE, 2007; SZELISKI and SHUM, 1997]. Szeliski and Shum [SZELISKI and SHUM, 1997] advocate this approach and propose extracting three-dimensional camera rotations directly. It is elaborated that this approach is significantly more robust with regards to image registration. Additionally, the computational complexity and thus the computation time is reduced. A similar approach is presented by Brown and Lowe [BROWN and LOWE, 2007], which, in contrast to [SZELISKI and SHUM, 1997], uses feature-based registration. A rotational homography can be obtained by performing singular value decomposition (SVD) [FAUGERAS and LUSTMAN, 1988; MALIS and VARGAS, 2007; VARGAS and MALIS, 2005]. Based on this foundation, a number of other methods have been proposed, such as sequential image stitching for mobile devices [XIONG and PULLI, 2009b], panorama painting for mobile devices [XIONG and PULLI, 2010], and super resolution panoramas [CAPEL and ZISSERMAN, 1998].

In all these methods, a panorama type is assumed to be given. Planar as well as rotational stitching pipelines generally only reject images which they consider to be outliers, but an evaluation of the panorama type is not part of the registration procedure. We address this problem in [HERBON et al., 2014b] and the integration into wood pile surveying methods is discussed in Chapter 4 (*Adaptive image stitching*).

3.4 3D reconstruction / structure from motion

Structure from motion (SfM) is defined as the reconstruction of a scene (structure) and camera movement (motion) from a set of images. Similar to human vision, the structure can be recovered by matching features across images [MOESLUND and GRANUM, 2001; TREUE et al., 1991; YOUNG et al., 1993]. Hence, a three-dimensional structure can be retrieved from two dimensional images. SLAM (simultaneous localization and mapping) is a similar task to SfM, with optimization toward real time pose estimation and mapping of the environment. It is mainly used for robot navigation [NÜCHTER, 2009]. Similar to image stitching methods, direct und feature-based SLAM and SfM methods are distinguished between [IRANI and ANANDAN, 2000; TORR and ZISSERMAN, 2000]. While direct methods aim to minimize a distance measure for each pixel of overlapping images, feature-based methods extract and

match features through which the relations between images can be established [TORR and ZISSERMAN, 2000]. In this thesis, only feature-based methods are used, since this approach is more suitable for a more general structure and motion setup, especially with regards to large baseline geometries (cases in which the camera translation is large), as shown by Torr and Zisserman [TORR and ZISSERMAN, 2000].

A variety of different feature detectors and descriptors have been proposed, which are suitable for structure from motion. It is vital to distinguish between detectors and descriptors, as detectors usually find corners (points with high peaks in the gradient domain), blobs, or regions, and descriptors extract a unique description of the region around a feature point which can be used for matching [TUYTELAARS and MIKOLAJCZYK, 2008]. Tuytelaars and Mikolajczyk [TUYTELAARS and MIKOLAJCZYK, 2008] list some of the most prominent feature detectors, which are: the Harris corner detector [HARRIS and STEPHENS, 1988], SU-SAN [SMITH and BRADY, 1997], Harris-affine [MIKOLAJCZYK and SCHMID, 2004], Hessian affine region detector [MIKOLAJCZYK and SCHMID, 2002], salient regions detector [KADIR and BRADY, 2001], intensity-based regions [TUYTELAARS and VAN GOOL, 2000], MSER [MATAS et al., 2004], and superpixels [MORI et al., 2004; REN and MALIK, 2003]. With these feature detectors as a foundation, a number of feature descriptors and efficient implementations have been developed.

Arguably the most prominent of these are SIFT [LOWE, 2004], PCA-SIFT [KE and SUK-THANKAR, 2004], SURF [BAY et al., 2008; BAY et al., 2006], FAST [ROSTEN and DRUMMOND, 2005; ROSTEN and DRUMMOND, 2006], ORB [RUBLEE et al., 2011], and recently A-KAZE [ALCANTARILLA et al., 2012; ALCANTARILLA et al., 2013]. In this thesis, SIFT features will be used, as they are well supported by many structure from motion pipelines. Alcantarilla et al. [ALCANTARILLA et al., 2013] demonstrate that SIFT performs well for many different data sets in terms of repeatability, although A-KAZE generally performs best. As long as images can be registered at all, which is usually the case for images from the *HAWKwood* database, the choice of feature descriptor in structure from motion is of less importance compared to the SfM method and non-linear refinement (bundle adjustment, BA).

Most structure from motion and SLAM pipelines work incrementally, which, according to Moulon et al. [MOULON et al., 2013a], means that an initial pair is chosen from which the first part of the scene and the initial camera poses are reconstructed. More images are added incrementally via resectioning. After each resectioning, bundle adjustment is performed, as proposed for example by Wu et al. [WU et al., 2011]. A number of incremental SfM approaches have been presented; some often used pipelines include [FARENZENA et al., 2009; GHERARDI et al., 2010; GHERARDI and FUSIELLO, 2010; MOULON et al., 2013a; SNAVELY et al., 2006; SNAVELY et al., 2008; WU, 2011; WU, 2013]. Moulon et al. [MOULON et al., 2013b] point out that the incremental approach leaves room for optimization with regards

to drift, which is induced through the incremental extension of the 3D model. Furthermore, the choice of the initial pair has great influence on the overall calibration result. The global approach of Moulon et al. [MOULON et al., 2013b], which computes relative global rotations and translations first, then performs bundle adjustment, is proven to be superior in terms of accuracy of the reconstruction, running time, and scalability. For global methods, the explicit choice of an initial pair is not necessary, which improves the reduction of drift. By using the trifocal tensor [HARTLEY and ZISSERMAN, 2004, pp. 365–407], feature matches are not only computed between image pairs, but between image triplets. This way, the drift is distributed more evenly across the scene; hence, better scalability is achieved. For these reasons, the proposed global pipeline seems to be a good fit for three-dimensional wood pile surveying and in Chapter 7 (*Three-dimensional surveying*) how it can be employed for surveying on mobile devices will be discussed.

All these methods, incremental or global, are feature-based and therefore only reconstruct a sparse set of 3D points (point cloud). To obtain a quasi-dense point cloud, patch-based multiple view stereo algorithms can be applied after reconstruction. Furukawa and Ponce [FURUKAWA and PONCE, 2007; FURUKAWA and PONCE, 2010] propose detecting uniformly distributed corner and blob features which are then matched across multiple images. For each feature in one image, all features in the other images which satisfy the epipolar constraint (they are close to the epipolar line in terms of an L2 norm), are considered to be potential matches. They are then triangulated and filtered to ensure photometric consistency. In the last step, patches are created, expanded, and filtered from the matched features. This way, a quasi-dense set of patches can be obtained, which aims to cover the entire object surface. In later parts of this thesis distinctions will be made between sparse and dense reconstructions, with sparse reconstruction referring to the initial SfM model and dense reconstruction meaning the reconstruction after quasi-dense surface patch expansion.

3.5   Wood pile surveying systems used in the wood industry

In the field of wood log surveying, some alternatives to photo-optical and photogrammetric surveying have been proposed. Laser-scanning of timber trucks is performed by Nylinder et al. [NYLINDER et al., 2008], where logs of different dimensions are scanned and the stems on the truck are reconstructed in 3D. Furthermore, a comparison between manual and automatic surveying shows that the laser scanning procedure tends to underestimate the volume by up to 5.0%. The systematic error is shown to be correlated to the log diameter. With decreasing diameter, the negative error increases. Funck et al. [FUNCK et al., 1993] propose to measure the roundness of wood logs with a laser scanning system. Wood grain can also be detected by laser light scattering, as shown in [SIMONAHO et al., 2004].

Since laser-scanning systems are expensive and difficult to transport, a number of photo-optical surveying systems have been established in the wood industry. The *Polterluchs*-System [JAEGER et al., 2014] from *Wahlers Forsttechnik* and *Visiosens* is able to count wood logs in a pile and document the results based on digital images [FORBRIG et al., 2012]. It does not survey the actual wood mass. Ullrich and Krätzschmar [ULLRICH and KRÄTZSCHMAR, 2012] from *iABG* introduce a wood log segmentation pipeline based on edge, color, and texture features, where wood logs are approximated by an ellipse. A measuring stick is used as a scale reference. No information is provided about the type of reconstruction model that is used. The proposed method can be used for random sampling for the determination of the wood log size distribution, but it is unclear how it could be employed for wood pile surveying.

Jørgensen [JØRGENSEN, 2008] from *Heidegesellschaft* proposes a pipeline which requires a wood pile to be divided into sections with a measuring stick as a scale reference. Digital images of these sections are taken and later evaluated on a desktop PC, where the images are preprocessed and then manually segmented. This approach exhibits two significant disadvantages. Firstly, the results are not computed on-site. This means that image acquisition errors lead to the failure of the entire surveying pipeline and the images have to be retaken. When considering that a person needs to drive into the woods again, locate the wood pile, and then take the pictures once more, it becomes clear that this is a very time-consuming task. Secondly, the wood pile is not reconstructed in its entirety, only a side view is available. Hence, the user cannot be sure that all photos actually belong to the wood pile, and the wood pile cannot be compared to a digital model when negotiating the sales price on-site.

In accordance with their patent application [SCHELLER et al., 2014], Foese et al. [BARKOWSKI, 2013; FOESE et al., 2012] from *AFoRS* offer a marker-based surveying method which measures individual sections of the wood pile. The overlapping images are stitched based on the detected markers. The surveying error (the definition is unspecified) is said to be less than 5%. While this error can be assumed to be sufficiently low for surveying purposes, the proposed method exhibits two problems. Firstly, applying markers to the wood pile is not only time-consuming, the physical properties of the wood log cut surfaces contaminate the markers with resin, which makes the markers sticky and increasingly difficult to handle. Secondly, the method requires an active internet connection. This is a constraint which is difficult to fulfill in rural areas, as the coverage is usually not sufficient for this task. Off-site processing is possible, but the same problems as encountered by *Heidegesellschaft* then apply.

Dralle A/S [DRALLE A/S, 2014] offers a vehicle-mounted camera system, as described by their patent [TARP-JOHANSEN and DRALLE, 2012], which consists of a stereo camera, a lighting unit, and a GPS sensor. The vehicle drives along the side of the wood pile and takes pictures of it. According to [LOHSE and HEUER, 2014], the repeatability is 2% (no confidence interval is given). Although this is no proof of the actual measurement accuracy,

it suggests that the system is reasonably precise. Because a stereo camera system is used, no scale reference is needed. The scale factor can be directly obtained from camera calibration information. The measurement evaluation is performed onboard the vehicle. In rough terrain, it can prove difficult to access certain wood piles with a vehicle, which is why the method is well-suited for large, easily accessible piles, but lacks flexibility and portability for small to medium off-road piles.

From this overview of the industrial surveying systems, it becomes clear, that none of them fulfills all the constraints that were found to be crucial in Section 2.1. The motivation for this thesis is therefore based on the non-existence of surveying methods which adequately meet all the requirements of modern forestry. As shown in Section 3.2 (*Recognition and segmentation of wood logs and clustered objects*), wood pile surveying is an active research area. This is due to the fact that a solution for the academic challenges imposed by wood pile surveying constraints is yet to be presented.

3.6   Related applications

In this chapter, the challenges and the related work of wood pile surveying have been discussed. The methods proposed in this thesis can be transferred to a number of other problems with different applications. These similar problems will be reviewed in this section with a brief outlook on possible future academic work.

Gutzeit and Lukas [GUTZEIT and LUKAS, 2013] provide a summary of applications where self-similar objects are being segmented. Four specific cases are listed, namely the segmentation of wood log faces, apples, fruit in general, and fish. The quality of food can be inspected by computer vision methods, as demonstrated by [BROSNAN and SUN, 2004]. A segmentation procedure, which is applied to apples, is shown by Bulanon et al. [BULANON et al., 2002], where a simple thresholding operation is used. In [LEEMANS et al., 1998] and [LEEMANS et al., 1999] a method is proposed to find defects on apples through a color-based segmentation approach. An apple recognition rate of 81% is achieved by the pipeline presented by Yongsheng et al. [YONGSHENG et al., 2009], which is based on the k-means algorithm. Tabb et al. [TABB et al., 2006] successfully segment between 85% and 96% of apples from video data by applying Gaussian mixture models. The detection of olives is addressed by Riquelme et al. [RIQUELME et al., 2008], who successfully find defects in different types of olives. An overview of computer vision methods which can be used to locate fruit on trees can be found in [JIMENEZ et al., 2000].

The detection and segmentation of fish is comparable to the surveying of wood piles. Although the location of fish is not stationary in a temporal sense, fish usually appear in a clustered fashion. Spampinato et al. [SPAMPINATO et al., 2010] demonstrate an approach with

which multiple instances of fish in a single image can be recognized. To achieve recognition, texture features and shape features are combined. This approach also seems feasible for the detection of wood logs. The correct detection rate is shown to be 92%. Huang et al. [HUANG et al., 2012; HUANG et al., 2013; HUANG and HUANG, 2014] also aim for the detection of fish based on a set of features, consisting of color, shape, and texture information. Li [LI, 2012] takes fish recognition one step further by proposing a method to segment parts of fish, such as the tail, the contour, or the curvature. The extraction of fish features is addressed in [CHUANG et al., 2014]. Fourier descriptors are used to locate anatomical parts of fish, from which a feature descriptor is obtained through a supervised learning approach. Appearance, location, and size are extracted for each body part.

A research field, which closely resembles wood log segmentation, is the recognition of cells. Like wood logs, cells appear in clusters of similar sizes and shapes, and segmentation in microscopic images is important for medical image processing. [WU et al., 1995] shows that the segmentation of cells in low-contrast images under uneven illumination and intensity conditions is possible. The cell is separated from the background and the cell boundary is extracted. A dataset of hand-segmented fluorescence microscopy images is provided by Coelho et al. [COELHO et al., 2009]. They implement and compare several computer vision methods which employ techniques such as thresholding, watershed segmentation, active masks, and region merging. The watershed algorithm by Lin et al. [LIN et al., 2003] was found to perform best on the test dataset.

Dima et al. [DIMA et al., 2011] compare state-of-the-art algorithms in cell segmentation. Active contours and convex energy functions are implemented by Bergeest and Rohr [BERGEEST and ROHR, 2011] and are tested on a set of microscopy images with different cell types, through which a quantitative evaluation and a comparison to previous approaches is undertaken. Cell recognition can also be achieved through template matching, as shown in [CHEN et al., 2013]. Hagwood et al. [HAGWOOD et al., 2012] perform pixel-based evaluations of classification and segmentation algorithms, such as Otsu-based thresholding, k-means clustering, Canny edge detection, and watershed segmentation. The segmentation of cell nuclei is addressed in [BERGEEST and ROHR, 2012], where level sets and convex energy functions are used in combination with active contour segmentation. The detection of melanocytes, which usually appear in spatially limited clusters, is important for the diagnosis of skin melanoma. Lu et al. [LU et al., 2013] examine histopathological images in which melanocytes are segmented through a double ellipse descriptor. This descriptor is specifically designed to detect biological features of melanocytes. Held et al. [HELD et al., 2011] show the feasibility of segmenting clustered macrophages in microscopic images through watershed and level set methods.

Many other methods in cell segmentation and recognition have been proposed, which

cannot all be discussed in the scope of this thesis. The pure number of publications shows the importance of academic work in this research area. Most methods are directly transferrable to wood log segmentation and vice versa, except for the fact that wood logs usually exhibit a non-uniform background. This characteristic means that many of the simpler cell segmentation methods fail and calls for some of the more advanced methods to be employed for this task.

Apart from the actual recognition of wood logs, some methods have recently been presented which go beyond wood pile surveying and aim to characterize individual wood logs through biometric features. The goal of these approaches is to create a photo-based tracking pipeline for wood logs. In a preliminary study [SCHRAML, 2013], the traceability of wood logs based on biometric features is investigated. For this purpose, log end images are taken and the annual ring pattern of the cross-section is described as a texture pattern through biometric features. The segmentation of the cross-section is a key aspect for the characterization of a wood log, as is the estimation of the pith (the biological center of a wood log). A technique for the robust estimation of the pith is reviewed in [SCHRAML and UHL, 2013]. It is based on different Fourier spectrum analysis methods. Peak analysis and principal component analysis give the best results for pith estimation.

The starting point for the description of a wood log is the segmentation of the cross-section. Unlike the related work discussed in Section 3.2 (*Recognition and segmentation of wood logs and clustered objects*), where segmentation procedures for clustered wood logs are shown, Schraml and Uhl [SCHRAML and UHL, 2014a] propose a method for the segmentation of the cross-section in an image, when only one cut surface is visible. This is achieved through a similarity-based region growing algorithm combined with intensity histograms as texture features and a histogram distance measure. To further characterize and describe the properties of a wood log, Schraml and Uhl [SCHRAML and UHL, 2014b] compute temporal and longitudinal variances of wood log cross-sections. These variances are then used as input for a biometric fingerprint matching technique, where the goal is to match images of the same wood log cross-section from different points in time. The annual ring pattern-based matching is also discussed in [SCHRAML et al., 2014], where biometric feature fusion is found to increase the robustness of the matching step. The combination of annual ring patterns and shape features ensures robust recognition across cross-section variations.

In this section it was shown that there exists a wide variety of related applications which are directly relevant to wood log surveying and vice versa. This increases the motivation for academic research in this area, since the contributions of this thesis are of importance to all of the aforementioned computer vision tasks.

Chapter 4

Adaptive image stitching

In this chapter, a novel method for adaptive image stitching is presented, which we originally published in [HERBON et al., 2014b]. The proposed method is applicable to the general stitching scenario and serves as the basis for the two-dimensional surveying approach of this thesis. Based on the original paper, the following sections in part summarize and extend the described method. The goal in the scope of this thesis is to take photos of a wood pile front surface from different viewpoints and fuse these images together, while determining whether or not the panorama was properly captured. Although image stitching itself is a well-studied research field, the stitching of wood pile front surfaces imposes additional constraints. It will be discussed what measures should be taken to ensure proper image acquisition and how existing approaches can be incorporated into the novel stitching method. If the panorama is valid according to the standards defined below, then it can be used as the input for a wood log detection and segmentation method. Figure 4.1 shows an example of the desired output, which was generated by the proposed approach.

Figure 4.1: Exemplary output of the proposed stitching pipeline, composed of eight input images.

4.1 Objective

When performing image stitching, the goal is to seamlessly fuse overlapping photos into a panoramic image. In Section 3.3 (*Panoramic image stitching*), it was shown that one out of two constraints for the scene geometry must be enforced in order to retain validity of the homographic registration model. In other words, the images must either all depict a common world plane or be taken during a pure rotation about the camera's optical center (which is a special case of the common world constraint). When registering images for planar stitching (common world plane constraint), eight degrees of freedom are necessary, which are encapsulated into a general homography matrix. For the registration of rotational panoramas, only three degrees of freedom are required, which greatly simplifies the stitching procedure. The full homographic stitching model can also be applied to rotational image stitching, but should be avoided if possible. The reason for this is that when applying an eight parameter model to a problem with three degrees of freedom, the solution is ambiguous and the pipeline becomes unstable [BROWN and LOWE, 2007; SHUM and SZELISKI, 2000; SZELISKI, 2006; SZELISKI and SHUM, 1997].

In the case of wood pile surveying, a user must not perform rotational image stitching, because a rotational model induces distortions towards the edges of the pile. The resulting panoramic image should rather come close to an orthogonal projection of the wood pile front surface, as this is the model in which measurements are applied in forestry surveying techniques. If one were able to approximate the camera translation based on image correspondences, the validity of the camera trajectory and thus the correct homographic model could be established.

The recovery of the camera trajectory is usually performed through multiple view auto calibration techniques, such as visual odometry, SLAM, or structure from motion. In the *Research challenges* section the constraints for multiple view reconstruction were discussed, which include the so-called planarity degeneracy. According to [HARTLEY and ZISSERMAN, 2004, pp. 295–296], the essential / fundamental matrix-based auto calibration requires that not all world points are coplanar. This means that a degenerate configuration occurs when all scene points lie in a common world plane, which is the case when taking pictures of wood pile front surfaces. For this reason, the camera trajectory cannot be recovered through essential matrix decomposition. In the next sections it will therefore be established how the camera translation and rotation can be approximated based on a set of homographies which are relating the input images.

Apart from validating the stitching model, the proposed method yields other advantages over conventional pipelines. Firstly, rotational stitching can be performed significantly faster when choosing a three parameter model over the eight DOFs model. Secondly, as discussed above, the stitching process becomes more stable when using the correct number of DOFs.

The third and most visually significant advantage is that a proper projection surface can be chosen, so as to minimize distortions. In the case of rotational stitching, a spherical or cylindrical mapping surface can be used, while a planar mapping surface works best for image sequences of planar scenes. This goal builds towards the statement of Szeliski in [SZELISKI, 2006], which reads as follows:

> "Automatically making this [composition surface] selection and smoothly transitioning between representations based on the extent of the panorama is an interesting topic for future research."
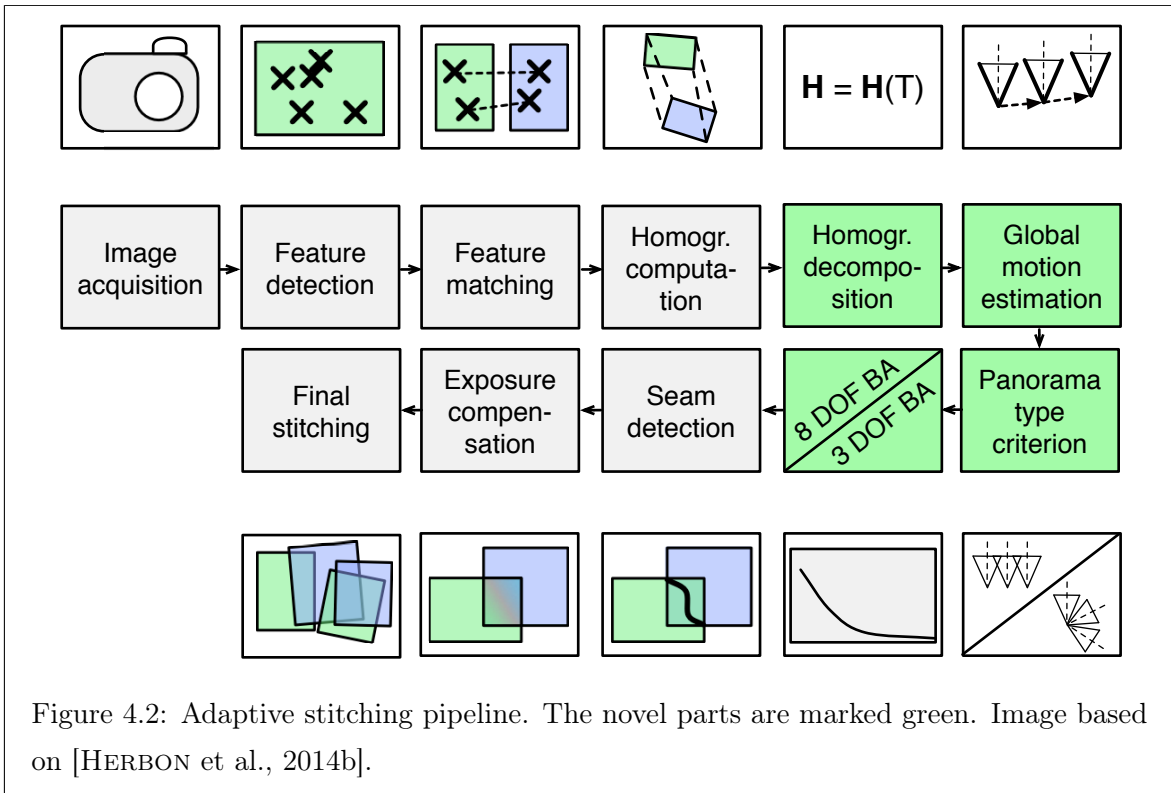
4.2 Adaptive stitching pipeline

Several approaches for image stitching pipelines have been proposed, most of which address rotational image stitching [BROWN and LOWE, 2007; SZELISKI and SHUM, 1997]. Based on these methods, a new adaptive approach is described in this section. The main contributions of the proposed method are:

1. Definition of the panorama type criterion (PTC)
2. Adaptive (planar and rotational) bundle adjustment
3. Selection of the correct mapping surface
4. Optimization for mobile devices

Figure 4.2 shows the different steps of the proposed pipeline, with the green boxes marking the novel parts that are not included in standard pipelines. An image sequence without additional information is taken as input. For each image the features are computed and feature description is performed. In the proposed approach, SIFT features are chosen for robustness, applicability to general image content, and wide availability, but any other feature-type can be used as well. The choice of feature may also be based on potential a priori knowledge of the content, where, e.g., blob-based features can be used for textureless content.

The feature-matching problem is solved by using FLANN [MUJA and LOWE, 2009; MUJA and LOWE, 2012] (Fast Library for Approximate Nearest Neighbors). From the feature matches, a homography for each image pair is computed through a RANSAC-based (Random Sample Consensus) approach [FISCHLER and BOLLES, 1981] if sufficient inliers are available. Outliers are rejected based on the geometric model and the homography is refined from inliers only, as proposed by Zhang and Kosecka [ZHANG and KOSECKA, 2006a; ZHANG and KOSECKA, 2006b]. A confidence level is computed for every homography, based on the number of inliers for the feature matches in relation to the number of outliers, which gives an indication of the validity of the homography. If the confidence does not exceed a certain

Figure 4.2: Adaptive stitching pipeline. The novel parts are marked green. Image based on [HERBON et al., 2014b].

threshold, the homography is rejected and the stitching process is only performed for the largest subset, in which the images were registered correctly.

The method so far is largely common practice in stitching pipelines, such as those proposed by [BROWN and LOWE, 2007; SZELISKI and SHUM, 1997]. After the homography computation, the novel adaptive stitching approach takes place. The first step is to perform homography decomposition based on the method provided by Borgstadt and Ferrier [BORGSTADT and FERRIER, 2001]. Each homography is decomposed into a rotational and a translational component, $R \in \mathbb{R}^{3\times3}$ and $t \in \mathbb{R}^{3\times1}$. The details of the parametrization and decomposition will be discussed in Section 4.4 (*Homography decomposition*). In the next step, a global camera motion estimation is computed based on the homography decomposition. Any camera can be used as the reference camera; in practice the first camera, or the camera in the center, of the sequence, are chosen.
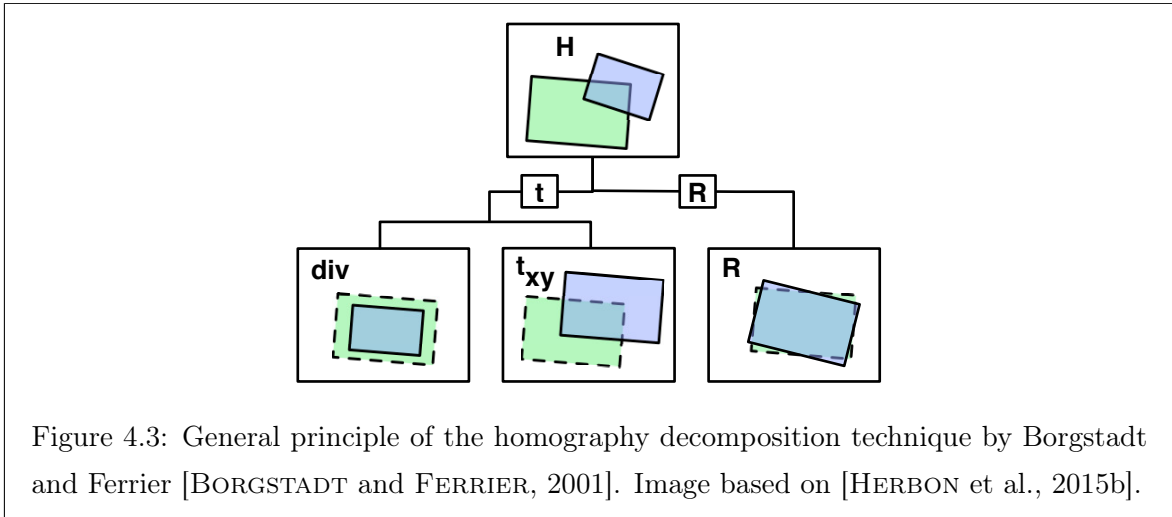
The main contribution of this approach is the definition of the panorama type criterion. Based on the global motion, the criterion determines whether a rotational or a planar panorama was captured. In the case of a rotational panorama, the panorama type criterion can account for small translational motions that are usually inevitable when capturing images on a handheld device. A detailed description will be given in Section 4.6 (*Panorama type criterion*). The last step of the novel approach is adaptive bundle adjustment. Depending on the decision of the panorama type criterion, a suitable model is chosen, which is then used

for nonlinear Levenberg-Marquardt optimization [LEVENBERG, 1944; MARQUARDT, 1963; MORÉ, 1978]. Concluding the stitching pipeline, standard seam detection, exposure compensation, and final stitching are performed. While the first two of these steps are independent of the panorama type, the final stitching result is strongly influenced by the chosen mapping surface, as explained by Szeliski [SZELISKI, 1996; SZELISKI, 2006]. Depending on the image acquisition geometry, either a planar or cylindrical / spherical mapping surface should be chosen, so as to minimize distortions. By making this distinction, a visually appealing, as well as a geometrically correct, result image can be obtained.

4.3 Optimization for mobile devices

Choosing a three parameter model for rotational image stitching simplifies the stitching process and reduces the computation time. Apart from this optimization, additional strategies can be followed to further enhance the stitching approach specifically for mobile devices. The choice of feature detector is an important aspect of the image registration step. In Section 3.3 (*Panoramic image stitching*), different feature detectors which can be used alternatively were discussed. Since the stitching process takes place directly on the mobile device, parallelization, through which feature detection and extraction can be performed during image acquisition, is possible. By considering this strategy, the choice of feature detector is relativized, as the extraction time becomes critical only if it takes more than a few seconds - the time which a user needs to capture consecutive frames. Additionally, the image resolution can be reduced for feature detection, while the end result can still be stitched in full resolution. By performing bundle adjustment, the quantization errors, which are introduced by scaling the image, can be largely compensated.

Many stitching algorithms [BROWN and LOWE, 2007; SZELISKI, 2006] assume an unordered image collection. While this assumption must be made for the general case, information about image acquisition can be incorporated into the registration process. This is especially useful if capturing and processing occur on the same device. Xiong and Pulli [XIONG and PULLI, 2009a; XIONG and PULLI, 2010] propose to perform sequential image stitching. Since the order of the images is known, the registration can be performed between consecutive pictures only. In other words, each image I_i must now only be registered with images I_{i-1} and I_{i+1}. The complexity for the panorama recognition problem is thus reduced from a quadratic problem ($O(n^2)$) to a linear problem ($O(n)$). Furthermore, global motion estimation can also be performed sequentially, similar to SLAM-based approaches, as will be shown in Section 4.5 (*Motion estimation*).

Figure 4.3: General principle of the homography decomposition technique by Borgstadt and Ferrier [BORGSTADT and FERRIER, 2001]. Image based on [HERBON et al., 2015b].

4.4 Homography decomposition

The panorama type criterion relies on accurate motion estimation. The basis for computing a global motion is the decomposition of the homography matrices. There exist several decomposition techniques, such as [BORGSTADT and FERRIER, 2001; FAUGERAS and LUSTMAN, 1988; MALIS and VARGAS, 2007; ZHANG and HANSON, 1996]. We have previously been successful in applying the method provided by Borgstadt and Ferrier [BORGSTADT and FERRIER, 2001], as shown in [HERBON et al., 2014b] and [HERBON et al., 2015b]. The goal of homography decomposition is to extract a rotational and a translational component from the homography matrix, which offers insight into the camera trajectory. Figure 4.3 shows the principle of this procedure, based on the connotation of Borgstadt and Ferrier. The initial step of the approach in [BORGSTADT and FERRIER, 2001] is to parametrize a given homography matrix H_k, as shown in Equation 4.1.

$$H_k = \begin{bmatrix} R_{k,1,1} & R_{k,1,2} & t_{k,1} \\ R_{k,2,1} & R_{k,2,2} & t_{k,2} \\ R_{k,3,1} & R_{k,3,2} & t_{k,3} \end{bmatrix} \tag{4.1}$$

The first two columns of H_k correspond to the first two columns of the rotation matrix R_k. This matrix R_k is the rotational component of the homography, as opposed to the translational component t_k. Since a homography only offers eight degrees of freedom, H_k must be normalized. Borgstadt and Ferrier propose to perform normalization by $H_k(3,3)$.

For this purpose, the translation vector t_k is defined as

$$t_k = \begin{bmatrix} t_{k,1} \\ t_{k,2} \\ t_{k,3} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \\ \frac{1}{div_k} \end{bmatrix} \tag{4.2}$$

where div_k is a measure for the scaling induced by the homography, called the divergence. This scaling factor is equivalent to a translation along the z-axis when the assumption of a pinhole camera model is made. Borgstadt and Ferrier provide a definition of div_k, for which the general, normalized homography of Equation 4.3 is presumed. The definition of div_k is then introduced as shown in Equation 4.4.

$$H_k = \begin{bmatrix} a_k & b_k & c_k \\ d_k & e_k & f_k \\ g_k & h_k & 1 \end{bmatrix} \tag{4.3}$$

$$div_k = \sqrt{\frac{a_k^2 + b_k^2 + d_k^2 + e_k^2 + g_k^2 + h_k^2}{2}} \tag{4.4}$$

The normalized parametrization of H_k is denoted as H_k' and is given by the following equation.

$$H_k' = \frac{1}{div_k} \begin{bmatrix} div_k R_{k,1,1} & div_k R_{k,1,2} & div_k x_k \\ div_k R_{k,2,1} & div_k R_{k,2,2} & div_k y_k \\ div_k R_{k,3,1} & div_k R_{k,3,2} & 1 \end{bmatrix} \tag{4.5}$$
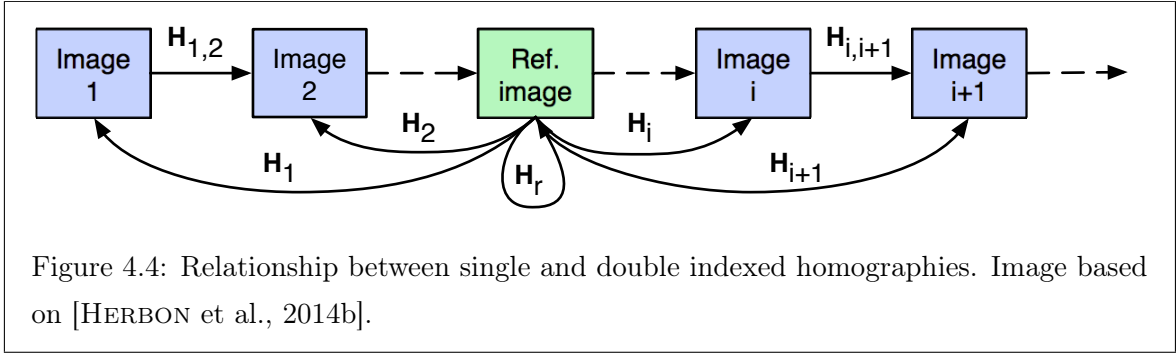
The definition of the displacement vector t_k alone is not sufficient for the approximation of the camera trajectory. Since H_k' only includes the first two columns of the rotation matrix R_k, the third column of R_k must be reconstructed based on the fact that rotation matrices are orthogonal. The columns of R_k are denoted as r_1, r_2, and r_3, so that $R_k = [r_1 \ r_2 \ r_3]$. The normalized rotation matrix, where each column vector is normalized through $|r_i| = 1$, is defined as $R_k' = [r_1' \ r_2' \ r_3']$. Due to the orthogonality property, r_3' can be defined as the cross product of r_1' and r_2', which completes the homography decomposition procedure.

$$r_3' = \frac{r_1' \times r_2'}{|r_1' \times r_2'|} \tag{4.6}$$

With the extraction of the rotational and translation component, estimation of the camera trajectory is possible. This will be shown in the next section.

4.5   Motion estimation

The motion estimation technique builds on the optimization for mobile devices that enables the registration to be constricted to consecutive frames. For this purpose consider images I_i

Figure 4.4: Relationship between single and double indexed homographies. Image based on [HERBON et al., 2014b].

with $i = 1 \ldots n$. Each consecutive image pair is related by a double indexed homography $H_{i,i+1}$, which relates images I_i and I_{i+1}. The establishment of a global motion requires a reference image I_r, which introduces a global coordinate system. Based on the double indexed homographies, a perspective transform in the global coordinate system that transforms the image I_i to the global coordinate system, must now be computed for each input image. This transform is denoted as the single indexed homography H_i. Figure 4.4 shows the relationship between the double and the single indexed homographies.

I_r defines the global coordinate system and thus $H_r = [I]$, with $[I]$ being the 3×3 identity matrix. The definition of the global, single indexed homographies is given by Equation 4.7, where they are defined as the product of consecutive double indexed homographies.

$$
H_i = \begin{cases}
\prod\limits_{j=i}^{r+1} H_{j-1,j} = H_{i-1,i} \cdot_{\ldots} \cdot H_{r,r+1} & \text{for } i > r \\[2ex]
\left[I \right] & \text{for } i = r \\[2ex]
\prod\limits_{j=i}^{r-1} H_{j,j+1} = H_{i,i+1} \cdot_{\ldots} \cdot H_{r-1,r} & \text{for } i < r
\end{cases}
\tag{4.7}
$$

The distinction between different cases becomes necessary due to the fact that matrix multiplications are not commutative. Three cases, namely $i > r$, $i = r$, and $i < r$, must be considered. The order in which matrices are multiplied must begin with the homography closest to the image of interest and must end with the homography that links to the reference image. This process will be referred to as homography chaining.

The homography decomposition produces a translational and a rotational component, from which the camera center can be recovered. In the pinhole camera model the ith camera center is defined, based on [HARTLEY and ZISSERMAN, 2004, p. 156], as shown in Equation 4.8.

$$
C_i = -R_i^{-1} \cdot t_i = -R_i^{-1} \cdot \begin{pmatrix} \frac{x_i}{div_i} \\ \frac{y_i}{div_i} \\ \frac{1}{div_i} \end{pmatrix}
\tag{4.8}
$$

The initial definition of the reference image was $H_r = [I]$, which is induced through $R_i = [I]$ and $t'_i = [0, 0, f]^T$, where f is the focal length of the camera. t'_i is then normalized by f, which results in $t_i = [0, 0, 1]^T$. The reference camera center, according to Equation 4.8, is therefore defined as shown by Equation 4.9.

$$C_r = -R_r^{-1} \cdot t_r = - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \tag{4.9}$$
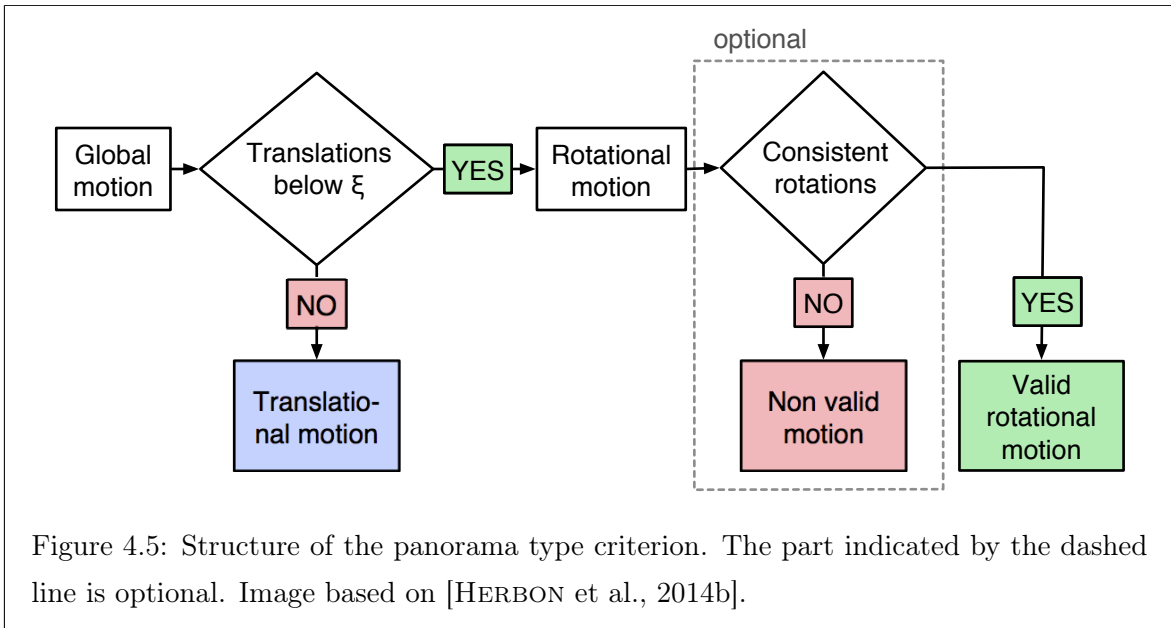
According to [HARTLEY and ZISSERMAN, 2004, p. 156] the approximation of the camera pose can be defined by Equation 4.10. Since normalized point correspondences are assumed, the intrinsic matrix K can be set to the identity matrix as $K = [I]$. Through the camera centers and camera poses, the displacement of the cameras, and thus the camera trajectory, can be obtained, which will be taken as input by the panorama type criterion in the next section.

$$P_i = K[R_i \mid t_i] = [R_i \mid t_i] \tag{4.10}$$

4.6   Panorama type criterion

The centerpiece of the proposed method is the panorama type criterion. The goal is to take as input the global motion approximation and then initiate a decision process, which results in one out of three possible outcomes: rotational panorama, planar panorama, or non-valid motion. From Figure 4.2 it becomes clear that, at this stage, the homography computation was successful, as was the global motion estimation. If the pipelines failed during these stages, then the image sequence is not eligible for panoramic image stitching. Instead, multiple view reconstruction can be performed, if sufficient overlap is given. In the previous steps, the homography decomposition method is applied to each of the single indexed homographies. The decomposition yields a translational and a rotational component. Figure 4.5 shows the decision structure of the panorama type criterion, where the first step is to validate the translational components.

Translational condition   An ideal rotational panorama should exhibit no translation, so that $t_i = \mathbf{0}$ for $i = 1 \ldots n$. In reality, panoramas taken with handheld cameras always include small translations. Shum and Szeliski [SHUM and SZELISKI, 1999] propose to restrict a user's movement during rotational panorama acquisition to a concentric circle. Based on this approach, the panorama type criterion allows for small translations, as long as a certain threshold ξ is not exceeded. For this purpose, the Euclidean distance from the reference camera center C_r is computed as $D_i = |C_r - C_i|$. If there exists some distance d_i with $d_i > \xi$,

Figure 4.5: Structure of the panorama type criterion. The part indicated by the dashed line is optional. Image based on [HERBON et al., 2014b].

then the panorama type must be planar, since, at this stage, all images have been registered successfully. The value for ξ depends on how permissive the panorama type criterion should be. A good heuristic for handheld devices, which was experimentally obtained, is $\xi = \frac{w_r}{f}$, with w_r being the reference image width and f being the focal length. Figure 4.6 depicts the translational condition with application to both kinds of panoramas.

Rotational condition If the translational threshold is not exceeded, then it is hypothesized that the panorama is rotational. In Section 4.3 (*Optimization for mobile devices*) it was discussed that the complexity can be greatly simplified by only stitching consecutive frames. For this reason, the rotational motion can be restricted to a single row panorama. The constraint is optional and if a single row panorama should be enforced, then this entails that the rotation direction must be consistent. This can be ensured by computing the angle between the optical axes of adjacent cameras. From Figure 4.7 it becomes clear that the optical axis of a camera in global coordinates can be computed via the optical center of the camera $\hat{p}'_{\text{oc},i} = [0, 0, -1, 1]^T$ (normalized homogeneous representation), which is the point where the optical axis passes through the image plane of the pinhole camera model. $\hat{p}'_{\text{oc},i}$ can be transformed to the global coordinate system through the inverse of the camera pose matrix P_i from Equation 4.10, as shown in Equation 4.11.

$$p_{\text{oc},i} = P_i^{-1} \cdot \hat{p}'_{\text{oc},i} \tag{4.11}$$

The direction of the optical axis vector o_i can then be computed through the subtraction of the camera center C_i (which is already in global coordinates) from $p_{\text{oc},i}$ and normalizing
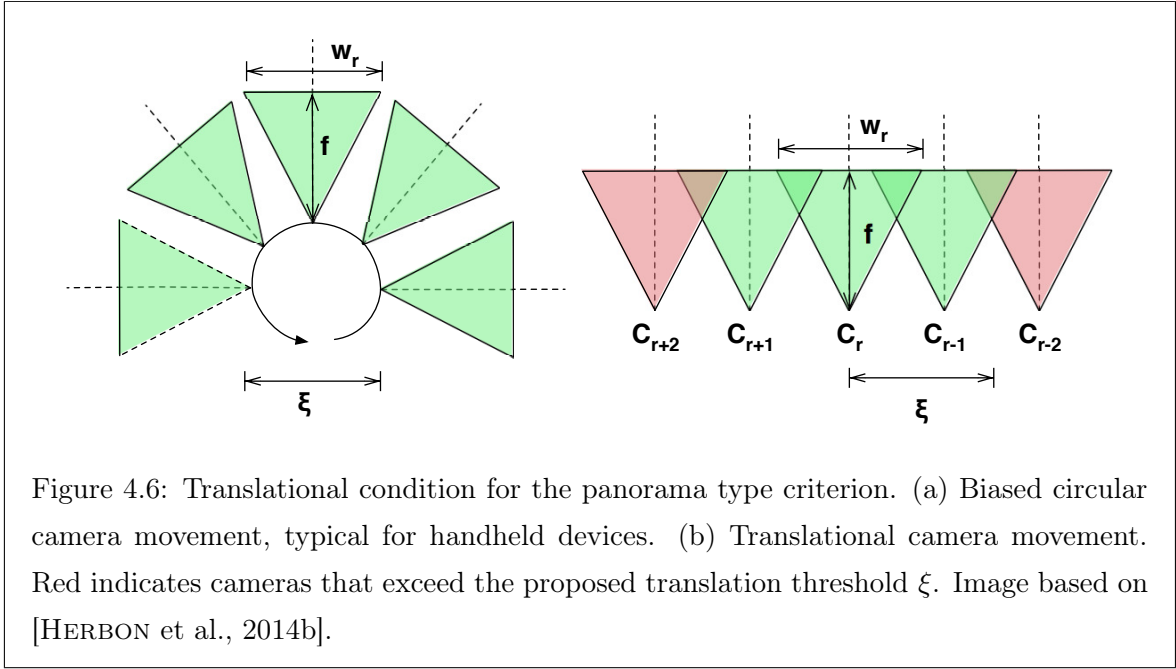
Figure 4.6: Translational condition for the panorama type criterion. (a) Biased circular camera movement, typical for handheld devices. (b) Translational camera movement. Red indicates cameras that exceed the proposed translation threshold ξ. Image based on [HERBON et al., 2014b].

by the norm of the resulting vector.

$$o_i = \frac{C_i - p_{\text{oc},i}}{|C_i - p_{\text{oc},i}|} \tag{4.12}$$

With the optical axis vector known for each camera, the angle θ between two of these vectors can be simply calculated via the inverse cosine of the dot product. $\theta_{i-1,i}$ denotes the angle between cameras $i-1$ and i.

$$\theta_{i-1,i} = acos(o_{i-1} \cdot o_i) \tag{4.13}$$

To ensure consistency between camera triplets for the rotational condition of the panorama type criterion, both of the following equations must hold true.

$$\theta_{i-1,i+1} > \theta_{i-1,i} \tag{4.14}$$

$$\theta_{i-1,i+1} > \theta_{i,i+1} \tag{4.15}$$

The rotational condition is valid, if the angle $\theta_{i-1,i+1}$ is larger than $\theta_{i-1,i}$ and $\theta_{i,i+1}$. This means that the rotation between consecutive frames must be smaller than the rotation between three frames. As soon as the rotation changes direction, $\theta_{i-1,i+1}$ will decrease and the rotational motion is considered to be non-valid. In this case, the panorama is dismissed, in the other case, the stitching pipeline continues.

Figure 4.7: Computation of the angles between the optical axes. Image based on [HERBON et al., 2014b].
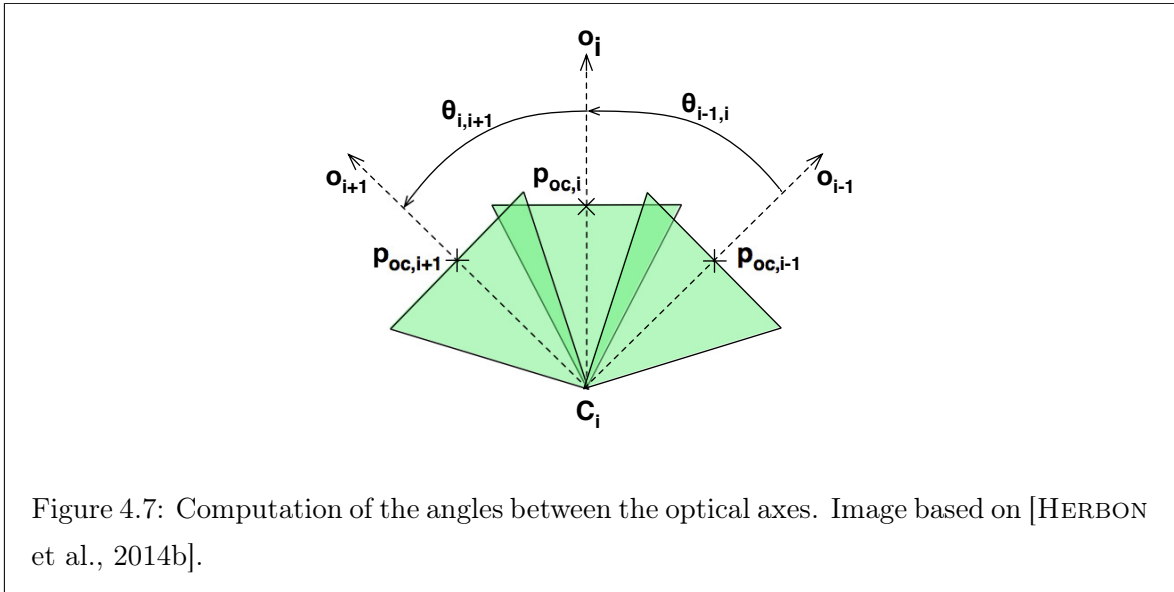
4.7 Adaptive stitching

4.7.1 Bundle adjustment

One of the major advantages of employing the panorama type criterion is the possibility of using adaptive bundle adjustment. In panoramic image stitching as well as in multiple view reconstruction, bundle adjustment is used for the optimization of intrinsic (e.g. principal point, focal length) and extrinsic (rotation and translation) parameters [TRIGGS et al., 2000]. For the proposed method, Levenberg-Marquardt optimization is performed [LEVENBERG, 1944; MARQUARDT, 1963; MORÉ, 1978], via the implementation of [BRADSKI, 2000]. As a cost function, the reprojection error, as defined by [HARTLEY and ZISSERMAN, 2004, p. 95] (Equation 4.16), is used, which the bundle adjustment aims to minimize. x and x' are perfectly matched points in the first and second image respectively, $\hat{x}$ and $\hat{x}'$ are the same points, transformed by the homography H, and D is the Euclidean distance. The selection of the reprojection error as a geometric cost function is based on the property that it accounts for uncertainties in both images.

$$\sum_j D(x_j, \hat{x}_j)^2 + D(x'_j, \hat{x}'_j)^2 \quad \text{with} \quad \hat{x}'_j = H_i \hat{x}_j \quad \forall j \tag{4.16}$$

Since distinctions are made between two panorama types, the implementation of the adaptive bundle adjustment must provide two suitable models with the proper number of degrees of freedom. In the case of a rotational model, bundle adjustment needs to optimize 3+3 DOFs, where the first three degrees account for intrinsic parameters (u, v of the principal point and the focal length f) and the second three degrees represent the possible rotation an-

gles. More parameters for the intrinsic matrix could be used, such as skew, but most cameras are sufficiently described by three parameters [HARTLEY and ZISSERMAN, 2004, p. 185]. As mentioned previously, the reduction of the extrinsic parameters to three degrees of freedom has been well studied, e.g. by [BROWN and LOWE, 2007; CHEN and KLETTE, 1999; SHUM and SZELISKI, 2000; SZELISKI, 2006; SZELISKI and SHUM, 1997]. The rotation matrix R can be decomposed to its rotation angles θ_1, θ_2, θ_3 by applying the Rodrigues formula [MEBIUS, 2007; RODRIGUES, 1816], which are then directly used for nonlinear optimization.

If the translational condition indicates a planar panorama, the second bundle adjustment model is used. It consists of a 3+8 parameter setup, where the three intrinsic parameters remain the same, but eight extrinsic degrees of freedom cover the parameters of the general homography matrix. The scale ambiguity is accounted for by enforcing a norm of 1. The results of the bundle adjustment procedure will be discussed in Section 4.8 (*Evaluation*).

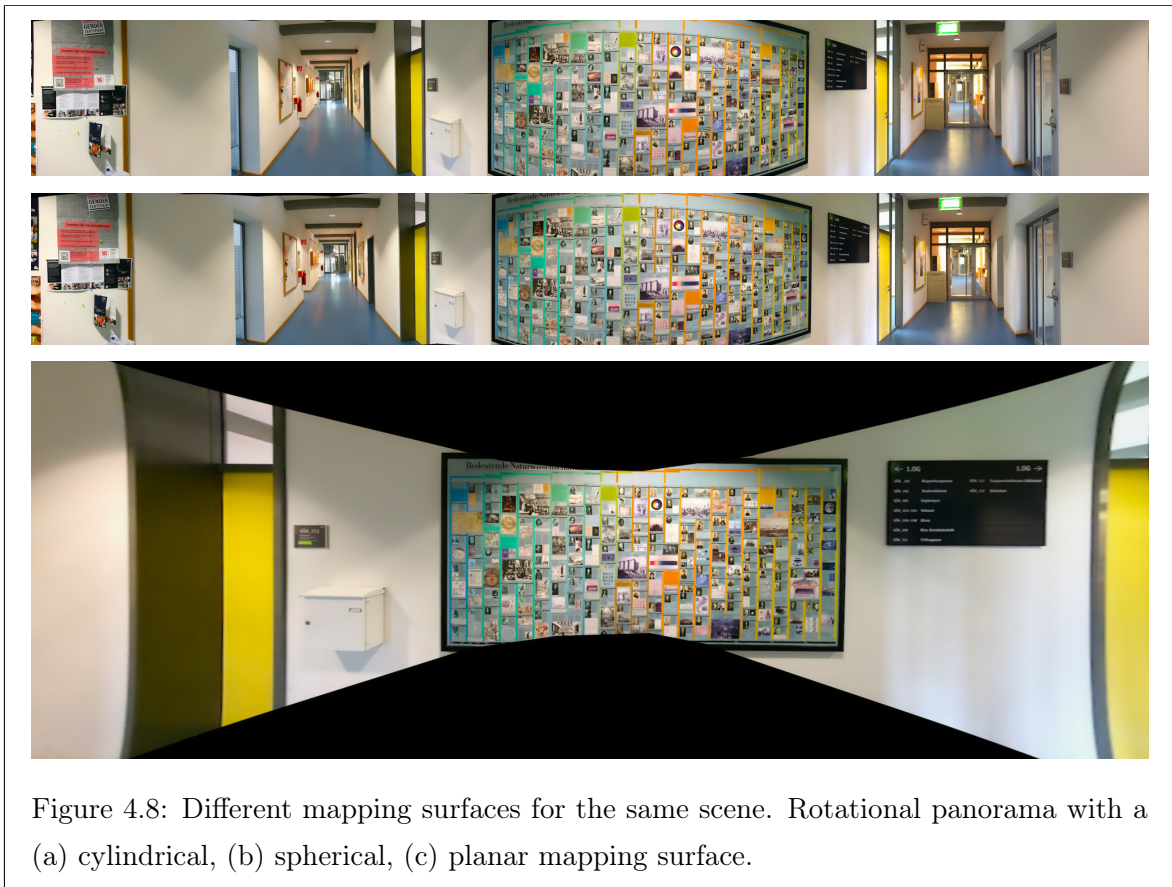### 4.7.2 Mapping surface selection

Figure 4.8: Different mapping surfaces for the same scene. Rotational panorama with a (a) cylindrical, (b) spherical, (c) planar mapping surface.

The second crucial part of the stitching process is the selection of the mapping surface (for details see, e.g., [SZELISKI, 2006]). Figure 4.8 (a)-(c) shows three examples of standard
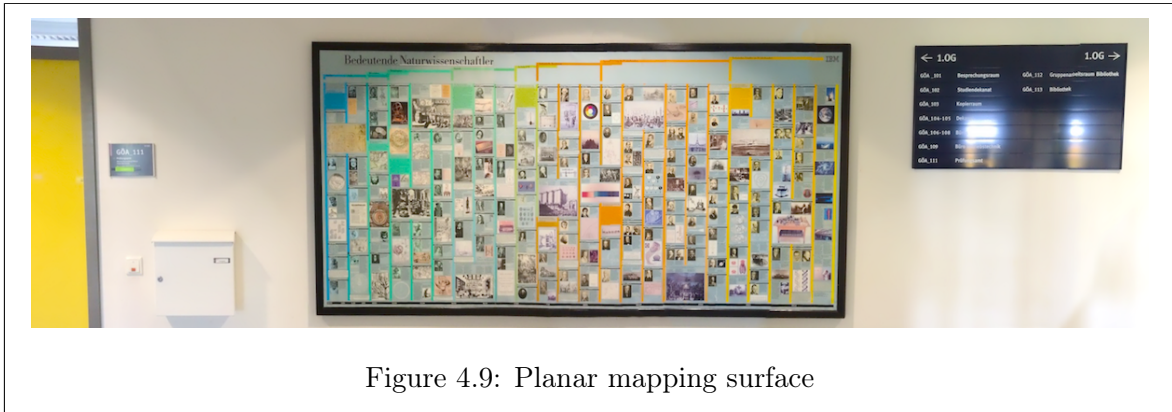
Figure 4.9: Planar mapping surface

mapping surfaces, namely a cylindrical, a spherical, and a planar model. All three images originate from the same input images. It becomes clear that the panorama is rotational, which is why generally a spherical mapping surface is appropriate. When mapping single row panoramas, a cylindrical model can also be employed. This type of warping induces some small distortions towards the top and the bottom, but the underlying mathematical model is easier to implement [SZELISKI and SHUM, 1997]. Figure 4.8 (c) emphasizes the necessity for an appropriate mapping surface. It can be seen that the planar mapping of a rotational panorama produces strong distortions. The parts to the far left and far right of the image were clipped in order to display the image at all. In contrast to the distorted rotational panorama, Figure 4.9 depicts the same scene captured as a planar panorama. Planar mapping is performed and it can be observed that the result is free from mapping distortions.

4.8   Evaluation

4.8.1   Results

The proposed method was tested on 50 data sets, of which 20 contain planar panoramas and 30 a rotational panorama, with between 5 and 15 images each. Some of these sets show the same scene and both types of camera motions are captured. This way, a direct analysis of the panorama type criterion is possible. The evaluation showed a 100% success rate in determining the correct panorama type for each of the 50 data sets. The correct bundle adjustment and mapping surface were chosen in all cases. Table 4.1 shows the average reprojection error after bundle adjustment for different numbers of input images with a resolution of 800x600 pixels and different panorama types. It can be observed that the reprojection error is low enough for a good quality result. The error is slightly larger for rotational panoramas, which is due to the fact that all images were captured by hand. This induces a small translation which causes a deviation from the purely rotational model.

In Section 4.1 (*Objective*) it was claimed that rotational stitching can be performed faster

images	planar	rotational
5	1.11px	1.27px
7	1.06px	1.28px
10	1.06px	1.31px
15	1.06px	1.30px

Table 4.1: Average reprojection error for 5, 7, 10, and 15 images in pixels. Table source: [HERBON et al., 2014b]

than planar stitching, due to fewer degrees of freedom. Tables 4.2 and 4.3 show the average bundle adjustment and overall processing times according to the number of input images, measured on an iPhone 5S smartphone. As expected, the bundle adjustment and the processing time increase with the number of images. The ratio between rotational and planar bundle adjustment is always between 3.6 and 4.0, independent of the number of images. This way it becomes clear, that the type of bundle adjustment also plays a significant role in the overall processing time. The total time for stitching does not include feature detection, as this step is performed during acquisition.

images	3+8 DOF	3+3 DOF	ratio
5	0.954s	0.283s	3.944
7	2.765s	0.712s	3.884
10	3.759s	1.617s	3.602
15	12.159s	3.449s	3.637

images	planar	rotational
5	2.83s	2.11s
7	5.39s	4.03s
10	7.77s	5.23s
15	14.56s	9.13s

Table 4.2: Bundle adjustment times for the 3+3 DOFs and 3+8 DOFs setups. Table source: [HERBON et al., 2014b]

Table 4.3: Overall processing time, excluding feature detection, for 5, 7, 10, and 15 images. Table source: [HERBON et al., 2014b]

Although the technical specifications of smartphones and other mobile devices gradually improve, one strongly limiting factor is the available memory. Image stitching can be very memory consuming, which is why the actual memory usage of the proposed method was monitored and the results for images with a resolution of 800x600 pixels are shown in Table 4.4. The largest number of images in the available data sets was 15, which leads to a peak memory usage of approximately 96MB. With a total memory of 1GB in the test device, the proposed method takes up only about 10% of the memory and is therefore well suited for mobile usage.

images	peak memory usage
5	49.6 MB
7	57.9 MB
10	72.8 MB
15	96.2 MB

Table 4.4: Average peak memory usage for 5, 7, 10, and 15 images. Table source: [Herbon et al., 2014b]

4.8.2 Remarks

The proposed method is generally applicable to panoramic imaging for both multiple and single row scenarios. The implementation of the approach builds on the complexity reduction, which is introduced through homography chaining, meaning that only consecutive images are registered. While this simplifies the computational complexity, it is still sufficient for most real world applications, especially wood pile surveying. Nevertheless, one might wish to capture panoramas with multiple rows. In this case only the image registration and the global motion computation need to be adjusted. The panorama type criterion remains valid, and can even be simplified by removing the necessity for the rotational condition.

The input data sets were all captured by a smartphone camera, which is the primary area of application for the proposed method. The threshold for the translational condition is designed for a fixed focal length of the camera. When stitching images with different focal lengths, the criterion can be adjusted if necessary. A minimum number of five images has been shown to work well for the application of the panorama type criterion. Fewer images exhibit the risk of failure, since the camera translation might be insufficient to accurately apply the translational condition of the panorama type criterion.

In both cases, rotational and translational, it is advisable to account for lens distortion. Especially planar panoramas suffer from these distortions, which can be removed be performing intrinsic calibration first, e.g., through the method by Zhang [Zhang, 2000]. Further visual results can be found in *Appendix A: Stitching results*. General scenes as well as wood piles are shown, with some scenes captured as both planar and rotational panoramas.

4.9 Summary and application to wood pile surveying

In this chapter an approach was presented, which was proven to be capable of distinguishing between rotational and translational panoramas. This problem has so far been rarely addressed in related work. The panorama type criterion provides the largest contribution and is designed to be integrated into existing approaches, either to validate a stitching model or,

as was also described in this chapter, for adaptive stitching. The camera trajectory is recovered through homography decomposition, which provides a rotational and a translational component. A rotational panorama is present when all camera translations do not exceed a specific threshold. This requirement is defined as the translational condition. Optionally, the rotational component may be used to evaluate the consistency of the camera rotation, in case the approach imposes the discussed complexity reduction for mobile devices.

The panorama can be stitched regardless of its type by choosing the proper degrees of freedom, performing adaptive bundle adjustment, and selecting a suitable mapping surface. For planar panoramas, a 3+8 DOF bundle adjustment model, combined with a planar projection surface, is used. Rotational panoramas can be optimized through a 3+3 DOF model, with a mapping onto a spherical or a cylindrical surface. 50 data sets were used to quantitatively and qualitatively evaluate the proposed approach. The results are shown in this chapter and in Appendix A. The bundle adjustment of rotational panoramas is, on average, three and a half to four times as fast as planar stitching.

For wood pile surveying, rotational stitching is not an option, because the image of a wood pile would become distorted towards the edges. An orthographic projection of the front surface conforms ideally to the definition of forestry surveying approaches. Planar stitching poses an attractive alternative, as the computational complexity is lower than that of a full 3D reconstruction. The front of a wood pile can usually be considered to be approximately planar, or quasi-planar. This constraint has been found to be sufficient for planar stitching, illustrated by the results shown in Appendix A.2. The panorama type criterion helps to guide a user on-site when capturing images of a wood pile. Non-valid planar panoramas can be dismissed quickly after the global motion estimation and the user can improve acquisition by moving the camera correctly. An interesting point for future research would be to perform motion estimation during acquisition for live feedback.

Chapter 5

Two-dimensional surveying

The image-based recognition of wood logs is one of the most important aspects of wood pile surveying. Many environmental influences affect the image acquisition and the visual appearance of the wood pile. A potential method for this task must therefore be able to deal with such varying conditions and still perform as well as possible. In [HERBON et al., 2014c] we presented an approach which shows promising results for wood log recognition. The current chapter is based on this original paper, while the evaluation results have been greatly extended to not only consider a larger number of test images, but also the impact of an extended parametrization.

5.1   Objective

This chapter describes an approach which recognizes wood logs in digital images and performs surveying based on the detection results. A wood pile, including its wood logs, can be geometrically described as a clustered object. Other examples of such clustered objects include fish, fruit, cells, and many more, as shown in Section 3.6 (*Related applications*). The proposed method can be viewed as a general scheme, which is refined by certain characteristics of roundwood logs. The objective of the recognition method is to correctly detect as many wood logs as possible (high true positive rate) with as few false matches as possible (low false positive rate). This method is then used as the basis for the two-dimensional surveying approach, where it is applied to panoramic images. These images are planar panoramas of wood pile front surfaces created through the approach presented in Chapter 4 (*Adaptive image stitching*).

In Section 2.2 (*Research challenges*) a number of obstacles were identified, which must be accounted for during wood log recognition. In this context internal and external parameters of the method are distinguished between. External parameters (such as illumination, noise, 3D rotation, and image resolution) alter the digital image before processing, while internal

parameters of the algorithm are evaluated in order to find the best performing parametrization. The results are assessed based on their true and false positive rate for both recognition and segmentation. Like all methods discussed in this thesis, special attention is given to processing time and memory consumption with regards to use on handheld devices.
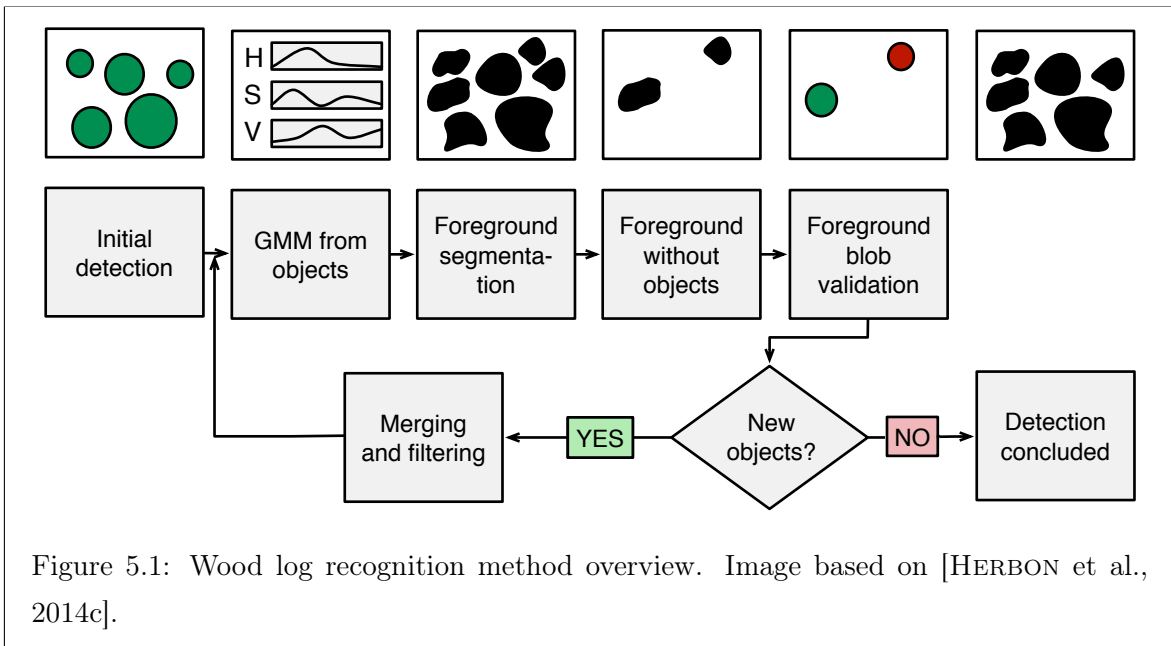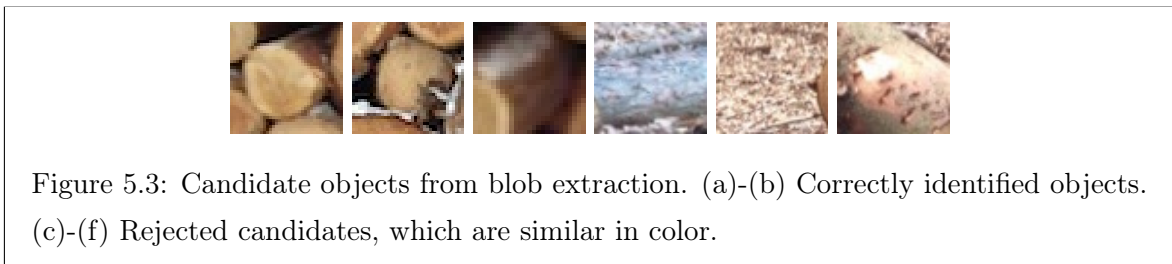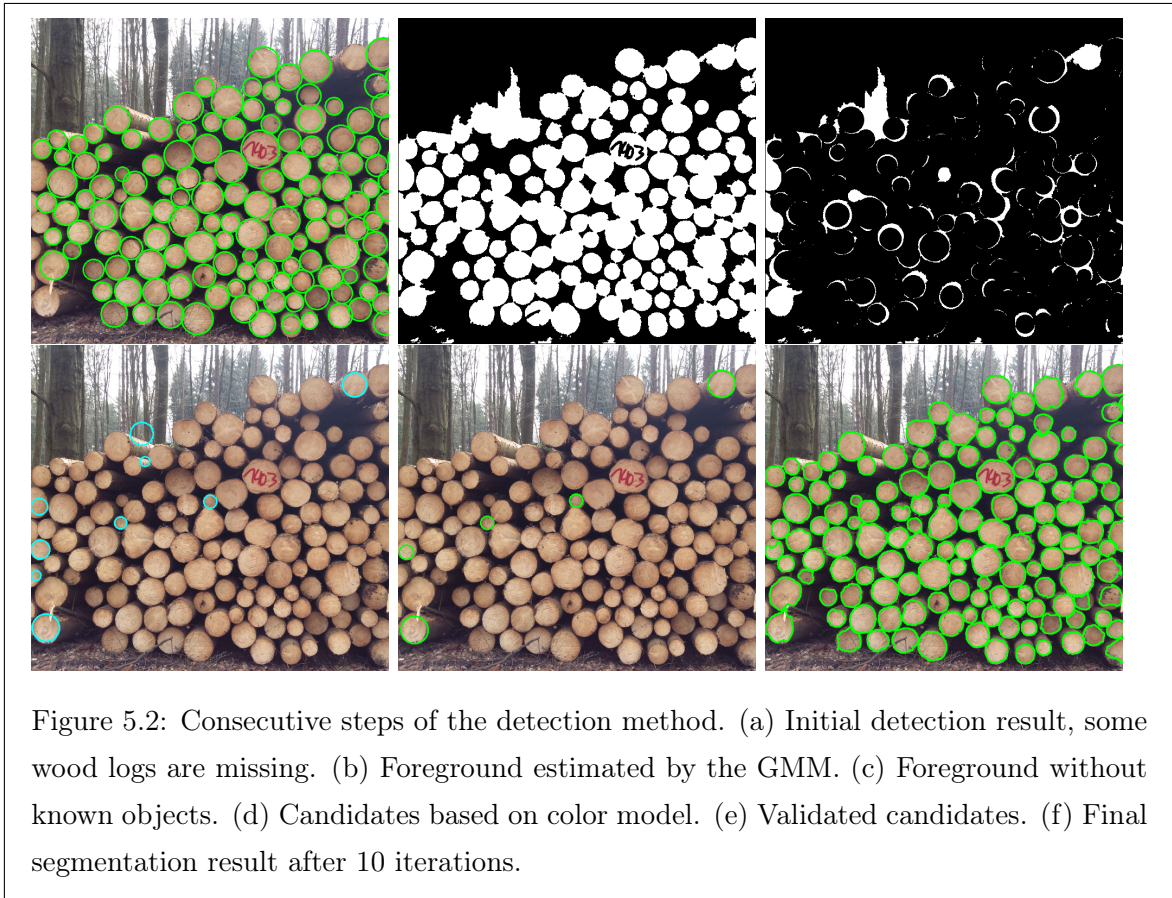
5.2 Detection and segmentation

Figure 5.1: Wood log recognition method overview. Image based on [HERBON et al., 2014c].

The general scheme of the proposed recognition method is shown in Figure 5.1. This approach is based on the method by Gutzeit and Voskamp [GUTZEIT and VOSKAMP, 2012], who propose to perform detection, create a statistical model, and recognize further objects within potential regions of interest. The method in this chapter optimizes and extends this approach, which will be shown in this section. The first step of the presented method is an initial detection of wood logs in the input image I_{in} (Figure 5.2(a)), which yields the initial object set S_{init}. One or more feature detectors (cascades of boosted weak classifiers), such as LBP, HOG, and Haar-like, are chosen for this task. All three types of detectors were trained with 8767 positive and 5161 negative samples that are not contained in the test data set of the *HAWKwood* database. All positive training images have a resolution of 24x24 pixels. The specific combination of detectors is subject to evaluation in Section 5.4.2 (*Image-based detection*). Since the objects of interest are not of a defined size, the detection procedure is implemented as a multi-scale approach. At each scale, the image is resized with a scale factor s_{detector}, whose influence will also be investigated in Section 5.4.2 (*Image-based detection*). In advance of the evaluation results it is important to consider that a larger number of scales

(smaller s_{detector}) results in a longer processing time. For this reason, s_{detector} should be chosen with regard to the targeted performance.

Figure 5.2: Consecutive steps of the detection method. (a) Initial detection result, some wood logs are missing. (b) Foreground estimated by the GMM. (c) Foreground without known objects. (d) Candidates based on color model. (e) Validated candidates. (f) Final segmentation result after 10 iterations.

Figure 5.3: Candidate objects from blob extraction. (a)-(b) Correctly identified objects. (c)-(f) Rejected candidates, which are similar in color.

Once the initial detection is complete, the objects are filtered in a way such that overlapping objects (more than 50%) are removed or merged. A Gaussian mixture model is computed based on the color information provided by the remaining objects [GUTZEIT and VOSKAMP, 2012]. Based on this color model a probability map is obtained, which is thresholded twice to generate the trimap T [GUTZEIT and VOSKAMP, 2012]. This trimap labels the pixels in the input image as foreground, background, or unknown. Unlike Gutzeit and Voskamp, who use graph cuts for segmentation, we perform a watershed transform. This is solely based on a

consideration of the processing time, which is significantly lower, while the results are similar. The trimap and the recognized objects are taken as input by the watershed transform, where the objects are used as markers. Concluding the application of the watershed transform, the binary foreground image I_{fg} is computed (Figure 5.2(b)). All initial objects are subtracted from the foreground to compute the difference image (Figure 5.2(c)).

Via a distance transform, blobs are extracted and stored as potential candidate objects S_{blob} (Figure 5.2(d)). The most important difference to [GUTZEIT and VOSKAMP, 2012] is the processing of these object candidates. While [GUTZEIT and VOSKAMP, 2012] suggests to view all blobs as correctly recognized objects, the method in this chapter performs a detailed validation for each candidate, which removes candidates, that are just similar to a wood log in color but that actually do not belong to the object cluster. Figure 5.3 emphasizes the necessity for such a discrimination, by showing examples of object candidates. The detailed validation is performed by decreasing the scale s_{detector} and running the detection scheme with the combination of Haar-like, HOG, and LBP features locally in the areas of interest (blobs). The validated objects $S_{\mathrm{blob,val}}$ (Figure 5.2(e)) are merged with the initial objects as $S_{\mathrm{cons}} = S_{\mathrm{init}} \cup S_{\mathrm{blob,val}}$, where S_{cons} are the consecutive objects with which processing continues. In the next step, the GMM is refined from S_{cons} and new candidates are extracted and validated. This is repeated until $S_{\mathrm{blob,val}} = \varnothing$, which means that no new objects were found. In this case, S_{cons} is the final detection result, and the last watershed segmentation is the final segmentation result (Figure 5.2(f)).

In summary, the proposed method builds on the approach by Gutzeit and Voskamp [GUTZEIT and VOSKAMP, 2012], but provides some significant improvements that help to optimize the detection result:

1. The proposed method offers a higher true positive rate, which is achieved through a combination of different feature detectors, instead of a single detector.
2. The true positive rate is further increased through an iterative scheme, composed of detection, statistical modeling, and merging steps.
3. The false positive rate is significantly lower, by performing individual validation of candidate objects.
4. An extended parametrization offers the possibility to adjust the algorithm to specific tasks.

5.3   Surveying

Building on the recognition and segmentation approach of the previous section and the stitching technique of Chapter 4, a method for two-dimensional surveying is derived in this section. The quantities of interest are the wood log count, the contour volume V_{c}, and the solid wood

Figure 5.4: Different measurements during surveying. (a) Measurement of the pile width. (b) Solid wood volume. (c) Smoothed contour volume.

volume V_s. The surveying process begins with the automatic recognition approach and then offers the user an interface for interactive editing. This way it can be assured that all wood logs were correctly marked with minimal effort for the user. To perform the measurement, the scale of the wood pile must be recovered. The most robust way to obtain the scale is to measure the width of the wood pile w_p in [m] and put it in relation to the width in the image w_{px} in [px]. The width of the pile is defined as the largest measurable length (see Figure 5.4(a)). Due to its geometric properties, a wood pile's width is always larger than its height and the width is thus always an unambiguously defined property. Note that the width is not measured purely horizontally, but also features a vertical component, which makes the pile image invariant to rotations about the z-axis. The scale s, which converts pixel lengths to a metric length is defined as shown by Equation 5.1.

$$s = \frac{w_p}{w_{px}} \tag{5.1}$$

In order to compute the solid cubic meters (Figure 5.4(b)), the number of wood pixels

$n_{\mathrm{px,wood}}$ must be converted to square meters, which corresponds to the solid wood area A_{s} of the front surface.

$$A_{\mathrm{s}} = n_{\mathrm{px,wood}} \cdot s^2 \tag{5.2}$$

The solid wood volume can then be simply derived by multiplying the result with the known log length l.

$$V_{\mathrm{s}} = l \cdot A_{\mathrm{s}} \tag{5.3}$$

The computation of the contour volume requires additional preprocessing. In the forestry standards for manual surveying [Bundesministerium für Ernährung, Landwirtschaft und Forsten, 1969; Deutscher Forstwirtschaftsrat e.V. und Deutscher Holzwirtschaftsrat e.V., 2014], the section volume method would be suitable for this task. Since no explicit definition exists for digital surveying, the formula for manual sectioning can be adjusted to work on a pixel level. In recapitulation, the contour volume is defined by Equation 2.8 as the product of the section width w_{s}, the number of sections n_{s}, and the average height, computed by averaging the section heights h_i with $i = 1 \ldots n_{\mathrm{s}}$. The number of sections can now be substituted by the horizontal number of pixels as $n_{\mathrm{s}} \to n_{\mathrm{px,horz}}$. The width of a section is $1\,\mathrm{px}$, thus $w_{\mathrm{s}} \to 1$, and the section height in [m] is substituted by the section height in pixels $h_i \to h_{i,\mathrm{px}}$. Width and height in pixels can be easily obtained through the image of the solid wood volume (Figure 5.4(b)). Equation 5.4 shows the computation of the metric contour area, which is then used in Equation 5.5 for the determination of the contour volume via the known length l.

$$A_{\mathrm{c}} = \sum_{i=1}^{n_{\mathrm{px,horz}}} h_{i,\mathrm{px}} \cdot s^2 \tag{5.4}$$

$$V_{\mathrm{c}} = A_{\mathrm{c}} \cdot l \tag{5.5}$$

The application of Equation 5.5 results in a rather unsteady contour. During manual measurement, smoothing is performed automatically, based on continuity assumptions. These adjustments need to be accounted for, in order to assure comparability of the results. In the last step the contour is therefore smoothed by a morphological closing operation of the binary contour area image. As a structuring element, a circle with the same diameter as the largest detected wood log is used (see Figure 5.4(c)).

5.4   Results

In this section the influence of internal and external parameters on detection and segmentation is studied. The evaluation of the recognition performance is conducted for both single images and panoramic images. Additionally, the pixel-based segmentation procedure is analyzed for single images. On the basis of these results, the 2D surveying approach is compared to manually obtained ground truth data. The *HAWKwood* database provides benchmarks for each of these scenarios, which can be found in categories *S.1-S.2* and *M.1-M.3*.

5.4.1   Evaluation metric

In a quantitative evaluation, the true and false positive rate (*tpr* and *fpr*) are the most important parameters. Their definition for clustered object recognition must be explicitly stated, because the false positive recognition results are not limited to a finite number of cases. Therefore, both quantities will be measured with respect to the known ground truth number of objects N_{gt}. In Equations 5.6 and 5.7, N_{tp} stands for true positive objects (correctly matched objects) and N_{fp} declares the number of objects which were classified as such, but actually do not exist.

$$tpr_{logs} = \frac{N_{tp}}{N_{gt}} \tag{5.6}$$

$$fpr_{logs} = \frac{N_{fp}}{N_{gt}} \tag{5.7}$$

The *HAWKwood* database provides wood pile images with ground truth. In the case of object recognition, ground truth consists of an object list with center locations and diameters. The ground truth number of wood logs N_{gt} is known from the size of the object list. When comparing the recognition result to ground truth, a matching metric needs to be defined. For clustered object recognition problems in two as well as in three dimension, the detection result usually consists of the object sizes and locations. Therefore, two conditions must be fulfilled (Equations 5.8 and 5.9), in order to establish a correct match.

$$D(c, c_{gt}) \leq r_{gt} \tag{5.8}$$

$$0.5r_{gt} \leq r \leq \frac{1}{0.5}r_{gt} \tag{5.9}$$

In the equations above c and c_{gt} are the object center for the recognized object and the center of the ground truth object respectively. D denotes the Euclidean distance and r / r_{gt} are the object and the ground truth radius. During evaluation of single images, only

objects that do not intersect the image border and that are no more than 25% occluded are considered.

The results for the wood log segmentation are characterized by their true and false positive rate on a pixel level (tpr_{pixel} and fpr_{pixel}). In some cases, the results are described by their precision and recall values. The term recall is a synonym for the true positive rate, while precision is also referred to as the positive predictive value (ppv). The definitions of these values are given by Equations 5.10, 5.11, and 5.12 [CORTES and MOHRI, 2004], where n is the number of pixels and the indices tp (true positive), fn (false negative), fp (false positive), and tn (true negative) indicate the pixel class. Except for cases in which results are compared to results obtained by state-of-the-art methods, the true and false positive rate will be used for evaluation, due to their intuitive nature. All shown error bars represent the $\pm s$ interval (empiric standard deviation), which enables a visualization of the mean as well as the scattering of the results.

$$tpr_{\text{pixel}} = \frac{n_{\text{tp}}}{n_{\text{tp}} + n_{\text{fn}}} \tag{5.10}$$

$$fpr_{\text{pixel}} = \frac{n_{\text{fp}}}{n_{\text{fp}} + n_{\text{tn}}} \tag{5.11}$$

$$ppv_{\text{pixel}} = \frac{n_{\text{tp}}}{n_{\text{tp}} + n_{\text{fp}}} \tag{5.12}$$

For surveying, three parameters are calculated for each set of results. The difference between a measurement X_i and the ground truth value μ_{gt} is defined as x_i. This is used to compute the average difference $\bar{x}$ (Equation 5.13), which is a measure for the bias of a surveying approach, where $\bar{x} = 0.0$ means that the approach is free of bias.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{n} (X_i - \mu_{\text{gt},i})}{n} \tag{5.13}$$

In order to quantize how scattered the measurements are in respect to the ground truth value, the empiric standard deviation s is used, which is defined by Equation 5.14.

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{5.14}$$

In some of the benchmarks, the same object is measured multiple times. In such cases, the repeatability is an important measure for analyzing the deviation of different measurements. Let k be the number of objects and n_j the number of measurements for the jth object. Then

the repeatability standard deviation s_{rep} is the average of the empiric standard deviations of repeated measurements of the same surveying objects.

$$s_{\text{rep}} = \frac{1}{k} \sum_{j=1}^{k} \left(\sqrt{\frac{1}{n_j - 1} \sum_{i=1}^{n_j} (x_i - \bar{x}_j)^2} \right) \tag{5.15}$$

5.4.2 Image-based detection

In this section, the influence of internal and external parameters on the wood log detection results are discussed. Internal parameters are such parameters that alter the algorithms internal procedures, while external parameters solely affect the input image, either in acquisition or post-processing. The goal of this section is to derive a set of standard internal parameters which can be used for further applications of the method, with respect to the influence of external parameters.

Noise and brightness   The first parameter covers robustness regarding changes in brightness and noise. For this purpose, a parametrizable model is defined, which changes the brightness of an input image and adds Gaussian noise through the parameter β. For the change in brightness, β is a multiplicative parameter, so that $\beta = 1$ does not change the brightness, while $\beta \ll 1$ and $\beta \gg 1$ induces strong alterations. The modified image intensity for a grayscale pixel $I_{i,j} \in [0,1]$ is therefore given by $I'_{i,j} = \beta I_{i,j}$. Additionally, insufficient or extensive lighting usually induces noise which is critical when considering the effect of direct sunlight or dim lighting conditions. Gaussian noise is defined as shown in Equation 5.16, which is added with a mean of $\mu = 0$, where $z = I'_{i,j}$ is the grayscale value of a pixel.
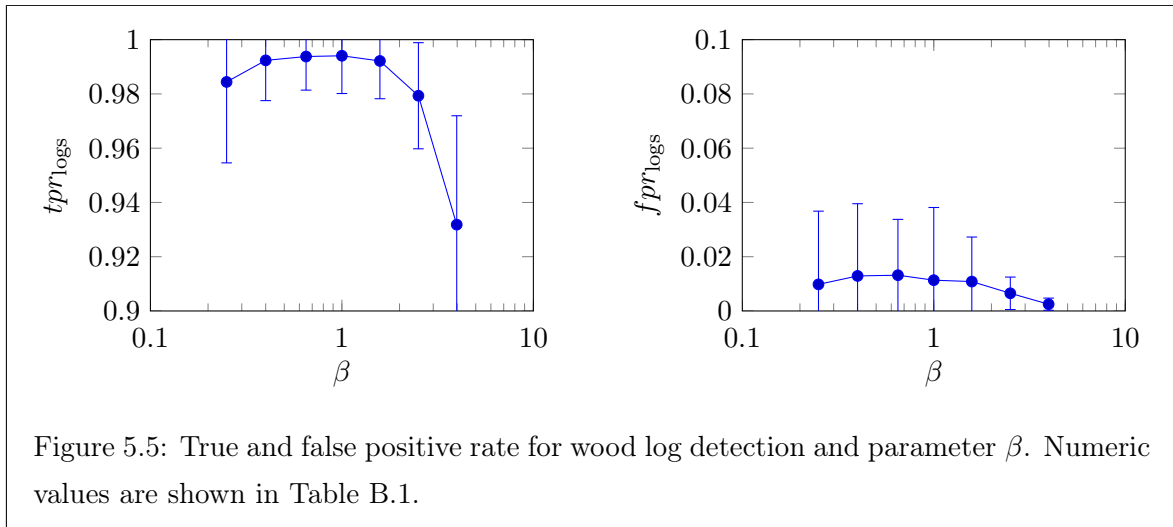
$$N_{\text{G}}(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z-\mu)^2}{2\sigma^2}} \tag{5.16}$$

The strength of the noise σ corresponds to the change in brightness, through a parametrization by β (Equation 5.17). This way, the image will exhibit more noise when changes in brightness are very high and no noise when the brightness remains unchanged.
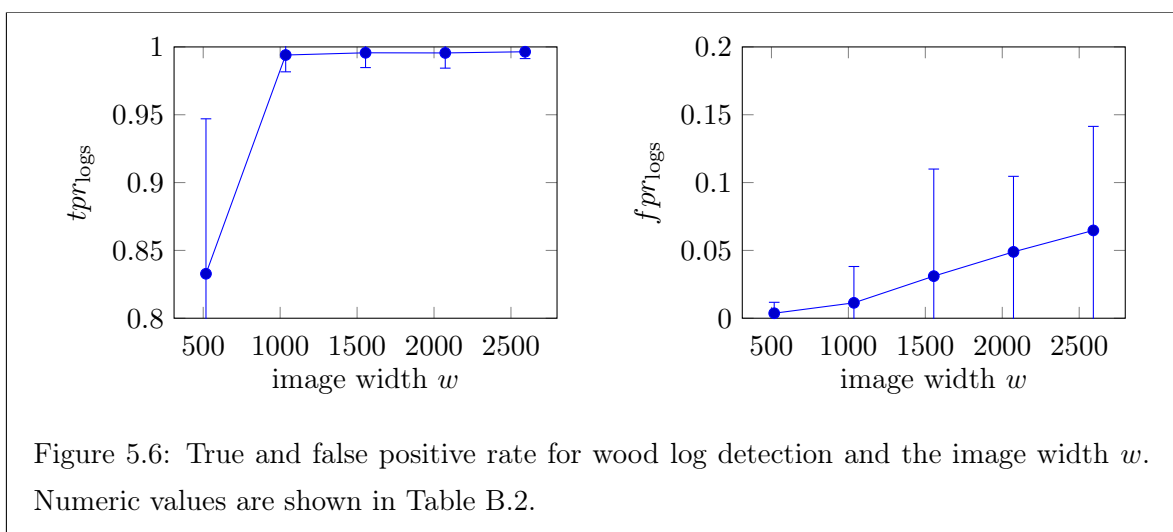
$$\sigma = 0.2 \cdot |\log_{10}(\beta)| \tag{5.17}$$

The results of the brightness and noise evaluation are displayed in Figure 5.5. It can be observed that the method is, in general, robust against such changes. An increase in brightness affects the true positive rate more strongly then a decrease. This is due to the fact that bright pixels of the wood logs are more likely to become oversaturated and thus prevent detection. With lower brightness, the false positive rate slightly increases. A very small or very large β results in fewer detected objects, which directly affects the false positive rate.

Figure 5.5: True and false positive rate for wood log detection and parameter β. Numeric values are shown in Table B.1.
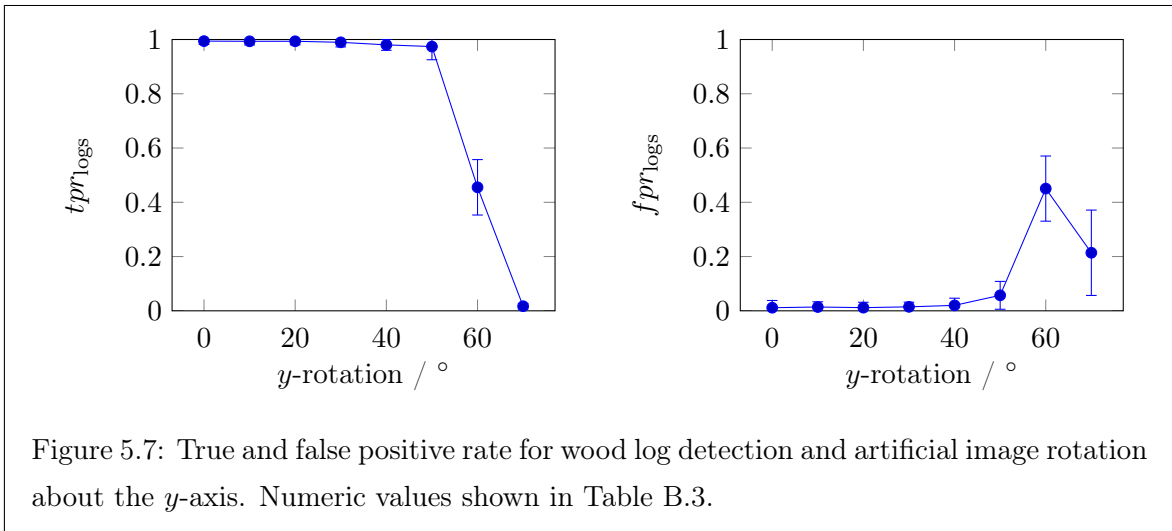
Image resolution All images are provided with a resolution of five megapixels or more. As the second external parameter, the effect of the image resolution is evaluated. Scaling is performed with constant aspect ratio $a = \frac{4}{3}$. Figure 5.6 shows that the effect of the image resolution is small within certain boundaries. In the range of $w = 2592\ldots 1036$ (with w being the image width) changes in the true positive rate are negligible. Only when the width drops to $w = 518$, does tpr_{logs} plummet. The false positive rate on the other hand, decreases linearly with smaller image sizes. Although detection is implemented as a multi-scale approach, detection seems to work best when the unknown objects' size is similar to the size of the training images. This behavior leads to the conclusion that the use of a lower resolution is preferable, as long as the critical threshold is not exceeded. For this reason, a resolution of 1036x774 px is used in further evaluations.

Figure 5.6: True and false positive rate for wood log detection and the image width w. Numeric values are shown in Table B.2.
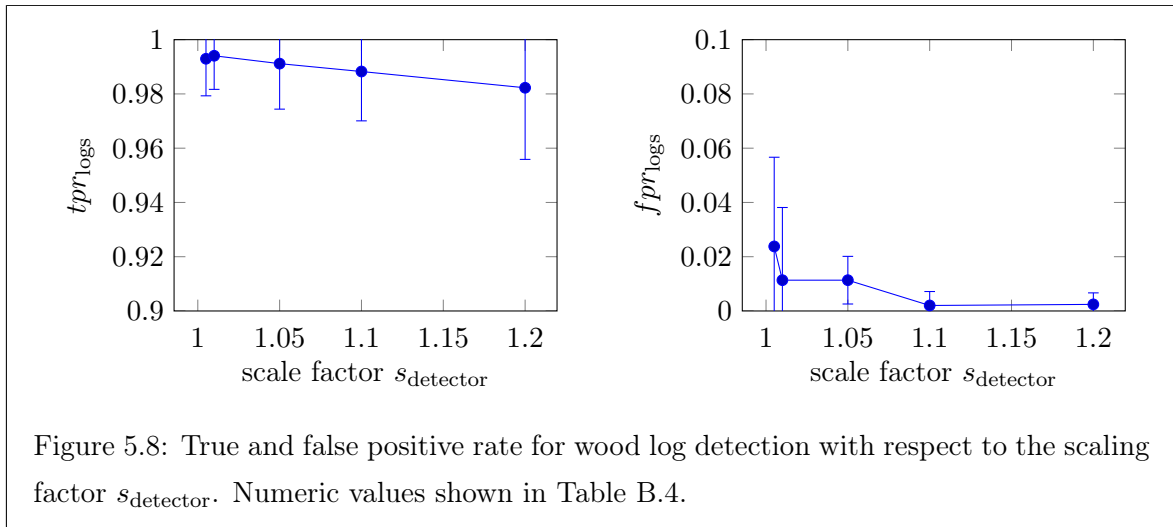
Rotation From empirical observations it became clear that inexperienced users sometimes have difficulties capturing images orthogonally to the wood pile front surface. The proposed methods must therefore provide a certain degree of rotation invariance about the x- and y-axis of the camera coordinate system. The experiments shown below are performed with a rotation about the y-axis. Until the rotation reaches $50°$, the fpr and tpr decrease only slightly. When rotating more than $50°$ the tpr goes towards zero while the fpr rises. In practice, rotations of no more than $10\text{-}20°$ are common when taking pictures with a handheld camera, even by inexperienced users. These results indicate, that the proposed method is rotation invariant in a way that is more than sufficient for practical purposes, when using handheld devices.

Figure 5.7: True and false positive rate for wood log detection and artificial image rotation about the y-axis. Numeric values shown in Table B.3.

Detector scale factor s_{detector} The scale factor s_{detector} was introduced in Section 5.3 (*Surveying*) as the factor between adjacent levels in the multi-scale detection step. Figure 5.8 shows that the tpr peaks at $s_{\text{detector}} = 1.01$. It is possible to sacrifice the high tpr for a lower fpr, when choosing $s_{\text{detector}} = 1.1$. A higher scale factor should not be used because the tpr decreases while the fpr remains approximately constant. A scale factor of $s_{\text{detector}} < 1.01$ leads to a significant increase of the fpr and should therefore also be avoided. In all further evaluations $s_{\text{detector}} = 1.01$ is used.

Feature detectors Three different feature detectors and their combinations are analyzed for the goal of optimizing the recognition task. Figure 5.9 shows results for Haar-like, HOG, and LBP features, and combinations thereof. The exact values can be found in Table B.5. The LBP detector offers the highest tpr and the HOG detector the lowest. Correspondingly, the fpr is the highest for LBP and the lowest for HOG. The combination of LBP and HOG features is especially interesting since it shows one of the highest true positive rates, while the false positive rate is only slightly increased in comparison to the LBP detector. If a high

Figure 5.8: True and false positive rate for wood log detection with respect to the scaling factor s_{detector}. Numeric values shown in Table B.4.

true positive rate, regardless of a large number of false positives, is desired, the combination of all three detectors is advisable. The standard setting of LBP and HOG is used in further experiments, as it offers a good compromise, with a tendency toward a high tpr.

Figure 5.9: True and false positive rate for different detectors. Numeric values shown in Table B.5.
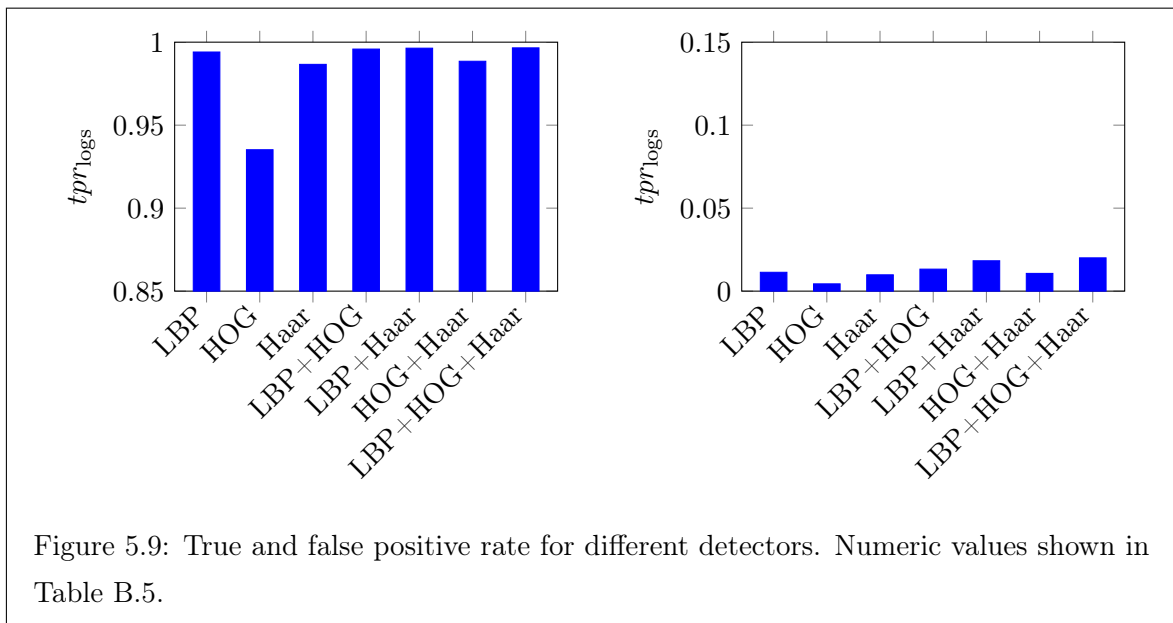
Iterations The iterative implementation of the proposed method optimizes the true and false positive rate at each iteration. Figure 5.10 shows the average true and false positive rate, where the tpr rises after the first iteration. Interestingly, the fpr is slightly lower after two iterations, which is due to the refinement of the statistical model. In all test cases convergence was reached after four iterations.

Figure 5.10: True and false positive rate for wood log detection after each iteration. Numeric values shown in Table B.6.
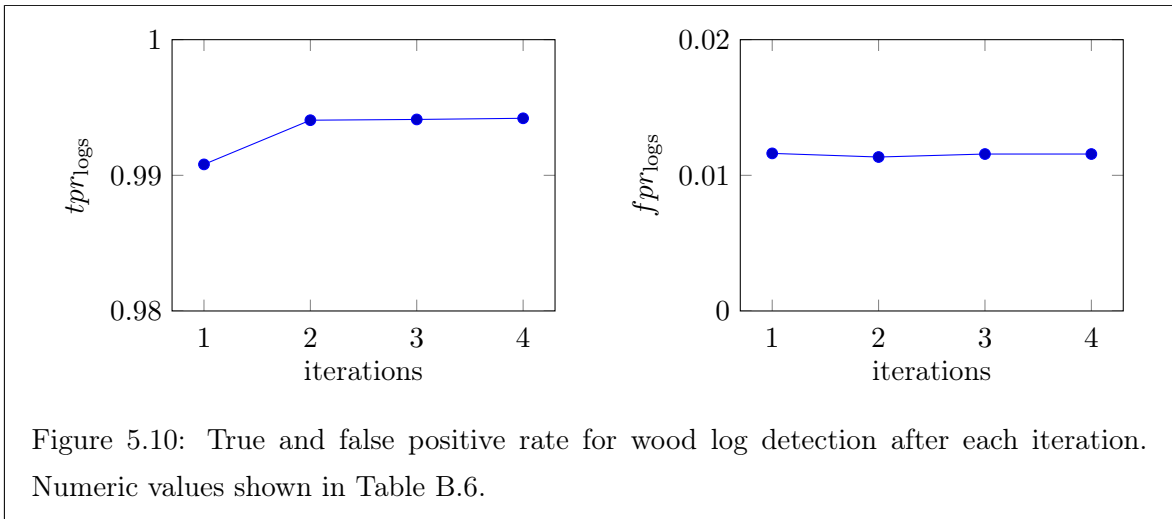
Standard parameters In all evaluations and in all future applications of the detection method (unless noted otherwise), the combination of HOG+LBP detectors will be used with a scale factor of $s_{\text{detector}} = 1.01$ and a detection resolution of 1036x774 px. In some cases it may be plausible to choose the combination of all three detectors, when the true positive rate is more important than the false positive rate. These standard parameters are used for the segmentation results in the next section, where they will be compared to manual ground truth segmentation.

Comparison to the state-of-the-art The only approach which is comparable to the proposed method is the one presented by Gutzeit and Voskamp [GUTZEIT and VOSKAMP, 2012]. The previous paragraphs have already shown that the combination of LBP and HOG features outperforms Haar-like features. The iterative implementation of the proposed method is able to increase the tpr while maintaining a constant fpr. For this reason, the novel method described in this chapter offers better results than the current state-of-the-art. Table 5.1 shows a brief comparison between the new method and a reimplementation of [GUTZEIT and VOSKAMP, 2012]. The significantly lower false positive rate originates from the detailed validation of candidate objects, which was proposed in Section 5.2.

detection	new method	[GUTZEIT and VOSKAMP, 2012]
true positive rate	**0.994**	0.986
false positive rate	**0.011**	0.070

Table 5.1: Comparison of the novel detection method to the reimplementation of [GUTZEIT and VOSKAMP, 2012]

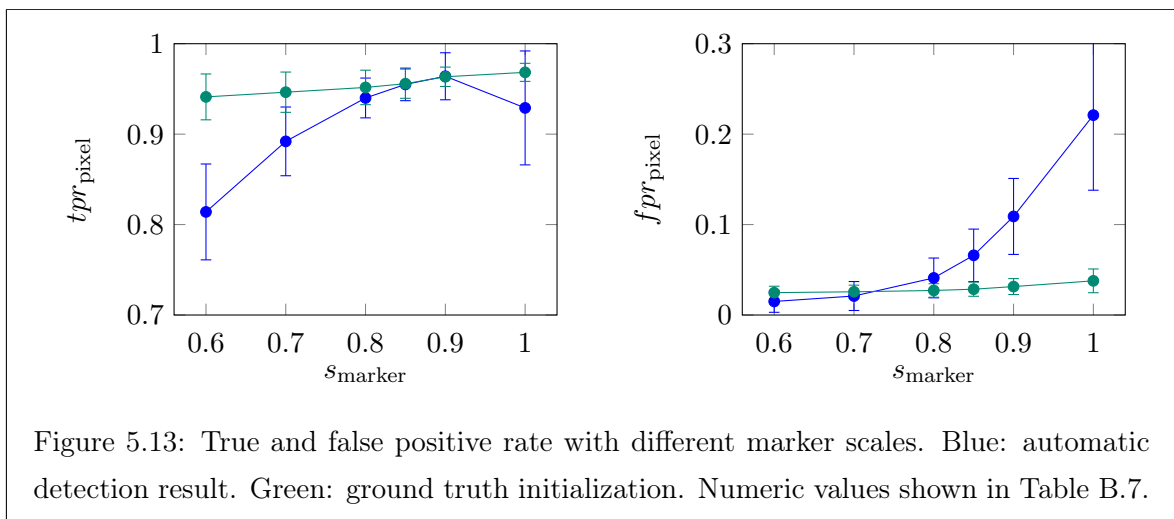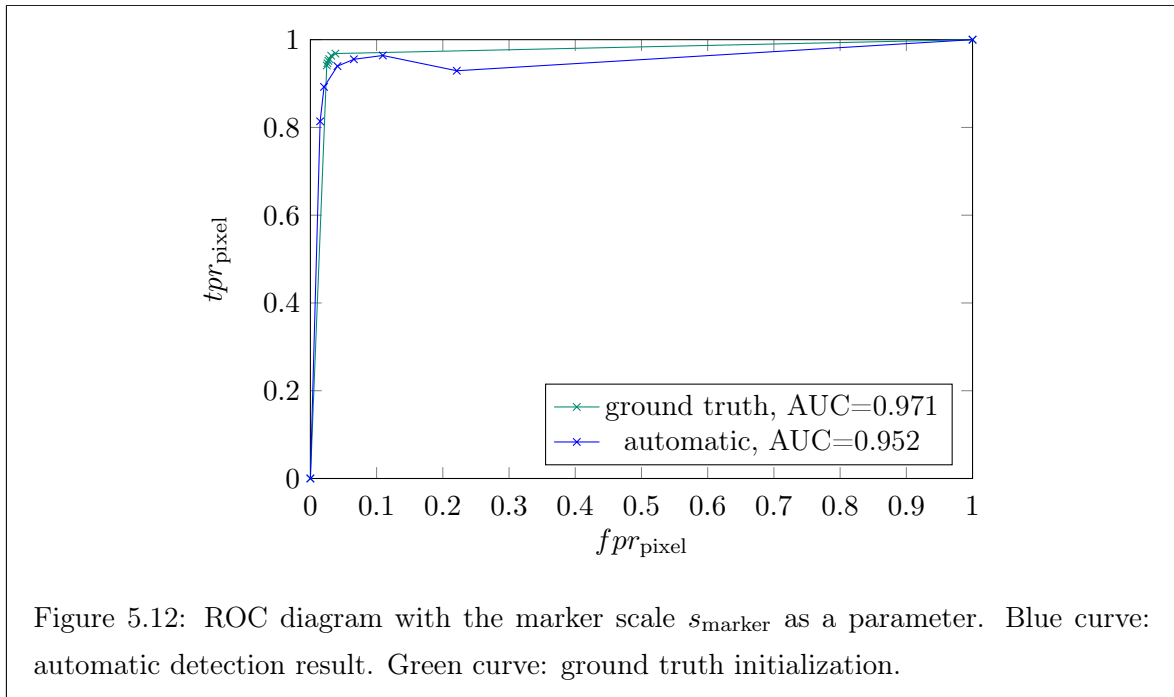### 5.4.3 Image-based segmentation

In this section, the segmentation procedure is evaluated. The *HAWKwood* database provides
20 images with ground truth pixel labeling. Two parameters are analyzed: the image resolu-
tion and the scaling of the markers for the watershed transform. The results are displayed as
ROC (receiver operating characteristic) diagrams [FAWCETT, 2004] which show the *tpr* over
the *fpr*. Additionally, the area under the curve (AUC) is given as a quality measure. An
AUC of 1 describes an ideal classifier, while an AUC of 0.5 indicates random guessing.

Marker scale   Marker-based watershed requires an initialization with known foreground.
The detected wood log objects are scaled down by the marker scale factor s_{marker} and used as
definite foreground (figure 5.11). The effect of this parameter is shown in the ROC diagram in
Figure 5.12. Both the result for automatically detected wood logs as well as an initialization
with ground truth objects are plotted. As expected, the ground truth initialization shows a
higher AUC, although the automatic AUC is still very high with a value of 0.952 (Figures
5.12 and 5.13). The segmentation procedure is therefore well-suited to the separation of wood
logs from background.

Figure 5.11: Foreground markers for watershed

The best performing scaling value in terms of the best combination of tpr_{segm} and fpr_{segm}
has been determined to be $s_{\mathrm{marker}} = 0.86$, as it is the value with the smallest geometric
distance to the point (0,1), which is the point in the upper left corner of an ROC diagram,
e.g., in Figure 5.12. An initialization with the automatically detected objects produces a dent
for $s_{\mathrm{marker}} = 1.0$. This is caused by the marker being larger than the actual object in some
cases and thus marking some background parts as foreground. The result is an instability
of the watershed procedure. A marker scaling factor of $s_{\mathrm{marker}} < 1.0$ is therefore advisable.
The marker scale value of $s_{\mathrm{marker}} = 0.86$, which will be used as the standard value for this
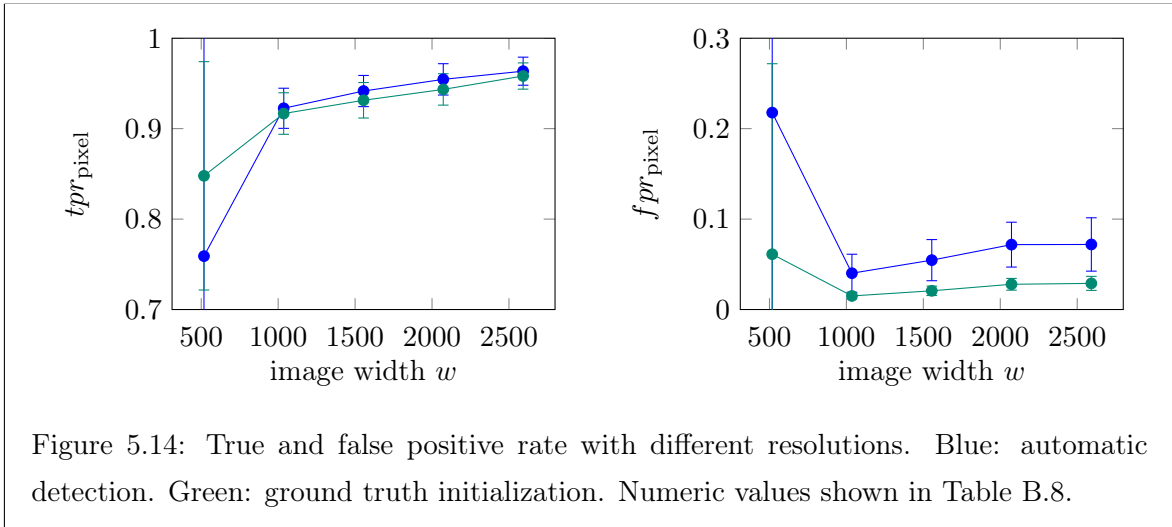parameter, produces the results shown in Table 5.2.

Figure 5.12: ROC diagram with the marker scale s_{marker} as a parameter. Blue curve: automatic detection result. Green curve: ground truth initialization.

Figure 5.13: True and false positive rate with different marker scales. Blue: automatic detection result. Green: ground truth initialization. Numeric values shown in Table B.7.

segmentation	tpr_{pixel}	fpr_{pixel}
automatic	0.964 (s=0.016)	0.072 (s=0.030)
ground truth	0.958 (s=0.015)	0.029 (s=0.008)

Table 5.2: Segmentation results for $s_{\mathrm{marker}} = 0.86$

Image resolution   For all segmentation procedures the full five megapixel resolution is used, which leads to the results presented in the previous paragraphs. For performance reasons, a reduction of the image resolution can be applied. Figure 5.14 shows the impact of the

resolution on the segmentation result. The true positive rate decreases significantly with a lower resolution, as does the false positive rate. When the image width falls under 1036 px the segmentation algorithm becomes unstable, as can be concluded from the strong increase of the standard deviation.

Figure 5.14: True and false positive rate with different resolutions. Blue: automatic detection. Green: ground truth initialization. Numeric values shown in Table B.8.

Comparison to the state-of-the-art Gutzeit and Voskamp [Gutzeit and Voskamp, 2012] tested their method under similar conditions. Different algorithms were implemented, the best of which performed with a precision of 0.878 (s=0.08) and a recall of 0.95 (s=0.0334). Table 5.3 shows the method proposed in this section (automatic results) in comparison to the method by Gutzeit and Voskamp. The new method performs best in terms of precision as well as recall, with significantly lower standard deviations. This speaks for a comparable performance of the watershed transform and graph cuts, although the impact of the higher automatic detection rate plays a role in these results.

segmentation	new method	[Gutzeit and Voskamp, 2012]
precision	**0.890** (s=0.029)	0.878 (s=0.080)
recall	**0.964** (s=0.016)	0.950 (s=0.033)

Table 5.3: Comparison of the novel segmentation method to [Gutzeit and Voskamp, 2012]

5.4.4 Two-dimensional surveying

On the basis of the 2D detection / segmentation and the panoramic image stitching described in the previous chapter, the results for the proposed two dimensional surveying method will be

discussed below. In general, the 2D recognition method is applied to panoramic images, and the contour and solid wood volume are computed as shown in Section 5.3 (*Surveying*). The input data sets are provided by the *HAWKwood* database, namely the multi-image benchmark (categories *M.1-M.3*).

Detection For the evaluation of the detection performance, the set of standard parameters from Section 5.4.2 is used and the method is applied to all 72 test images from the *HAWKwood M.1* benchmark. Table 5.4 shows the results of the proposed method based on manually segmented ground truth. The true positive rate is 0.9% lower than the detection result for single images. This is caused by two factors. Firstly, panoramic images build on a planarity constraint, which requires all points (in this case all wood log front surfaces) to lie in a common plane. In reality this constraint is only approximately fulfilled and wood logs, which lie outside of the common plane, are subject to seam detection and artifact removal which complicates the detection procedure. Secondly, the wood logs on the border of the pile tend to be more difficult to detect, since the contrast to background pixels is different to wood logs within the pile.

detection	tpr_{logs}	fpr_{logs}
$\bar{x}$	0.985	0.020
s	0.015	0.027

Table 5.4: Detection results for panoramic images as the initial step of the 2D surveying approach

Both of these factors also negatively affect the false positive rate, which is 0.9% higher than for single images. Despite the slightly lower values in comparison to single images, the detection rates show promising results for real world use, and work well under real world conditions, reflected by the wide variety of input images.

Solid wood volume The solid wood volume and the contour volume are computed based on the wood log objects after interactive editing. The edited objects are then used as the input for the segmentation method. This way, the actual surveying performance can be measured independently of the detection result. The *HAWKwood M.2* benchmark (34 data sets) serves as the data basis for the results shown in Table 5.5. The results are given as the average difference to ground truth, which in this case is the result of manual surveying by using the complete midpoint diameter inventory (see Section 2.5 (*Manual surveying methods*)). On average, the solid wood volume is underestimated by 2.6% with a standard deviation of 3.8%. The slight underestimation of the solid wood volume is due to the different color and texture of the bark, compared to the actual wood. This makes it difficult to segment the wood log

including its bark. The repeatability standard deviation of 2.3% shows that the method generally produces results within a reasonable margin of error, when performing repeated measurements of the same of object.

	$\bar{x}$	s	s_{rep}
V_{s}	-0.026	0.038	0.023

Table 5.5: Solid wood volume results (difference to ground truth) for panoramic images after interactive editing

Contour volume The solid wood volume result serves as the basis for the calculation of the contour volume. The *HAWKwood M.3* benchmark offers ground truth for all 117 data sets, which was obtained through the section volume method (see Section 2.5 (*Manual surveying methods*)). When comparing the contour volume results of Table 5.6 to the solid wood volume segmentation, it can be observed that the average difference is lower, while the standard deviation is slightly higher. Since the *HAWKwood* database does not offer synthetic data for this benchmark, it cannot be ruled out that the imprecise section volume method contributes to the higher variance, which is also indicated by the low repeatability standard deviation. Nevertheless, the results indicate that the method is indeed usable in practice.

	$\bar{x}$	s	s_{rep}
V_{c}	-0.019	0.046	0.024

Table 5.6: Contour volume results (difference to ground truth) for panoramic images after interactive editing

5.5 Summary

This chapter introduced a novel two-dimensional wood pile surveying approach. State-of-the-art wood log recognition and segmentation methods were improved for the goal of optimizing the number of correctly detected wood logs, while simultaneously significantly reducing the number of false positives. Based on single images of wood pile front surfaces, the influence of different parameters on the detection process was analyzed and the method was found to be robust against noise, changes in brightness, varying image acquisition locations, and image resolutions. A set of parameters was derived from the evaluation which best suits the requirements for this method and which is used in further applications of the method within this thesis. The proposed method outperforms the state-of-the-art method in terms

of the true and false positive rate in wood log detection, as well as in precision and recall for pixel-based segmentation. The recognition method can also be applied to other problems where the goal is to detect and segment clustered objects. Applications in which the proposed method might be useful were discussed in Section 3.6 (*Related applications*). These include the detection of fruit, cells, or animals. For the purpose of wood pile surveying, a pipeline was proposed, in which the wood log recognition method serves as the basis for the determination of the contour and the solid wood volume. By generating a panoramic image, as described in Chapter 4 (*Adaptive image stitching*), and by using the pile width as a scale, a wood pile can be measured in its entirety, while providing visual results for a user.

Chapter 6

Front surface segmentation

The surveying approach proposed in the previous chapter is built on the assumption that surveying can be performed based on object recognition. While this has been proven to be true for the case study in this thesis, the question arises whether alternatives to detection-based wood pile surveying exist. In this chapter, an approach will be presented which aims to segment the front surface of a wood pile solely based on geometric constraints. Additionally, applications of this method to autonomous vehicle navigation are discussed. We initially published this method in [HERBON et al., 2015b] and the following chapter is based on the original publication.

6.1 Objective

Motivation   The approach of the previous section implicitly constrains the input images to exhibit detectable wood log front surfaces. In all test cases provided by the *HAWKwood S.1* benchmark, the proposed algorithm was in general able to recognize these wood logs. In some very rare cases wood log detection might not be possible. This may be caused by the geometric properties or environmental factors. Figure 6.1 shows two examples of wood piles which potentially cause problems for the detection and segmentation method. Since the cut surfaces of the wood logs are largely covered in mud, color and texture based approaches are bound to fail. Nevertheless, the geometric properties of the wood pile are conserved.

In Section 2.3 (*Wood pile as a geometric meta-object*), the geometric properties of a wood pile were discussed. Based on this knowledge, a method is proposed in this chapter, which aims for the segmentation of wood pile front surfaces without object detection. The front surface is modeled as a so-called quasi-planar surface S_Q, which is a surface that can be approximated by a plane, called the principal plane, but also exhibits local non-planarities. In man-made or artificial environments quasi-planar surfaces are common, since many structures can be roughly approximated by a three-dimensional plane but would not be considered strictly pla-

Figure 6.1: Examples of challenging wood piles where detection may fail

nar. While all points on a planar surface generally satisfy a coplanarity constraint, points on a quasi-planar surface do not. The segmentation of quasi-planar surfaces has many applications, such as panoramic image stitching, autonomous vehicle navigation, auto-calibration, and image registration. The proposed method is therefore not only applicable to wood piles but considers these applications as well. Quasi-planar surfaces are especially of interest in the field of robot navigation, where obstacle detection and 3D modeling are important aspects. In off-road navigation, surfaces such as dirt roads and unpaved gravel roads can often be approximated by quasi-planar surfaces, as can be seen from e.g. [CHILIAN and HIRSCHMÜLLER, 2009; WELLINGTON and STENTZ, 2004; WELLINGTON and STENTZ, 2006].

Related work   There exist some approaches for the segmentation of planar surfaces, but all of these methods consider strictly planar surfaces only. Even small non-planarities lead to erroneous results. Some early approaches include the method by Sinclair and Blake [SINCLAIR and BLAKE, 1996], which takes five point correspondences as inputs in order to perform plane detection. An extension of this approach to stereo images is performed by Fornland and Schnorr [FORNLAND and SCHNORR, 1997]. A non-coplanarity constraint is enforced for the point correspondences. Ohnishi and Imiya [OHNISHI and IMIYA, 2005; OHNISHI and IMIYA, 2006] use a sparse optical flow field to extract a low-resolution dominant plane from a robot mounted camera. Optical flow is also used by Zucchelli et al. [ZUCCHELLI et al., 2002], where multiple planes can be detected through a least squares maximum likelihood estimator. An interesting approach, with similar applications as the method in this chapter, has been presented by Pears and Liang [PEARS and LIANG, 2001], who argue that using the fundamental matrix as a geometric relationship between images with strong planar regions is likely to fail when monocular image acquisition is performed, as degeneracy poses a difficult challenge. They instead propose a homography-based navigation approach, where the homography is

calculated from corner features. All feature points on the dominant plane are grouped to a bounding polygon. This shape is used to define the foreground color model (dominant plane) and background color model for segmentation purposes. The approach suffers as soon as the dominant plane exhibits concave regions or a non-uniform color distribution. Pears and Liang also enforce a strong coplanarity constraint, which is not able to adequately model the geometric properties of a wood pile front surface. The approach presented in this chapter is related to what Irani et al. [IRANI et al., 1998] propose as a method for multiple-view reconstruction based on a plane+parallax formulation. This enables view synthesis by determining a point's displacement between two views based on a real or virtual planar surface in the image. For this, the computation of the essential or fundamental matrix is not necessary (and sometimes not possible due to degeneracy), because a homography sufficiently describes the scene geometry.

Goal   The 2D segmentation in the previous chapter aims for the extraction of the solid wood volume before the contour volume is derived from the segmentation result. The goal for the method in this chapter is to perform a segmentation of a dominant quasi-planar surface directly, based on its geometric properties. In the case of wood pile surveying, the front surface is segmented as a whole, which instantly yields the contour volume instead of using the individual wood logs. The results of this approach will be compared to the two-dimensional segmentation method. This task is performed through homographic registration of an image pair and the use of block matching to recover depth information. The result of the block matching step is used for the segmentation of the quasi-planar surface. In the next section, this method will be described in detail.

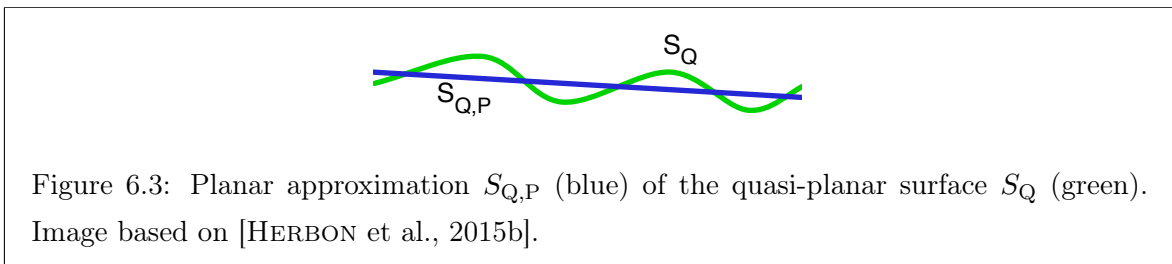## 6.2   Segmentation of quasi-planar surfaces

Overview   The proposed method works by performing a homography-based registration of two input images. Figure 6.2 shows a diagram which gives an overview of the proposed approach. The input images are weakly constrained to show the same quasi-planar surface S_Q, which in this case is a wood pile front surface, but can be any other quasi-planar surface as well. The homography is then computed and parametrized and the so-called principal flow vector (PFV) is extracted from it. Both input images are aligned through the homography matrix and rotated so that the principal flow vector aligns with the x-axis. Semi-Global-Matching (SGM) is performed, based on the principal flow lines (similar but not equivalent to the epipolar lines), which are the lines parallel to the PFV. The depth map produced by the SGM is then segmented through an expectation maximization algorithm.

Figure 6.2: Front surface segmentation method overview. Image based on [HERBON et al., 2015b].
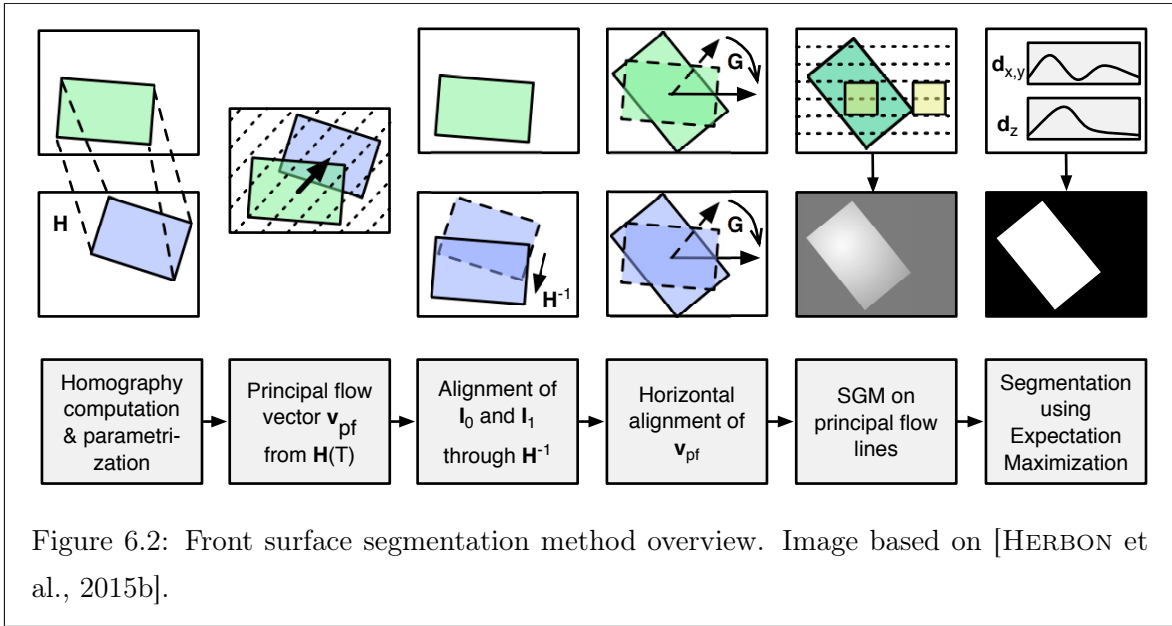
Contributions   The approach described in this chapter offers four novelties which enable the segmentation of quasi-planar surfaces:

1. The principal flow vector v_{pf} of a planar homography is introduced.
2. It is shown how the principal flow vector can be used for inter-image rectification.
3. It is demonstrated how the search space for block matching can be defined based on the principal flow vector, as opposed to the epipolar lines.
4. Two methods for the segmentation of the depth map, which are based on expectation maximization, are described.

Projective relations   The approach for quasi-planar surface segmentation described in this section takes as input two overlapping images of a wood pile front surface. The proposed method asserts that each image pair I_0 and I_1 depicts the same planar surface S_{P} or quasi-planar surface S_{Q} from slightly different viewpoints. For every quasi-planar surface there exists a strictly planar approximation, called the principal plane, which will be denoted as $S_{\mathrm{Q,P}}$ (Figure 6.3).

Figure 6.3: Planar approximation $S_{\mathrm{Q,P}}$ (blue) of the quasi-planar surface S_{Q} (green). Image based on [HERBON et al., 2015b].

The input images are related by a planar homography H, based on $S_{\mathrm{Q,P}}$, with eight degrees of freedom (in contrast to affine or rotational homographies with six or three degrees of freedom respectively) [BORGSTADT and FERRIER, 2001]. Two image points in the planar surface approximation, $p_0 = [x, y, 1]^T$ in I_0 and $p_1 = [x, y, 1]^T$ in I_1, are subject to the following relation:

$$p_1 = Hp_0 \tag{6.1}$$

Figure 6.3 shows that the first step of the proposed method is the computation of the homography H. This is achieved through the extraction of SIFT features and then by performing the well-known RANSAC-based registration approach. H is first computed from all feature matches and is then refined from inliers only. During the computation of H, the reprojection error is used as an error metric, which was defined in Equation 4.16. The camera intrinsic matrix is known from prior calibration. For images I_0 and I_1, the quasi-planar surface approximation $S_{\mathrm{Q,P}}$ is implicitly defined by the homography matrix H. The actual pose of the surface cannot by recovered by this approach, due to the planarity degeneracy. The pose could be recovered if at least one point was known that is not coplanar to the points on $S_{\mathrm{Q,P}}$, which cannot be guaranteed in general. For the proposed method only the homography is of importance, not the pose of the plane itself, since the homography is sufficient for inter-image rectification and depth map computation, as will be shown in the following paragraphs.
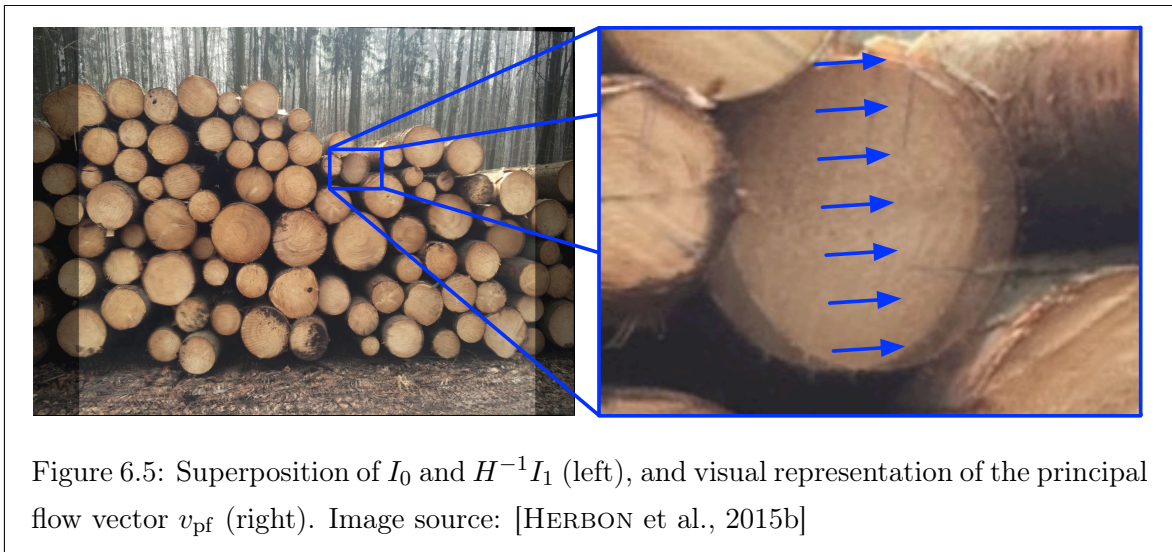
Principal flow vector v_{pf} The definition of the principal flow vector is based on the homography decomposition by Borgstadt and Ferrier [BORGSTADT and FERRIER, 2001], which was described in detail in Section 4.4 (*Homography decomposition*). In summary, a homography H can be decomposed into a translational component t and a rotational component R. The principal flow vector is defined as the projection of the camera translation onto the principal plane. The normal vector n of the principal plane is, by definition, parallel to the z-axis of I_0 as $n = [0, 0, 1]^T$. The homogeneous principal flow vector $\hat{v}_{\mathrm{pf}}$ is then given by Equation 6.2, where $\hat{v}_{\mathrm{pf}}$ is defined as the projection of t onto $S_{\mathrm{Q,P}}$. Based on Figure 4.3, the derivation of the principal flow vector is visualized in Figure 6.4.

$$\hat{v}_{\mathrm{pf}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \hat{t} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{6.2}$$

Inter-image rectification   As the next step in Figure 6.2, inter-image rectification is performed, which in this case transforms image I_1 to align with image I_0 (Figure 6.5, left). After rectification, points on the principal plane align perfectly (in theory), while points not on

Figure 6.4: Principal flow vector from homography decomposition in extension of Figure 4.3
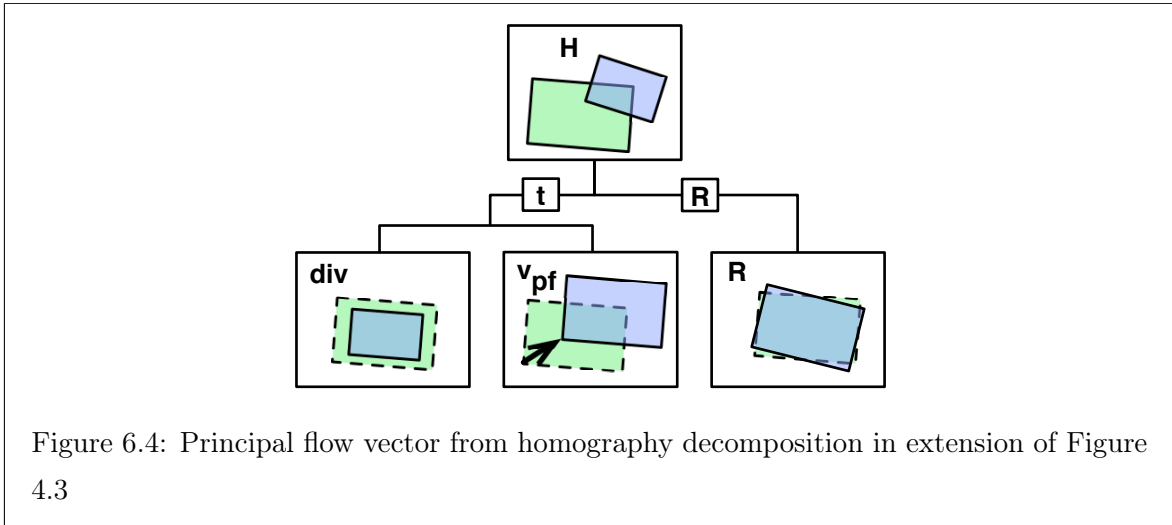
the principal plane (but still on the quasi-planar surface) undergo a translation in the same direction as v_{pf}. A visual representation in image space is given by Figure 6.5, which shows v_{pf} in detail. It can be seen that a wood log face, that is slightly behind the principal plane, exhibits a larger disparity than other wood logs.

Figure 6.5: Superposition of I_0 and $H^{-1}I_1$ (left), and visual representation of the principal flow vector v_{pf} (right). Image source: [HERBON et al., 2015b]

Search space alignment   The search space for corresponding pixels on the quasi-planar surface in the rectified images is always along the principal flow vector, since it is that vector which defines the translation of a pixel on S_{Q} during rectification. Block matching is performed on a horizontal search space, therefore v_{pf} must be rotated to align with the x-axis, defined in homogeneous coordinates as $\hat{x} = [1, 0, 1]$. The angle φ between the x-axis and the principal flow vector is given by Equation 6.3 as the *arccos* of the dot product of $\hat{x}$ and $\hat{v}_{\mathrm{pf}}$, where 1

must be subtracted due to homogeneous coordinates.

$$\varphi = \arccos(\hat{x} \cdot \hat{v}_{\mathrm{pf}} - 1) \tag{6.3}$$

The angle φ can now be used to compute the rotation matrix G, which transforms both images so that the desired search space along v_{pf} becomes purely horizontal. In Equation 6.4, c_x and c_y are the camera's optical center, which can be either obtained from intrinsic calibration or approximated by the image center.

$$G = T_c R_z(\varphi) T_c^{-1} = \begin{bmatrix} 1 & 0 & \frac{c_x}{2} \\ 0 & 1 & \frac{c_y}{2} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -\cos(\varphi) & \sin(\varphi) & 0 \\ -\sin(\varphi) & -\cos(\varphi) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -\frac{c_x}{2} \\ 0 & 1 & -\frac{c_y}{2} \\ 0 & 0 & 1 \end{bmatrix} \tag{6.4}$$

The final preprocessing step for block matching consists of combining the rotation matrix G with the inverse of the homography matrix and thus creating transform matrices M_0 and M_1 for images I_0 and I_1 respectively. This way, rectification and search space alignment can be combined into a single operation, thus minimizing distortions and processing time. The images transformed by matrices M_0 and M_1 are denoted as I_0' and I_1'.

$$M_0 = G \tag{6.5}$$

$$M_1 = GH^{-1} \tag{6.6}$$

Block matching   The image pair I_0' and I_1' is now eligible for block matching. For the proposed method, the block matching algorithm by Hirschmüller [HIRSCHMÜLLER, 2005] is used. Figure 6.6 shows the superposition of I_0' and I_1' on the left and the disparity map DM on the right. Obviously depth can only be recovered in overlapping regions. It is very important to note that the disparity values for the obtained depth map do not represent the distance from the camera center but the distance from the principal plane. In this case, this information is more useful than the distance from the camera center, because the goal of this method is to extract the dominant plane. From Figure 6.6(b) it can also be seen that many regions of the background are not detected by the block matching algorithm. This is because the search space is only aligned for the quasi-planar surface. If one wishes to detect background also, epipolar rectification would be a preferable choice, but planar degeneracy may prevent the computation of an epipolar rectification in the monocular case. For quasi-planar surface segmentation, the non-detected background is actually very beneficial, since it can be quickly marked as not belonging to $S_{\mathrm{Q,P}}$.

Through a disparity map that provides a distance measure from the dominant plane, the dimensionality of the problem is reduced, since the pose of the plane in 3D does not need to be

Figure 6.6: (a) Superposition of I_0' and I_1', (b) Disparity map DM obtained by applying [HIRSCHMÜLLER, 2005]. Image source: [HERBON et al., 2015b].
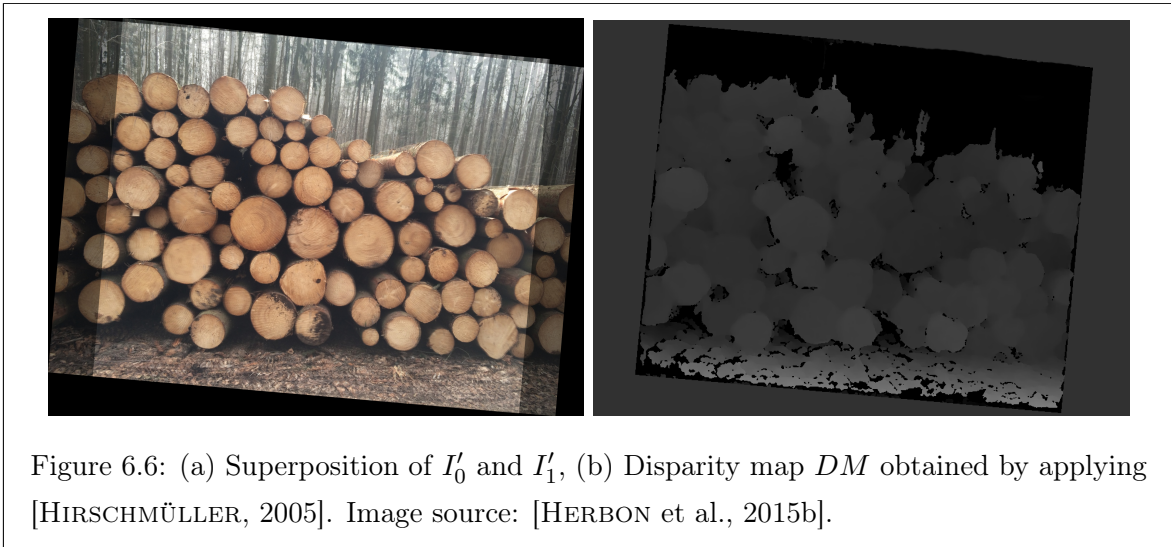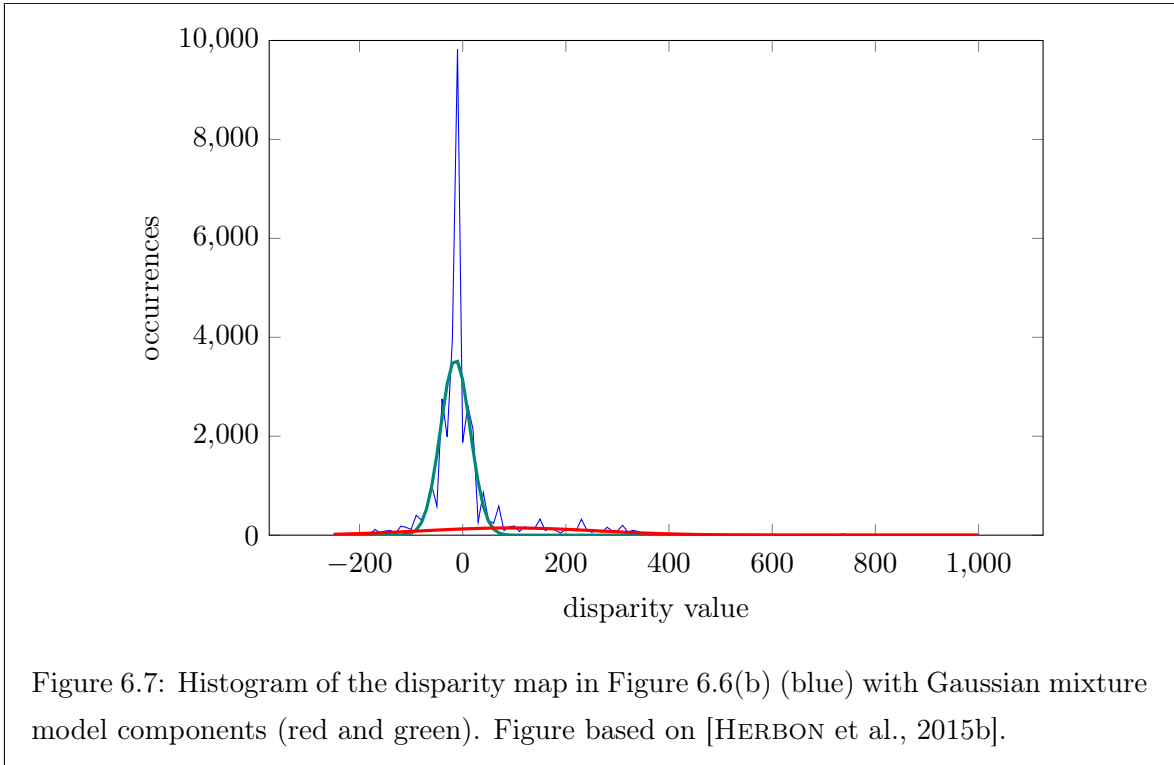
considered. All perfectly overlapping pixels lead to a disparity of zero. Instead of performing plane detection in three dimensions, globally consistent thresholding can be performed directly on the disparity map without determining the pose of the plane, which will be the subject of the next paragraph. This is based on the knowledge that disparities for pixels on the quasi-planar surface will be close to (but not necessarily equal to) zero.

Histogram modeling   The disparity map offers a depth value for each pixel $DM(x_i, y_i) = z_i$, where the pictures overlap. This overlapping region in images I_0' and I_1' is defined as the region of interest (ROI) $J = I_0' \cap I_1'$. The goal of the segmentation procedure is now to find a meaningful segmentation for the disparity values in J. The diagram in Figure 6.7 shows the histogram of the disparity map of Figure 6.6 (blue), which is typical for the class of images where there is one dominant quasi-planar surface present. It can be seen that most values are close to 0 and that occasionally scattered values with $z \gg 0$ or $z \ll 0$ are present.

This histogram can be approximated by a mixture of Gaussian distributions, where the number of distributions is n_{GMM}. For $n_{\mathrm{GMM}} = 1$ the distribution is called unimodal, $n_{\mathrm{GMM}} = 2$ defines a bimodal distribution. For a bimodal distribution, as shown in Figure 6.7, the distribution with the mean closest to zero represents pixels on the quasi-planar surface. The second component models background pixels and noise. Equation 6.7 shows the definition of the ith distribution $d_{\mathrm{GMM},i}$, with mean μ_i, standard deviation σ_i, and weight w_i.

$$d_{\mathrm{GMM},i}(x) = w_i e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} \tag{6.7}$$

The GMM is obtained via expectation maximization, as described by Dempster et al. [DEMPSTER et al., 1977]. The model can then be used for segmentation in two different ways, as will be shown in the next paragraphs.
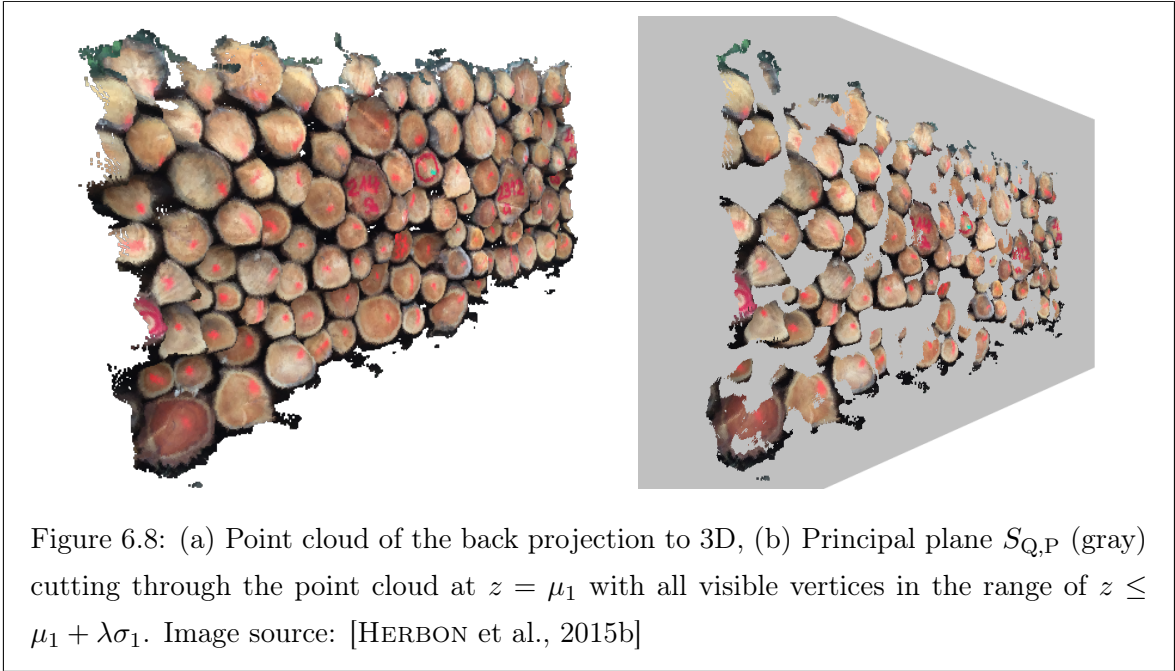
Figure 6.7: Histogram of the disparity map in Figure 6.6(b) (blue) with Gaussian mixture model components (red and green). Figure based on [HERBON et al., 2015b].

σ-**based segmentation** Each component of the Gaussian mixture model includes a mean μ_i and σ_i. The quasi-planar surface is described by μ_1 and σ_1. These parameters define a value range Ω, as shown by Equation 6.8. Disparity values within this range of $\mu_1 - \lambda\sigma_1 \dots \mu_1 + \lambda\sigma_1$ are considered inliers, while all other points are outliers.

$$\Omega = [\mu_1 - \lambda\sigma_1, \mu_1 + \lambda\sigma_1] \tag{6.8}$$

The quasi-planar surface segmentation can be characterized by this range, which is dependent on both σ and a scaling parameter λ. Since σ would be zero for an ideally planar surface, σ will be larger for quasi-planar surfaces with many pixels not on the principal plane. σ is thus a measure for the planarity of the surface. In order to perform optimal thresholding the parameter λ is analyzed and the true and false positive rate of the segmentation result are evaluated in Section 6.3 (*Results*). In Figure 6.8, the 3D result of the thresholding procedure can be seen, where only points in front of the principal plane $S_{\mathrm{P,Q}}$ are shown on the right.

Top value segmentation The sigma-based segmentation is a parametric model, which takes as input the histogram approximation by the GMM and the scaling parameter λ. When handling scenes with unknown content, a parameterless model might be preferable, as opposed to an explicit parameter choice. Such a model can be well-established for a bimodal histogram approximation with two Gaussian distributions, $d_{\mathrm{GMM,1}}$ (foreground) and $d_{\mathrm{GMM,2}}$

Figure 6.8: (a) Point cloud of the back projection to 3D, (b) Principal plane $S_{\mathrm{Q,P}}$ (gray) cutting through the point cloud at $z = \mu_1$ with all visible vertices in the range of $z \leq \mu_1 + \lambda\sigma_1$. Image source: [HERBON et al., 2015b]
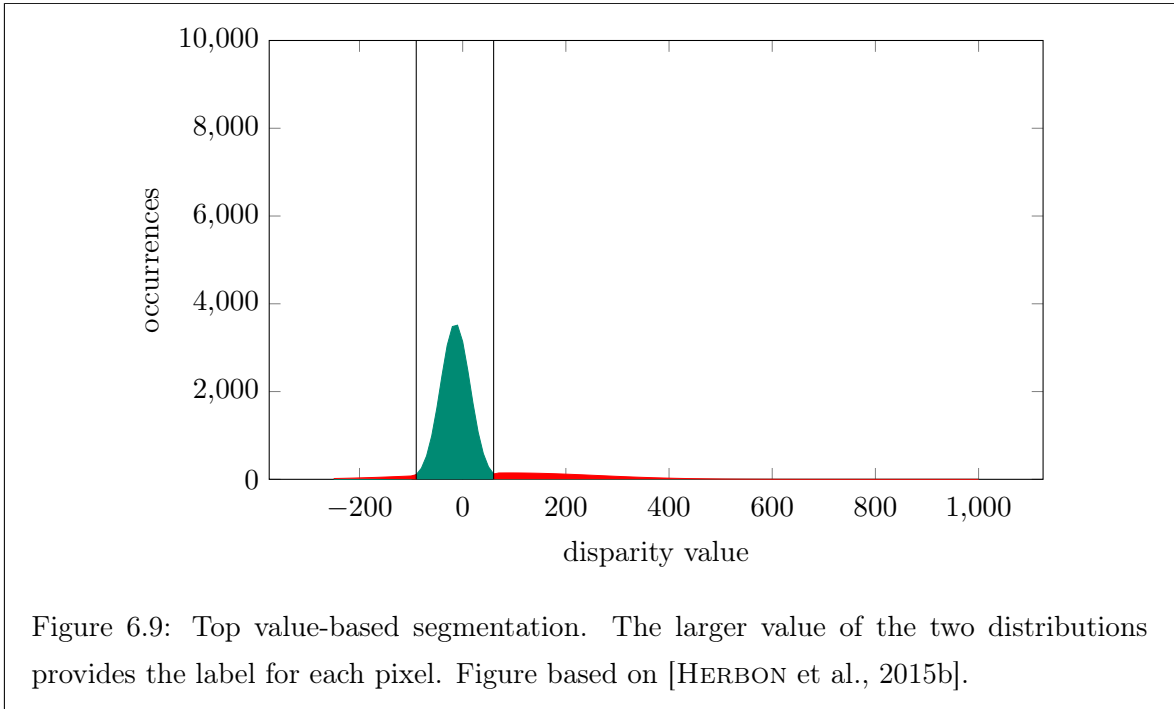
(background). A pixel classification for each pixel $p_{x,y}$ can be defined based on the two distributions, as shown by Equation 6.9.

$$p_{x,y} = \begin{cases} 1 & \text{for} \quad d_{\mathrm{GMM},1}(DM(x,y)) \geq d_{\mathrm{GMM},2}(DM(x,y)) \\ 0 & \text{for} \quad d_{\mathrm{GMM},1}(DM(x,y)) < d_{\mathrm{GMM},2}(DM(x,y)) \end{cases} \tag{6.9}$$

This method is comparable to the method proposed by Huang and Chau [HUANG and CHAU, 2008], i.e. chooising the threshold as the average of the means. Since in some cases two thresholds, a lower and an upper bound, are needed, the method by Huang and Chau is extended by considering the histogram value of the disparity map at each given pixel location $d_i(DM(x_i, y_i))$. In the case that the value of the first distribution is larger or equal to the value of the second distribution, the pixel is marked as foreground. All pixels which do not satisfy this condition are labeled as background. Figure 6.9 emphasizes this technique, where background regions are marked red and foreground regions green. Thresholds are indicated by vertical lines.

Key aspects in image acquisition The block matching algorithms require a certain degree of texture in order to match blocks of pixels correctly. If large textureless regions are present in the image, the block matching procedure might fail due to ambiguous matching results. Failure in the block matching step leads to "holes" in the disparity map. Such holes can be filled through interpolation if they are small enough, but larger holes might not be compensated adequately.

The principal flow vector was defined as the projection of the homography's translational

Figure 6.9: Top value-based segmentation. The larger value of the two distributions provides the label for each pixel. Figure based on [HERBON et al., 2015b].

component t onto the quasi-planar surface $S_{Q,P}$. From this definition it becomes clear that a degenerate exists for $t = [0, 0, div_k]$, which leads to $v_{pf} = \mathbf{0}$. In other words, the translation is purely in the z-direction. This case can be avoided by imposing the constraint shown in Equation 6.10. In the case of wood pile surveying and robot navigation, this constraint is usually fulfilled, since the user generally moves in parallel to the wood pile and a robot moves in parallel to the floor, which induces $x_k \gg 0$.

$$\sqrt{x_k^2 + y_k^2} \gg div_k \tag{6.10}$$

The proposed method is optimized for a dominant surface segmentation, which implies that a large, single plane is present in the images. The approach can be extended to detect multiple planes by applying k-means clustering [LLOYD, 1982]. This case is not relevant for wood pile segmentation and is therefore not discussed in the scope of this thesis.

6.3   Results

In this section, the segmentation results of the proposed method and the different thresholding approaches are compared. Two different databases are used for evaluation, one for quantitative and one for qualitative analysis. In order to assess the potential of the method for integration into the photogrammetric wood pile surveying approach, the results are compared to those obtained by two-dimensional surveying in Chapter 5.

6.3.1 Dominant Plane Database

The first database, called *Dominant Plane Database* [HERBON, 2014a], will be used for a quantitative analysis. It contains 15 uncalibrated images pairs, which each show a dominant planar or quasi-planar surface. Manual ground truth is provided for each image, so that the segmentation results can be compared. All images were taken by a consumer level digital or smartphone camera. Both the σ-based and the top value segmentation are applied to the provided images.

σ-based segmentation   As discussed in Section 5.4.3, the area under the curve is a quality measure for a segmentation algorithm. The goal is to achieve perfect segmentation with an AUC=1. Figure 6.10 shows the ROC diagram for the unimodal ($n_{\mathrm{GMM}} = 1$) and the bimodal ($n_{\mathrm{GMM}} = 2$) σ-based segmentation. It can be seen that the bimodal distribution is better suited for segmentation, since its AUC is higher than the AUC for unimodal segmentation. For the actual application of the proposed method, a cutoff point p_{co} must be defined. In this case, p_{co} is defined as the point with the smallest Euclidean distance to (0,1), as explained in Section 5.4.3.

Figure 6.10: ROC diagram for the unimodal (blue) and bimodal (green) σ-based segmentation and parameter λ. Figure based on [HERBON et al., 2015b].
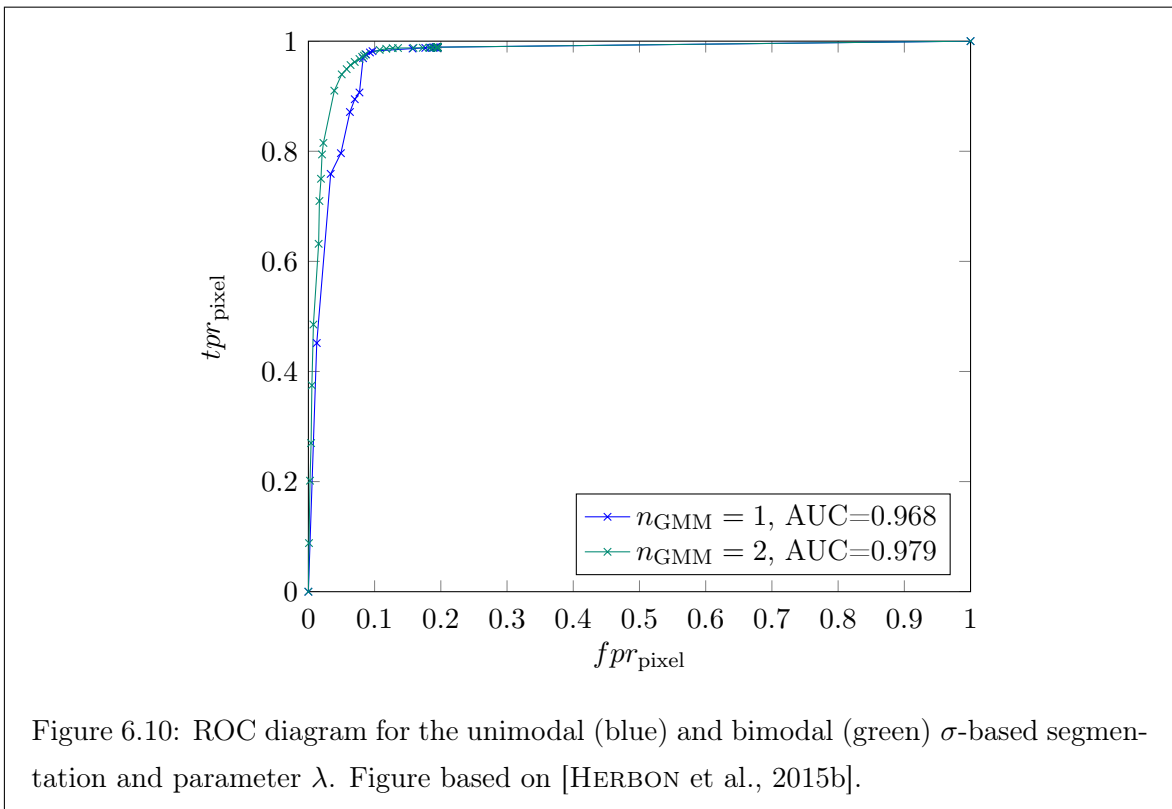
Table 6.1 shows results for the σ-based segmentation at the optimal cutoff points (indicated through the parameter λ). In the case of a bimodal distribution, λ is larger than for the

n	λ	tpr_{pixel}	fpr_{pixel}	AUC
1	0.7	0.969 (s=0.033)	0.083 (s=0.061)	0.968
2	3.0	0.940 (s=0.046)	0.049 (s=0.055)	**0.979**

Table 6.1: σ-based segmentation results for the *Dominant Plane Database* benchmark. Table source: [HERBON et al., 2015b]

unimodal distribution (3.0 vs. 0.7). This shows that, on average, the unimodal distribution features a larger σ and thus includes more uncertainty than actually necessary to model the quasi-planar surface. In addition to the higher AUC, it therefore seems preferable to use the bimodal distribution.

Top value segmentation   Since the top value segmentation is a non-parametric model, a meaningful ROC diagram cannot be obtained. Instead, the top value segmentation already provides a segmentation, which is ideal with respect to the bimodal histogram approximation. The results in Table 6.2 are very similar to the results of the bimodal σ-based segmentation.

	tpr_{pixel}	fpr_{pixel}
DPDB	0.943 (s=0.046)	0.053 (s=0.058)

Table 6.2: Top value segmentation results. Table based on [HERBON et al., 2015b].

6.3.2   *HAWKwood* database

The goal of the method proposed in this chapter is similar to the goal of the segmentation procedure in Chapter 5 (*Two-dimensional surveying*). In contrast to the 2D segmentation based on detected objects, the approach in this chapter segments the surface as a whole and not just individual wood logs. This behavior is desired but poses a challenge for method comparison. The *HAWKwood S.3* benchmark offers images with ground truth only for the actual wood logs (solid wood) and not for the entire front surface (contour volume). According to [VAN LAAR and AKÇA, 2007, p. 88] the conversion from the contour volume to the solid wood volume is usually done by a factor (stacking coefficient) of $f_{\text{s,c}} = 0.68 \ldots 0.7$. This means that when comparing the results of the quasi-planar with the 2D segmentation results, the false positive rate is expected to exhibit an increase of ∼30-32%. Some of the qualitative results can be seen in Figure 6.11.

Table 6.3 shows that the fpr is indeed increased by approximately 30%. In comparison to the 2D segmentation method, the quasi-planar surface segmentation shows a higher tpr for the σ-based segmentation with $n_{\text{GMM}} = 2$ and for the top value segmentation. The 2D segmentation performs better in terms of the false positive rate and also exhibits a much lower

Figure 6.11: Qualitative *HAWKwood S.3* results for two data sets (left / right column). First row: registered input image pair and horizontal alignment of principal flow vector. Second row: disparity map. Third row: segmentation result using the top value approach. Fourth row: segmentation result using the σ-based approach with $n_{\mathrm{GMM}} = 2$ and $\lambda = 3.0$. Image source: [HERBON et al., 2015b]
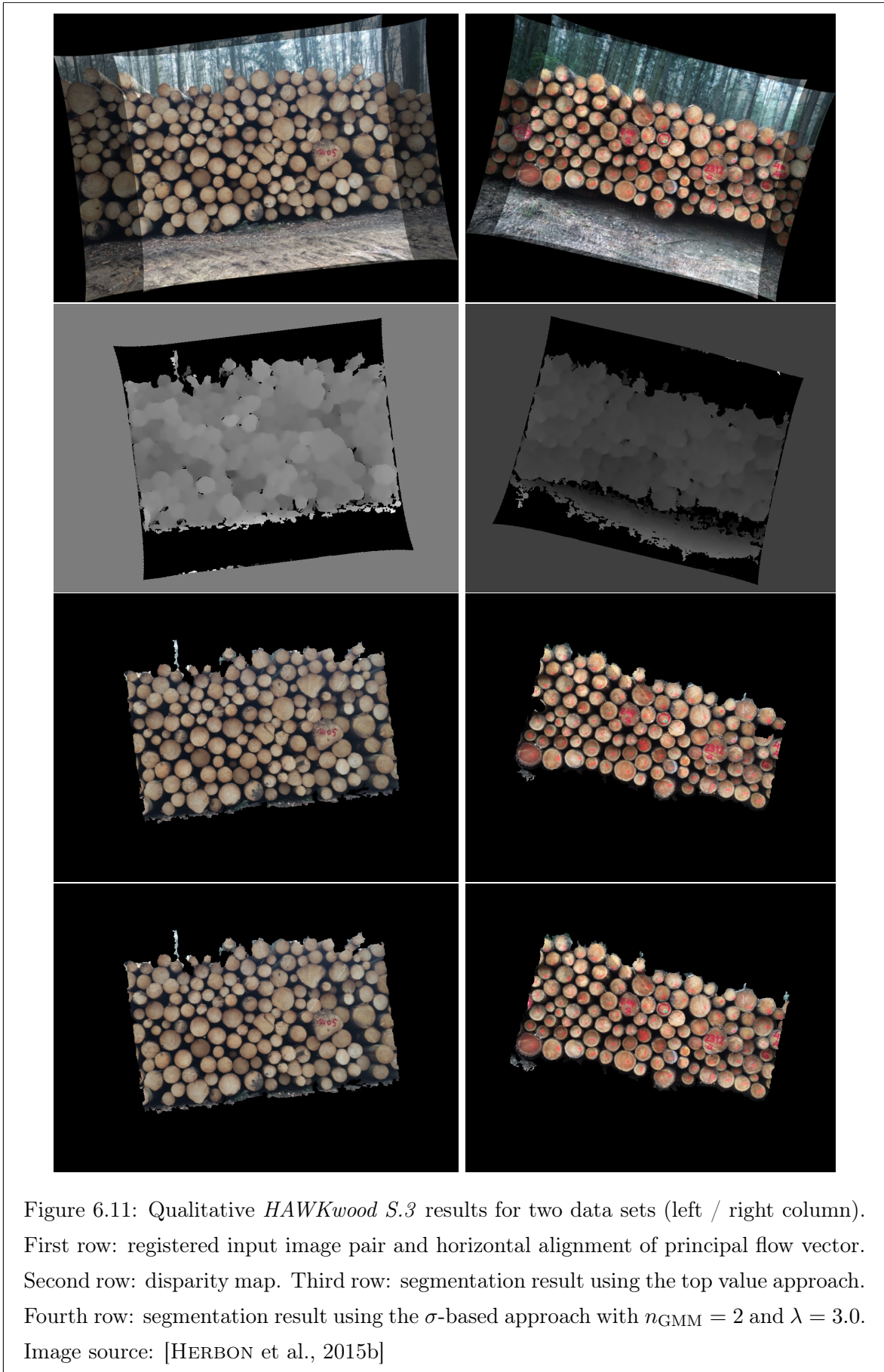
segmentation method	tpr_{pixel}	fpr_{pixel}
σ-based ($n_{\text{GMM}} = 1$, $\lambda = 0.9$)	0.920 (s=0.072)	0.320 (s=0.093)
σ-based ($n_{\text{GMM}} = 2$, $\lambda = 3.0$)	0.980 (s=0.013)	0.381 (s=0.111)
top value based	0.965 (s=0.034)	0.345 (s=0.108)
2D segmentation (sect. 5.4.3)	0.958 (s=0.015)	0.029 (s=0.008)

Table 6.3: Results for the *HAWKwood S.3* benchmark and comparison to 2D segmentation. Table partly based on [HERBON et al., 2015b].

empiric standard deviation for this parameter. Although the quasi-planar surface detection method shows promising results, the 2D segmentation, performs at least as well.

6.4 Summary and further applications

Further applications In addition to wood pile segmentation, the proposed method can be applied to a variety of problems. All scenarios, where the goal is to detect and segment a planar or quasi-planar surface, can potentially benefit from this approach. One of the most important areas for such a method is autonomous rough terrain navigation. Often roads are unpaved in rural areas and gravel roads occur frequently. Some approaches have been presented to perform terrain classification (e.g. [HÄSELICH et al., 2013]) and many approaches exist for road segmentation (e.g. [KONG et al., 2010; KUHNL et al., 2011; ZHANG and NAGEL, 1994]), although the homographic approach, including local non-planarities in the surface, is rarely considered [OKUTOMI et al., 2002]. The proposed method imposes only very weak constraints on the scene geometry. For this reason, it is well-suited for many types of off-road navigation problems.

An example for the application of the proposed method to a rough terrain scene is shown in Figure 6.12. One of the input images is given on the top left and a 3D reconstruction can be seen on the top right of the Figure, which emphasizes the non-planarity of the scene. On the bottom row, the segmentation for a strictly planar model (left) and the quasi-planar model (right) are shown. In the strictly planar segmentation, many parts of the road are not correctly segmented. Only parts that happen to lie on the principal plane are marked as foreground. The quasi-planar segmentation includes the non-planar parts of the road and still labels bushes and vegetation on the side of the road as obstacles.

Summary In this chapter it was proven that wood pile segmentation can in principle be performed without recognition of individual wood logs. The proposed method is able to segment a quasi-planar surface from a set of overlapping images. This is achieved through homographic registration, planar rectification, and block matching. The principal flow vector,

Figure 6.12: Segmentation results for rough terrain navigation. (a) Left input image. (b) Side view of the 3D reconstruction. It can be seen that the surface is quasi-planar, but not strictly planar. (c) Result for strictly planar segmentation. Many parts of the road are incorrectly labeled as obstacles (red). (d) Segmentation result of our method with correctly segmented obstacles on the side of the road (top left and top right). Image source: [HERBON et al., 2015b]

defined as the projection of the translational homography component onto the principal plane, is used to perform search space alignment for block matching, so that Semi-Global-Matching can be used to extract a disparity map. This disparity map does not show the distance from a pixel to the camera's optical center, it is a measure of the pixel's distance from the principal plane. The method can be applied to wood pile segmentation and any other class of problems where the goal is to find and segment a quasi-planar surface. For instance, off-road navigation can benefit from this approach. If it is desired to only detect regions that very likely belong to the foreground (road) then the parametrization can be changed to benefit the true negative rate. Alternatively, the true positive rate can be increased by the parametrization to detect more locally non-planar regions.

Based on the distance information, two distinct segmentation methods have been proposed, which include the parametric σ-based segmentation and the non-parametric top value segmentation. Both approaches were quantitatively evaluated through the *Dominant Plane Database*. The parametric approach showed a high AUC of 0.979 ($tpr_{\mathrm{pixel}} = 0.940$, $fpr_{\mathrm{pixel}} = 0.049$) for

$n_{\mathrm{GMM}} = 2$; the non-parametric segmentation resulted in $tpr_{\mathrm{pixel}} = 0.943$ and $tpr_{\mathrm{pixel}} = 0.053$. These values emphasize the applicability of the method to quasi-planar surface segmentation. Qualitative evaluation was performed for the *HAWKwood S.3* benchmark, which also showed promising results that were similar to those of the dominant plane benchmark. In comparison to the 2D segmentation step, it became clear that, while the results are well-suited for the general case of quasi-planar surface segmentation, the special case of wood pile segmentation can be comparably solved by the 2D segmentation procedure introduced in Chapter 5 (*Two-dimensional surveying*), which additionally provides individual wood log contours. The direct computation of the contour volume via the quasi-planar surface can be used as a fallback, should the rare case occur in which wood logs cannot be detected. An interesting point for future research would be the combination of the quasi-planar surface segmentation with wood log detection. The foreground estimation step in the wood log recognition pipeline could be amended by the method proposed in this chapter. Such an approach could even be integrated into the stitching pipeline, where homographic registration is done in any case.

Chapter 7

Three-dimensional surveying

The centerpiece of this dissertation is the three-dimensional photogrammetric surveying approach, which will be presented in this section. The method is based on our prior work, namely [HERBON et al., 2014a] and [HERBON et al., 2015a], and a preliminary study [OTTE, 2014], all of which address the three-dimensional object detection and the photogrammetric surveying approach. The first publication describes in detail the recognition of wood logs and other clustered objects in 3D space, based on structure from motion, 2D object detection, and quadric filtering. The second paper focuses on surveying of the contour volume and the solid wood volume based on the detection result and interactive editing. Many of the approaches and considerations of the previous chapters are either included in the method proposed in this chapter or used for comparing results.

7.1   Objective

So far in this dissertation, a two-dimensional surveying technique which uses either single or panoramic images, performs wood log recognition and segmentation, and derives the wood volume through a scale factor has been proposed. The main motivation for the two-dimensional approach was the use of the quasi-planarity property of the wood pile front surface, through which a two-dimensional front surface model could be established. In Chapter 6, an initial approach for the description of the wood pile as a three-dimensional geometric object was introduced, which was based on the regulatory a priori knowledge [BUNDESMINISTERIUM FÜR ERNÄHRUNG, LANDWIRTSCHAFT UND FORSTEN, 1969; DEUTSCHER FORSTWIRTSCHAFTSRAT E.V. UND DEUTSCHER HOLZWIRTSCHAFTSRAT E.V., 2014] of the wood pile front surface being quasi-planar. The goal of the current chapter is to explore in which ways a three-dimensional surveying approach can be beneficial for wood pile surveying and to investigate whether there are possible pitfalls when applying such a method. In comparison to the two-dimensional approach, the proposed 3D method does not have to make assumptions

about the planarity of the wood pile, but requires a sophisticated reconstruction model.

7.2 Method

7.2.1 Overview

In this section, the initialization and the 3D surveying method itself will be introduced briefly. Important algorithmic and implementational aspects will be discussed in detail over the course of the following sections.

Initialization Unlike the methods presented previously in this dissertation, the three-dimensional surveying approach requires an initialization by preprocessing data with the previously described methods (e.g., wood log recognition) and with external methods (lens distortion correction and SfM). Figure 7.1 shows that the method takes as input a set of images with high overlap and a scale reference. As discussed before, this scale reference is usually the pile width w_{p}, but other references will be considered as well. In addition to these inputs, the intrinsic calibration of the cameras is provided. This was obtained through prior calibration by the method of Heikkila and Silvén [HEIKKILA and SILVÉN, 1997]. The correction of the lens distortion is performed through the approach by Zhang [ZHANG, 2000]. This step could be skipped and be performed during bundle adjustment, but performing distortion correction prior to image registration was shown to yield more robust results [SAWHNEY and KUMAR, 1999]. The reason for this is that feature matching for structure from motion is based on the epipolar constraint, which requires images with little or no distortion. For each of the input images, wood logs are detected through the recognition method of Chapter 5. In this step, slightly different parameters are used, in order to optimize the processing time for mobile devices, which is the subject of Sections 7.5.1 (*Conceptual and parametrical optimizations*) and 7.5.2 (*Processing time and memory consumption*). No manual interaction takes place at this point, which means the detection results show a $fpr \geq 0$ and a $tpr \leq 1$. At the same time as object detection, structure from motion is computed. The result of the structure from motion process is a set of calibrated cameras and a sparse point cloud, both of which are used for the three-dimensional object detection step.

Object detection and surveying Based on the initialization step, the proposed method (as shown in Figure 7.2) performs an initial three-dimensional object detection, which outputs the wood log cut surfaces in 3D space. The three-dimensional wood logs can be used to compute a concave hull of the wood pile (contour area / volume). The scale of the 3D reconstruction is recovered through the scale reference, which, unless noted otherwise, is the pile width. During visualization, the orientation of the wood pile becomes important.

Figure 7.1: Initialization of the photogrammetric surveying approach

Figure 7.2: Photogrammetric surveying approach. Image based on [HERBON et al., 2015a].
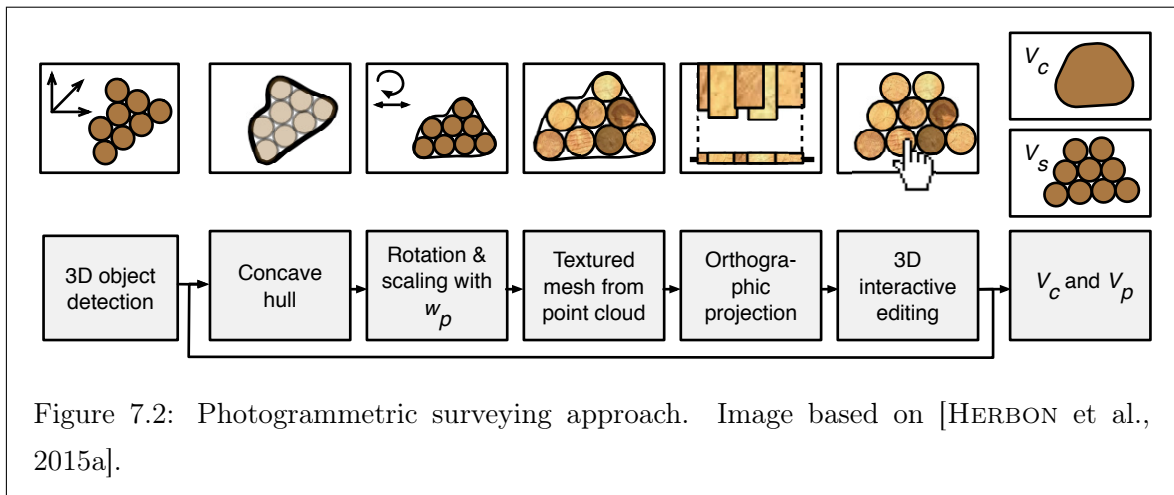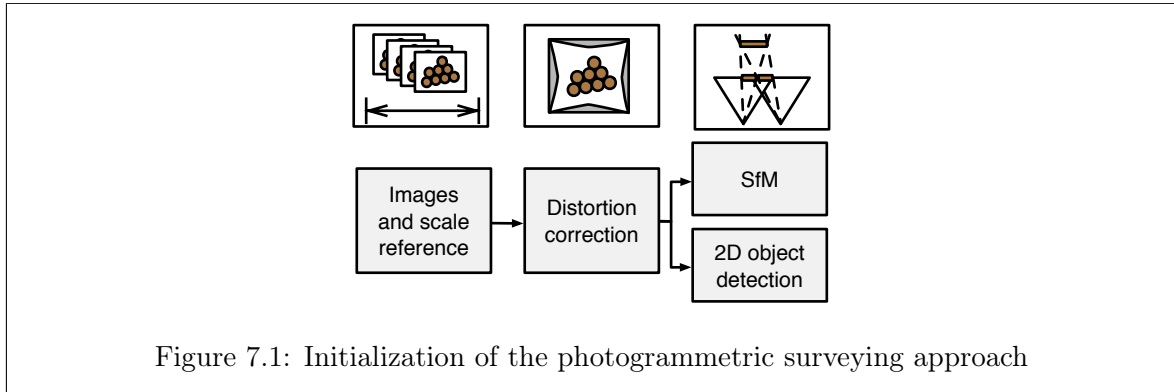
The model is rotated based on the principal axes (obtained via PCA or RANSAC plane fitting) so that it looks natural to a viewer. Similarly, while the texturing of the wood pile is not necessarily important for surveying purposes, it is important for interactive editing to emphasize whether or not a wood log was correctly detected. For the purpose of interactive editing, the wood pile must be projected onto an image plane. Orthographic projection, as opposed to a perspective projection, is used, in order to minimize distortions. Remaining wood logs can be added and non-existent logs can be removed during this step. Should a user modify the model this way, the computation is rerun. If no changes are made, the contour volume and the solid wood volume are computed.

Contributions   The proposed method provides a number of contributions, the most significant of which are:

1. The proposed method is the first mobile, photogrammetric wood pile surveying approach.
2. Clustered objects are recognized via 2D detection and reprojection to 3D.

3. The objects' locations in 3D are recovered through a novel quadric filtering and k-nn (k nearest neighbors) approach.

4. The location of a detected object can be tracked over multiple frames based on the location approximation.

5. The wood and contour volume are recovered based on the 3D objects' locations.

6. All algorithms are specifically optimized for mobile devices.

7. Results of extensive testing are provided to prove the feasibility of the proposed method.

8. It is shown how metric surveying via RTK-GPS can be performed.

7.2.2   Image acquisition

The topic of planar degeneracy has been discussed in Chapters 4 and 6. Decker et al. [DECKER et al., 2008] elaborate that a degenerate case occurs when multiple solutions can be found which are mathematically correct but which do not necessarily correctly map reality. Such cases include the possibility of all points lying in a small area in the image or all points lying on a common plane (planar degeneracy, a so-called H-degenerate configuration [CHUM et al., 2005]). The wood pile front surface poses a potential threat for planar degeneracy. During image acquisition it is therefore crucial to include a substantial amount of the scene which does not belong to the wood pile front surface.

Figure 7.3 shows how images of wood piles can be captured for this purpose. The image on the left most likely causes a degenerate configuration, since not many off-plane points are included. On the right hand side, the image includes a certain amount of ground pixels, which can be used as non-coplanar points during essential matrix estimation. In conclusion, the three-dimensional surveying approach inevitably reduces the space in an image, which can be used for wood log recognition, but it is for the same reason that the wood pile can be reconstructed in 3D.

7.2.3   Structure from motion

In Section 3.4 (*3D reconstruction / structure from motion*) different approaches for multiple view reconstruction were discussed. The approach by Moulon et al. [MOULON et al., 2013b] has shown promising results. The details of multiple view reconstruction exceed the scope of this thesis, therefore only an overview will be given. In contrast to the common incremental methods, [MOULON et al., 2013b] is a global approach, which achieves a higher precision with regard to the ground truth camera locations in addition to a lower computation time. Since a global solution is computed, the explicit choice of a starting pair is obsolete. The approach works on image triplets, unlike incremental methods, which only use image pairs and resectioning. This requires input images for this approach to show a comparatively high
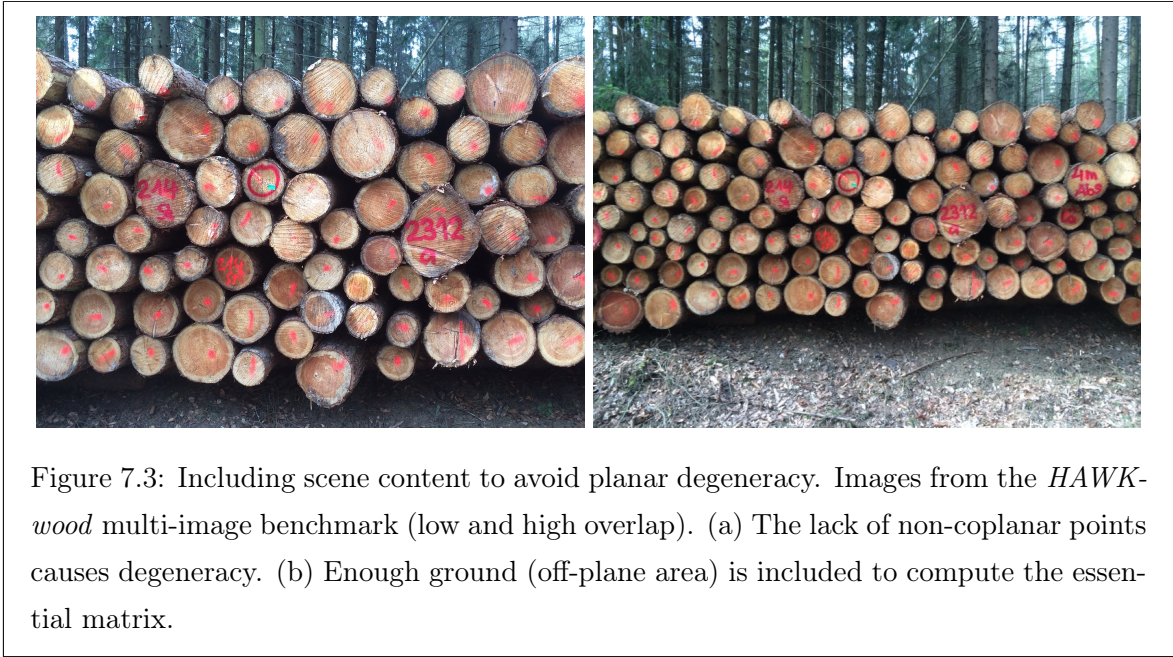
Figure 7.3: Including scene content to avoid planar degeneracy. Images from the *HAWK-wood* multi-image benchmark (low and high overlap). (a) The lack of non-coplanar points causes degeneracy. (b) Enough ground (off-plane area) is included to compute the essential matrix.

overlap. As its name implies, structure from motion provides two results, the structure of the scene and the motion of the camera. The structure of the scene is provided as a sparse point cloud, which consists of triangulated feature matches. For each of the input images I_i the corresponding camera pose is provided as the 4x4 pose matrix P_i in the pinhole camera model, shown in Equation 7.1 [HARTLEY and ZISSERMAN, 2004, p. 156].

$$P = K[R|t] = K[R| - RC] \tag{7.1}$$

In the above equation, K is the camera calibration matrix, R is the rotation of the camera coordinate system, t is the translation of the world coordinate system in camera coordinates, and C represents the camera center in world coordinates. Structure from motion only recovers the scene and camera trajectory up to scale and unknown translation with respect to a real world coordinate system. The term world coordinate system refers to a fixed coordinate system with arbitrary scale and arbitrary translation with respect to a real world reference frame. Both scale and translation can be recovered by registering the camera motion with global coordinates, which were obtained, e.g., through a global navigation satellite system (GNSS). This approach is discussed in Section 7.3 (*Scale reference*), as part of the scale reference selection.

7.2.4 3D object detection

Patch definition The detection of wood log cut surfaces is one of the most crucial parts of the surveying approach. For each of the input images $I_0 \ldots I_{n-1}$, two-dimensional object

Figure 7.4: 3D object detection approach. Image based on [HERBON et al., 2014a].
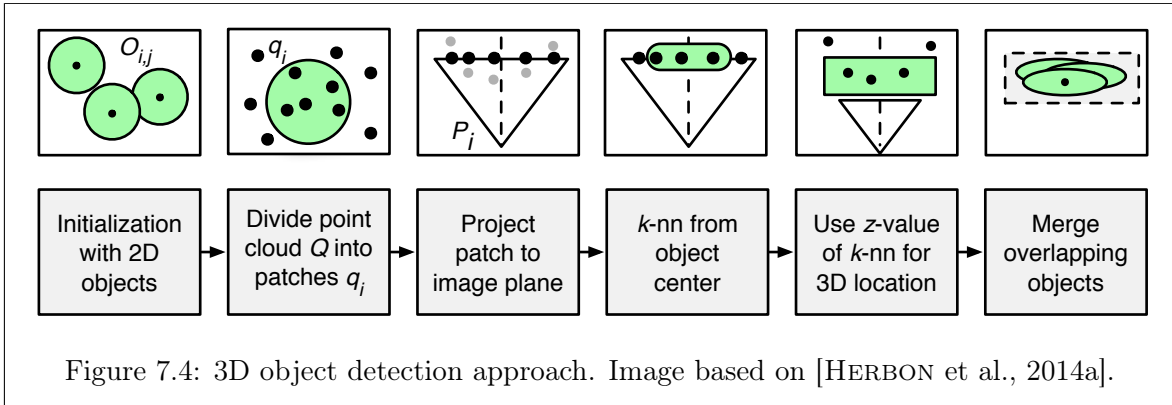
detection is performed, as described in Chapter 5. The recognized objects are contained in the object set O. For an object $O_{i,j}$, the image index is denoted by i, and j defines the object within this specific image. Each object consists of a 2D location (x,y) and a parameter r, which indicates the radius, thus $O_{i,j} = \{x_{i,j}, y_{i,j}, r_{i,j}\}$.

The point cloud (*structure*), as the result of structure from motion, is denoted as Q, which is a set of vertices in $\mathbb{R}^3$. Each vertex Q_m is defined as $Q_m = \{x_m, y_m, z_m\}$. In order to recover the 3D location approximation of an object, in this case a wood log cut surface, the point cloud must be divided into patches q, where each patch q_i is a subset of the point cloud Q (see Figure 7.4). The goal is to extract a patch from the point cloud for each of the detected 2D objects $O_{i,j}$. Figure 7.5(a) shows how a patch (dark points) can be defined through a cone (dashed line), which starts at the camera center and whose radius is defined by a detected object (green). The cone can be thought of as the projection of the detected object on the image plane back to 3D space. All vertices, which lie within the cone, are considered to belong to the patch q.

Figure 7.5: Definition of a patch through quadric filtering. The dashed line defines the cone, black vertices are kept after filtering. Image source: [HERBON et al., 2014a]
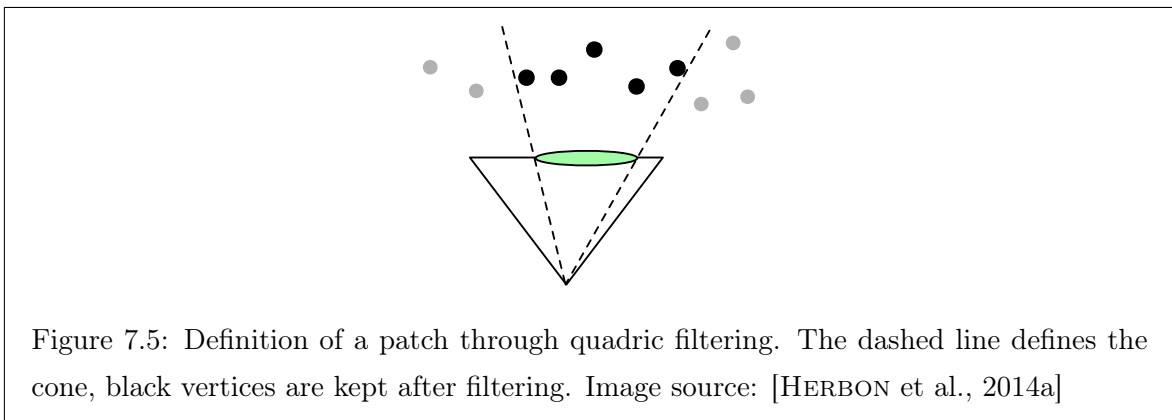
When assuming that a camera pose $P_i = [I]$, where $[I]$ is the 4x4 identity matrix, a circular

cone can be defined as a quadric, shown by Equation 7.2 [ZWILLINGER, 2011, pp. 213-215].

$$x^T A x + 2b^T x + c = 0 \tag{7.2}$$

The parameters of Equation 7.2 are defined as $A \in \mathbb{R}^{3 \times 3}$, $b \in \mathbb{R}^3$, $x \in \mathbb{R}^3$, and $c \in \mathbb{R}$ with

$$A = diag\left(\frac{1}{r_{i,j}^2}, \frac{1}{r_{i,j}^2}, -\frac{1}{f_i^2}\right) \tag{7.3}$$

$$b = \mathbf{0} \qquad c = 0 \tag{7.4}$$

where f_i is the focal length of camera P_i and x is a vertex in $\mathbb{R}^3$. This quadric definition of the circular cone can now be used to obtain all vertices of the point cloud within this cone. Since the cone is not limited to the space in front of the camera, an additional constraint must be imposed, which rejects points behind the camera. All remaining vertices form the patch q_i.

Depth approximation In order to determine the location of the wood log object along the z-axis of the cone, the vertices of q_i are projected onto the image plane (Figure 7.6). A vertex $v = \{v_x, v_y, v_z\}$ from the patch q_i with its corresponding camera pose P_i is transformed in a way that the camera pose becomes the identity matrix. The transformed camera matrix P_i' can be obtained through multiplication with the inverse of the camera matrix as $P_i' = P_i^{-1} \cdot P_i = [I]$. Correspondingly, the transformed vertex is $v' = P_i^{-1} \cdot v$. The projection of v onto the image plane is denoted as $w = \{w_x, w_y\}$ in $\mathbb{R}^2$ with $w_x = f_i \frac{v_x}{v_z}$ and $w_y = f_i \frac{v_y}{v_z}$.

Figure 7.6: (a) Projection of a patch to the image plane. The dashed lines indicate the projection rays. (b) Visualization of the average z-value of the patch.
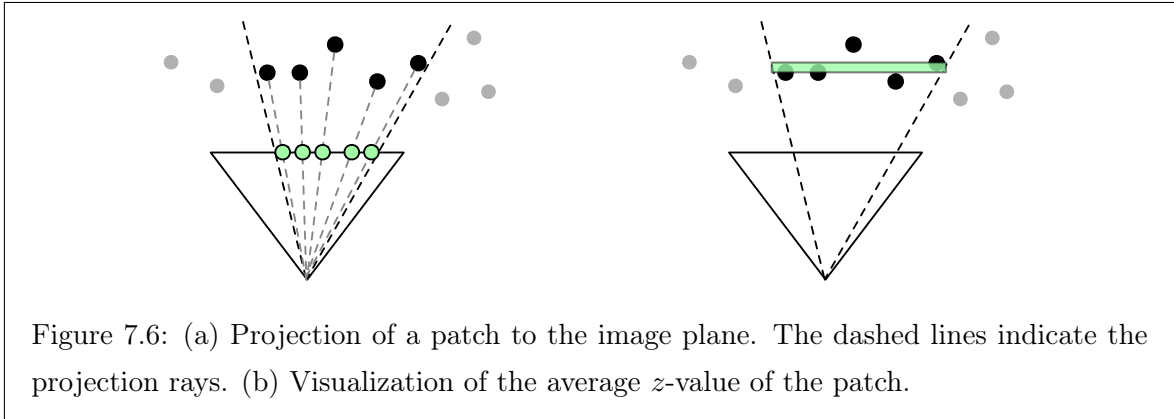
All projected vertices are scanned for statistical outliers (with regard to the z-axis), which are excluded in the process. The k nearest vertices are selected that show the smallest Euclidean distance from the object center. Out of these k vertices, the average z-value is computed and used as the approximation of the object center's z-value, which concludes the location approximation.

Object orientation Molton et al. [MOLTON et al., 2004] propose to view a patch around a matched image feature as locally planar. This definition can be extended for the purpose of multi-vertex patch definition, as to view the vertices of a patch q_i as locally planar. Due to this local planarity property, it is possible to determine a surface normal for each patch with a RANSAC plane-fitting approach, implemented by Rusu and Cousins [RUSU and COUSINS, 2011] based on [FISCHLER and BOLLES, 1981]. For this approach, at least three vertices must be contained in a patch, in order to define a plane. This constraint is not always met. Structure from motion usually only performs sparse feature matching, and the point cloud is thus also sparse with no information about the vertex density. The parameter k for the k-nearest neighbors approach is selected based on geometric distance information. Should the case occur that $k < 3$, fewer than three vertices of the patch are selected. In this scenario, the radius r must be increased until $k \geq 3$ and the surface normal of the patch can be approximated.
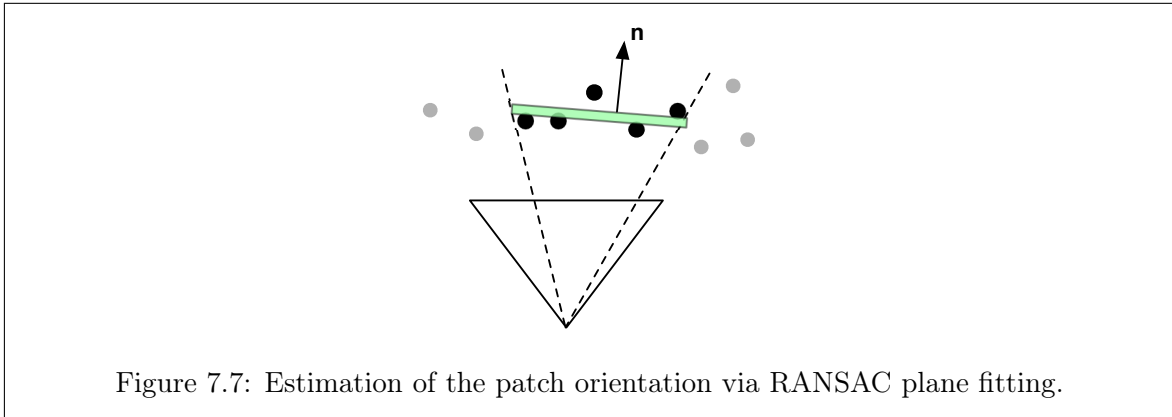
Figure 7.7: Estimation of the patch orientation via RANSAC plane fitting.

When performing wood log detection, a priori information can be optionally incorporated into the surface normal estimation. It is known that a wood pile front surface is quasi-planar. For the recovery of the patch normal, local quasi-planarity is sufficient, as opposed to global quasi-planarity of the entire pile front surface. This way, it can be determined whether the computed surface normal of the wood log (Figure 7.7) is approximately parallel to the surface normal of the surrounding wood pile front surface.

Object merging So far, the objects $O_{i,j}$ in each individual image have been located in 3D. The single indexed set $O_i = \{x_i, y_i, z_i, r_i\}$ denotes the set of global, three-dimensional objects, which can be obtained by merging the objects in $O_{i,j}$. Since the input images show significant overlap, the same object will most likely have been detected multiple times. These multiple detections must be merged in order to determine the correct number of wood logs but also to properly compute the wood volume. In the merging step, the Euclidean distance between neighboring objects O_a and O_b is computed, denoted as $D(O_a, O_b)$. An object O_a is merged

with all objects O_i with $i = 0 \dots n$ and $i \neq a$ if the following conditions hold true:

$$D(O_a, O_i) < \max(r_a, r_i) \tag{7.5}$$

$$0.8 \leq \frac{r_a}{r_i} \leq \frac{1}{0.8} \tag{7.6}$$

This means that the Euclidean distance between the objects' centers must be smaller than the larger radius. By enforcing this condition, it can be assured that all objects which lie within another object's bounds are considered to be the same object. Additionally, the ratio between the radii must not be smaller than 0.8 or larger than $\frac{1}{0.8}$.

An advantage of the multiple view recognition approach is its capability to filter outliers based on the number of images in which an object was detected. This number will be referred to as the minimum number of occurrences n_{occ}. In the merging step, n_{occ} is utilized in a way that only objects are kept which were detected in at least n_{occ} images. At the same time, this implies that at least n_{occ} images must be taken of a particular wood log, in order for it to be recognizable. The impact of this parameter will be studied in the result Section 7.4.1 (*Wood log recognition*).

7.2.5   Wood pile reconstruction

In the previous section, the wood logs were reconstructed in 3D space. The current section builds on this result by reconstructing the wood pile model from the individual logs. The 3D model of detected objects exhibits an arbitrary rotation and scale. The first step for pose correction of the wood pile model is the computation of the concave hull, based on the vertices of the reconstructed objects. Moreira and Santos [MOREIRA and SANTOS, 2007] propose a k-nearest neighbors approach for concave hull calculation, which takes as a parameter k, the number of adjacent vertices that should be considered. For the goal of contour volume estimation, k is iteratively refined from an initial guess of $k = 3$ until a single concave hull is found.

Considering the arbitrary rotation of the wood pile model, the goal is to perform a rotation correction, so that the wood pile looks natural to an observer. This is not necessary for the surveying procedure but it enhances the identification of digital and real wood logs. The rotation about the x- and y-axis is corrected by performing the aforementioned RANSAC plane fitting for the entire wood pile front surface. This can be done if the surveying object, such as the wood pile front surface, can be described as a quasi-planar surface. Alternatively, a principal component analysis of the vertex distribution can be used as a fallback, which yields the eigen vectors Υ_1, Υ_2, and Υ_3 of the pile, where Υ_1 is the eigen vector with the largest eigen value. By using Υ_2 and Υ_3 for orientation correction, similar results can be achieved

Figure 7.8: Final wood pile reconstruction result. Image source: [HERBON et al., 2015b].

with the advantage of the method being applicable to the general case. In the process, the x-y-plane of the wood pile is aligned to its canonical position, which is the x-y-plane in the coordinate system of the 3D model.

The rotation of the wood pile about the z-axis is not as trivial as would be intuitively expected. In most cases, the width vector (largest measurable distance) of the pile, which corresponds approximately to Υ_1, aligns well with the x-axis. However in some cases, the wood pile is located on a mountain side so that one end is at a higher or lower altitude than the other end and thus a rotation about the z-axis is introduced. Again, for surveying, this orientation is not important, but it aids in optimizing the interactive editing. Szeliski [SZELISKI, 2006] points out that an observer has an expectation of the object's or image's y-axis being aligned to the gravity vector. Υ_1 of the pile could be used for this as an approximation, by rotating the model so that Υ_1 is orthogonal to the gravity vector (which in this case is the y-axis). An approach which is more natural for an observer is discussed by Szeliski [SZELISKI, 2006]. It is theorized that a user, when taking pictures, intuitively holds his camera in parallel to the horizon, which implies that the image's y-axis is aligned to gravitation vector. When considering multiple cameras, e.g., in panoramic image stitching or structure from motion, this task can be formulated as a least squares problem. The result of this formulation yields a global rotation matrix that transforms (in this case) the 3D model in a way that it is correctly aligned to the gravity vector. [SZELISKI, 2006]

The problem of scale recovery will be addressed in Section 7.3 (*Scale reference*). For now it can be assumed that the scale was robustly computed and the wood pile model can be used for metric surveying. At the same time, the depth of the pile is known and the wood logs' cut surfaces are extended accordingly. The final reconstruction model is obtained by performing texture mapping from the images onto the detected 3D objects. The camera, whose optical axis aligns best with an object's surface normal, is used for the extraction of the texture. This way, distortions can be minimized and the initial reconstruction of the wood pile front surface

is concluded, as shown in Figure 7.8.

7.2.6 Interactive editing

Before computing the contour volume and the solid wood volume, it must be ensured that all wood logs were correctly detected and that no false positives are visible. This is especially important for the scaling of the model, since the far left and far right wood log are used for this task. A failure to mark these logs results in incorrect scaling of the model. The goal in this section is to provide an interface for a user to check the reconstruction and detection result. In the previous section, rotation correction was conducted. This makes the wood pile model eligible for projection onto the x-y-plane. To minimize distortions, orthographic projection is used for this task, with the resulting image being used as a visual representation for user input. A user can freely move, scale, add, and delete objects. The process of deleting objects is trivial, but adding objects to the model requires additional computational effort. Similar to the object detection step, quadric filtering is performed when specifying a location at which an object should be inserted. In Section 7.2.4 (*3D object detection*), this process is described in detail. When scaling an object, the 3D location remains constant. In the case that an object is translated, the z-coordinate is recomputed through the quadric filtering approach. From Figure 7.2, it can be seen that if changes were made by a user, the 3D model is reconstructed again, based on the new 3D objects. This process is run until no further changes are made and surveying can commence.

7.2.7 Surveying

Contour volume The contour volume in this three-dimensional surveying approach is computed based on the concave hull of the point cloud. In contrast to the two-dimensional surveying approach, the orthographic projection of the wood pile front surface is used instead of the panoramic image. This yields the advantage that the model for the wood pile has been reconstructed more accurately through structure from motion and that only little distortions from perspective projection are introduced. The contour volume is defined through the projection of the front surface onto the x-y-plane, which produces the contour area A_c. Equation 7.7 shows Green's theorem [CAUCHY, 1846] (implementation by [RUSU and COUSINS, 2011]), which is used to efficiently compute the area surrounded by the contour, where C_{obj} is the object's contour.

$$A_c = \iint dA = \int_{C_{obj}} (f(x,y)dx - g(x,y)dy) \tag{7.7}$$

The resulting contour is smoothed via a morphological operation to ensure comparability to manual measurements, as proposed in Section 5.3. In Equation 5.5 it was shown that the
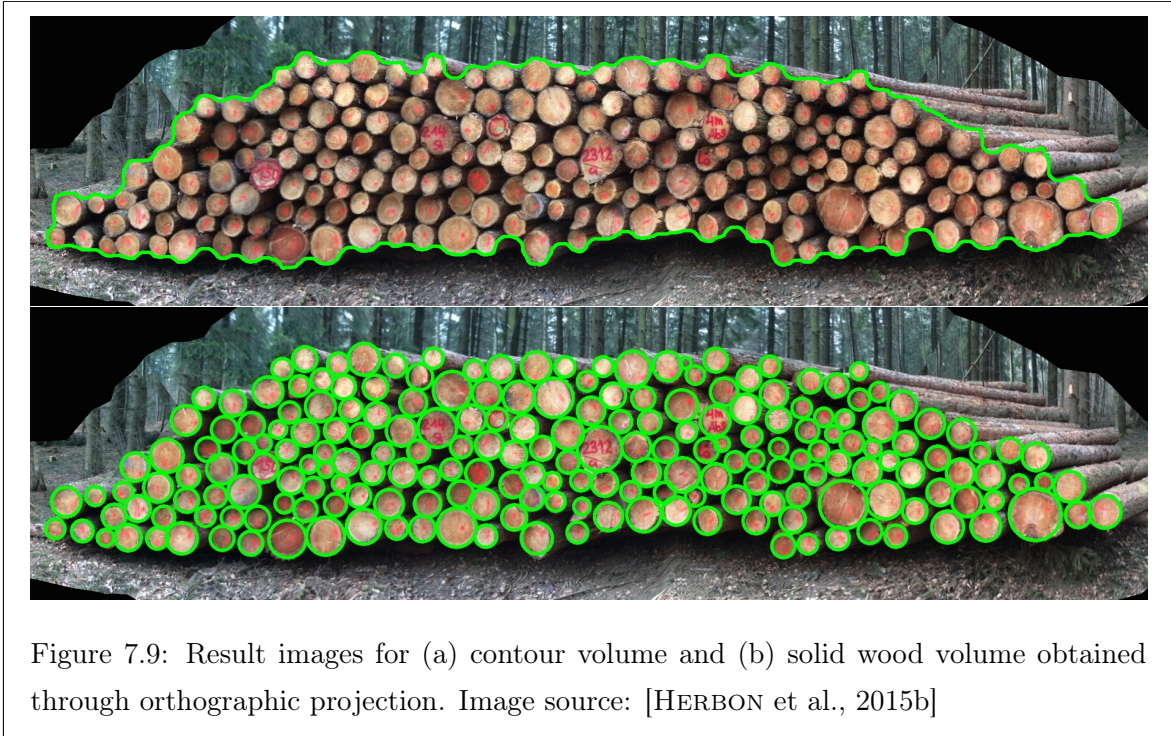
Figure 7.9: Result images for (a) contour volume and (b) solid wood volume obtained through orthographic projection. Image source: [HERBON et al., 2015b]

contour volume can now be obtained by multiplying the contour area A_c with the known wood log length l. A graphical representation can be seen in Figure 7.9(a).

Solid wood volume   The solid wood volume for the three-dimensional approach is not computed on a pixel level as in the two-dimensional approach, but directly on the 3D objects. The wood log faces in 3D are given by their circular approximation. In Section 2.3.1 (*Attributes of roundwood logs*), it was shown that the circular approximation of a wood log is generally accepted to be accurate enough. The solid wood area can thus be defined as the sum of all wood log faces, as shown in Equation 7.8. The solid wood volume is computed based on Equation 5.3, as previously discussed in Chapter 5. The result is shown in Figure 7.9(b).

$$V_\mathrm{s} = \sum_{i=1}^{N} \pi r_i^2 \tag{7.8}$$

7.3   Scale reference

7.3.1   Reference length

The recovery of the scale after 3D reconstruction has great impact on the surveying result. In Chapter 5 (*Two-dimensional surveying*), the wood pile width was used as a reference. The same referencing method will be used in most cases of the three-dimensional surveying. For
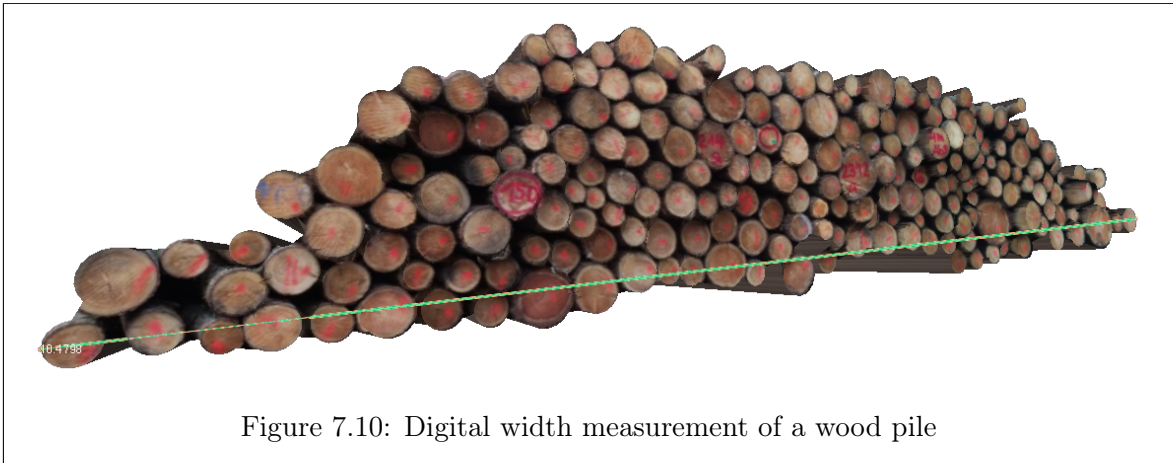
Figure 7.10: Digital width measurement of a wood pile

this, the largest measurable distance of the wood pile is used, which is defined by the outer ends of the wood logs on the left and right side (Figure 7.10). Table 7.1 shows the results of an extensive repeatability test, in which the width was measured 200 times by a total of seven different people, for wood piles with widths of between 4.1m and 35.3m. It can be seen that a very low repeatability standard deviation (defined in Equation 5.15) is achieved, which supports the case for using the pile width as a scale reference. However, due to knowledge about the camera locations, a couple of alternatives to width-referencing will be discussed in the next sections.

	measurements	$\bar{s}_{rep}$
pile width repeatability	200	0.12%

Table 7.1: Repeatability for width measurements

7.3.2   First alternative: monocular sensor fusion

Elmenreich [ELMENREICH, 2002] defines sensor fusion as the combination of sensory data for the purpose of enhancing the data. Some of the advantages according to Elmenreich are a higher robustness, an extended temporal and spatial coverage, an increase in the confidence of the data (which implies lower uncertainty and less ambiguity), and an increase in resolution. The term monocular sensor fusion refers to the use of monocular cameras, as opposed to stereo cameras. While a stereo camera can be calibrated to produce metric depth images, monocular cameras do not provide a scale of the depicted content. In theory, the acceleration and gyro sensors of a smartphone can be used to determine a metric scale of a camera movement in addition to the image sequence of the monocular camera. Mukai and Ohnishi [MUKAI and OHNISHI, 1999] show that the acceleration of a device, obtained from a noise-free acceleration-gyro sensor, can be integrated as shown by Equation 7.9, in order to obtain the metric velocity

of the camera. v_{G}^{W} is the velocity from the acceleration-gyro sensor, R is the rotation from the base world coordinate system, $\mathbf{a}^{W}$ is the sensor acceleration and $\mathbf{g}^{B}$ is the gravitation vector.

$$v_{\mathrm{G}}^{W}(t) = R(t)\left\{\int_{t_0}^{t}\left\{R^{-1}(\tau)\mathbf{a}^{W}(\tau) - \mathbf{g}^{B}\right\}d\tau + v^{B}(t_0)\right\} \tag{7.9}$$

In reality, the acceleration sensor data includes noise, which leads to a significant drift. If the velocity is integrated again, in order to determine the camera position, this drift is amplified. For that reason, simple integration of the accelerometer data does not work adequately in real world scenarios. Some research has been done to combine accelerometer and gyro data with the camera trajectory from a sequence of images, such as [JUNG and TAYLOR, 2001; MUKAI and OHNISHI, 1999], the work by Nützi et al. [NÜTZI et al., 2011], and the approach by Davison [DAVISON, 2003].

Mukai and Ohnishi [MUKAI and OHNISHI, 1999] propose an approach for shape recovery of a known object. The camera locations are fused with accelerometer and gyro data to obtain the unknown scale factor. This is achieved by minimizing a function which includes the scale factor and the velocity from both the image sequence and the acceleration sensor. Results for the scale factor are reported for simple and complex camera motions. In the simple scene, the scale factor was underestimated by up to 20%. In the complex case, a 94-frame sequence (5.64 seconds) was used. According to the authors, the scale could not be recovered reliably. A similar method is shown by [STRELOW and SINGH, 2002], who report position errors of up to 25cm with unknown trajectory length.

Davison [DAVISON, 2003] proposes to use an object with known scale for initialization, such as a sheet of paper. The goal of this approach is long-term navigation accuracy; the topic of scale recovery is not discussed in detail. Nevertheless, using a reference object seems like a straightforward approach, which is similar to the use of the pile width. An offline method is proposed by Jung and Taylor [JUNG and TAYLOR, 2001], who aim to perform trajectory fitting between inertial and vision data. A small set of key frames is extracted from an image sequence for which structure from motion is computed. For a 6.1m camera trajectory of an omnidirectional video sequence, an average position drift of 4.5cm with a maximum drift of 16.4cm was achieved. For another sequence with large rotational motions, a maximum drift of 22.1cm (average 8.2cm) is reported. For a conventional, non-omnidirectional video sequence of eleven second duration (and unknown camera displacement), an average displacement error of 1.05cm with a maximum error of 3.35cm is given.

A more complex approach is presented by Nützi et al. [NÜTZI et al., 2011], with the implementation of Weiss et al. [WEISS et al., 2011] being based on this method. The authors compare the method by Jung and Taylor [JUNG and TAYLOR, 2001] and a novel approach, where an Extended Kalman Filter (EKF), which fuses inertial and vision data, is used. According to Nützi et al. the approach by Jung and Taylor provides a strongly varying accuracy

in scale estimation, reaching from less than 5% to more than 50%. The Extended Kalman Filter results for real data are shown as diagrams where the convergence of the scale λ is given over time. After approximately 30 seconds, the scale converges, but still jitters in a range of about 15%. Experiments performed within the scope of this thesis showed that even on the same image sequence as used by the authors, the results are highly dependent on noise estimation parameters and a correct initialization. The scale estimation, as reported by the authors, could be achieved in the best case, but even slight deviations from the initialization led to large differences in the estimated scale.

Out of the discussed methods, the approach by Nützi et al. seems to be the most promising. Although these methods achieve good results in their respective fields, the scale estimation is not suitable for wood pile surveying. The reason for this is twofold. Firstly, it seems excessive for both computation time and consumption of resources to capture a video sequence of several minutes and process a 3D reconstruction on a mobile device. While it is possible to do so, the results are far less accurate when compared to using a reference object or length, which can be seen from the reported scale and position errors. Even in the best case, the scale results are not accurate enough. Secondly, the scale in [NÜTZI et al., 2011], although it converges, still varies over time due to jitter, which makes it impossible to determine an adequate cutoff point.

7.3.3   Second alternative: high precision global positioning

Related work   The location determination via global positioning with mobile receivers has significantly improved over the last few years. GPS, as the first global positioning system, was deployed in the 1970s [EL-RABBANY, 2002, pp. 78–80]. Today, the Russian GLONASS system is also available, while several other positioning systems are currently being established, such as the European GALILEO, the Indian IRNSS, the Japanese QZSS, and the Chinese COMPASS [BAUER, 2011]. The accuracy of GPS is described to be between 10m and 22m under perfect conditions [KAHMEN, 2005, p. 315; BAUER, 2011, pp. 218–224], which is not accurate enough for scale estimation of objects smaller than a few kilometers. This problem can be overcome by using correctional signals [WANNINGER, 2006] in addition to the pure satellite-based navigation. Centimeter and even millimeter accuracy can be achieved when using RTK-GPS (Real Time Kinematic) [WANNINGER, 2006]. RTK uses correctional signals from local reference stations, which are received via a cellular data connection. In most of the related work, the fusion of structure from motion, SLAM, or visual odometry with RTK-GPS data is only addressed as a means to compare the recovered camera positions to ground truth, instead of incorporating these measurements for scale estimation (e.g. [AGRAWAL and KONOLIGE, 2007; CIVERA et al., 2009; KONOLIGE et al., 2011; ROYER et al., 2005]). In contrast, [HWANG et al., 2012] propose to process RTK data directly on a smartphone, thus

enabling high precision location estimation on mobile devices. A similar approach is followed by [KIM et al., 2013] for gravity surveying.

Method   The fusion of structure from motion with RTK-GPS for the purpose of scale recovery is rarely addressed in related work. Nevertheless, the approach seems promising as an alternative to using a reference length. The idea behind the approach is to perform the structure from motion reconstruction of the wood pile, as discussed in Section 7.2.3, and then register the camera positions with the RTK-GPS locations for each frame. As opposed to using camera translations as priors in structure from motion, this way the scale can be recovered implicitly without a reference length, and the results can be compared to those obtained by using the width as the reference. Moulon and Duisit [MOULON and DUISIT, 2014] provide an implementation based on [HARALICK and SHAPIRO, 1992], which computes a transform (rotation, translation, and scale) of two point sets. Jian and Vemuri [JIAN and VEMURI, 2011] formulate this task as an optimization problem, where the goal is to find a transformation T, which transforms a point set S to align with a point set M. The optimization can be performed in a least-squares sense (with the disadvantage of being sensitive to outliers) or by using an M-estimator [HUBER, 1981].

In the case of wood pile surveying, M is the set of ground truth camera locations, obtained through RTK-GPS and S contains the camera centers after computing the structure from motion results. The point set registration algorithm yields the transform T, which is not only applied to the cameras from SfM but also to the point cloud representing the scene, i.e. the fully-reconstructed wood pile. This way, the wood pile volume can be computed with a metric scale based on the 3D model, without having to explicitly scale the model with a reference length.

Challenges   In contrast to pure GNSS location estimation RTK uses a cellular data connection which requires mobile internet. As discussed in Section 2.1 (*Environmental constraints*), mobile internet coverage is very poor in most rural areas. This makes it difficult, if not impossible, for an RTK receiver to communicate with a cell phone tower, thus complicating the reception of the correctional signal. Additionally, the satellite coverage suffers significantly when the sky is occluded, e.g., when covered by tree branches. This problem occurs frequently when performing measurements in or near a forest. Similar to buildings in urban areas, multipass noise occurs when mountains are close to the GNSS receiver [PETROVSKI, 2014, p. 229]. For RTK, this leads to a lower accuracy and thus disqualifies the particular measuring site. For this reason, only a limited set of test cases is discussed below, since it is very difficult to find wood piles which qualify for both a good cellular reception and a good satellite signal without occlusions of the sky. Nevertheless, the results give an impression of the suitability of this technique.

Width measurement comparison   The first experiment is concerned with the comparison of the scale factors obtained by a) the width measured with a tape measure and b) the width measured via RTK-GPS. The conventional measurement by hand was conducted by measuring the far left and far right end of the pile. The exact same positions were located via RTK-GPS with an accuracy (as given by the receiver) of less than 1cm (usually 0.7-0.8mm) for latitude and longitude and less than 2cm for height. Three different wood piles could be found, which satisfy the conditions described in the previous paragraph, with widths of between 3m and 7m. A total of 21 measurements were performed.

	$\bar{x}$	s
scale difference manual width / RTK-GPS	-0.002m (0.027%)	0.041m (0.746%)

Table 7.2: Absolute and relative difference between RTK-GPS width measurements and manual width measurements via tape measure.

Table 7.2 shows the results for these measurements. The absolute as well as the relative errors are very small, with the average error being only 0.027% with an empiric standard deviation of 0.746%. From these results it becomes clear that both referencing techniques only differ insignificantly. Hence, the width measurement via tape measure as well as the width measurement via RTK-GPS are interchangeable in real world use. The highly precise RTK-GPS measurement provides an advantage, when the wood pile geometry deviates strongly from a quasi-planar surface, since the width might not be easily measurable by tape measure.

Scale factor comparison   The fusion of RTK-GPS and structure from motion data is done as explained above, by the method of Moulon and Duisit [MOULON and DUISIT, 2014] based on [HARALICK and SHAPIRO, 1992], which means that the camera locations are registered to the RTK-GPS locations and transformed accordingly, including a scaling operation. The computed scale factor is compared to the manually measured scale factor by reference length (tape measure) for a total of 14 data sets. Two different structure from motion techniques, global [MOULON et al., 2013b] and incremental [MOULON et al., 2013a], are used to compare the scale factors, as shown in Figure 7.11. The relative differences between the computed scale factors are given, including their respective empiric standard deviation.

scale factor	$\bar{x}$	s
global [MOULON et al., 2013b]	0.5%	4.3%
incremental [MOULON et al., 2013a]	3.3%	4.5%

Table 7.3: Relative difference between the scale factor via manually-measured reference length and scale factor via RTK-GPS for Figure 7.11

Figure 7.11: Comparison of the scale factor via reference length (tape measure) and scale factor via RTK-GPS. The relative error between the RTK-GPS length and the reference length is given. Numeric values are shown in Table 7.3.
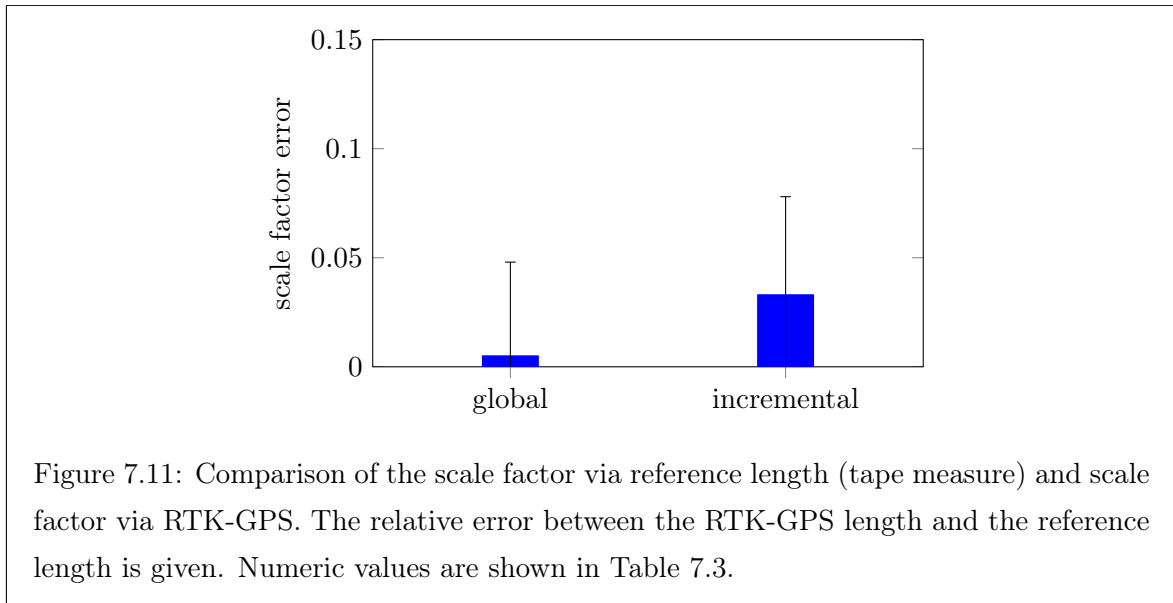
The referencing via RTK-GPS yields on average a slightly higher scale factor than the scaling via reference length. While the empiric standard deviations are approximately equal, the global method shows a much lower average offset (0.5% vs. 3.3%). This, again, speaks for the higher robustness of the global method with regard to drift.

In order to assess the quality of the 3D reconstruction and its influence on the scale factor determination, the average absolute difference of the SfM camera positions in comparison to the RTK-GPS locations were determined (Figure 7.12). In accordance with [MOULON et al., 2013b], it can be said that the global approach is more precise in terms of the average camera position error and the empiric standard deviation of the positions. The differences of the reconstructed camera locations to the ground truth locations explain the deviations in the scale factor estimation. With increasing quality of the structure from motion reconstruction, the scale factor estimation becomes more precise with regard to offset and standard deviation.

camera position error	$\bar{x}$	s
global [MOULON et al., 2013b]	0.016m	0.008m
incremental [MOULON et al., 2013a]	0.026m	0.019m

Table 7.4: Camera position error (SfM vs. RTK-GPS) for the global and the incremental method, as shown in Figure 7.12

Conclusion Summarizing the insights of this section, the scale reference can be most accurately determined through the wood pile width. In comparison to ground truth, obtained via RTK-GPS, the width measurement only exhibits a 2mm (0.026%) offset with an empiric
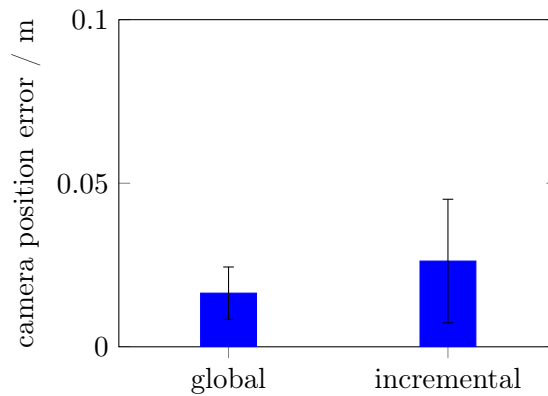
Figure 7.12: Camera position error (SfM vs. RTK-GPS) for the global and the incremental method. Numeric values are shown in Table 7.4.

standard deviation of 41mm (0.746%) on average. The first alternative scale referencing approach via sensor fusion is not accurate enough for real world use, since stable conditions are hard to achieve. Even in best case scenarios, a scale deviation of up to 15% is possible. The alternative scale estimation via RTK-GPS shows promising results. While the offset and standard deviation are approximately one order of magnitude higher than manual width measurement, this type of scale estimation could in theory be done without additional hardware, assuming that future mobile devices could be equipped with RTK-GPS sensors, while on-device processing of RTK-GPS data is already possible. The fusion of camera locations is dependent on the quality of the 3D reconstruction result. Using the global method provides a more accurate scale estimation than the incremental method, which is due to the position errors of the individual cameras of the structure from motion result.

7.4   Surveying results

In this section, the results for the proposed photogrammetric surveying approach are discussed. The data basis for the evaluation is the multi-image benchmark of the *HAWKwood* database. A total of 246 data sets are provided; 40 of these sets are synthetic, which means that the input images were digitally rendered. For all synthetic data sets, the number of wood logs, the contour volume, and the solid wood volume are known. Out of the 206 real data sets, the contour volume via the manual section volume method is provided for all data sets, 71 sets include the solid wood volume obtained via the complete diameter inventory, and the number of wood logs is known for 147 data sets. Four of the data sets could not be reconstructed via structure from motion. The reason for this is a mixture of too little overlap and planar degeneracy.

Three different cases are analyzed, to obtain an impression of the method's performance under varying conditions. Firstly, the regular structure from motion results are used as input. Since the point cloud in this case only contains few vertices, this will be referred to as the sparse case. Secondly, the same point cloud is post-processed by applying a multiple view stereopsis algorithm, as proposed by Furukawa and Ponce [FURUKAWA and PONCE, 2007]. The result of this method is a point cloud with more vertices and thus a higher vertex density, which will be called the dense case. Thirdly, synthetic data sets are used to make a reliable statement in noise free conditions about the accuracy of the proposed methods.

7.4.1   Wood log recognition

Performance for different data sets   Figure 7.13 shows the results for the true and false positive wood log detection rate. The true and false positive rate are defined in the same way as for 2D surveying, which is by dividing the true and false positive number of logs by the ground truth number of logs (see Section 5.4.1 (*Evaluation metric*), Equations 5.6 and 5.7). Interestingly, there is no significant difference between the sparse and the dense true positive rate. The false positive rate on the other hand is more than twice as high for the dense data sets. From this it can be concluded that the multiple view stereopsis algorithm introduces artifacts and noise which make wood log localization more difficult. This stands in contrast to the intuitive expectation that a higher vertex density provides a more accurate result. Indeed, the very precise vertices, which are the result of the global structure from motion pipeline, are better suited for object localization.

The true and false positive rate for the sparse synthetic data sets outperform the real data sets. This is not surprising, since the synthetic data sets provide intrinsic calibration information, no lens distortion, and are generally free from noise or compression artifacts. In other words, under ideal conditions the proposed three-dimensional wood log recognition approach works almost perfectly. For real data sets with significantly worse input data, only a slight decrease in performance is visible, which advocates the robustness of the method with regard to image quality and extrinsic parameters, such as the camera trajectory.

wood log recognition	$\bar{x}_{\text{tpr}}$	s_{tpr}	$\bar{x}_{\text{fpr}}$	s_{fpr}
sparse	0.988	**0.013**	**0.007**	**0.013**
dense	**0.991**	0.014	0.015	0.019
synthetic	0.999	0.003	0.003	0.005
2D surveying (sec. 5.4.4)	0.985	0.015	0.020	0.027

Table 7.5: Wood log recognition results for the three-dimensional approach in comparison to the 2D method for Figure 7.13

Figure 7.13: True and false positive wood log recognition rates for different data sets. Numeric values shown in Table 7.5.
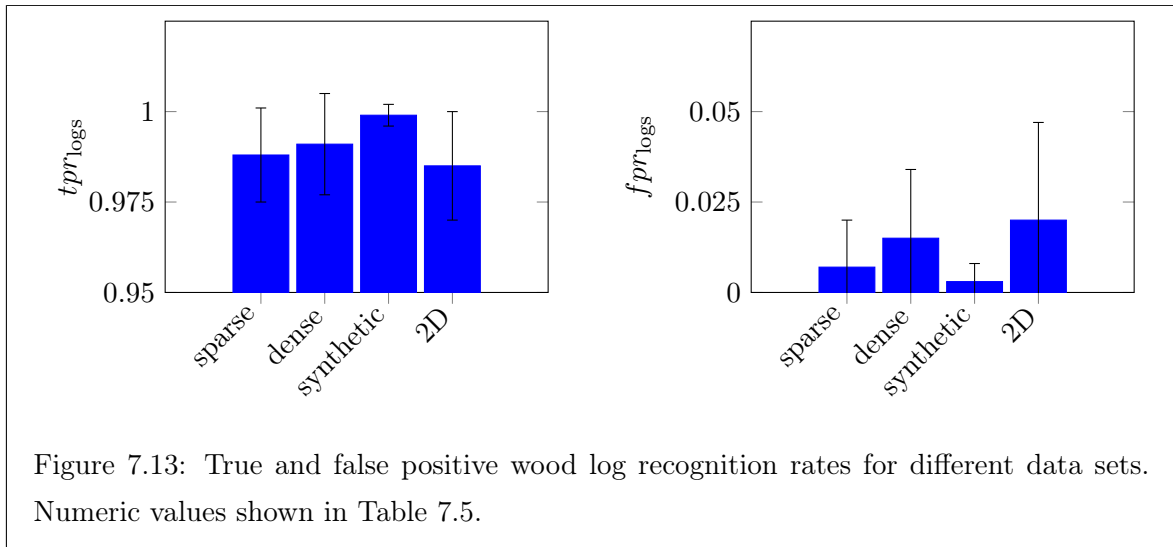
Comparison to 2D surveying   The comparison between the real data sets for the two-dimensional and the three-dimensional approach in Table 7.13 shows that the 3D approach outperforms the 2D approach for all measured quantities. The true positive rate is higher for the photogrammetric approach, with a lower empiric standard deviation. The most noticeable difference is the false positive rate, which is approximately three times smaller for the sparse data sets. The use of the 3D approach is thus very beneficial for wood log recognition.

Number of images per wood pile   An important aspect of the discussed approach is the number of input images. The true and false positive rate, with respect to the number of images for the same synthetic wood piles, are shown in Figure 7.14. As expected, the true positive rate increases with the number of images and the false positive rate decreases. From this fact the conclusion can be drawn that, although the changes are not very drastic, a too large or a too small number of images is not beneficial for the method's performance. At a certain point the true positive rate does not increase when adding more images but the false positive rate is amplified. As a rule of thumb, approximately 2-3 images per meter of the pile width provide the best results.

number of images	12	25	50	100
$\bar{x}_{\text{tpr}}$	0.997	0.999	1.000	1.000
s_{tpr}	0.003	0.002	0.002	0.002
$\bar{x}_{\text{fpr}}$	0.000	0.001	0.003	0.009
s_{fpr}	0.000	0.002	0.004	0.004

Table 7.6: True and false positive detection rate for different numbers of images, as shown in Figure 7.14
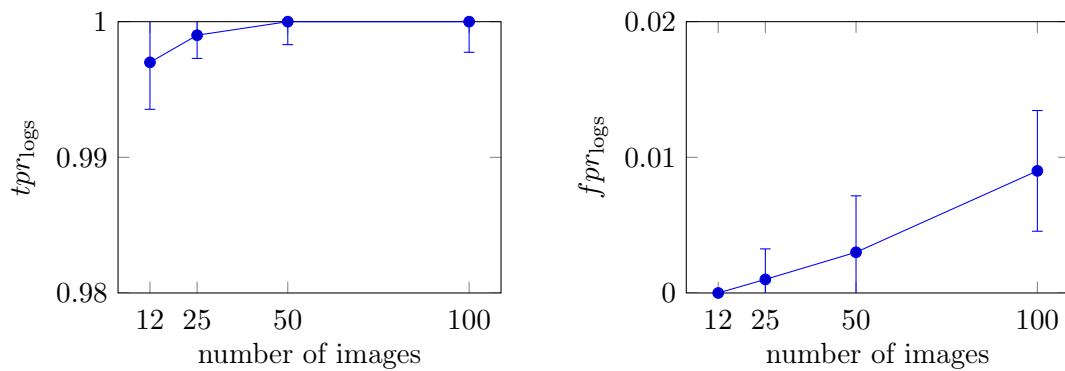
Figure 7.14: True and false positive rate with respect to the number of images for the same wood pile. Numeric values shown in Table 7.6.

Minimum number of occurrences During the object merging step, the minimum number of occurrences can be specified. The impact of this parameter is shown in Figure 7.15 for the sparse real data sets. A reasonable parameter range has been found to be $n_{occ} = 1 \ldots 3$. For $n_{occ} = 3$, the false positive rate only decreases slightly while the true positive rate decreases almost linearly. $n_{occ} = 2$, which was used in all other experiments, seems like a good parameter choice, since the true positive rate is still comparably high while the false positive rate is significantly lower.
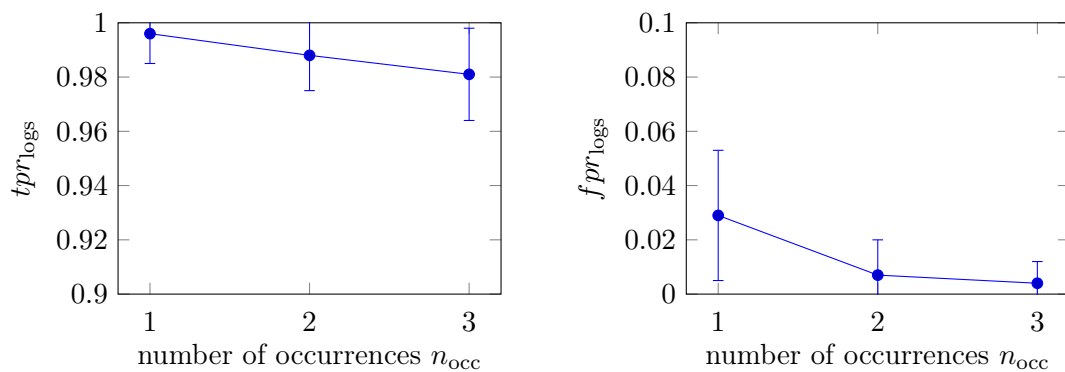
Figure 7.15: True and false positive rate with respect to wood log occurrences. Numeric values shown in Table 7.7.

occurrences	1	2	3
$\bar{x}_\text{tpr}$	0.996	0.988	0.981
s_tpr	0.011	0.013	0.017
$\bar{x}_\text{fpr}$	0.029	0.007	0.004
s_fpr	0.024	0.013	0.008

Table 7.7: True and false positive rate for different numbers of occurrences, as shown in Figure 7.15.

7.4.2 Solid wood volume

The solid wood volume for the three-dimensional approach is computed as shown by Equation 7.8. This is done after interactive editing, so that all wood logs are correctly marked as such and no false positives are present. The result of the solid wood volume surveying are shown in Table 7.8. Interestingly, the three-dimensional result exhibits a larger offset compared to the 2D surveying approach. This is caused by the merging step, in which the size of an individual wood log is averaged. In some cases, wood logs are smaller due to projective distortions of the input image or insufficient detection of the bark and thus the log size is decreased. An alternative could be to use the median diameter.

The standard deviation is slightly better for the 3D approach, but most importantly, the 3D results show a lower repeatability standard deviation, as defined in Section 5.4.1 (*Evaluation metric*). This speaks for the robustness of the method in terms of repeatability, as it outperforms the two-dimensional approach, which means that the method produces stable results when surveying the same wood pile multiple times. The same conclusion can be drawn from synthetic data. While these data sets show a slightly higher offset, the repeatability standard deviation is very small.

solid wood volume	$\bar{x}$	s	s_rep
sparse	-0.056	**0.034**	**0.020**
dense	-0.078	0.037	0.021
synthetic	-0.070	0.032	0.013
2D surveying (sec. 5.4.4)	**-0.026**	0.038	0.023

Table 7.8: Solid wood volume results for the three-dimensional approach in comparison to the two-dimensional results

7.4.3 Contour volume

The contour volume surveying outperforms the 2D approach in all categories. There exist only very slight differences for the dense and the sparse models, which shows that the use of a multiple view stereopsis algorithm as a post-processing step does not provide a significant benefit, but does negatively affect the processing time. Similar to the solid wood volume, the synthetic results are very close to the real data results, with a lower repeatability standard deviation than the 2D method. Again, this indicates a high robustness of the method with regards to producing the same values for the same wood pile.

contour volume	$\bar{x}$	s	s_{rep}
sparse	0.015	0.029	**0.013**
dense	**0.014**	**0.025**	0.018
synthetic	0.011	0.029	0.021
2D surveying (sec. 5.4.4)	-0.019	0.046	0.024

Table 7.9: Contour volume results for the three-dimensional approach in comparison to the two-dimensional results

7.5 Use on handheld devices

So far, the algorithmic details and results of the proposed three-dimensional approach have been discussed. Since the method is designed to work on handheld devices, the considerations for this implementation need to be elaborated. The goal of this dissertation is a proof of concept for photogrammetric surveying. An extensive optimization for the method's computational performance is beyond this work's scope. Nevertheless, some adjustments for the algorithms and parameters, which will be discussed in this section, are necessary for the method to work at all. On the algorithmic side, the choice of methods is crucial for the overall performance. For each of these methods, a number of parameters additionally influence the processing time and the memory consumption. The algorithm adjustments for mobile devices and parameter choices will be discussed in Section 7.5.1 (*Conceptual and parametrical optimizations*). As part of the proof of concept, the processing time on the mobile device will be briefly discussed in Section 7.5.2 (*Processing time and memory consumption*). All experiments are performed on an iPhone 6 smartphone, which features an A8 processor, 1GB RAM, and an 8 megapixel camera.

7.5.1 Conceptual and parametrical optimizations

Parallelization Beginning at the initialization of the 3D approach, the image-based detection algorithm extracts wood logs from each input photo. For three-dimensional surveying, the wood log recognition can make use of the multi-core CPU architecture of most modern smartphones. This enables the wood log detection to be performed in parallel to image acquisition. The processing can therefore start before a user is done with data acquisition. The same concept enables the extraction of image features (e.g., SIFT) as well as feature matching for structure from motion during image acquisition. A high number of CPU cores can potentially speed up the processing significantly.

After feature matching, the structure from motion pipeline is invoked. The beginning of this step implies that no further images will be acquired. During the processing of the 3D reconstruction, the scale reference in the form of the wood pile width can be measured by a user, thus utilizing the time to gather information, while the method is computing the preliminary results. When the structure from motion is complete, the results are scaled according to the scale reference and interactive editing beings (see Figure 7.2).

Wood log recognition optimization A conceptual aspect of the 2D surveying is the optimization of the true and false positive rate. In contrast to the 2D detection results, the 3D approach uses a different detector scale factor, with $s_{\mathrm{detector}} = 1.05$ instead of $s_{\mathrm{detector}} = 1.01$. While the true and false positive rate were shown to be better for a lower scale factor (Figure 5.8), the processing time increases strongly for a lower factor, which can be seen in Figure 7.16(a). For $s_{\mathrm{detector}} = 1.05$, in comparison to $s_{\mathrm{detector}} = 1.01$, the *tpr* is lower by 0.3% with constant *fpr*.

To compensate for this, instead of using a higher detector scale factor, the iterative approach optimizes the detection result, while the processing time is only increased slightly. Figure 7.16(b) shows that the second iteration adds approximately 50% of the additional processing time, while the increase from additional iterations is marginal. Accordingly, Figure 5.10 confirms that the *tpr* is increased after the second iteration by 0.3%. It therefore seems preferable, with regard to the processing time, to use a lower detector scale factor and compensate for it by employing the iterative scheme.

s_{detector}	**1.005**	**1.01**	**1.05**	**1.10**	**1.20**
processing time / s	14.37	3.82	0.31	0.20	0.18

Table 7.10: Processing time for single image wood log detection and the scale factor s_{detector} for Figure 7.16

Figure 7.16: Processing time for single image wood log detection with respect to s_{detector} and iterations. Numeric values shown in Tables 7.10 and 7.11.
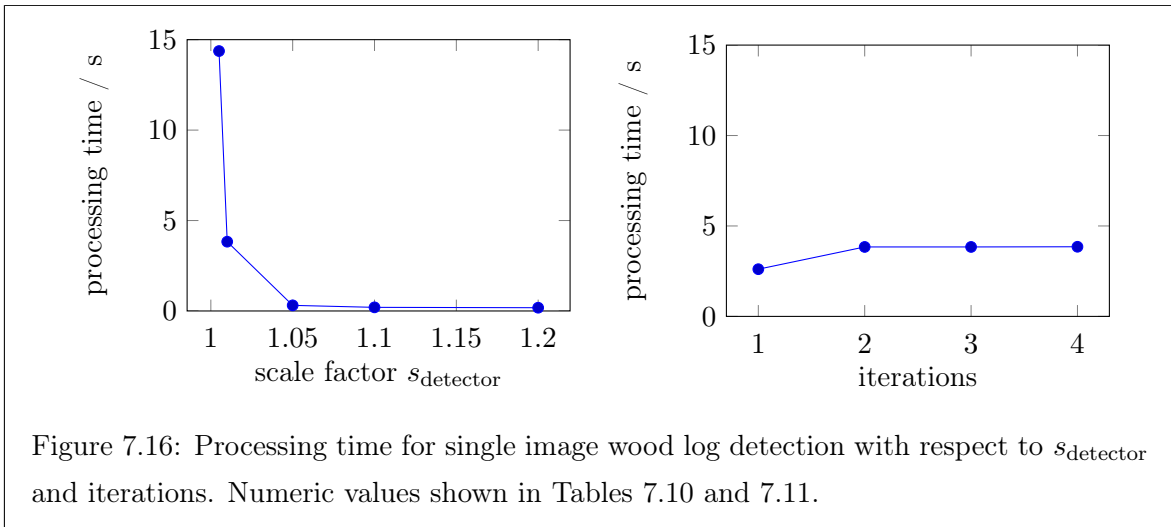
iterations	1	2	3	4
processing time / s	2.61	3.82	3.84	3.83

Table 7.11: Processing time for single image wood log detection with $s_{\text{detector}} = 1.01$ with respect to the number of iterations for Figure 7.16

Feature matching complexity reduction The global structure from motion method by Moulon et al. [MOULON et al., 2013b] can be modified to specify which images should be matched. When capturing an image sequence of a wood pile, it can be formulated as a priori information that an image will only overlap with the k adjacent images of the image sequence. A similar conjecture was made for the adaptive stitching approach (see Section 4.3 (*Optimization for mobile devices*)). As previously stated, the matching problem can be reduced from a quadratic problem $O(n^2)$ to a linear problem $O(n)$ only when the k adjacent images of the image sequence are matched. Especially when capturing large wood piles, this complexity reduction is beneficial for the processing time. Beyond the computational complexity, applying this a priori knowledge also increases the stability of the matching step, since a number of false image matches can be eliminated beforehand.

Relinquishment of multi-view stereopsis In Section 7.4 (*Surveying results*), the use of a multiple view stereopsis algorithm was analyzed. It was shown that the advantages, if any, are negligible. In some cases the results were even better for the sparse data sets, especially in terms of the false positive rate. Therefore, this step is not included in the final version of the proposed approach. All further evaluations are conducted based on the sparse data sets.

Detector combination   The use of different detectors leads to different processing times. The diagram in Figure 7.17 shows that the HOG detector only takes about 0.5 seconds for a single image, while the LBP and Haar-like detectors both need approximately 4 seconds to compute the result. When choosing a detector combination with regard to the processing time, the HOG detector can be safely combined with any of the other two, whereas the combination of Haar-like and LBP features should be avoided if possible, if the goal is a low computation time. In consideration of the detection performance (Figure 5.9), the choice of LBP+HOG does not only show a decent processing time, but also provides the best detection results.

Figure 7.17: Average processing time for a single images with different detectors
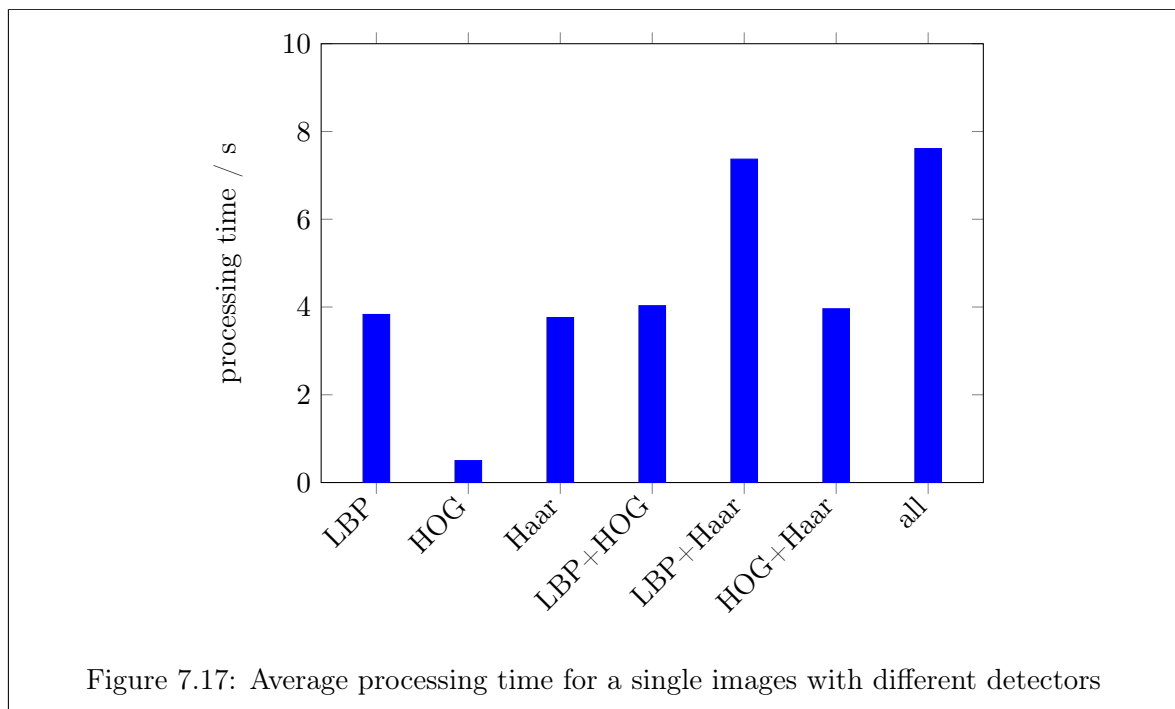
7.5.2   Processing time and memory consumption

For the proof of concept for the proposed photogrammetric method, the processing time and memory consumption must be assessed. In this evaluation, the synthetic data sets are used because only four different image counts exist for these sets. This way, the processing time and memory consumption of the entire implementation can be shown in respect to the number of input images. Current smartphones are equipped with at least one gigabyte of memory, therefore the goal for memory consumption should be to use less than that. Between 12 and 100 images per set are used. Images are scaled to 800x600 pixels, which is still well within the range for a good wood log detection rate. During structure from motion, the matching between images is limited to $j = i - 5 \ldots i + 5$ and $j \neq i$, where j is the index of an image

which should be matched to the current image I_i.

Figure 7.18 shows the average processing time for the entire surveying approach. Note that the time for manual editing is not included, since it can vary depending on the user. As expected, the processing time increases with the number of images. The shown processing times vary from under a minute for 12 images to about 25 minutes for 100 images. In all of the real *HAWKwood* data sets, a maximum of 60 images was used, while small wood piles can be covered by fewer than 10 images. The photogrammetric processing is still significantly faster than manual surveying. Nevertheless, the method shows potential for computational optimization.

Figure 7.18: Processing time for different number of images. The error bars indicate the observed maximum and minimum processing time.
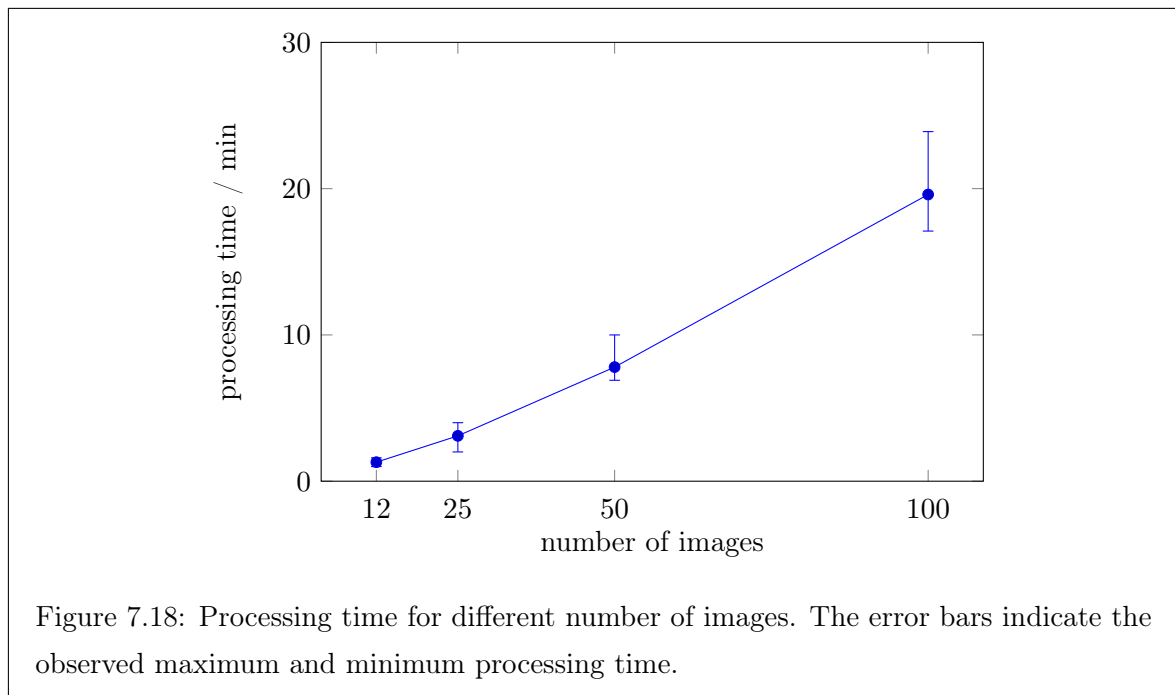
Crucial for the use on handheld devices is the peak memory consumption of the method. In Figure 7.19, the average and peak memory usage are shown. With a maximum of approximately 400MB for a set with 100 images, the memory consumption is well within the physical limits. The slope of the curve indicates that a minimum of 120MB is needed for the entire pipeline, including interactive editing. Beyond that, memory consumption increases linearly, which advocates good scalability of the proposed method. Certainly, the peak memory usage can be reduced by writing files to the disk. In this case, negative effects on the computational performance should be expected, but it seems possible to achieve constant, instead of linear, memory consumption.

Figure 7.19: Average and peak memory consumption for different number of images. The error bars indicate the observed maximum and minimum values.
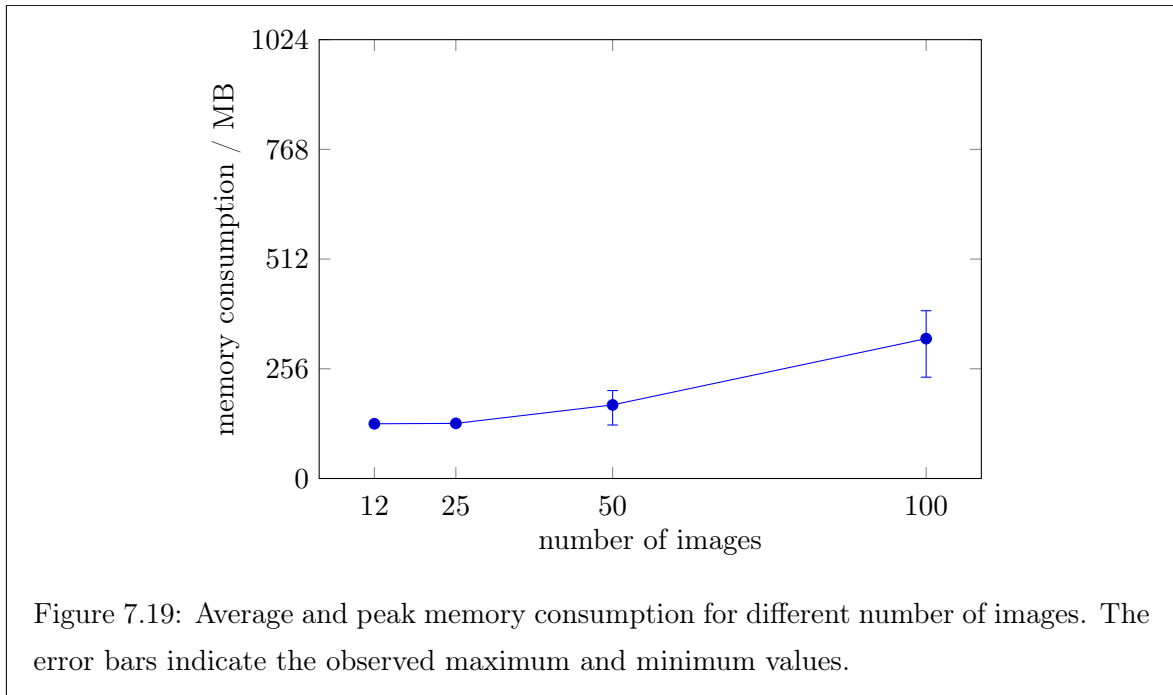
7.6   Summary and future work

Summary   The proposed photogrammetric surveying method works by performing structure from motion in parallel with two-dimensional object detection from a set of images. In the case study for this thesis, the images were taken by a smartphone camera or were rendered synthetically. A novel approach for three-dimensional object detection was proposed, which uses 3D reconstruction combined with detected 2D objects. Quadric filtering is performed, in order to approximate the 3D location and size of an object and to track objects across multiple frames. Based on the 3D objects, a digital model of the wood pile is established, which is used for the determination of the solid wood volume and the contour volume.

Scaling is by default performed through a reference length, in this case the wood pile width. It was shown that this size can be determined very accurately. Alternatively, referencing via high-precision global positioning is possible. The results for this referencing method were shown to be comparable, although the technical limitations (cellular and satellite reception) proved to be limiting factors in terms of availability. The scaled and rotated wood pile is orthographically projected onto an image plane, as to allow for interactive editing of the recognized wood log front faces. An evaluation based on 246 real and synthetic data sets was performed and the results were compared to ground truth and to the results obtained by the 2D surveying method.

Conclusion   In almost all aspects the photogrammetric method outperforms the two-dimensional approach. Firstly, the reconstruction model is not limited to a planar or quasi-planar surface, which is beneficial for the reduction of artifacts which occur, e.g., in the panoramic images of the 2D approach. Secondly, the information about the camera locations can be employed for an improved wood log detection result by filtering false positives and by recognizing wood logs from different viewpoints. A drawback of this method is caused by the limitations of essential matrix estimation in epipolar geometry. The wood pile front surface, which was shown to be geometrically describable as a quasi-planar surface, is very close to a degenerate case, called an H-degenerate configuration or planar degeneracy. In four of the data sets, the reconstruction failed partly because of this constraint. In most cases, capturing sufficient off-plane foreground is enough to avoid the degenerate configuration, provided that the off-plane foreground shows a certain degree of texture for feature extraction.

The surveying results show that the photogrammetric method produces a higher offset than the two-dimensional method for the solid wood volume, but the empiric standard deviation and repeatability standard deviation are lower. The contour volume exhibits a lower offset, lower standard deviation, and lower repeatability standard deviation than the two-dimensional method. This also points to the conclusion that the method provides stable results in terms of repeatability and thus gives very similar results for multiple measurements of the same wood pile. An advantage of the photogrammetric approach is the possibility for alternative referencing. Although the scale estimation with RTK-GPS referencing is approximately one order of magnitude less precise than referencing via size reference (see Tables 7.2 and 7.3), the results are still promising for devices which may in the future be equipped with a high accuracy receiver. The camera position registration with RTK-GPS is advantageous when the outer logs cannot be accurately detected or measured, since only the camera locations and not the scene content influences the scale estimation. The structure from motion result also impacts the recovery of the scale through the accuracy of the camera trajectory. A higher camera position accuracy results in a lower scale drift.

It was shown that the optimizations for mobile devices are necessary for the method to run with a reasonable memory consumption. For a high number of images, the processing time can take several minutes. Although normally no more than 50 to 60 images are needed, the computation time could surely be improved. Nevertheless, time savings in comparison to manual surveying can be achieved, given that manual surveying for large piles can take several hours. In conclusion, the proof of concept was successful and it was shown that photogrammetric wood pile surveying on mobile devices is indeed possible. Some limitations and constraints exist, such as constraints for image acquisition, but the method is feasible in general. With a possible integration of high precision GNSS receivers, the photogrammetric method potentially works without any additional hardware on a single handheld device.

Future work   The proof of concept within this thesis was successful. A couple of interesting topics for future research arise from the specific details of the implementation and the surveying results. The photogrammetric approach outperforms the 2D surveying method in all categories but the offset in the solid wood volume. This is most likely caused by the merging step of the three-dimensional objects. For a lower offset, an alternative merging procedure could be used, e.g., the median instead of the average wood log size could be used for higher robustness in terms of diameter estimation. Another factor, which complicates the diameter estimation in the two-dimensional object recognition step, is the bark of the wood logs. In Section 2.2 (*Research challenges*) it was shown that different types of trees exhibit significantly different bark consistencies. This poses a challenge for the circular approximation of the wood log cut surface and thus influences the solid wood volume result. It could be useful to differentiate between different types of wood, in order to model the appearance of the bark appropriately. Additional a priori knowledge would be needed for the method to benefit from a more precise wood log face segmentation.

The alternative scale estimation via high precision global positioning is currently used as an interchangeable alternative for the reference size. This was chosen explicitly, in order to enable good comparability. In recent years it has been proposed to use position priors in structure from motion [CARCERONI et al., 2006; IRSCHARA et al., 2011; MAURER et al., 2012; POLLEFEYS et al., 2008]. This could be an interesting topic for future research, since the scale would be recovered automatically, while simultaneously improving the calibration result. Especially in the proposed global structure from motion approach [MOULON et al., 2013b], the individual camera translations are estimated in a distinct processing step. Either this step could be skipped, or the translation estimation results can be validated by incorporating high accuracy position priors.

Another method for scale and structure recovery could be to use stereo or depth images. In both cases the scale is known, either from calibration of the stereo camera, or from calibration of the laser / structured light / time of flight sensor. Since, to this date, no commercially available smartphone with such sensors exists, this approach seems like an interesting field for future research. In the case of using data from such 3D sensors, the wood log recognition task could be performed directly in three dimensions and location priors for the wood log cut surfaces could be used for an optimization of the recognition procedure. With the presented monocular approach as a basis, it has been proven that photogrammetric wood pile surveying is indeed possible. As with all such approaches, further optimization of specific tasks is possible, but the proof of concept remains unaffected.

Chapter 8

Conclusion

8.1 Summary

Forestry industry and wood pile surveying   The forestry industry is a very active industrial area. The field of wood surveying in particular shows a comparatively low degree of automation and therefore offers a high potential for the application of computational assistance. The goal of this dissertation has been to prove that photogrammetric wood pile surveying can be performed on a handheld device without an internet connection and with little or no use of additional hardware.

With the increasingly wide availability of smartphones and tablets, such devices are a natural choice for on-site, low-cost surveying. In most cases it is neither practical nor feasible to carry large and heavy equipment to surveying sites. Alternative techniques were discussed, such as surveying with a vehicle mounted stereo camera or with offline and cloud-based computation. Practical considerations made it abundantly clear that such methods are not optimally suited to fulfill the requirements which arise in practice. Handheld devices on the other hand, which are potentially small enough to fit in a jacket pocket, can considerably increase the efficiency of wood pile surveying. The foundation for this is high computational power, combined with the increasing quality of built-in photo cameras and means for user-friendly interaction. The offline surveying constraint can be met with this technique, which is very important, considering the virtually non-existent coverage with mobile internet and given that results must be available immediately at the measuring site.

Geometric wood pile description   A problem rarely addressed in related work is the geometric description of a wood pile. Since a roundwood pile can be characterized as a clustered object, it seems natural to define the wood pile itself as meta object, consisting of individual wood logs, but also having distinctive properties based on the geometric distribution of those logs. Such properties include the solid wood volume, the contour volume, the unambiguously

determinable wood pile width, the front of the pile as a quasi-planar surface, the defined depth of the pile, the wood quality, and the tree species. This information serves as a priori knowledge for the proposed methods, but also defines a number of subproblems and inherent challenges which the proposed methods must overcome.

Adaptive stitching The novel adaptive image stitching technique is able to provide an image of the entire wood pile front surface, to which the object recognition method can be applied. This method, referred to as the two-dimensional surveying technique, makes use of the geometric description of the wood pile as a meta-object. One of the wood pile's most important properties is the quasi-planarity of the front surface. While quasi-planarity poses a difficult challenge for other methods through the induction of a degenerate configuration, planar image stitching is not only able to overcome the degeneracy, but can make use of this property as geometric a priori information.

As with most stitching techniques, applications are very broad. Medical imaging (stitching of microscopic photos) as well as robot navigation can incorporate the adaptive property of the proposed approach. A consumer-level application is especially interesting, considering the executability on mobile devices. With the capability of stitching planar as well as rotational panoramas, a higher degree of flexibility is achieved.

Two-dimensional surveying Over the course of this dissertation it was shown that wood pile surveying can be done in different ways, where each approach offers distinct characteristics and advantages. The two-dimensional, image-based wood log recognition method was proven to not only be useful as a standalone algorithm but can also be incorporated into other approaches. The algorithm's performance in regards to wood log detection and surveying was tested on a large number of images. It was shown that the new method outperforms the published state-of-the-art method. The determination both of the solid wood volume and the contour volume were accomplished with a comparatively low drift and low repeatability standard deviation, which makes this method applicable in practice.

A number of interesting observations were made, e.g., that a higher resolution is not always beneficial for wood log recognition or that a higher scale factor increases the true positive detection rate, but at the same time exponentially increases the processing time. Apart from wood log recognition, a number of problems to which this method can potentially be applied were identified. In all cases where objects are clustered and in which the objects' properties can be described in a consistent way, the proposed method can be used for enhanced recognition and segmentation. Examples include food processing, recognition of fish, and medical imaging, e.g., the detection of cells in microscopic images.

Front surface segmentation Although very rare, the detection of wood logs can be challenging in some special cases. As an alternative it was proposed to perform direct segmentation of the wood pile front surface, based on its quasi-planarity property and the introduction of the principal flow vector, combined with a block matching method. The evaluation showed that the results are comparable to recognition-based segmentation, but the latter shows a significantly better repeatability standard deviation. This means that the quasi-planar surface segmentation can best be incorporated into the other wood pile surveying pipelines to improve results by providing an alternative segmentation approach, instead of serving as a standalone algorithm. Furthermore, it was shown that the proposed method has applications to autonomous vehicle navigation, especially for rough terrain navigation, which would be an interesting topic for future research.

Photogrammetric surveying The core of this dissertation is the photogrammetric surveying approach. Building on the object recognition method and a multiple view reconstruction algorithm, the wood log cut surfaces are located in 3D and used as the basis for three-dimensional surveying. A novel quadric-filtering approach was introduced, which is able to locate an object in a structure from motion result, by back-projecting it onto the scene structure. This enables the tracking across multiple frames and aids in the removal of outliers and false positives. A user can freely interact with the generated model and can thus assure that the detection and surveying results are indeed correct. The photogrammetric surveying outperforms the two-dimensional method in almost all aspects. An important issue for practical applications is the processing time. With devices of the current generation, processing time is in the range of minutes. This is still significantly faster than manual surveying and at the same time exhibits a high potential for computational optimization.

Different scale estimation techniques were discussed, such as a manual reference size via pile width, sensor fusion, and high precision global positioning. The sensor fusion approach proved to be inadequate for this purpose, whereas the global positioning and the reference size showed good results. While the reference size in the form of the pile width was proven to be a very accurate scale reference, the fusion with global positioning data yields the distinct advantage of being unaffected by the correct recognition of the outer wood logs.

On the other hand, the quality of the multiple view reconstruction strongly affects the robustness of this estimation technique. In principle, this method can be applied to the same problems as the 2D recognition approach wherever objects appear in a clustered fashion and where their appearance can be described unambiguously. Examples include the detection and location estimation of pedestrians via multiple surveillance cameras, recognition of cells, or the supervision of food processing procedures.

8.2   Contributions

The main contribution of this dissertation is the proof of concept for photogrammetric wood pile surveying. To achieve this goal, a number of subproblems had to be solved. In the previous chapters, the individual scientific contributions of this thesis were listed, the most important of which are the following:

1. Description of the wood pile as a geometric meta-object
2. Availability of the *HAWKwood* database as a public benchmark
3. Definition of the panorama type criterion leading to adaptive image stitching
4. State-of-the-art image-based wood log recognition through an iterative scheme
5. Recognition-independent front surface segmentation based on the definition of the principal flow vector
6. A novel photogrammetric surveying approach, based on multiple view reconstruction and quadric-filtering
7. Assessment of scale estimation techniques
8. Methodical and parametrical optimization for mobile devices
9. Extensive evaluation based on the *HAWKwood* database

8.3   Future work

All proposed methods are designed for the surveying and recognition of the most important types of wood, which are spruce, pine, beech, and oak. It was discussed that trees of these species potentially exhibit quite different appearances. A thorough distinction between the different types, especially for wood log recognition, is an interesting approach for future research. It seems likely that such an approach could generate even better results. The same holds true for the segmentation of bark for individual roundwood logs. With an a priori known model for bark, segmentation can potentially be made more precisely.

For each method, a certain constraint for image acquisition is imposed, e.g., that panoramic images should overlap a certain degree or that wood log recognition is best performed with little camera rotation with respect to the wood pile front surface. When aiming for real world applications, the importance of the usability should not be underestimated. For practical purposes, a live feedback during image acquisition would be beneficial, especially for inexperienced users. This requires a high processing power during acquisition, which at the moment is already used for preprocessing. Future generations of handheld devices are expected to offer higher computational capabilities, which will potentially solve this problem. The incorporation of a SLAM algorithm seems feasible for this task, since, with a certain degree of accuracy, the camera trajectory and the structure can be used for self localization and orientation ad-

justment with respect to a wood pile or other measuring objects. In recent years, the use of three-dimensional image acquisition techniques has gradually gained importance. Multiple approaches exist for this goal, e.g., time of flight sensors, laser scanners, stereo cameras, or structured light sensors. Assuming that future handheld devices could be equipped with such image acquisition systems, a natural next step is surveying based on pure three-dimensional data. The same considerations concerning feedback during acquisition should be taken into account for the identification of future research tasks.

Based on the promising results obtained in this thesis, the proof of concept for the recognition and surveying of clustered objects (with wood piles as a case study) was shown to be successful. This serves as a foundation for the future research tasks discussed above, which are expected to enhance and improve wood pile surveying even more, through improved usability during image acquisition and by incorporating detailed a priori models into the surveying method. Furthermore, the novel solutions can be applied to a number of other surveying and recognition problems, thus underlining the applicability of this dissertation's contributions to future research in many areas beyond wood pile surveying.

Appendix A

Stitching results

This appendix shows exemplary results of the stitching method in Chapter 4 (*Adaptive image stitching*), divided into the categories "general scenes" and "wood piles". For some general scenes, both rotational and planar panoramas were captured and the results are shown in the same Figure.

A.1   General scenes

Figure A.1: School wall scene. (a) Planar panorama stitched from 9 images, (b) rotational panorama also stitched from 9 images. Image source [HERBON et al., 2014b]

Figure A.2: Book shelf scene. (a) Planar panorama stitched from 8 images, (b) rotational panorama stitched from 9 images. Image source [HERBON et al., 2014b]

Figure A.3: Rotational conference table scene, composed of 15 images. Image source [HERBON et al., 2014b]

A.2   Wood piles

Figure A.4: Example of a stitched wood pile front surface

Figure A.5: Example of a stitched wood pile front surface

Figure A.6: Example of a stitched wood pile front surface

Figure A.7: Example of a stitched wood pile front surface

Figure A.8: Example of a stitched wood pile front surface

Figure A.9: Example of a stitched wood pile front surface

Figure A.10: Example of a stitched wood pile front surface

Figure A.11: Example of a stitched wood pile front surface

Appendix B

Additional 2D results

B.1 Segmentation visualization

Figure B.1: Detection progress. a) Enlarged input image, b) initial detection result, c) ROI (black) for candidate logs, d) candidates (blue) and validated objects (green) after first iteration, e) final detection result after 4 iterations, f) final segmentation result.
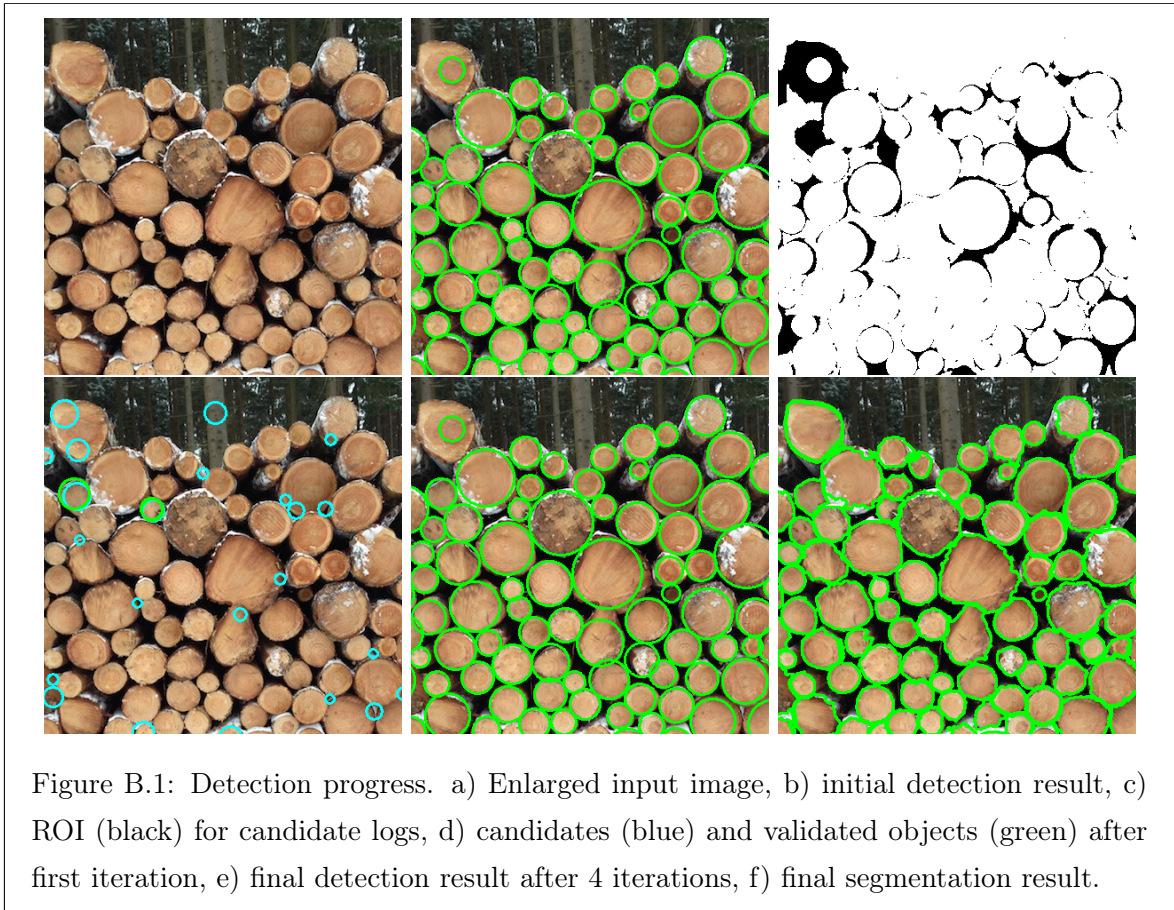
B.2 Image-based detection results

β	0.25	0.40	0.65	1.00	1.58	2.51	3.98
$\bar{x}_\text{tpr}$	0.984	0.992	0.994	0.994	0.992	0.979	0.932
s_tpr	0.030	0.015	0.012	0.014	0.014	0.020	0.040
$\bar{x}_\text{fpr}$	0.010	0.013	0.013	0.011	0.011	0.007	0.002
s_fpr	0.027	0.027	0.021	0.027	0.016	0.006	0.002

Table B.1: True and false positive rate for wood log detection and parameter β for Figure 5.5

w / px	2592	2073	1555	1036	518
$\bar{x}_\text{tpr}$	0.996	0.996	0.996	0.994	0.833
s_tpr	0.005	0.011	0.011	0.012	0.114
$\bar{x}_\text{fpr}$	0.065	0.049	0.031	0.013	0.004
s_fpr	0.077	0.056	0.079	0.027	0.008

Table B.2: True and false positive rate for wood log detection and the image width w for Figure 5.6

θ / °	0	10	20	30	40	50	60	70
$\bar{x}_\text{tpr}$	0.994	0.993	0.993	0.990	0.980	0.974	0.455	0.017
s_tpr	0.012	0.013	0.013	0.017	0.020	0.049	0.102	0.010
$\bar{x}_\text{fpr}$	0.011	0.014	0.012	0.015	0.020	0.057	0.451	0.214
s_fpr	0.027	0.019	0.020	0.017	0.026	0.051	0.120	0.157

Table B.3: True and false positive rate for wood log detection and artificial rotation about the y-axis for Figure 5.7

$\mathbf{s}_{\text{detector}}$	**1.005**	**1.01**	**1.05**	**1.10**	**1.20**
$\bar{x}_{\text{tpr}}$	0.993	0.994	0.991	0.988	0.982
s_{tpr}	0.014	0.012	0.017	0.018	0.026
$\bar{x}_{\text{fpr}}$	0.024	0.011	0.011	0.002	0.002
s_{fpr}	0.033	0.027	0.009	0.005	0.004

Table B.4: True and false positive rate for wood log detection and scale factor s_{detector} for Figure 5.8

detect.	**LBP**	**HOG**	**Haar**	**LBP+HOG**	**LBP+Haar**	**HOG+Haar**	**all**
$\bar{x}_{\text{tpr}}$	0.994	0.935	0.987	0.996	0.996	0.988	0.997
s_{tpr}	0.012	0.031	0.025	0.010	0.012	0.017	0.010
$\bar{x}_{\text{fpr}}$	0.011	0.004	0.010	0.013	0.018	0.011	0.020
s_{fpr}	0.027	0.007	0.016	0.034	0.027	0.016	0033

Table B.5: True and false positive rate for wood log detection and different detectors for Figure 5.9

iterations	**1**	**2**	**3**	**4**
$\bar{x}_{\text{tpr}}$	0.991	0.994	0.994	0.994
s_{tpr}	0.018	0.012	0.013	0.013
$\bar{x}_{\text{fpr}}$	0.012	0.011	0.012	0.012
s_{fpr}	0.027	0.027	0.026	0.026

Table B.6: True and false positive rate for wood log detection after each iteration for Figure 5.10

B.3 Image-based segmentation results

s_{marker}	0.60	0.70	0.80	0.85	0.90	1.00
$\bar{x}_{\mathrm{tpr}}$ (aut.)	0.814	0.892	0.940	0.955	0.964	0.929
s_{tpr} (aut.)	0.053	0.038	0.022	0.018	0.026	0.063
$\bar{x}_{\mathrm{fpr}}$ (aut.)	0.015	0.021	0.041	0.066	0.109	0.221
s_{fpr} (aut.)	0.012	0.016	0.022	0.029	0.042	0.083
$\bar{x}_{\mathrm{tpr}}$ (gt)	0.941	0.946	0.952	0.956	0.963	0.968
s_{tpr} (gt)	0.025	0.022	0.019	0.016	0.011	0.010
$\bar{x}_{\mathrm{fpr}}$ (gt)	0.025	0.026	0.027	0.029	0.031	0.038
s_{fpr} (gt)	0.007	0.007	0.007	0.008	0.009	0.013

Table B.7: True and false positive rate for pixel-based segmentation with automatic detection, ground truth detection, and s_{marker} as a parameter for Figure 5.13

w	2592	2073	1555	1036	518
$\bar{x}_{\mathrm{tpr}}$ (aut.)	0.964	0.955	0.942	0.923	0.759
s_{tpr} (aut.)	0.016	0.017	0.017	0.022	0.246
$\bar{x}_{\mathrm{fpr}}$ (aut.)	0.072	0.072	0.055	0.040	0.218
s_{fpr} (aut.)	0.030	0.025	0.023	0.021	0.378
$\bar{x}_{\mathrm{tpr}}$ (gt)	0.958	0.943	0.931	0.917	0.848
s_{tpr} (gt)	0.015	0.017	0.020	0.023	0.126
$\bar{x}_{\mathrm{fpr}}$ (gt)	0.029	0.028	0.021	0.015	0.061
s_{fpr} (gt)	0.008	0.007	0.005	0.003	0.211

Table B.8: True and false positive rate with automatic detection, ground truth detection, and different image resolutions for Figure 5.14

Acyronyms

Acronym	Description
AUC	Area under the curve
AWU	Annual work unit
BA	Bundle Adjustment
CMDI	Complete midpoint diameter inventory
DOF	Degree of freedom
EKF	Extended Kalman filter
EM	Expectation maximization
FLANN	Fast library for approximate nearest neighbors
FOV	Field of view
GMM	Gaussian mixture model
GPS	Global positioning system
GNSS	Global navigation satellite system
HOG	Histograms of oriented gradients
k-nn	k nearest neighbors
LBP	Local binary patterns
OCR	Optical character recognition
OS	Operating system
PCA	Principal component analysis
PTC	Panorama type criterion
PFV	Principal flow vector
RANSAC	Random sample consensus
RTK	Real time kinematic
ROC	Receiver operating characteristic
ROI	Region of interest
SfM	Structure from motion
SIFT	Scale invariant feature transform
SBIR	Sketch-based image retrieval

SLAM   <u>S</u>imultaneous <u>l</u>ocalization <u>a</u>nd <u>m</u>apping

SGM   <u>S</u>emi <u>g</u>lobal <u>m</u>atching

SVD   <u>S</u>ingular <u>v</u>alue <u>d</u>ecomposition

SVM   <u>S</u>upport <u>v</u>ector <u>m</u>achine

Symbols

Latin symbol	Description
a	Aspect ratio
A_c	Contour area of the front surface of a wood pile in [m^2]
A_s	Solid wood area of the front surface of a wood pile in [m^2]
$\mathbf{a}^W$	Acceleration from an accelerometer
c	Object center
c_x	Optical center (x coordinate)
c_y	Optical center (y coordinate)
C	Camera center
C_obj	An object's contour
c_gt	Ground truth object center
d	Measured midpoint diameter of a wood log in [cm]
D	Euclidean distance
div	Divergence
d_u / d_l	Measured upper / lower diameter of a wood log in [cm]
d_GMM	Distribution in a Gaussian mixture model
f	Focal length
f_mppx	Factor for the conversion from [m] to [px] in [$\frac{\mathrm{m}}{\mathrm{px}}$]
fpr	False positive rate
fpr_logs	False positive rate for wood logs
fpr_pixel	False positive rate for pixels
fpr_segm	False positive rate for segmentation
$f_\mathrm{s,c}$	Stacking coefficient / correctional factor for the calculation of $V_\mathrm{s,c}$ from V_c
$\mathbf{g}^B$	Gravity vector
g_m	Midpoint cross-section area of a wood log in [m^2]
g_u / g_l	Cross-section area at the upper / lower cross-section of a wood log in [m^2]

h_i	Height of the ith section of a wood pile in [m]
$h_{i,\mathrm{px}}$	Height of the ith section of a wood pile in [px]
H	Homography matrix
I_i	ith image
I_{fg}	Foreground image
I_{in}	Input image
l	Defined wood log length in [m]
l_s	Length of a wood log section in [m]
J	Region of interest, where two images overlap
K	Camera intrinsic matrix
N	Number of wood logs in a roundwood pile
N_{fp}	False positive number of wood logs
N_{gt}	Ground truth number of objects
N_{tp}	True positive number of wood logs
n	Number of pixels
n_{GMM}	Number of components in a Gaussian mixture model
n_{occ}	Number of occurrences
$n_{\mathrm{px,wood}}$	Number of wood pixels
n_{s}	Number of sections
o	Vector from the camera center to the optical center
O	Object set containing 2D objects for multiple images
P	Camera pose matrix
p_{co}	Cutoff point
p_{oc}	Location of the optical center
$p_{x,y}$	Pixel with x- and y-coordinates
Q	Point cloud
r	Object radius
R	Rotation matrix
r_{gt}	Ground truth object radius
s	Empiric standard deviation / scale factor for the reconstructed model
s_{rep}	Repeatability standard deviation
s_{detector}	Scale factor for multi-scale detection
S_{blob}	Candidate objects from blobs
$S_{\mathrm{blob,val}}$	Validated objects from blobs
S_{cons}	Consecutive objects
S_{init}	Initial set of detected objects
S_{P}	Planar surface

S_Q	Quasi-planar surface
$S_\mathrm{Q,P}$	Planar approximation of a quasi-planar surface
t	Translation vector
T	Trimap
tpr	True positive rate
tpr_logs	True positive rate for wood logs
tpr_pixel	True positive rate for pixels
tpr_segm	True positive rate for segmentation
u	Location of a camera's optical center in x-direction
v	Location of a camera's optical center in y-direction
V_c	Contour volume of a roundwood pile in $[\mathrm{m}^3]$
$V_\mathrm{c,gt}$	Ground truth contour volume in $[\mathrm{m}^3]$
v_G^{W}	Velocity from a gyro sensor
V_s	Solid wood volume of a roundwood pile in $[\mathrm{m}^3]$
V_stem	Stem volume
$V_\mathrm{s,f}$ / $V_\mathrm{s,b}$	Solid wood volume of a roundwood pile measured from the front / back in $[\mathrm{m}^3]$
w	(Image) width
w_p	Width of a roundwood pile in $[\mathrm{m}]$
w_px	Width of a roundwood pile in $[\mathrm{px}]$
w_r	Reference image width
w_s	Width of a section in $[\mathrm{m}]$
x	Location in x-direction / two- or three-dimensional point
$\bar{x}$	Mean of a sample
y	Location in y-direction

Greek symbol	Description
β	Parameter for noise and brightness
σ	Standard deviation
θ	Angle between two optical center vectors
ξ	Translation threshold used in the translational condition of the panorama type criterion
Υ	Eigen vector

List of Figures

List of Tables

Bibliography

Agarwala, A., M. Agrawala, M. Cohen, D. Salesin, and R. Szeliski (2006). "Photographing long scenes with multi-viewpoint panoramas". In: *ACM SIGGRAPH 2006 Papers.* SIGGRAPH '06. New York, NY, USA: ACM, pages 853–861.

Agrawal, M. and K. Konolige (2007). "Rough terrain visual odometry". In: *Proceedings of the International Conference on Advanced Robotics (ICAR).* Volume 25, pages 28–30.

Ahonen, T., A. Hadid, and M. Pietikainen (2006). "Face description with local binary patterns: Application to face recognition". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 28.12, pages 2037–2041.

Ahonen, T., M. Pietikainen, A. Hadid, and T. Maenpaa (2004). "Face recognition based on the appearance of local regions". In: *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on.* Volume 3. IEEE, pages 153–156.

Alcantarilla, P. F., A. Bartoli, and A. J. Davison (2012). "KAZE features". In: *Computer Vision–ECCV 2012.* Springer, pages 214–227.

Alcantarilla, P. F., J. Nuevo, and A. Bartoli (2013). "Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces". In: *Proceedings of the British Machine Vision Conference.* BMVA Press, pages 13.1–13.11.

Altinay, D. and A. Bradley (2011). "An Evaluation of Multi-resolution Microscope Slide Scanning Algorithms". In: *Digital Image Computing Techniques and Applications (DICTA), 2011 International Conference on*, pages 319–324.

Appleton, B., A. Bradley, and M. Wildermoth (2005). "Towards Optimal Image Stitching for Virtual Microscopy". In: *Digital Image Computing: Techniques and Applications, 2005. DICTA '05. Proceedings 2005*, pages 44–44.

Barkowski, R. (2013). "Automatisierte Fotogrammetrische Rohholz-Vermessungs-Service AF-oRS". In: *AFZ DerWald* 21, page 34.

Battiato, S., G. Gallo, G. Puglisi, and S. Scellato (2007). "SIFT features tracking for video stabilization". In: *Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on.* IEEE, pages 825–830.

Bauer, M. (2011). *Vermessung und Ortung mit Satelliten: Globales Navigationssatellitensystem (GNSS) und andere satellitengestützte Navigationssysteme*. Wichmann. ISBN: 978387-9074822.

Bay, H., A. Ess, T. Tuytelaars, and L. Van Gool (2008). "Speeded-up robust features (SURF)". In: *Computer vision and image understanding* 110.3, pages 346–359.

Bay, H., T. Tuytelaars, and L. Van Gool (2006). "Surf: Speeded up robust features". In: *Computer Vision–ECCV 2006*. Springer, pages 404–417.

Bayrische Staatsforsten (2013). *Standards und allgemeine Anforderungen an Holzerntemaßnahmen bei den Bayerischen Staatsforsten für Unternehmer*. Bayern.

Belongie, S., J. Malik, and J. Puzicha (2001). "Matching shapes". In: *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. Volume 1. IEEE, pages 454–461.

Bentley, J. L. (1975). "Multidimensional binary search trees used for associative searching". In: *Communications of the ACM* 18.9, pages 509–517.

Bergeest, J.-P. and K. Rohr (2011). "Fast Globally Optimal Segmentation of Cells in Fluorescence Microscopy Images". In: *Medical Image Computing and Computer-Assisted Intervention MICCAI 2011*. Edited by G. Fichtinger, A. Martel, and T. Peters. Volume 6891. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pages 645–652. ISBN: 978-3-642-23622-8.

Bergeest, J.-P. and K. Rohr (2012). "Efficient globally optimal segmentation of cells in fluorescence microscopy images using level sets and convex energy functionals". In: *Medical image analysis* 16.7, pages 1436–1444.

Beucher, S., M. Bilodeau, and X. Yu (1990). "Road segmentation by watershed algorithms". In: *PROMETHEUS Workshop, Sophia Antipolis, France*.

Beucher, S. and M. Bilodeau (1994). "Road segmentation and obstacle detection by a fast watershed transformation". In: *Intelligent Vehicles' 94 Symposium, Proceedings of the*. IEEE, pages 296–301.

Beucher, S. and C. Lantuéjoul (1979). "Use of watersheds in contour detection". In: *Workshop on image processing, real-time edge and motion detection*.

Bleyer, M. and M. Gelautz (2007). "Graph-cut-based stereo matching using image segmentation with symmetrical treatment of occlusions". In: *Signal Processing: Image Communication* 22.2, pages 127–143.

Bombosch, F., L. Brekerbohm, and D. Sohns (2011a). *Acquiring round timber piles*. WO Patent App. PCT/EP2010/058,851.

Bombosch, F., L. Brekerbohm, and D. Sohns (2011b). *Erfassung von Rundholzpoltern Detection of log yards*. DE Patent App. DE200,910,027,449.

Borgstadt, J. and N. J. Ferrier (2001). "Visual servoing: path interpolation by homography decomposition". In: *IEEE International Conference on Robotics and Automation, 2001. Proceedings 2001 ICRA.* Volume 1, pages 723–730.

Bort, U., G. Mahler, and C. Pfeil (1989). "Vermessung von Fixlängen". In: *Versuchsberichte 1999.* Freiburg: Forstliche Versuchs- und Forschungsanstalt Baden-Württemberg.

Bourque, E., G. Dudek, and P. Ciaravola (1998). "Robotic sightseeing-a method for automatically creating virtual environments". In: *Robotics and Automation, 1998. Proceedings. 1998 IEEE International Conference on.* Volume 4. IEEE, pages 3186–3191.

Boykov, Y. and M.-P. Jolly (2001). "Interactive graph cuts for optimal boundary & region segmentation of objects in ND images". In: *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on.* Volume 1. IEEE, pages 105–112.

Boykov, Y. and V. Kolmogorov (2001). "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision". In: *Energy minimization methods in computer vision and pattern recognition.* Springer, pages 359–374.

Boykov, Y. and V. Kolmogorov (2003). "Computing geodesics and minimal surfaces via graph cuts". In: *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on.* IEEE, pages 26–33.

Boykov, Y. and V. Kolmogorov (2004). "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 26.9, pages 1124–1137.

Boykov, Y., O. Veksler, and R. Zabih (2001). "Fast approximate energy minimization via graph cuts". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 23.11, pages 1222–1239.

Bradski, G. (2000). "The OpenCV Library". In: *Dr. Dobb's Journal of Software Tools.*

Brosnan, T. and D.-W. Sun (2004). "Improving quality inspection of food products by computer vision—-a review". In: *Journal of Food Engineering* 61.1, pages 3–16.

Brown, M. and D. G. Lowe (2007). "Automatic panoramic image stitching using invariant features". In: *International journal of computer vision* 74.1, pages 59–73.

Brox, T. and J. Weickert (2004). "Level set based image segmentation with multiple regions". In: *Pattern recognition.* Springer, pages 415–423.

Bulanon, D., T. Kataoka, Y. Ota, and T. Hiroma (2002). "Automation and Emerging Technologies: A Segmentation Algorithm for the Automatic Recognition of Fuji Apples at Harvest". In: *Biosystems Engineering* 83.4, pages 405–412. ISSN: 1537-5110.

Bundesministerium für Ernährung, Landwirtschaft und Forsten (1969). *Verordnung über gesetzliche Handelsklassen für Rohholz vom 31. Juli 1969.* Bonn.

Bundesministerium für Ernährung und Landwirtschaft (BMEL) (2014). *Der Wald in Deutschland - Ausgewählte Ergebnisse der dritten Bundeswaldinventur*. Bundesministerium für Ernährung und Landwirtschaft (BMEL).

Caetano, T. S. and D. A. C. Barone (2001). "A probabilistic model for the human skin color". In: *Image Analysis and Processing, 2001. Proceedings. 11th International Conference on*. IEEE, pages 279–283.

Capel, D. and A. Zisserman (1998). "Automated mosaicing with super-resolution zoom". In: *1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1998. Proceedings*. Pages 885–891.

Carceroni, R., A. Kumar, and K. Daniilidis (2006). "Structure from motion with known camera positions". In: *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*. Volume 1. IEEE, pages 477–484.

Carson, C., S. Belongie, H. Greenspan, and J. Malik (2002). "Blobworld: Image segmentation using expectation-maximization and its application to image querying". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24.8, pages 1026–1038.

Carvalho, F. D., B. A. B. Correia, R. Davies, F. C. Rodrigues, and J. C. A. Freitas (1993). "Image processing system for the measurement of timber truck loads". In: *Proc. SPIE International Conference on Manufacturing Automation* 1713, pages 86–92.

Caselles, V., R. Kimmel, and G. Sapiro (1997). "Geodesic active contours". In: *International journal of computer vision* 22.1, pages 61–79.

Cauchy, A. (1846). "Considérations nouvelles sur les intégrales définies qui s' étendent à tous les points d'une courbe fermée, et sur celles qui sont prises entre des limites imaginaires". In: *CR* 23, pages 689–704.

Chan, T. F. and L. A. Vese (2001). "Active contours without edges". In: *Image processing, IEEE transactions on* 10.2, pages 266–277.

Chen, C., W. Wang, J. A. Ozolek, and G. K. Rohde (2013). "A flexible and robust approach for segmenting cell nuclei from 2D microscopy images using supervised learning and template matching". In: *Cytometry Part A* 83.5, pages 495–507.

Chen, C.-Y. and R. Klette (1999). "Image Stitching - Comparisons and New Techniques". In: *Computer Analysis of Images and Patterns*. Edited by F. Solina and A. Leonardis. Volume 1689. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pages 615–622.

Chilian, A. and H. Hirschmüller (2009). "Stereo camera based navigation of mobile robots on rough terrain". In: *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*. IEEE, pages 4571–4576.

Chuang, M.-C., J.-N. Hwang, and K. Williams (2014). "Supervised and Unsupervised Feature Extraction Methods for Underwater Fish Species Recognition". In: *Computer Vision for Analysis of Underwater Imagery (CVAUI), 2014 ICPR Workshop on.* IEEE, pages 33–40.

Chum, O., T. Werner, and J. Matas (2005). "Two-view geometry estimation unaffected by a dominant plane". In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on.* Volume 1. IEEE, pages 772–779.

Civera, J., O. G. Grasa, A. J. Davison, and J. Montiel (2009). "1-point RANSAC for EKF-based structure from motion". In: *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/-RSJ International Conference on.* IEEE, pages 3498–3504.

Coelho, L. P., A. Shariff, and R. F. Murphy (2009). "Nuclear segmentation in microscope cell images: a hand-segmented dataset and comparison of algorithms". In: *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on.* IEEE, pages 518–521.

Cortes, C. and M. Mohri (2004). "AUC optimization vs. error rate minimization". In: *Advances in neural information processing systems* 16.16, pages 313–320.

Cristianini, N. and J. Shawe-Taylor (2000). *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods.* Cambridge University Press. ISBN: 9780521780193.

Da, F. and H. Zhang (2010). "Sub-pixel edge detection based on an improved moment". In: *Image and Vision Computing* 28.12, pages 1645–1658.

Dahl, A. B., M. Guo, and K. H. Madsen (2006). *Scale-Space and Watershed Segmentation for Detection of Wood Logs.* Technical report. Informatics and Mathematical Modelling, Technical University of Denmark.

Dalal, N. and B. Triggs (2005). "Histograms of oriented gradients for human detection". In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on.* Volume 1. IEEE, pages 886–893.

Davis, R. (1990). *Log measuring method and apparatus.* US Patent 4,913,551.

Davison, A. J. (2003). "Real-time simultaneous localisation and mapping with a single camera". In: *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on.* IEEE, pages 1403–1410.

de Leóna, G. C. and L. P. Uranga-Valenciaa (2013). "Theoretical evaluation of Huber and Smalian methods applied to tree stem classical geometries". In: *Bosque* 34.3, pages 311–317.

Decker, P., D. Paulus, and T. Feldmann (2008). "Dealing with degeneracy in essential matrix estimation". In: *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on.* IEEE, pages 1964–1967.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). "Maximum likelihood from incomplete data via the EM algorithm". In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38.

Deutscher Forstwirtschaftsrat e.V. und Deutscher Holzwirtschaftsrat e.V. (2014). *Rahmenvereinbarung für den Rohholzhandel in Deutschland (RVR)*. Freiburg / Berlin.

Dietz, P. (1985). "Holzernte in Mitteleuropa". In: *Handbuch Holzwirtschaft*. Deutscher Betriebswirte Verlag GmbH.

Dima, A. A., J. T. Elliott, J. J. Filliben, M. Halter, A. Peskin, J. Bernal, M. Kociolek, M. C. Brady, H. C. Tang, and A. L. Plant (2011). "Comparison of segmentation algorithms for fluorescence microscopy images of cells". In: *Cytometry Part A* 79.7, pages 545–559.

Dralle A/S (2014). *sScale^{TM}- Measure, track and trade*. Venlighedsvej 4, DK-2970 Hoersholm, Denmark.

Dralle, K. (2005). *A system for grading of industrial wood*. WO Patent App. PCT/DK2005/-000,093.

Dralle, K. and J. Mads (2004). *A method and a system for automatic measurement and tracking of logs, industrial wood and boards*. WO Patent App. PCT/DK2003/000,561.

Dralle, K. and M. Tard-Johansen (2010). *Method and a system for automatic measurement and tracking of logs, industrial wood and boards*. US Patent 7,660,433.

Eitz, M., K. Hildebrand, T. Boubekeur, and M. Alexa (2011). "Sketch-based image retrieval: Benchmark and bag-of-features descriptors". In: *Visualization and Computer Graphics, IEEE Transactions on* 17.11, pages 1624–1636.

Elmenreich, W. (2002). "Sensor fusion in time-triggered systems". PhD thesis. Technische Universität Wien.

European Commission and Eurostat (2012). *Agriculture, fishery and forestry statistics : main results - 2010-11*. Luxembourg: Publications office of the European Union. ISBN: 978-92-79-25431-4.

European Commission and Eurostat (2013). *Agriculture, forestry and fishery statistics*. Luxembourg: Publications office of the European Union. ISBN: 978-92-79-33005-6.

Farenzena, M., A. Fusiello, and R. Gherardi (2009). "Structure-and-motion pipeline on a hierarchical cluster tree". In: *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*. IEEE, pages 1489–1496.

Faugeras, O. D. and F. Lustman (1988). "Motion and structure from motion in a piecewise planar environment". In: *International Journal of Pattern Recognition and Artificial Intelligence* 2.03, pages 485–508.

Fawcett, T. (2004). "ROC graphs: Notes and practical considerations for researchers". In: *Machine learning* 31, pages 1–38.

Feng, X., M. Pietikainen, and A. Hadid (2005). "Facial expression recognition with local binary patterns and linear programming". In: *Pattern Recognition And Image Analysis C/C of Raspoznavaniye Obrazov I Analiz Izobrazhenii* 15.2, page 546.

Fink, F. (2004). "Foto-optische Erfassung der Dimension von Nadelrundholzabschnitten unter Einsatz digitaler, bildverarbeitender Methoden". PhD thesis. Universitätsbibliothek Freiburg.

Fischler, M. A. and R. C. Bolles (1981). "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography". In: *Commun. ACM* 24.6, pages 381–395.

Foese, B., E. Gutzeit, M. Müller, and J. Voskamp (2012). "AFoRS-Automatisierter Fotogrammetrischer Rohholz-Vermessungs-Service." In: *GIL Jahrestagung*, pages 91–94.

Forbrig, A., K. Büchler, R. Hofmann, J. Morat, L. Nick, D. Ruppert, U. Seeling, and G. Weise (2012). "Visiosens-Polterzähler". In: *FTI Mitgliederzeitschrift des KWF* 7+8, page 20.

Forgy, E. W. (1965). "Cluster analysis of multivariate data: efficiency versus interpretability of classifications". In: *Biometrics* 21, pages 768–769.

Fornland, P. and C. Schnorr (1997). "A robust and convergent iterative approach for determining the dominant plane from two views without correspondence and calibration". In: *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on.* IEEE, pages 508–513.

Freistaat Sachsen Staatsbetrieb Sachsenforst und GIS-Dienst GmbH (2012). *Methode zur Bestimmung der Stückzahl, des Volumens und der Güteanteile von in Form von Poltern gelagerten Holzstämmen.* DE Patent 202,012,009,846.

Frommhold, H. (2013). *Holzsortierung und -vermessung.* Technical report 4. Hochschule für nachhaltige Entwicklung Eberswalde.

Funck, J. W., J. B. Forrer, D. A. Butler, C. C. Brunner, and A. G. Maristany (1993). "Measuring surface roughness on wood: a comparison of laser-scatter and stylus-tracing approaches". In: *Proc. SPIE* 1821, pages 173–184.

Furukawa, Y. and J. Ponce (2007). "Accurate, Dense, and Robust Multi-View Stereopsis". In: *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on.* IEEE, pages 1–8.

Furukawa, Y. and J. Ponce (2010). "Accurate, dense, and robust multiview stereopsis". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 32.8, pages 1362–1376.

Gehrig, S. K. and U. Franke (2007). "Improving stereo sub-pixel accuracy for long range stereo". In: *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on.* IEEE, pages 1–7.

Gherardi, R., M. Farenzena, and A. Fusiello (2010). "Improving the efficiency of hierarchical structure-and-motion". In: *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on.* IEEE, pages 1594–1600.

Gherardi, R. and A. Fusiello (2010). "Practical autocalibration". In: *Computer Vision–ECCV 2010*. Springer, pages 790–801.

Gläser, H. (1955). "Die photographische Methode zur Festgehaltsermittlung von Schichtholz". In: *Holz-Zentralblatt* 47, page 545.

Grau, V., A. Mewes, M. Alcaniz, R. Kikinis, and S. K. Warfield (2004). "Improved watershed transform for medical image segmentation using prior information". In: *Medical Imaging, IEEE Transactions on* 23.4, pages 447–458.

Greenspan, H., J. Goldberger, and I. Eshet (2001). "Mixture model for face-color modeling and segmentation". In: *Pattern Recognition Letters* 22.14, pages 1525–1536.

Greig, D., B. Porteous, and A. H. Seheult (1989). "Exact maximum a posteriori estimation for binary images". In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 271–279.

Guglhör, W. (1994). "Vermessung - ein Dauerbrenner? Kontrollmaß für den Forstunternehmer". In: *Forst & Technik* 6, pages 10–13.

Gutzeit, E. and U. von Lukas (2013). "Methodik zur Segmentierung organischer Objekte einer gleichartigen Gruppe in Farbbildern". In: *19. Workshop Farbbildverarbeitung - Berlin*, pages 135–145.

Gutzeit, E., S. Ohl, A. Kuijper, J. Voskamp, and B. Urban (2010). "Setting Graph Cut Weights for Automatic Foreground Extraction in Wood Log Images." In: *VISAPP (2)*, pages 60–67.

Gutzeit, E., S. Ohl, J. Voskamp, A. Kuijper, and B. Urban (2011). "Automatic wood log segmentation using graph cuts". In: *Computer Vision, Imaging and Computer Graphics. Theory and Applications*. Springer, pages 96–109.

Gutzeit, E. and J. Voskamp (2012). "Automatic segmentation of wood logs by combining detection and segmentation". In: *Advances in Visual Computing*. Springer, pages 252–261.

Haar, A. (1910). "Zur Theorie der orthogonalen Funktionensysteme". In: *Mathematische Annalen* 69.3, pages 331–371.

Hadid, A. and M. Pietikäinen (2009). "Combining appearance and motion for face and gender recognition from videos". In: *Pattern Recognition* 42.11, pages 2818–2827.

Hagwood, C., J. Bernal, M. Halter, and J. Elliott (2012). "Evaluation of segmentation algorithms on cell populations using CDF curves". In: *Medical Imaging, IEEE Transactions on* 31.2, pages 380–390.

Hahn, H. K. and H.-O. Peitgen (2000). "The skull stripping problem in MRI solved by a single 3D watershed transform". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2000*. Springer, pages 134–143.

Hannuksela, J., P. Sangi, J. Heikkila, X. Liu, and D. Doermann (2007). "Document image mosaicing with mobile phones". In: *Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on*. IEEE, pages 575–582.

Haralick, R. and L. Shapiro (1992). *Computer and robot visión*. Computer and Robot Vision 1. Addison-Wesley Pub. Co. ISBN: 9780201108774.

Harris, C. and M. Stephens (1988). "A combined corner and edge detector." In: *Alvey vision conference*. Volume 15. Manchester, UK, page 50.

Hartley, R. and A. Zisserman (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press. ISBN: 9781139449144.

Harville, M., G. Gordon, and J. Woodfill (2001). "Foreground segmentation using adaptive mixture models in color and depth". In: *Detection and Recognition of Events in Video, 2001. Proceedings. IEEE Workshop on*. IEEE, pages 3–11.

Häselich, M., M. Arends, N. Wojke, F. Neuhaus, and D. Paulus (2013). "Probabilistic terrain classification in unstructured environments". In: *Robotics and Autonomous Systems* 61.10, pages 1051–1059.

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

Held, C., J. Wenzel, P. Ralf, R. Lang, and T. Wittenberg (2011). "Segmentierung von Makrophagen in Fluoreszenzbildern mittels Fast Marching Level Set Verfahren". In: *Bildverarbeitung für die Medizin 2011*. Springer, pages 129–133.

Herbon, C. (2014a). *Dominant Plane Database*. `http://193.175.106.27/~cherbon/dominantplanedb.pdf`. Tech Report.

Herbon, C. (2014b). "The HAWKwood Database". In: *CoRR* abs/1410.4393. URL: `http://arxiv.org/abs/1410.4393`.

Herbon, C., B. Otte, K. Tönnies, and B. Stock (2014a). "Detection of Clustered Objects in Sparse Point Clouds Through 2D Classification and Quadric Filtering". In: *Pattern Recognition*. Springer, pages 535–546.

Herbon, C., B. Otte, K. Tönnies, and B. Stock (2015a). "Mobile 3D Wood Pile Surveying". In: *Proceedings of the 14th IAPR International Conference on Machine Vision Applications, Tokyo, Japan*. IEEE.

Herbon, C., G. Schumann, K. Tönnies, and B. Stock (2015b). "Detection and Segmentation of Quasi-Planar Surfaces Through Expectation Maximization under a Planar Homography Constraint". In: *Proceedings of the 12nd Canadian Conference on Computer and Robot Vision*. IEEE.

Herbon, C., K. Tönnies, and B. Stock (2014b). "Adaptive planar and rotational image stitching for mobile devices". In: *Proceedings of the 5th ACM Multimedia Systems Conference*. ACM, pages 213–223.

Herbon, C., K. Tönnies, and B. Stock (2014c). "Detection and segmentation of clustered objects by using iterative classification, segmentation, and Gaussian mixture models and application to wood log detection". In: *Pattern Recognition.* Springer, pages 354–364.

Hirschmüller, H. (2005). "Accurate and efficient stereo processing by semi-global matching and mutual information". In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on.* Volume 2. IEEE, pages 807–814.

"HOMA Holzmeßanweisung : Bestimmungen über die Ausformung, Messung und Sortenbildung des Holzes in den deutschen Forsten vom 1. April 1936" (1936). In: *Dt. Holz-Anzeiger.*

Hornung, A. and L. Kobbelt (2006). "Hierarchical volumetric multi-view stereo reconstruction of manifold surfaces based on dual graph embedding". In: *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on.* Volume 1. IEEE, pages 503–510.

Hrabar, S. and G. Sukhatme (2003). "Omnidirectional vision for an autonomous helicopter". In: *Robotics and Automation, 2003. Proceedings. ICRA'03. IEEE International Conference on.* Volume 1. IEEE, pages 558–563.

Hu, R., R. Shi, I.-f. Shen, and W. Chen (2007). "Video stabilization using scale-invariant features". In: *Information Visualization, 2007. IV'07. 11th International Conference.* IEEE, pages 871–877.

Hu, R., M. Barnard, and J. Collomosse (2010). "Gradient field descriptor for sketch based retrieval and localization". In: *Image Processing (ICIP), 2010 17th IEEE International Conference on.* IEEE, pages 1025–1028.

Hu, R. and J. Collomosse (2013). "A performance evaluation of gradient field HOG descriptor for sketch based image retrieval". In: *Computer Vision and Image Understanding* 117.7, pages 790–806.

Huang, P. X., B. J. Boom, and R. B. Fisher (2012). "Hierarchical classification for live fish recognition". In: *BMVC student workshop paper.*

Huang, P. X., B. J. Boom, and R. B. Fisher (2013). "Underwater Live Fish Recognition Using a Balance-Guaranteed Optimized Tree". English. In: *Computer Vision ACCV 2012.* Edited by K. M. Lee, Y. Matsushita, J. M. Rehg, and Z. Hu. Volume 7724. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pages 422–433. ISBN: 978-3-642-37330-5.

Huang, X. and P. X. Huang (2014). "Balance-Guaranteed Optimized Tree with Reject option for live fish recognition". PhD thesis. University of Edinburgh.

Huang, Z.-K. and K.-W. Chau (2008). "A new image thresholding method based on Gaussian mixture model". In: *Applied Mathematics and Computation* 205.2, pages 899–907.

Huber, P. J. (1981). *Robust statistics.*

Husch, B., T. Beers, and J. Kershaw (2002). *Forest Mensuration.* Ecology (John Wiley and Sons). John Wiley & Sons. ISBN: 9780471018506.

Hwang, J., H. Yun, Y. Suh, J. Cho, and D. Lee (2012). "Development of an RTK-GPS Positioning Application with an Improved Position Error Model for Smartphones". In: *Sensors* 12.10, pages 12988–13001.

Ilumets, A. (2013). *Method and a device for accurate measuring of volume of timber*. EP Patent App. EP20,110,180,283.

Irani, M. and P. Anandan (2000). "About direct methods". In: *Vision Algorithms: Theory and Practice*. Springer, pages 267–277.

Irani, M., P. Anandan, and D. Weinshall (1998). "From reference frames to reference planes: Multi-view parallax geometry and applications". In: *Computer Vision—ECCV'98*. Springer, pages 829–845.

Irschara, A., C. Hoppe, H. Bischof, and S. Kluckner (2011). "Efficient structure from motion with weak position and orientation priors". In: *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*. IEEE, pages 21–28.

Jaeger, D., M. Opferkuch, S. Schönherr, and T. Wagner (2014). *Integration of high-quality harvester data and new log scaling technology for efficient control of wood flow in German wood supply chains*. Precision Forestry Symposium.

Jähne, B. (2005). *Digitale Bildverarbeitung*. Springer London, Limited. ISBN: 9783540273844.

Jian, B. and B. C. Vemuri (2011). "Robust point set registration using gaussian mixture models". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 33.8, pages 1633–1645.

Jiang, K., Q.-M. Liao, and S.-Y. Dai (2003). "A novel white blood cell segmentation scheme using scale-space filtering and watershed clustering". In: *Machine Learning and Cybernetics, 2003 International Conference on*. Volume 5. IEEE, pages 2820–2825.

Jimenez, A., R. Ceres, J. Pons, et al. (2000). "A survey of computer vision methods for locating fruit on trees". In: *Transactions of the ASAE-American Society of Agricultural Engineers* 43.6, pages 1911–1920.

Jørgensen, E. R. and L. Kristiansen (2008). "Digitale Fotovermessung von Industrieholz und Abschnitten". In: *AfZ-DerWald*, pages 284–285. URL: http://www.heidegesellschaft.de/media/Artikel_AfZ.pdf.

Jørgensen Erik Røj und Kristiansen, L. (2008). "Digitale Fotovermessung von Industrieholz und Abschnitten". In: *AFZ-DerWald*.

Jung, S.-H. and C. J. Taylor (2001). "Camera trajectory estimation using inertial sensor measurements and structure from motion results". In: *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. Volume 2. IEEE, pages II–732.

Kadir, T. and M. Brady (2001). "Saliency, scale and image description". In: *International Journal of Computer Vision* 45.2, pages 83–105.

Kahmen, H. (2005). *Angewandte Geodäsie: Vermessungskunde*. De-Gruyter-Lehrbuch. Walter de Gruyter. ISBN: 9783110184648.

Kakumanu, P., S. Makrogiannis, and N. Bourbakis (2007). "A survey of skin-color modeling and detection methods". In: *Pattern recognition* 40.3, pages 1106–1122.

Kass, M., A. Witkin, and D. Terzopoulos (1988). "Snakes: Active contour models". In: *International journal of computer vision* 1.4, pages 321–331.

Kauffmann, C. and N. Piche (2008). "Cellular automaton for ultra-fast watershed transform on GPU." In: *ICPR*, pages 1–4.

Kauppinen, M. (1993). *Stacking method and equipment for measuring the timber volume and other dimensions of a stack of logs*. US Patent 5,267,018.

Ke, Y. and R. Sukthankar (2004). "PCA-SIFT: A more distinctive representation for local image descriptors". In: *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. Volume 2. IEEE, pages II–506.

Kecman, V. (2001). *Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models*. A Bradford book. Bradford Book. ISBN: 9780262112550.

Kim, J., Y. Lee, S. Cha, C. Choi, and S. Lee (2013). "Development of a Network RTK Positioning and Gravity-Surveying Application with Gravity Correction Using a Smartphone". In: *Sensors* 13.7, pages 8879–8894.

Knyaz, V. and A. Maksimov (2014). "Photogrammetric Technique for Timer Stack Volume Control". In: *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 1, pages 157–162.

Knyaz, V. A. and Y. V. Vizilter (2001). "Method for 3D noncontact measurements of cut trees package area". In: *Intelligent Systems and Smart Manufacturing*. International Society for Optics and Photonics, pages 276–285.

Köhler, B. (2005). *Konzepte Der Statistischen Signalverarbeitung*. Springer London, Limited. ISBN: 9783540275091.

Kolmogorov, V. and R. Zabih (2001). "Computing visual correspondence with occlusions using graph cuts". In: *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. Volume 2. IEEE, pages 508–515.

Kolmogorov, V. and R. Zabih (2002). "Multi-camera scene reconstruction via graph cuts". In: *Computer Vision—ECCV 2002*. Springer, pages 82–96.

Kong, H., J.-Y. Audibert, and J. Ponce (2010). "General road detection from a single image". In: *Image Processing, IEEE Transactions on* 19.8, pages 2211–2220.

Konolige, K., M. Agrawal, and J. Sola (2011). "Large-scale visual odometry for rough terrain". In: *Robotics Research*. Springer, pages 201–212.

Kramer, H. and A. Akça (2008). *Leitfaden zur Waldmesslehre*. Sauerländer. ISBN: 97837939-08807.

Kuhnl, T., F. Kummert, and J. Fritsch (2011). "Monocular road segmentation using slow feature analysis". In: *Intelligent Vehicles Symposium (IV), 2011 IEEE*. IEEE, pages 800–806.

Langford, M., A. Fox, and R. Smith (2012). *Langford's Basic Photography: The Guide for Serious Photographers*. Taylor & Francis. ISBN: 9781136096693.

Lee, K.-Y., Y.-Y. Chuang, B.-Y. Chen, and M. Ouhyoung (2009). "Video stabilization using robust feature trajectories". In: *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, pages 1397–1404.

Leemans, V., H. Magein, and M.-F. Destain (1998). "Defects segmentation on 'Golden Delicious' apples by using colour machine vision". In: *Computers and Electronics in Agriculture* 20.2, pages 117–130.

Leemans, V., H. Magein, and M.-F. Destain (1999). "Defect segmentation on 'Jonagold'apples using colour vision and a Bayesian classification method". In: *Computers and Electronics in Agriculture* 23.1, pages 43–53.

Levenberg, K. (1944). "A method for the solution of certain problems in least squares". In: *Quarterly of applied mathematics* 2, pages 164–168.

Li, C., R. Huang, Z. Ding, J. Gatenby, D. N. Metaxas, and J. C. Gore (2011). "A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI". In: *Image Processing, IEEE Transactions on* 20.7, pages 2007–2016.

Li, Y. (2012). "Fish Component Recognition". Master's thesis. University of Edinburgh.

Liang, J., D. DeMenthon, and D. Doermann (2006). "Camera-Based Document Image Mosaicing". In: *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*. Volume 2, pages 476–479.

Lin, G., U. Adiga, K. Olson, J. F. Guzowski, C. A. Barnes, and B. Roysam (2003). "A hybrid 3D watershed algorithm incorporating gradient cues and object models for automatic segmentation of nuclei in confocal image stacks". In: *Cytometry Part A* 56.1, pages 23–36.

Litvin, A., J. Konrad, and W. C. Karl (2003). "Probabilistic video stabilization using Kalman filtering and mosaicing". In: *Electronic Imaging 2003*. International Society for Optics and Photonics, pages 663–674.

Lloyd, S. (1982). "Least squares quantization in PCM". In: *Information Theory, IEEE Transactions on* 28.2, pages 129–137.

Lohse, I. and O. Heuer (2014). *OptischeVermessung von Rundholzpoltern im Gelände*. Technical report. PTB Physikalisch-Technische Bundesanstalt Braunschweid und Berlin.

Lowe, D. G. (1999). "Object recognition from local scale-invariant features". In: *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*. Volume 2. Ieee, pages 1150–1157.

Lowe, D. G. (2001). "Local feature view clustering for 3D object recognition". In: *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. Volume 1. IEEE, pages I–682.

Lowe, D. G. (2004). "Distinctive image features from scale-invariant keypoints". In: *International journal of computer vision* 60.2, pages 91–110.

Lu, C., M. Mahmood, N. Jha, and M. Mandal (2013). "Automated segmentation of the melanocytes in skin histopathological images". In: *Biomedical and Health Informatics, IEEE Journal of* 17.2, pages 284–296.

MacQueen, J. et al. (1967). "Some methods for classification and analysis of multivariate observations". In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. California, USA, pages 281–297.

Maler, G. (1997). "Zur Bedeutung des Maßes für Waldbesitz und Holzwirtschaft". In: *AFZ/Der Wald* 15, pages 809–812.

Malis, E. and M. Vargas (2007). *Deeper understanding of the homography decomposition for vision-based control*. Technical report. INRIA.

Marquardt, D. W. (1963). "An algorithm for least-squares estimation of nonlinear parameters". In: *Journal of the Society for Industrial & Applied Mathematics* 11.2, pages 431–441.

Matas, J., O. Chum, M. Urban, and T. Pajdla (2004). "Robust wide-baseline stereo from maximally stable extremal regions". In: *Image and vision computing* 22.10, pages 761–767.

Mattheck, C. et al. (1991). *Trees: the mechanical design*. Springer-Verlag.

Maurer, M., M. Rumpler, A. Wendel, C. Hoppe, A. Irschara, and H. Bischof (2012). "Georeferenced 3d reconstruction: Fusing public geographic data and aerial imagery". In: *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, pages 3557–3558.

Mebius, J. E. (2007). "Derivation of the Euler-Rodrigues formula for three-dimensional rotations from the general formula for four-dimensional rotations". In: *arXiv preprint math-/0701759*.

Meyer, R. (1995). "Aufbau eines EDV-gestützten Messystems für forstliche Anwendungen am Beispiel der Volumenbestimmung von Holzpoltern". PhD thesis. Universität Göttingen.

Mikolajczyk, K. and C. Schmid (2002). "An affine invariant interest point detector". In: *Computer Vision—ECCV 2002*. Springer, pages 128–142.

Mikolajczyk, K. and C. Schmid (2004). "Scale & affine invariant interest point detectors". In: *International journal of computer vision* 60.1, pages 63–86.

Mikolajczyk, K. and C. Schmid (2005). "A performance evaluation of local descriptors". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 27.10, pages 1615–1630.

Ministerium für Umwelt, Forsten und Verbraucherschutz Rheinland-Pfalz (2013). *Rundholz Vermessungsanweisung RV-A*. Rheinland-Pfalz.

Moeslund, T. B. and E. Granum (2001). "A survey of computer vision-based human motion capture". In: *Computer Vision and Image Understanding* 81.3, pages 231–268.

Mohan, A., C. Papageorgiou, and T. Poggio (2001). "Example-based object detection in images by components". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 23.4, pages 349–361.

Molton, N., A. J. Davison, and I. Reid (2004). "Locally Planar Patch Features for Real-Time Structure from Motion." In: *BMVC*, pages 1–10.

Moré, J. J. (1978). "The Levenberg-Marquardt algorithm: implementation and theory". In: *Numerical analysis*. Springer, pages 105–116.

Moreira, A. and M. Y. Santos (2007). "Concave hull: A k-nearest neighbours approach for the computation of the region occupied by a set of points". In: *Proceedings of the 2nd International Conference on Computer Graphics Theory and Applications*.

Moreno, P., A. Bernardino, and J. Santos-Victor (2009). "Improving the SIFT descriptor with smooth derivative filters". In: *Pattern Recognition Letters* 30.1, pages 18–26.

Mori, G., X. Ren, A. A. Efros, and J. Malik (2004). "Recovering human body configurations: Combining segmentation and recognition". In: *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. Volume 2. IEEE, pages II–326.

Moulon, P. and B. Duisit (2014). *openMVG Documentation - Read the Docs*. URL: https://media.readthedocs.org/pdf/openmvg/latest/openmvg.pdf.

Moulon, P., P. Monasse, and R. Marlet (2013a). "Adaptive structure from motion with a contrario model estimation". In: *Computer Vision–ACCV 2012*. Springer, pages 257–270.

Moulon, P., P. Monasse, and R. Marlet (2013b). "Global Fusion of Relative Motions for Robust, Accurate and Scalable Structure from Motion". In: *Computer Vision (ICCV), 2013 IEEE International Conference on*. IEEE, pages 3248–3255.

Muja, M. and D. G. Lowe (2009). "Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration". In: *International Conference on Computer Vision Theory and Application VISSAPP'09)*. INSTICC Press, pages 331–340.

Muja, M. and D. G. Lowe (2012). "Fast Matching of Binary Features". In: *Proceedings of the 2012 Ninth Conference on Computer and Robot Vision*. CRV '12. Washington, DC, USA: IEEE Computer Society, pages 404–410. ISBN: 978-0-7695-4683-4.

Mukai, T. and N. Ohnishi (1999). "Sensor Fusion of a CCD Camera and an Acceleration-Gyro Sensor for the Recovery of Three-Dimensional Shape and Scale". In: *IEEE Proceedings of the Second International Conference of Information Fusion*, pages 221–228.

Ng, H., S. Ong, K. Foong, P. Goh, and W. Nowinski (2006). "Medical image segmentation using K-means clustering and improved watershed algorithm". In: *Image Analysis and Interpretation, 2006 IEEE Southwest Symposium on*. IEEE, pages 61–65.

Noonpan, V. and R. Chaisricharoen (2013). "Wide area estimation of piled logs through image segmentation". In: *Communications and Information Technologies (ISCIT), 2013 13th International Symposium on*. IEEE, pages 757–760.

Nosrati, M. and G. Hamarneh (2014). "Local Optimization Based Segmentation of Spatially-Recurring, Multi-Region Objects With Part Configuration Constraints". In: *Medical Imaging, IEEE Transactions on* 33.9, pages 1845–1859.

Nüchter, A. (2009). *3D Robotic Mapping: The Simultaneous Localization and Mapping Problem with Six Degrees of Freedom*. Springer Tracts in Advanced Robotics. Springer. ISBN: 9783540898832.

Nützi, G., S. Weiss, D. Scaramuzza, and R. Siegwart (2011). "Fusion of IMU and vision for absolute scale estimation in monocular SLAM". In: *Journal of intelligent & robotic systems* 61.1-4, pages 287–299.

Nylinder, M., T. Kubénka, and M. Hultnäs (2008). "Roundwood measurement of truck loads by laser scanning". In: *Field study at Arauco pulp mill Nueva Aldea*, pages 1–9.

Ohnishi, N. and A. Imiya (2005). "Dominant plane detection using optical flow and independent component analysis". In: *Brain, Vision, and Artificial Intelligence*. Springer, pages 487–496.

Ohnishi, N. and A. Imiya (2006). "Dominant plane detection from optical flow for robot navigation". In: *Pattern Recognition Letters* 27.9, pages 1009–1021.

Ojala, T., M. Pietikainen, and D. Harwood (1994). "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions". In: *Proceedings of the 12th IAPR International Conference on Pattern Recognition, 1994. Vol. 1 - Conference A: Computer Vision & Image Processing*. Volume 1, 582–585 vol.1.

Ojala, T. and M. Pietikäinen (1999). "Unsupervised texture segmentation using feature distributions". In: *Pattern Recognition* 32.3, pages 477–486.

Ojala, T., M. Pietikäinen, and D. Harwood (1996). "A comparative study of texture measures with classification based on featured distributions". In: *Pattern recognition* 29.1, pages 51–59.

Ojala, T., M. Pietikainen, and T. Maenpaa (2002). "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24.7, pages 971–987.

Okutomi, M., K. Nakano, J. Maruyama, and T. Hara (2002). "Robust estimation of planar regions for visual navigation using sequential stereo images". In: *Robotics and Automa-*

tion, 2002. Proceedings. ICRA'02. IEEE International Conference on. Volume 4. IEEE, pages 3321–3327.

Osher, S. and J. A. Sethian (1988). "Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations". In: *Journal of computational physics* 79.1, pages 12–49.

Otsu, N. (1975). "A threshold selection method from gray-level histograms". In: *Automatica* 11.285-296, pages 23–27.

Otte, B. (2014). "Dreidimensionale Objekterkennung durch zweidimensionale Klassifikation am Beispiel von Stirnflächen von Holzstämmen". Supervisors: B. Stock and C. Herbon. Master's thesis. HAWK - Hochschule für angewandte Wissenschaft und Kunst.

Papula, L. (2008). *Mathematik für Ingenieure und Naturwissenschaftler - Band 3*. Mathematik für Ingenieure und Naturwissenschaftler / Lothar Papula. Vieweg + Teubner.

Paragios, N. and R. Deriche (2002). "Geodesic active regions and level set methods for supervised texture segmentation". In: *International Journal of Computer Vision* 46.3, pages 223–247.

Pears, N. and B. Liang (2001). "Ground plane segmentation for mobile robot visual navigation". In: *Intelligent Robots and Systems, 2001. Proceedings. 2001 IEEE/RSJ International Conference on.* Volume 3. IEEE, pages 1513–1518.

Permuter, H., J. Francos, and I. Jermyn (2006). "A study of Gaussian mixture models of color and texture features for image classification and segmentation". In: *Pattern Recognition* 39.4, pages 695–706.

Petrovski, I. (2014). *GPS, GLONASS, Galileo, and BeiDou for Mobile Devices: From Instant to Precise Positioning.* Cambridge University Press. ISBN: 9781139952453.

Phung, S. L., A. Bouzerdoum, and D. Chai Sr (2005). "Skin segmentation using color pixel classification: analysis and comparison". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 27.1, pages 148–154.

Plattform Forst & Holz (2012). *Erster Bestandteil der RVR auf dem Weg.* Press release.

Plattform Forst & Holz (2014). *RVR nach siebenjährigem Verhandlungsmarathon unterzeichnet.* Press release.

Pollefeys, M., D. Nistér, J.-M. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S.-J. Kim, P. Merrell, et al. (2008). "Detailed real-time urban 3d reconstruction from video". In: *International Journal of Computer Vision* 78.2-3, pages 143–167.

Price, B. L., B. Morse, and S. Cohen (2010). "Geodesic graph cut for interactive image segmentation". In: *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on.* IEEE, pages 3161–3168.

Prodan, M. (1965). *Holzmesslehre.* Sauerländer.

El-Rabbany, A. (2002). *Introduction to GPS: The Global Positioning System*. Artech House mobile communications series. Artech House. ISBN: 9781580531832.

Rahman, S., S. Yella, and M. Dougherty (2011). "Image processing technique to count the number of logs in a timber truck". In: *Proceedings of the IASTED Conference on Signal and Image Processing, USA*.

Ren, X. and J. Malik (2003). "Learning a classification model for segmentation". In: *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*. IEEE, pages 10–17.

Riquelme, M., P. Barreiro, M. Ruiz-Altisent, and C. Valero (2008). "Olive classification according to external damage using image analysis". In: *Journal of Food Engineering* 87.3, pages 371–379.

Rodrigues, O. (1816). *Correspondance sur l'École impériale Polytechnique: Avril 1804 - Janvier 1816*. Correspondance sur l'École impériale Polytechnique: Avril 1804 - Janvier 1816 3. Bernard, pages 361–385.

Roerdink, J. B. and A. Meijster (2000). "The watershed transform: Definitions, algorithms and parallelization strategies". In: *Fundamenta Informaticae* 41.1, pages 187–228.

Rosten, E. and T. Drummond (2005). "Fusing points and lines for high performance tracking". In: *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*. Volume 2. IEEE, pages 1508–1515.

Rosten, E. and T. Drummond (2006). "Machine learning for high-speed corner detection". In: *Computer Vision–ECCV 2006*. Springer, pages 430–443.

Rother, C., V. Kolmogorov, and A. Blake (2004). "Grabcut: Interactive foreground extraction using iterated graph cuts". In: *ACM Transactions on Graphics (TOG)*. Volume 23. ACM, pages 309–314.

Rousson, M. and N. Paragios (2002). "Shape priors for level set representations". In: *Computer Vision—ECCV 2002*. Springer, pages 78–92.

Royer, E., J. Bom, M. Dhome, B. Thuilot, M. Lhuillier, and F. Marmoiton (2005). "Outdoor autonomous navigation using monocular vision". In: *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*. IEEE, pages 1253–1258.

Rublee, E., V. Rabaud, K. Konolige, and G. Bradski (2011). "ORB: an efficient alternative to SIFT or SURF". In: *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, pages 2564–2571.

Rundholzsortierungsvorschrift - RSV 88 - Vorschrift über die Sortierung, Vermessung und Kennzeichnung von Rundholz durch die Forstbehörden des Landes Nordrhein-Westfalen - RSV 88 - RdErl. des Ministers für Umwelt, Raumordnung und Landwirtschaft vom 1.10.1988 (1988). Nordrhein-Westfalen Ministerium für Umwelt, Raumordnung und Landwirtschaft.

Russ, J. (2011). *The Image Processing Handbook, Sixth Edition*. CRC Press. ISBN: 97814398-40634.

Rusu, R. B. and S. Cousins (2011). "3d is here: Point cloud library (pcl)". In: *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, pages 1–4.

Sachße, T. (2003). "Untersuchungen zu Fehlerquellen bei der Volumenermittlung von Kiefern-Langholzabschnitten im Wald - Vergleich der Aufnahmeverfahren mit der Werksvermessung". Master's thesis. München: Fachhochschule Eberswalde.

Sauter, U. H., S. Verhoff, and J. Dehning (2012). "Die Gütesortierung von Rohholz nach der Forst-HKS - eine Option für Deutschland?" In: *AFZ-DerWald* 14, pages 758–762.

Sawhney, H. S. and R. Kumar (1999). "True multi-image alignment and its application to mosaicing and lens distortion correction". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 21.3, pages 235–243.

Schapire, R. E. and Y. Singer (1999). "Improved boosting algorithms using confidence-rated predictions". In: *Machine learning* 37.3, pages 297–336.

Scheller, M., J. Voskamp, E. Gutzeit, and S. Mader (2014). *Verfahren zur automatisierten Mengenermittlung von beliebig großen Holzpoltern mittels eines Fotohandys*. DE Patent 102,012,017,878.

Schöttle, R., C. Pfeil, and M. Bacher (1999). "Hochmechanisierte Holzernte im Steilhang - Vorkonzentration durch Raupenharvester mit anschließender Seilkranrückung". In: *Versuchsberichte 1999*. Freiburg: Forstliche Versuchs- und Forschungsanstalt Baden-Württemberg.

Schraml, R., J. Charwat-Pessler, A. Petutschnigg, and A. Uhl (2014). *Robustness of biometric wood log traceability using digital log end images*. Technical report. Department of Computer Sciences, University of Salzburg.

Schraml, R. and A. Uhl (2013). "Pith estimation on rough log end images using local fourier spectrum analysis". In: *Proceedings of the 14th Conference on Computer Graphics and Imaging (CGIM'13), Innsbruck, AUT*.

Schraml, R. and A. Uhl (2014a). "Similarity based cross-section segmentation in rough log end images". In: *Artificial Intelligence Applications and Innovations*. Springer, pages 614–623.

Schraml, R. and A. Uhl (2014b). "Temporal and longitudinal variances in wood log cross-section image analysis". In: *IEEE International Conference on Image Processing*.

Schraml, R. (2013). "TreeBio - Preliminary study on traceability of tree logs using digital log end images". Master's thesis. Paris-Lodron-Universität Salzburg.

Sethian, J. A. (1999). *Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*. Volume 3. Cambridge university press.

Seto, K. (2011). *Administration method of lumber*. US Patent App. 13/006,635.

Shen, Y., P. Guturu, T. Damarla, B. P. Buckles, and K. Namuduri (2009). "Video stabilization using principal component analysis and scale invariant feature transform in particle filter framework". In: *Consumer Electronics, IEEE Transactions on* 55.3, pages 1714–1721.

Shojaii, R., J. Alirezaie, and P. Babyn (2005). "Automatic lung segmentation in CT images using watershed transform". In: *Image Processing, 2005. ICIP 2005. IEEE International Conference on.* Volume 2. IEEE, pages II–1270.

Shum, H.-Y. and R. Szeliski (1999). "Stereo reconstruction from multiperspective panoramas". In: *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on.* Volume 1. IEEE, pages 14–21.

Shum, H.-Y. and R. Szeliski (2000). "Systems and experiment paper: Construction of panoramic image mosaics with global and local alignment". In: *International Journal of Computer Vision* 36.2, pages 101–130.

Simonaho, S.-P., J. Palviainen, Y. Tolonen, and R. Silvennoinen (2004). "Determination of wood grain direction from laser light scattering pattern". In: *Optics and Lasers in Engineering* 41.1, pages 95–103.

Sinclair, D. and A. Blake (1996). "Quantitative planar region detection". In: *International Journal of Computer Vision* 18.1, pages 77–91.

Smith, S. M. and J. M. Brady (1997). "SUSAN—a new approach to low level image processing". In: *International journal of computer vision* 23.1, pages 45–78.

Snavely, N., S. M. Seitz, and R. Szeliski (2006). "Photo tourism: exploring photo collections in 3D". In: *ACM transactions on graphics (TOG)* 25.3, pages 835–846.

Snavely, N., S. M. Seitz, and R. Szeliski (2008). "Modeling the world from internet photo collections". In: *International Journal of Computer Vision* 80.2, pages 189–210.

Spampinato, C., D. Giordano, R. Di Salvo, Y.-H. J. Chen-Burger, R. B. Fisher, and G. Nadarajan (2010). "Automatic fish classification for underwater species behavior understanding". In: *Proceedings of the first ACM international workshop on Analysis and retrieval of tracked events and motion in imagery streams.* ACM, pages 45–50.

Statistisches Bundesamt (2014). *Land- und Forstwirtschaft, Fischerei.* Wiesbaden: Publications office of the European Union.

Staudenmaier, J. (2012). *Verfahren zur einzelstammweisen Volumen- und Konturermittlung von Rundholz am Beispiel von Nadel-Stammholzabschnitten : Jörg Staudenmaier.* Schriftenreihe Freiburger forstliche Forschung. Forstliche Versuchs- und Forschungsanstalt. ISBN: 9783933548559.

Steckhan, D., T. Bergen, T. Wittenberg, and S. Rupp (2008). "Efficient large scale image stitching for virtual microscopy". In: *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 4019–4023.

Steckhan, D. and D. Paulus (2010). "A quadratic programming approach for the mosaicing of virtual slides that incorporates the positioning accuracy of the microscope stage". In: *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 72–77.

Stiftung Warentest (2011). "Schlusslicht E-Plus". In: *test* 08/2011.

Strelow, D. and S. Singh (2002). "Optimal motion estimation from visual and inertial measurements". In: *Applications of Computer Vision, 2002.(WACV 2002). Proceedings. Sixth IEEE Workshop on*. IEEE, pages 314–319.

Swain, M. J. and D. H. Ballard (1992). "Indexing via color histograms". In: *Active Perception and Robot Vision*. Springer, pages 261–273.

Szeliski, R. (1996). "Video mosaics for virtual environments". In: *Computer Graphics and Applications, IEEE* 16.2, pages 22–30.

Szeliski, R. (2006). "Image Alignment and Stitching: A Tutorial". In: *Foundations and Trends in Computer Graphics and Vision* 2.1.

Szeliski, R. and H.-Y. Shum (1997). "Creating full view panoramic image mosaics and environment maps". In: *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*. SIGGRAPH '97. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., pages 251–258.

Tabb, A. L., D. L. Peterson, and J. Park (2006). "Segmentation of apple fruit from video via background modeling". In: *ASABE paper*.

Tai, Y.-W., J. Jia, and C.-K. Tang (2005). "Local color transfer via probabilistic segmentation by expectation-maximization". In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. Volume 1. IEEE, pages 747–754.

Tan, X. and B. Triggs (2010). "Enhanced local texture feature sets for face recognition under difficult lighting conditions". In: *Image Processing, IEEE Transactions on* 19.6, pages 1635–1650.

Tarp-Johansen, M. and K. Dralle (2012). *A method and a system for automatic measurement and tracking of logs, industrial wood and boards*. EP Patent App. EP20,120,155,369.

Tek, F. B., A. G. Dempster, and I. Kale (2005). "Blood cell segmentation using minimum area watershed and circle radon transformations". In: *Mathematical Morphology: 40 Years On*. Springer, pages 441–454.

Tian, X. and G. E. Murphy (1997). "Detection of trimmed and occluded branches on harvested tree stems using texture analysis". In: *Journal of Forest Engineering* 8.2, pages 65–78.

Toennies, K. (2012). *Guide to Medical Image Analysis: Methods and Algorithms*. Advances in Computer Vision and Pattern Recognition. Springer. ISBN: 9781447127505.

Tomasi, C. and T. Kanade (1991). *Detection and tracking of point features*. School of Computer Science, Carnegie Mellon Univ. Pittsburgh.

Tönnies, K. (2005). *Grundlagen der Bildverarbeitung.* I, Informatik. Pearson Studium. ISBN: 9783827371553.

Torr, P. H. and A. Zisserman (2000). "Feature based methods for structure and motion estimation". In: *Vision Algorithms: Theory and Practice.* Springer, pages 278–294.

Treue, S., M. Husain, and R. A. Andersen (1991). "Human perception of structure from motion". In: *Vision research* 31.1, pages 59–75.

Triggs, B., P. McLauchlan, R. Hartley, and A. Fitzgibbon (2000). "Bundle Adjustment - A Modern Synthesis". In: *Vision Algorithms: Theory and Practice.* Edited by B. Triggs, A. Zisserman, and R. Szeliski. Volume 1883. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pages 298–372. ISBN: 978-3-540-67973-8.

Tuytelaars, T. and K. Mikolajczyk (2008). "Local invariant feature detectors: a survey". In: *Foundations and Trends® in Computer Graphics and Vision* 3.3, pages 177–280.

Tuytelaars, T. and L. J. Van Gool (2000). "Wide Baseline Stereo Matching based on Local, Affinely Invariant Regions." In: *BMVC.* Volume 412.

Ullrich, J. and E. Krätzschmar (2012). "Die Fotooptische Holzvermessung - Bildanalyse findet Baumstamm". Arbeitsgruppe Automation in Kartographie, Photogrammetrie und GIS. URL: http://www.ikg.uni-hannover.de/aga/fileadmin/aga/documents/pdf-files_aga2012/b3_3_ullrich_kraetzschmar.pdf.

Umesh Adiga, P. and B. Chaudhuri (2001). "An efficient method based on watershed and rule-based merging for segmentation of 3-D histo-pathological images". In: *Pattern Recognition* 34.7, pages 1449–1458.

Uyttendaele, M., A. Criminisi, S. B. Kang, S. Winder, R. Szeliski, and R. Hartley (2004). "Image-based interactive exploration of real-world environments". In: *Computer Graphics and Applications, IEEE* 24.3, pages 52–63.

Van Laar, A. and A. Akça (2007). *Forest Mensuration.* Managing Forest Ecosystems. Springer. ISBN: 9781402059919.

Vargas, M. and E. Malis (2005). "Visual servoing based on an analytical homography decomposition". In: *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC'05. 44th IEEE Conference on.* IEEE, pages 5379–5384.

Vese, L. A. and T. F. Chan (2002). "A multiphase level set framework for image segmentation using the Mumford and Shah model". In: *International journal of computer vision* 50.3, pages 271–293.

Viola, P. and M. Jones (2001). "Rapid object detection using a boosted cascade of simple features". In: *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on.* Volume 1. IEEE, pages I–511.

Vogiatzis, G., C. Hernández, P. H. Torr, and R. Cipolla (2007). "Multiview stereo via volumetric graph-cuts and occlusion robust photo-consistency". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29.12, pages 2241–2246.

Vogiatzis, G., P. H. Torr, and R. Cipolla (2005). "Multi-view stereo via volumetric graph-cuts". In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on.* Volume 2. IEEE, pages 391–398.

Wanninger, L. (2006). "Netz-RTK". In: *GPS und GALILEO. Methoden, Lösungen und neueste Entwicklungen. Beiträge zum* 66, pages 59–69.

Wegelaar, R. (1997). "Rundholzvermessung mit Harvestern". In: *AFZ/Der Wald* 15, pages 809–812.

Weinstein, R. S., A. R. Graham, L. C. Richter, G. P. Barker, E. A. Krupinski, A. M. Lopez, K. A. Erps, A. K. Bhattacharyya, Y. Yagi, and J. R. Gilbertson (2009). "Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future". In: *Human pathology* 40.8, pages 1057–1069.

Weiss, S., D. Scaramuzza, and R. Siegwart (2011). "Monocular-SLAM–based navigation for autonomous micro helicopters in GPS-denied environments". In: *Journal of Field Robotics* 28.6, pages 854–874.

Wellington, C. and A. Stentz (2004). "Online adaptive rough-terrain navigation vegetation". In: *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on.* Volume 1. IEEE, pages 96–101.

Wellington, C. and A. Stentz (2006). "Learning predictions of the load-bearing surface for autonomous rough-terrain navigation in vegetation". In: *Field and Service Robotics.* Springer, pages 83–92.

West, P. (2009). *Tree and Forest Measurement.* Biomedical and Life Sciences. Springer. ISBN: 9783540959663.

Whichello, A. P. and H. Yan (1998). "Document image mosaicing". In: *Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on.* Volume 2. IEEE, pages 1081–1083.

Willmann, M. (2009). *Verfahren zur Mengenermittlung von Industrieholzsortimenten.* DE Patent 10,031,410.

Wu, C. (2011). *VisualSFM: A visual structure from motion system.* URL: http://ccwu.me/vsfm/.

Wu, C. (2013). "Towards linear-time incremental structure from motion". In: *3DTV-Conference, 2013 International Conference on.* IEEE, pages 127–134.

Wu, C., S. Agarwal, B. Curless, and S. M. Seitz (2011). "Multicore bundle adjustment". In: *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on.* IEEE, pages 3057–3064.

Wu, K., D. Gauthier, and M. D. Levine (1995). "LIVE CELL IMAGE SEGMENTATION". In: *IEEE Transactions on Biomedical Engineering* 42.1.

Xiong, Y. and K. Pulli (2009a). "Sequential image stitching for mobile panoramas". In: *7th International Conference on Information, Communications and Signal Processing, 2009. ICICS 2009.* Pages 1–5.

Xiong, Y. and K. Pulli (2010). "Fast image stitching and editing for panorama painting on mobile phones". In: *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 47–52.

Xiong, Y. and K. Pulli (2009b). "Sequential image stitching for mobile panoramas". In: *Information, Communications and Signal Processing, 2009. ICICS 2009. 7th International Conference on.* IEEE, pages 1–5.

Yang, M.-H. and N. Ahuja (1999). "Gaussian mixture model for human skin color and its application in image and video databases". In: *Proc. SPIE: Storage and Retrieval for Image and Video Databases VII.* Volume 3656, pages 458–466.

Yella, S. and M. Dougherty (2013). "Automatically Detecting the Number of Logs on a Timber Truck". In: *Journal of Intelligent Systems* 22.4, pages 417–435.

Yongsheng, S., L. Gang, and G. Rui (2009). "Segmentation algorithm for green apples recognition based on K-means algorithm". In: *Transactions of the Chinese Society for Agricultural Machinery* 40.Suppl 1, pages 100–104.

Young, M. J., M. S. Landy, and L. T. Maloney (1993). "A perturbation analysis of depth perception from combinations of texture and motion cues". In: *Vision research* 33.18, pages 2685–2696.

Yu, X., S. Beucher, and M. Bilodeau (1992). "Road tracking, lane segmentation and obstacle recognition by mathematical morphology". In: *Intelligent Vehicles' 92 Symposium., Proceedings of the.* IEEE, pages 166–172.

Yuen, D. and B. MacDonald (2002). "Natural landmark based localisation system using panoramic images". In: *ICRA '02. IEEE International Conference on Robotics and Automation, 2002. Proceedings.* Volume 1, pages 915–920.

Zhang, J. and H.-H. Nagel (1994). "Texture-based segmentation of road images". In: *Intelligent Vehicles '94 Symposium, Proceedings of the*, pages 260–265.

Zhang, W. and J. Kosecka (2006a). "A New Inlier Identification Scheme for Robust Estimation Problems." In: *Robotics: Science and Systems.*

Zhang, W. and J. Kosecka (2006b). "Image based localization in urban environments". In: *3D Data Processing, Visualization, and Transmission, Third International Symposium on.* IEEE, pages 33–40.

Zhang, Y., B. Pang, Y. Yu, and C. Zhang (2013). "Automatic microscopy image stitching based on geometry features". In: *Image and Signal Processing (CISP), 2013 6th International Congress on.* Volume 2, pages 927–931.

Zhang, Z. (2000). "A flexible new technique for camera calibration". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22.11, pages 1330–1334.

Zhang, Z. and A. R. Hanson (1996). "3D Reconstruction Based on Homography Mapping". In: *In ARPA Image Understanding Workshop*, pages 0249–6399.

Zhao, G., M. Barnard, and M. Pietikainen (2009). "Lipreading with local spatiotemporal descriptors". In: *Multimedia, IEEE Transactions on* 11.7, pages 1254–1265.

Zhao, G. and M. Pietikainen (2007). "Dynamic texture recognition using local binary patterns with an application to facial expressions". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29.6, pages 915–928.

Zheng, J. Y. (2003). "Digital route panoramas". In: *IEEE Multimedia* 10.3, pages 57–67.

Zheng, Q. and R. Chellappa (1995). "Automatic feature point extraction and tracking in image sequences for arbitrary camera motion". In: *International journal of computer vision* 15.1-2, pages 31–76.

Zhu, Q., K.-T. Cheng, C.-T. Wu, and Y.-L. Wu (2004). "Adaptive learning of an accurate skin-color model". In: *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on.* IEEE, pages 37–42.

Zhu, Q., M.-C. Yeh, K.-T. Cheng, and S. Avidan (2006). "Fast human detection using a cascade of histograms of oriented gradients". In: *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on.* Volume 2. IEEE, pages 1491–1498.

Zhu, X., L. Yang, and A. Waibel (2000). "Segmenting hands of arbitrary color". In: *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on.* IEEE, pages 446–453.

Zoghlami, I., O. Faugeras, and R. Deriche (1997). "Using geometric corners to build a 2D mosaic from a set of images". In: *Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1997*, pages 420–425.

Zucchelli, M., J. Santos-Victor, and H. I. Christensen (2002). "Multiple Plane Segmentation Using Optical Flow." In: *BMVC.* Volume 2, pages 313–322.

Zwillinger, D. (2011). *CRC Standard Mathematical Tables and Formulae, 32nd Edition (Discrete Mathematics and Its Applications).* CRC Press. ISBN: 1439835489.