Department of Computer Science
Series of Publications A
Report A-2020-11

Multi-Projective Camera-Calibration, Modeling, and Integration in Mobile-Mapping Systems

Ehsan Khoramshahi

Doctoral dissertation, to be presented for public examination with the permission of the Faculty of Science of the University of Helsinki, in Exactum auditorium CK112, on the 14th of December, 2020 at 14:00 o'clock.

University of Helsinki
Finland

Supervisor
Eija Honkavaara, Finnish Geospatial Research Institute, Finland
Arto Klami, University of Helsinki, Finland
Petri Myllymäki, University of Helsinki, Finland

Pre-examiners
Petri Rönnholm, Aalto University, Finland
Jan Dirk Wegner, ETH Zürich, Switzerland

Opponent
Janne Heikkilä, University of Oulu, Finland

Custos
Petri Myllymäki, University of Helsinki, Finland

Contact information

Department of Computer Science
P.O. Box 68 (Pietari Kalmin katu 5)
FI-00014 University of Helsinki
Finland

Email address: info@cs.helsinki.fi
URL: http://cs.helsinki.fi/
Telephone: +358 2941 911

ISSN 1238-8645
ISBN 978-951-51-6845-0 (paperback)
ISBN 978-951-51-6846-7 (PDF)
Helsinki 2020
Unigrafia

Multi-Projective Camera-Calibration, Modeling, and Integration in Mobile-Mapping Systems

Ehsan Khoramshahi

Department of Computer Science
P.O. Box 68, FI-00014 University of Helsinki, Finland
ehsan.khoramshahi@helsinki.fi;
https://www.mv.helsinki.fi/home/khoramsh/

PhD Thesis, Series of Publications A, Report A-2020-11
Helsinki, December 2020, 85+107 pages
ISSN 1238-8645
ISBN 978-951-51-6845-0 (paperback)
ISBN 978-951-51-6846-7 (PDF)

Abstract

Optical systems are vital parts of most modern systems such as mobile mapping systems, autonomous cars, unmanned aerial vehicles (UAV), and game consoles. Multi-camera systems (MCS) are commonly employed for precise mapping including aerial and close-range applications. Aerial photogrammetry has been recently progressed in many directions including real-time applications, and classification.

In the first part of this thesis a simple and practical calibration model and a calibration scheme for multi-projective cameras (MPC) is presented. The calibration scheme is enabled by implementing a camera test field equipped with a customized coded target as FGI's camera calibration room. The first hypothesis was that a test field is necessary to calibrate an MPC. Two commercially available MPCs with 6 and 36 cameras were successfully calibrated in FGI's calibration room. The calibration results suggest that the proposed model is able to estimate parameters of the MPCs with high geometric accuracy, and reveals the internal structure of the MPCs. The first hypothesis was proven by a concise assessment of results.

In the second part, the applicability of an MPC calibrated by the proposed approach was investigated in a mobile mapping system (MMS). The second hypothesis was that a system calibration is necessary to achieve high geometric accuracies in a multi-camera MMS. The MPC model was updated

to consider mounting parameters with respect to GNSS and IMU. A system calibration scheme for an MMS was proposed. The results showed that the proposed system calibration approach was able to produce accurate results by direct georeferencing of multi-images in an MMS. Results of geometric assessments suggested that a centimeter-level accuracy is achievable by employing the proposed approach. The high-accuracy results proved the second hypothesis. A novel correspondence map is demonstrated for MPCs that helps to create metric panoramas.

In the third part, the problem of real-time trajectory estimation of a UAV equipped with a projective camera was studied. The main objective of this part was to address the problem of real-time monocular simultaneous localization and mapping (SLAM) of a UAV. A customized multi-level pyramid-matching scheme based on propagating rectangular regions was proposed. The cost of matching was reduced by employing the proposed approach, while the quality of matches was preserved. An angular framework was discussed to address the gimbal lock singular situation. The results suggest that the proposed solution is an effective and rigorous monocular SLAM for aerial cases where the object is near-planar. The performance of the proposed algorithm suggests that it achieves time and precision goals of a real-time UAV trajectory estimation problem.

In the last part, the problem of tree-species classification by a UAV equipped with two hyper-spectral an RGB cameras was studied. The objective of this study was to investigate different aspects of a precise tree-species classification problem by employing state-of-art methods. Different combinations of input data layers were classifies to find the best combination of features. A 3D convolutional neural-network (3D-CNN) and a multi-layered perceptron (MLP) were proposed and compared for the classification task. Both classifiers were highly successful in their tasks, while the 3D-CNN was superior in performance. The classification result was the most accurate results published in comparison to other works. A combination of hyper-spectral and RGB data was demonstrated as the best data model, however, RGB data was shown as an inexpensive and efficient data-layer for many classification applications.

Computing Reviews (2012) Categories and Subject Descriptors:

Computing methodologies → Computer graphics → Image manipulation → Image processing
Computing methodologies → Computer graphics → Image manipulation → Computational photography
Computing methodologies → Modeling and simulation → Model development and analysis → Modeling methodologies

General Terms:
Design, Experimentation, Theory

Additional Key Words and Phrases:
Structure from Motion, Stereo Vision, Multi-Camera Calibration, Computer Vision, Monocular SLAM, Bundle Block Adjustment, Photogrammetry, Probabilistic Modeling, Coded Target, Automation, Calibration Room, Classification, Deep neural network, Convolutional neural network

Acknowledgements

I am grateful to Dr. Eija Honkavaara who trusted me, encouraged me, and supervised me in my scientific activities in FGI. Her excellent guidance and support pushed me through my scientific activities. Without her support, I would have faced so many difficult situations that would possibly lead to a failure.

I wish to express my sincerest thanks to my wife and my colleague Mrs. Somayeh Nezami for supporting me during these years. She was a wonderful source of help whenever needed. A common scientific language helped us to solve so many difficult situations that we faced.

I am grateful to Dr. Petri Myllymäki, Dr. Juha Hyyppä, and Dr. Arto Klami for supporting and supervising me. I also express my gratitude towards my co-authors for joint publications, colleagues in FGI for discussions, Dr. Harri Kaartinen and Dr. Antero Kukko for joyful scientific collaborations, and anonymous reviewers for commenting on our articles. I am especially grateful to Dr. Antonio Maria Garcia Tommaselli and Dr. Mariana Batista Campos for their valuable contribution and comments on Article **II**. I am also grateful to Finnish Geospatial Research Institute (FGI), National Land Survey of Finland (NLS), and University of Helsinki that financially supported me during my PhD.

Finally, and most importantly, I express my love and gratitude towards my family: Seyed Hadi, Masoumeh, Seyed Mohammad, Fatemeh Sadat, Seyed Ali, Seyedeh Kosar, and Noora Sadat Khoramshahi for their love and support.

Helsinki, November 2020
Ehsan Khoramshahi

List of Publications

This thesis consists of four original research papers that are referenced as Articles **I-V** throughout this thesis. Here you find a list of papers with a short description concerning the contributions of each author. Reprints of Articles **I-IV** are included at the end of the thesis.

Article I
Modelling and Automated Calibration of a General Multi-Projective Camera. *Photogrammetric Record*, Feb. 2018, Ehsan Khoramshahi, Eija Honkavaara.

Author roles: I made the literature review, design, and implementation. Dr. Honkavaara supervised the work and provided me important guidance and comments.

Article II
Accurate Calibration Scheme for a Multi-Camera Mobile Mapping System. *MDPI remote-sensing*, Dec. 2019, Khoramshahi, Ehsan, Mariana Batista Campos, Antonio Maria Garcia Tommaselli, Niko Viljanen, Teemu Mielonen, Harri Kaartinen, Antero Kukko, and Eija Honkavaara.

Author roles: I made the literature review, design, and implementation. Dr. Campus, Mr. Viljanen, Dr. Tommaselli, and Dr. Honkavaara contributed by providing comments. Dr. Kaartinen and Dr. Kukko made the mobile mapping system and performed the observation sessions. Mr. Viljanen helped with data preprocessing steps.

Article III
An Image-Based Real-Time Georeferencing Scheme for a UAV Based on a New Angular Parametrization. *MDPI Remote-Sensing*, Sep. 2020, Ehsan Khoramshahi, Raquel Oliveira, Niko Koivumäki, and Eija Honkavaara.

Author roles: The general design was done by me and Dr. Oliveira and Dr. Honkavaara through discussion sessions. I made the literature review, main implementation, validation and testing, and initial draft creation. Dr. Oliveira and Dr. Honkavaara helped by providing important comments. Mr. Koivumäki contributed in data collection, and photogrammetric data processing.

Article IV
Tree Species Classification of Drone Hyperspectral and RGB Imagery with Deep Learning Convolutional Neural Networks. *MDPI remote-sensing*, Feb. 2020, Somayeh Nezami, Ehsan Khoramshahi, Olli Nevalainen, Ilkka Pölönen, Eija Honkavaara.

Author roles: Mrs. Nezami, Dr. Nevalainen, and Dr. Honkavaara initiated discussion about the problem. Me and Mrs. Nezami created the first draft. Me and Mrs. Nezami made the literature review. I proposed the structure for the 3D-CNN and implemented the method. Mrs. Nezami trained the network and classified the data. Dr. Pölönen contributed in the original text of 3D-CNN part. Dr. Nevalainen, Dr. Pölönen, and Dr. Honkavaara contributed by providing valuable comments.

Contents

Chapter 1

Introduction

Optical systems have recently enjoyed a significant outburst in technology and applications [1–3]. Many relatively inexpensive and practical optical sensors are available nowadays as computer-vision solutions [4]. Modern optical system such as multi-fisheye cameras, multi-projective cameras (MPC), high-resolution projective cameras, rotating-head panoramic cameras, and hyper-spectral cameras are vital part of almost any smart solution such as 360 cameras [5], autonomous cars [6], unmanned aerial vehicles (UAV) [7], mobile mapping systems [8], and modern game consoles [9]. The optical sensors are efficient and practical measurement units.

An optical system is efficiently employed in a smart solution when a) its internal structure is precisely estimated, and b) accurate synchronization and calibration mechanisms are employed to ensure its consistency with respect to other sensors. The internal structure of an optical system is usually formulated as abstract models that capsulate realities of image sensors and optics.

Important aspects of an optical solution include geometric [9-12] and radiometric calibration [13], geometric accuracy assessment [14–16], synchronization with other sensors [17, 18], and real-time and post processing aspects [19–22]. Geometric calibration consists of pre-calibration and self-calibration of interior orientation parameters of a sensor, and relative orientation of internal parts with respect to a fixed frame, or a local unit. Pre-calibration includes employing designed objects such as calibration plates or calibration test fields to estimate unknown internal parameters of a camera. Real-time processing steps include estimating the output parameters that are expected as the result of a measurement mission. These parameters could be e.g. sensor positions and orientations, uncertainty statements about the estimated parameters, or classification labels.

Employing an optical system requires efficient camera modeling. A camera model is a suitable mathematical form that represents physical reality of lenses and sensors [23]. Specialized camera models are designed to address specific conditions that a group of cameras encounters. For example, a projective camera model enables to appropriately address the distortion parameters caused by curvatures of a lens [24]. It also takes into account the relative position of a lens with respect to its sensor by appropriate parameters such as focal length or principal point parameters. This model is typically unable to address other sensors that are structurally different, such as fisheye or multi-cameras. Camera models are therefore defined for a group of sensors that are fundamentally similar in structure, lens and sensor.

In a smart system with complex optics such as an autonomous car, a comprehensive plot could be imagined based on sensing units (sensors), data-processing units (algorithms), and acting units (actors) [25]. Sensors are important first-level input units that are considered to work in a suitable way to generate necessary data for processing units such as localization methods, and classifiers. In this plot, a data-processing task may need inputs from more than one sensor; therefore, all sensors including cameras, lidars, global navigation satellite systems (GNSS), and inertial sensors should be synchronized and calibrated with respect to each other, a local frame, or their parent system. In such a system, raw sensor data is analyzed and converted to high-level information such as classification labels, conclusions, or system actions.

An important aspect of a smart system with complex optics relates to real-time processing [20]. In many modern photogrammetric applications (such as aerial object tracking, real-time aerial mapping, traffic monitoring, aerial fire monitoring, etc.) the final goal is to tackle a problem in real-time. For an application such as tree classification, the processing starts from geometric and radiometric calibration of cameras and other sensors. Real-time aspects include online transmission of data and results, task management, direct georeferencing, structure from motion (sfm) and real-time trajectory estimation, and on-board and cloud-based processing. Post processing aspects include ground control point measurement, Bundle Block Adjustment (BBA), dense point cloud generation, orthomosaic generation, and classification.

The objective of this thesis is to study different aspects of rigorous solutions for fundamental subtasks in the aforementioned comprehensive plot, particularly geometric data processing, real-time processing, and analytics. The geometric modeling is the first focus of this work. Here, solutions for

different aspects of the plot, from calibrating an MPC and sensor synchronization, to real-time trajectory estimation and classification are rigorously proposed and accurately tested.

Image processing is the main tool to achieve important objectives of this work. Image processing is the art and science of dealing with images of objects, animals, peoples, trees, locations, etc., and extract useful information from captured images [26]. Indisputably, image processing is a key player in modern technologies and one of the most interesting adventures since the invention of digital computing. Image processing turned into a domain to develop algorithms that perform similar tasks that our brains have been able to perform for thousands of years [27]. Image processing includes a wide spectrum of routines from basic image enhancements such as geometric or radiometric enhancements e.g. sharpening, blurring, histogram equalization, to more complex tasks such as geometric and radial calibration of cameras, and classification.

We create and employ a test field to estimate internal parameters of a camera. A custom coded-target is proposed to build the test field. The main question that is answered in this respect is how to create an "easy-to-detect" and "efficient" coded-target for single and multi-projective cameras. Parameters such as accuracy, good visibility (minimum maximum distance, and usability for a specific field of view), embedded identifier, embedded scale, and an automatic reading capability are key factors to consider.

Once we have a camera calibration room, we can use it to estimate internal structure of a camera. In this thesis we consider multi-project camera calibration and modeling. Multi cameras are optical systems consisting of a set of internal cameras that bond together in a rigid frame [28]. Multi cameras are usually designed to either cover larger areas that a single camera is unable to cover, or measure several spectral channels of the electromagnetic spectrum to produce multi-spectral or hyper-spectral mosaics. There are three most common forms of multi cameras: a) multi-fisheye, b) multi-projective, and c) mixed types. Multi-fisheye cameras are usually equipped with two or more cameras with fisheye lenses to produce super-wide images (usually 360 degrees images). MPCs have a simple structure that make them desirable for many modern applications. The next question that is answered by this thesis is "how to employ the proposed camera calibration room equipped with the planned coded-target to calibrate a single camera or an MPC?". Challenges in this regard are the applicability of a designed room for a camera class, room size and its structure, distribution of targets on the room's surface, initial estimation of target positions, un-

certainty statement for estimated targets, and addressing singularities that arises in this problem.

After calibrating an MPC, the sensor-synchronization aspect is the next challenge. A custom calibration scheme for a mobile mapping system is proposed by including mounting parameters of an MPC with respect to GNSS and IMU. Direct georeferencing is employed to demonstrate geometric capabilities of the proposed model. Next, an efficient plot for creating metric panoramas is proposed.

A standard application of a pre-calibrated camera is in aerial mapping. Despite the availability of GNSS, image-based trajectory estimation is still a valuable alternative source specially when GNSS is unable to provide reliable positional information due to factors such as unavailability of satellites or multi-path error [29]. Real-time trajectory estimation of a UAV by a pre-calibrated single camera (SC) is known as monocular simultaneous localization and mapping (SLAM) [30]. This problem involves addressing sub tasks such as fast and reliable image matching, fast network initialization, and network optimization. This thesis contributes to monocular SLAM problem by proposing small solutions that address different aspects of it.

A bigger scheme to consider is dealing with a real-time photogrammetric platform calibrated to perform image interpretation tasks such as classification. When an accurate camera trajectory is acquired, we can employ it to perform tasks such as accurate localization of 3D point, dense point cloud generation, and accurate classification of objects. One such application relates to the problem of tree-species classification by a UAV that is equipped with RGB cameras and hyper spectral sensor [21]. This thesis contributes to this problem by investigating different aspects of an airborne tree-species classification problem. We propose a 3D Convolutional Neural Network (3D-CNN) classifier and compare it to a Multi-Layered Perceptron (MLP). We study the structure of the proposed 3D-CNN model along with different combinations of data layers to suggest the most efficient structure for the network and best set of features for the classification.

1.1 Research questions and hypotheses

In this section, a list of 8 research questions and 2 hypotheses are stated. The research questions are answered in Chapter 5, and the hypothesis are elaborated in Chapter 6.

- **RQ1**: What are the important parameters to consider when designing a coded target? How to achieve high sub-pixel accuracy, embedded coded-target, and embedded scale-bar in a coded target?

- **RQ2**: How to build a camera calibration room in an easy and efficient way for single and multi-cameras? How to accurately estimate the structure of the camera calibration room?

- **RQ3**: How to calibrate a multi-projective camera in an efficient and easy way? What are the challenges in this regard?

- **RQ4**: How to integrate a calibrated multi-projective camera in a mobile mapping system? What are the challenges and opportunities?

- **RQ5**: How to compile metric panoramas from multi-projective images?

- **RQ6**: How to address the gimbal-lock singularity in Jacobian matrix of a BBA?

- **RQ7**: What are the challenges to build a real-time photogrammetry system? How to address real-time challenges efficiently? What are the software limitations?

- **RQ8**: How to integrate deep learning to a UAV-based multisensorial photogrammetric mapping system? What are the challenges and opportunities?

- **HT1**: A camera test field is required to geometrically calibrate a multi-projective camera.

- **HT2**: A system calibration is essential to acquire high geometric accuracies from a mobile mapping system equipped with a multi-projective camera, a GNSS and an IMU.

We will return to these research questions and hypotheses in Chapter 5.

1.2 Contributions

The main contributions of this thesis are the following:

1. A novel code target and a calibration room A novel coded-target was introduced based on parameters such as sub-pixel accuracy, embedded identifier, embedded scale, and ease of automatic detection (Article **I**). A calibration room was equipped with the proposed coded-target for camera calibration (Article **I**).

2. Modelling and automated calibration of a general multi-projective camera: The proposed coded-target was employed to build a calibration room for single and multi-cameras (Article **I**). A novel calibration scheme for multi-projective cameras was finally proposed (Article **I**). The model was successfully demonstrated to find the internal structure of two multi-projective cameras with 36 (Article **I**) and 6 projective cameras (Article **II**).

3. An accurate calibration scheme for a multi-camera mobile mapping system: The multi-projective model that was proposed in Article **I** was improved in Article **II** to accept lever-arm vector parameters and boresight angles with respect to GNSS receiver and IMU sensor. A sparse BBA scheme was proposed in Article **II**. Direct geo-referencing was investigated based on the proposed model.

4. 3D mapping by employing a mobile mapping system consisting of multi-projective cameras: The proposed calibration scheme described in Article **I** was employed to calibrate a multi-projective camera. The mobile mapping calibration scheme described in Article **II** was employed to calibrate an MMS. The image points of the calibrated multi-camera in the MMS were employed to estimate positions of 3D object points in Article **II**. The global quality of 3D object point positioning was assessed by check points (Article **II**). The role of intersection geometry on the quality of 3D object point positioning was investigated (Article **II**).

5. Metric panoramic generation for multi-projective cameras A novel scheme for creating non-stitching panoramas for multi-projective cameras was proposed and discussed in Article **II**. Stitching panoramas had a lack of metric information that was addressed by introducing a contribution and a correspondence map (Article **II**). The contribution map offered the possibility to optimally design a multi-projective camera. The correspondence map offered a scheme for fast compilation of panorama for multi-projective cameras (Article **II**).

6. A new angular parametrization of BBA to address the gimbal lock singularity: A new angular parametrization of BBA based on spherical rotation coordinate system was presented to address the gimbal lock singularity (Article **III**).

7. Real-time georeferencing of UAV images: The sparse BBA scheme that was proposed in Articles **I** and **II** was employed for trajectory estimation of a UAV. A monocular SLAM solution was proposed in

Article **III**. The novel parametrization of sparse BBA (Articles **I** and **II**) based on spherical rotation coordinate system was presented in Article **III**. A modified network creation scheme was presented to address the robustness of network creation, while respecting its time constraints. A multi-level matching approach based on scale invariant feature transform was proposed to preserve high-frequency key-points in a real-time scheme (Articles **I**, **II** and **III**).

8. Tree-species classification by hyperspectral and RGB imagery: we studied different aspects of a tree-species classification by a UAV that was equipped with an RGB camera and a hyper-spectral sensor (Article **III**). A convolutional neural network was proposed to perform the classification task. The proposed classifier was compared with a multi-layered perceptron classifier. Different combinations of data layers were studied to find the most efficient set of features.

9. Efficiency of computations in a photogrammetric solution: A futuristic photogrammetric solution needs to be efficient in terms of sensor integrity, computation time, RAM usage, task splitting between edge and cloud, and classification aspects such as time and accuracy. Efficiency was the cross-cutting theme in all the research of this thesis (Articles **I-IV**).

1.3 Structure of this thesis

This thesis is organized as six chapters. The first chapter is introduction that contains a general overview, research questions and contributions. The second chapter concerns a brief overview of the background methods that are used in this work. A literature review of state-of-art methods is also provided in this chapter. The third chapter concerns a short introduction to implementation details of the proposed approaches of Articles **I-IV**. More details could be found in the original articles. In the fourth chapter results are organized and briefly presented. A brief discussion about results is organized as answers to the research questions in the fifth chapter that followed by a summary and conclusion in the sixth chapter. Abbreviations and synonyms used in this work are listed in Appendix A.

Chapter 2

Background and literature review

In this part, the background schemes of this work and a recent literature review are briefly presented. By the background schemes, we mean more of the basics of the problems that are addressed, as well as highlights on their usability in actual applications, and less about technical details. The research papers at the end of this thesis contains more technical details. The chapter starts with a brief introductory part that is followed by a mathematical background and literature review.

2.1 Cameras and multi-cameras

A modern optical system is a combination of digital sensors, electric boards, lenses, wiring, power resource, and storage that is designed to measure parts of the electromagnetic spectrum. Modern optical systems are inseparable parts of complex solutions such as UAVs, or autonomous cars.

A projective camera (or a single-frame camera) is a simple optical system consisting a rectangular sensor and a set of lenses. The basic assumption about a projective camera is that each point in its focal plane could be linked to the corresponding 3D object point through a projective transformation [31]. A linear correspondence is assumed after applying corrections for lens distortions and other sources of errors. A projective transformation is a non-linear eight parameter transformation that is employed to represent a perspective geometry. A distortion model helps to move from a projective camera to a pinhole camera.

A pinhole camera is a hypothetical distortion-free camera box with an infinitely small focal point [32]. A hypothetical ray puts a unique mark on its sensor when traveling through its focal point, therefore we can assume a perfect linear relationship between its image points, the focal point, and

the corresponding 3D object points. An interior orientation model is a mathematical form that helps to transfer distorted image points of a projective camera to their corresponding distortion-free pinhole coordinates. It takes the geometry and built imperfection of the lens and sensor into a suitable non-linear relationship with adjustable parameters called interior orientation parameters (IOP).

Panoramic cameras are an important type of optical systems to capture wide and super wide images. These cameras are efficient tools to capture more of the surrounding environment of any moving object. Panoramic cameras are usually employed in a wide variety of instruments such as 360 cameras, mapping instruments, robots, autonomous cars.

Multi-cameras are a sub-category of panoramic cameras. On the positive side, multi-cameras are inexpensive, easy to build, and geometrically solid and stable, mainly because of their intrinsic sturdy design, when compared to other wide-view systems such as rotating-head cameras. On the negative side, calibrating multi-cameras is a challenging task that needs specialized hardware and software. An MCS has a number of cameras are mounted fixed on a rigid frame that could be a metal bar, or a plastic frame with the purpose to concentrate on regions of interest. In today's market, a wide range of multi-cameras could be found, from inexpensive panoramic cameras to high-accuracy complex systems such as oblique aerial cameras.

An important mapping system that recently benefits from MCS is an MMS that is usually a collection of subsystems such as GNSS, inertial measurement unit (IMU), and imaging sensors such as cameras, lidars, and odometer sensors, integrated in a common moving platform mainly used for navigation and geometric and radiometric data acquisition [33].

A useful step regarding sensor modeling of panoramic cameras is to categorize them based on common properties and similarities. A categorization for panoramic imaging could be e.g. found in [23] where 4 groups of mirror-based rotating-head, scanning 360, stitched, and near 180 are listed. Panoramas from stitching are usually generated for non-metric applications such as visualization; stitched panoramas are therefore outside the focus of metric applications. we can present a simpler categorization based on common geometric properties for most of panoramic cameras as: a) rotating head, b) multi-fisheye, c) multi-projective, and d) catadioptric cameras [24]. This categorization helps to concentrate on a specific group with similarities that make a uniform modeling feasible.

The first category (rotating-head) consists of cameras that are equipped with a linear CCD array. These cameras are usually equipped with an arm that captures several shots from different angles. A panorama is finally

compiled by merging several images. EYESCAN M3 and SpheroCam are two examples of this class. EYESCAN M3 was jointly developed by DLR and KST, and SpheroCam was developed by SpheronVR AG. There are motorized rotating heads that equivalently employ a simple mechanism with a projective camera and a rotating arm. Few examples of this class are Roundshot metric, GigaPan EPIC pro, Phototechnik AG, or piXplorer 500. Usually a synchronized mechanism between camera shutter and step motors ensure by a central processing unit. An important factor when dealing with these cameras relates to the problem of putting the optical center of the projective camera on the center of rotation. Under this physical calibration limit, a high-accuracy high-resolution metric panorama is accessible in this class [3].

In this class, usually one or two step motors that are synchronously work to rotate the central projective camera to a given direction and orientation. Then, few shots of designed directions will be captured and combined to generate a high-resolution panorama. Color blending is the technique that is commonly used here to soften the edges that will otherwise affect the quality of the final panoramas. Compiling a panoramic image with this class requires few seconds to few minutes to compile depending to the output resolution, therefore, this class is not suitable for real-time applications such as mobile mapping. Important applications of this class could be mentioned e.g. in 3D modeling, artistic photography, high-precision surveying [10, 23, 34–36] and recording and classification of cultural heritage [1, 37].

The second category (multi-fisheye cameras) consist of a dual fisheye camara configuration that is mounted on a rigid frame. This configuration is suitable for portable and compact 360 imaging photography. Examples of this class include Ricoh Theta S [38] and Samsung Gear 360 [39, p. 360]. Fisheye cameras were studies e.g. in [11, 40, 41]. Dual-fisheye configuration was studies by [12, 38, 42]. For surveying application, low to medium accuracy is achievable for this class of cameras by employing non-linear weighted BBA. Advantages of this class of cameras include simplicity of design, compactness, relatively low-data-capturing rate. These advantages are desirable specially for application such as medium-accuracy mobile mapping [4, 39, 43–45], or 3D visualization where high capturing speed and simplicity are leading factors. On the other hand, Dual fisheye cameras have some limitations. We may mention e.g. non-perspectiveness, considerable distortions, considerable variation in scale and illumination between scenes, and nonuniformity of spatial resolution over an image. These factors affect the quality of panoramas that are generated in this category.

Often these panoramas are at a lower quality in comparison to the first and third category.

The third category (multi-projective cameras or MPC in short) contains a set of projective cameras that are mounted fixed to a rigid structure. These cameras are designed to address specific imaging, navigation, or surveying challenges. In multi-projective cameras, a fast-synchronized shutter mechanism is necessary to ensure simultaneous capturing of images from all cameras with minimum possible delay. Examples of this class include Panono with 36 projective camera, Ladybug v.5 with 6 cameras by FLIR, oblique airborne cameras with 2 or more oblique projective cameras such as Leica RCD 30, Ultracam Osprey mark 3, WELKIN-C1, Leica city-mapper 2, or Foxtech 3DM-Mini. The most basic approach to compile a panorama for this class is by employing automatic image stitching techniques. In contrast to the panoramas of the first and second category, these panoramas are mainly used for artistic and visualization purposes. The metric characteristics of these panoramas are lost due to errors that image stitching produces. A metric panoramic compilation scheme for this class is feasible by employing the multi-projective sensor information. High data capturing rate of this class makes it useful for various applications e.g. in visualization, cinema, artistic photography, land and aerial surveying, mobile mapping, and navigation.

The fourth category contains catadioptric cameras. This category contains complex optical systems such as shaped mirrors, spherical and aspherical lenses, and hyperbolic and elliptical mirrors [46]. A prototype of a dual catadioptric camera was proposed by [47, p. 360].

2.2 Sensor modeling (Articles I-III)

In this part, the mathematical background for most part of this thesis is elaborated. The first part of this section, the interior orientation model that was developed for this work is discussed. This model is followed by a description about collinearity and co-planarity conditions. Then a novel rotational frame work is described.

2.2.1 Collinearity and interior orientation model

In micro scale, we can safely assume that a light ray is travelling in a straight line. Our basic assumption fails when the scale of a problem increases, that is mainly due to the distortions that occur by traveling of light in a non-empty space. The role of gravity on distorting the light's path is

also considerable in macro scales. A micro scale problem simplifies by this assumption when dealing with an optical system modeling [32].

An ideal mathematical situation in an optical image system assumes by considering an object point, a focal center, and a corresponding image point collinear (Figure 1). The real situation on a micro scale is however different from mainly because of the effects such as imperfectness in lens built, effect on noise, and imperfectness in sensor built. Interior orientation helps to remove those effect by introducing an appropriate non-linear transformation [48]. We may first assume that we deal with an ideal camera with an ideal image coordinates that makes a perfect connection between to an abject point its focal point and a corresponding image point through collinearity equation:

$$\mathbf{X} = \lambda.\mathbf{R}_{(\omega,\phi,\kappa)}.\mathbf{x} + (\mathbf{X}_0)_t, \tag{1}$$

where $\mathbf{R}_{(\omega,\phi,\kappa)}$ is a right-handed 3×3 rotation matrix of the camera at time (t) with Euler angles (ω, ϕ, κ). The order of applying rotations are from end to start. $(\mathbf{X}_0)_t$ is the 3D position of camera. In Equation 1, x is pinhole image coordinates, $\mathbf{X}$ is the corresponding object point, and λ is the unknown scale. This scale could be removed by dividing first two equations by z:

$$x = -\frac{\mathbf{M}_1.(\mathbf{X} - (\mathbf{X}_0)_t)}{\mathbf{M}_3.(\mathbf{X} - (\mathbf{X}_0)_t)}, y = -\frac{\mathbf{M}_2.(\mathbf{X} - (\mathbf{X}_0)_t)}{\mathbf{M}_3.(\mathbf{X} - (\mathbf{X}_0)_t)}, \tag{2}$$

where $\mathbf{M} = \mathbf{R}^{(-1)}$. We name image coordinates of Equations 1 and 2 pinhole coordinates throughout this thesis. The word pinhole come from the fact that in an ideal setting (Figure 1), the lens is assumed as a point.

We may now convert the distorted image coordinates (Figure 2) into undistorted pinhole coordinates that satisfies Equations 1 and 2. For this purpose, we may first remove the shift of principal point (vertical projection of focal point on the image plane)

$$x_1 = \frac{(x)_{t_1} - PP_x}{f}, y_1 = \frac{(y)_{t_1} - PP_y}{f}, \tag{3}$$

then remove radial and tangential distortions by

$$r^2 = \sqrt{x_1^2 + y_1^2}, Rad = (1 + K_1.r^2 + K_2.r^4 + K_3.r^6), \tag{4}$$

$$(x_n)_{t_1} = x_1.Rad + 2.P_1.x_1.y_1 + P_2.(r^2 + 2(x_1)^2) - \delta.x_1 + \lambda.y_1, \tag{5}$$

$$(y_n)_{t_1} = y_1.Rad + 2.P_2.x_1.y_1 + P_1.(r^2 + 2(y_1)^2) + \lambda.x_1. \tag{6}$$

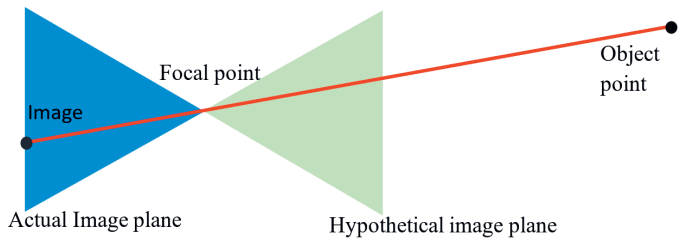

Figure 1: Collinearity condition in and ideal pinhole camera.

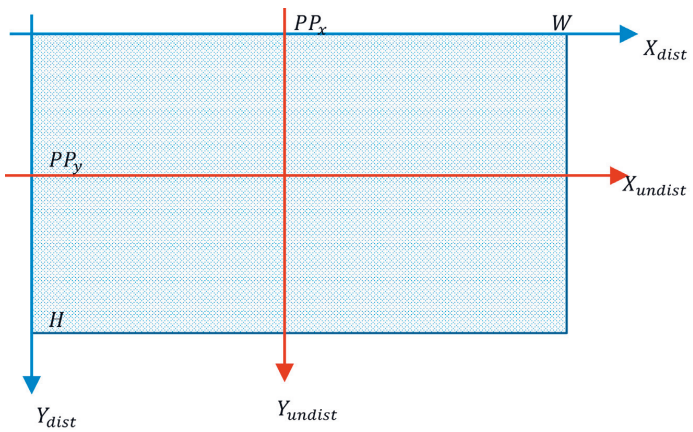

Figure 2: Distorted and undistorted (pinhole) image coordinates.

2.2.2 Coplanarity

The coplanarity equation is expressed as the condition that an object point, two focal centers of images that see the object, and the object point's corresponding image points are on the same plane [31] (coplanarity in Figure 3).

Figure 3 demonstrates two images that can see an object point. In this figure, we may assume a local right-hand-side Cartesian coordinate system for each image such that its center lies on the focal center of image (F_i). X and Y axis are parallel to the pinhole image axis, and Z perpendicular to the image plane. We set the focal distance equal to one. In this way, image coordinates of the form (x, y) will be corresponding to a 3D coordinates $(x, y, 1)$ in the local coordinate system of their parent image. A 3D point in a local coordinate system of an image could be transformed to the local

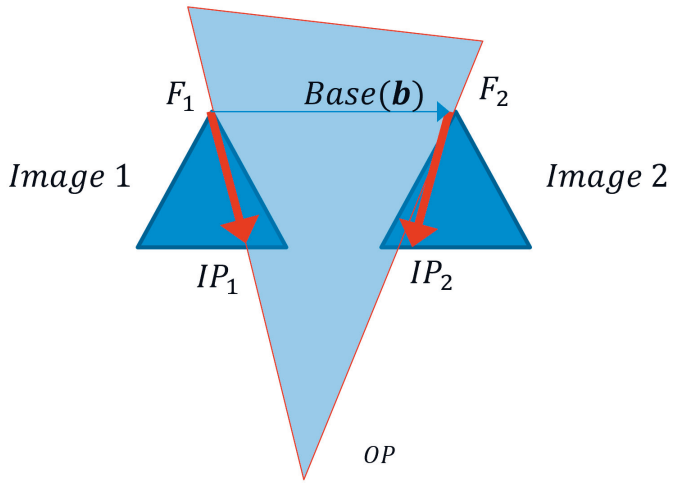

Figure 3: Side view of two (mirror) images that look downward to an object point. The gray triangle that is bounded by red lines represent the coplanarity condition in and ideal pinhole camera.

coordinate system of its neighboring images, if transformation parameters (3 rotations, 3 translations, and one scale) are known. This relationship could be written as

$$(x)_j = \lambda.\mathbf{R}_{ij}.(\mathbf{x})_i + (\mathbf{X}_0)_{ij}, \tag{7}$$

where $(\mathbf{x})_i$ is an arbitrary point in coordinate system i, $\mathbf{R}_{ij}$ is the relative rotation matrix between two coordinate systems, $\mathbf{b} = (\mathbf{X}_0)_{ij}$ is the relative shift, and λ is the scale factor.

If we assume that both local coordinate systems have the same scale, then $\lambda = 1$. Coplanarity between $(\mathbf{x}_1)_1$, $(\mathbf{x}_2)_1$, and $(\mathbf{b})_1$ is written as

$$(\mathbf{x}_1)_1.[(\mathbf{x}_2)_1 \times (\mathbf{b})_1] = 0, \tag{8}$$

where (.) is inner vector product, and $(\times)$ is cross vector product. Finally, Equation 8 boils down to

$$(\mathbf{x}_1)_1.[(\mathbf{R}_{12}.(\mathbf{x}_2)_2) \times (\mathbf{b})_1] = 0. \tag{9}$$

By considering $\mathbf{E} = [(\mathbf{X}_0)_t]_x.\mathbf{R}\mathbf{R}_t$, Equation 9 could be rewritten in matrix form

$$(\mathbf{x}_1)'_1.\mathbf{E}.(\mathbf{x}_2)_2 = 0, \tag{10}$$

where $(\mathbf{x}_1)_1$, and $(\mathbf{x}_2)_2$ are (3×1) matrices, and $\mathbf{E}$ is the 3×3 essential matrix. Therefore, the essential matrix is an appropriate tool to establish 3D relationship between two corresponding image points of an object point. To estimate an essential matrix, at-least few image to image correspondences should be known. There are well-studied algorithms that can estimate the essential (or a fundamental) matrix by having few image point correspondences. A brief list includes 8-point [49], 7-point [50], normalize 8-point [51], 5-point [52], and 6-point [53]. Nister's 5 point algorithm [52] is among the most robust algorithms when the focal length is approximately known.

2.2.3 Multi-projective camera

A sensor model for a multi-projective camera should include parameters concerning physical reality of the camera, e.g. interior orientation of projective cameras or relative orientations among them. The first models that proposed for the multi-cameras involved stereo camera calibration [54–56]. One of the first attempts to employ relative orientation constraints was made by [54] that proposed a system with a stereo camera and a GNSS receiver to localize 3D object points by spatial intersection of image points. It was later that [55] employed relative orientation stability constraints in a BBA. Another early attempt was made by [56] who employed a moving target with fixed-length for stereo-camera calibration. Then, other researchers proposed solutions for more complex calibration scenarios, e.g. [57] proposed a calibration scheme for a multi-camera with four projective cameras. System stability analysis was employed by [58] for a multi-camera system calibration (MCS). Moreover, they studied the optimum number of images necessary for a time-efficient major or minor multi-camera calibration. An MCS_3 of a thermal camera and a stereo camera was calibrated by [59] by using distance constraints. Variation of IOPs/EOPs of MPC was studied by [28]. A calibration scheme and a toolbox based on a chess pattern was proposed by [60]. They employed this toolbox to calibrate a stereo camera, and an MPC with 4 side-looking cameras. One limitation of their proposed approach was related to a requirement that at-least small portions of pattern should be visible by each camera.

2.3 Calibration (Articles I-III)

Camera calibration is an important way to estimate parameters of an interior orientation model. Camera calibration is an optimization process to estimate IOPs of a camera through a statistical model. Camera calibration could be performed by employing fixed objects such as a calibration plate

or a calibration room, or through a self-calibrating optimization process that also estimates the IOPs beside other parameters such as trajectory of the camera and position of object points.

A camera calibration room (or a calibration test field) is a suitable space that is designed to estimate intrinsic parameters of a camera or a class of cameras [10, 11, 41, 61]. A camera calibration room acts as a rigid body with known geometry that accelerates the process of camera calibration. A-priori known structure of a camera calibration room is usually estimated up to an arbitrary coordinate system and a scale. Usually scale bars with known lengths are employed to estimate the correct scale of a calibration room [36]. A calibration room is usually covered by coded targets, scale bars, fixed surveying structures, and suitable lighting sources. A coded target is an object with known geometry that is designed to be easily detectable inside images by a computer algorithm [14].

Many sensor modeling research works have been done based on employing a camera calibration room, e.g. [62] employed a camera calibration room with installed rectangular coded-targets to register bands of a tunable Fabry–Pérot interferometer (FPI) camera. A calibration field was developed by [63] to estimate parameters of a rotating head camera. The calibration room was named as "ETH Zurich Panoramic Test field". A calibration test field with 200 object points was developed in university of TU Dresden for panoramic cameras [15]. By using AURUCO coded targets [64], a calibration field was proposed by [65]. This calibration filed was later employed to calibrate catadioptric, multi-camera, and fisheye cameras. A geometric calibration model is proposed for MPCs e.g. by [66, 67]. The calibration model is required to achieve high-accuracies for tasks such as surveying and mapping, texturing, visualization, and panoramic compilation. A calibration scheme was proposed by [68] for catadioptric cameras.

To design a camera calibration room, usually parameters such as field of view of the target camera, and Ground Sampling Distance (GSD) should be taken into account. Physical constraints such as availability of limited locations for a calibration room usually plays a negative role. However, this limitation could be to-some-degree compensated by a proper consideration of the scale parameter of the printed targets. A suitable distribution of coded targets also improve the situation of improper locations.

A coded target is an important building block of a camera calibration room. A coded target is a printed pattern of geometric objects such as circles, ellipses, rectangles, lines that have distinguishable shapes or corners

that could be precisely located in images. Many coded-targets have been designed and employed by photogrammetric and computer-vision societies.

Image processing provides necessary tools to detect edges, lines and circles, holes and more complex shapes in images. Operators such as edge detectors, blob detectors, and Hough transform are commonly used in localizing geometric shapes [26]. Circles are one of the easiest geometric shapes to detect in images. A circle (or a blob) is usually projected into a rotated ellipse because of the perspective geometry of a camera. Image distortions (radial distortion and tangential distortion as well as non-perspectiveness) could produce even strange non-elliptical projections of circles in images which affect the process of automatic detection. In comparison to other geometry shapes, blobs are usually detectable inside an image with sub-pixel accuracy, since boundary points strengthen the process of localizing a center point. A non-symmetric combination of circles is a good nominate to create a rotation-invariant coded-target.

2.4 Image-based scene reconstruction (Article III)

Image-based scene reconstruction is a large category of algorithms that aim to estimate a) shape (structure) of an object from its projections in an image or a set of images, and (optionally), and b) status of images taken from the object. Shape from shading, shape from silhouettes, and structure from motion (sfm) are three sub classes of image-based scene reconstruction [69]. In this thesis we equivalently use "image-based scene reconstruction" and sfm, since the other two sub-categories are outside the scope.

Structure from motion includes topics such as sparse and dense image-matching (automatic tie-point extraction), and network estimation. Network estimation (or camera trajectory estimation) is a part of the sfm problem that concerns estimating status of images when a sufficient set of tie-points between them are given. This problem could be simplified in several ways, e.g. prior information about sensor parameters reduce the number of unknowns, or navigational observations from GNSS and IMU could be employed as initial estimations of image orientations and positions if estimation of lever-arm vector and boresight angles exist.

Automatic tie-point extraction is generally known as feature image matching problem that is a central component of many computer-vision algorithms such as panoramic generation, or sfm. Distinctive presentation of image points as key points has significantly helped these problems, e.g. [70] proposed a method to extract relatively invariant key points based on applying difference of gaussian operator (DoG) on an image to detect ex-

tremums. In this approach, a scale space was generated by applying DoG operator. Stable key points were localized as extremums of the scale space. Then a set of orientation were extracted and assigned to each localized key point to uniquely represent it. This method was named as Scale Invariant Feature Transform (SIFT). This method was initially observed as a successful image-registration method, although it had some drawbacks such as high computation time, low success rate when affinity changes were large, or a considerable amount of required random-access memory (RAM) to operate on a normal personal computer.

Many researchers followed this path to address those drawbacks, e.g. [71] proposed employing principle component analysis (PCA) on patches of normalized gradients to produce descriptors that were more compact and distinctive. An implementation on graphical processing unit was proposed by [72] to address timing aspects. This implementation was called SIFT-GPU. Affinity was addressed by adding six-parameters affine transform to the original implementation by [73] as ASIFT. Box filters was proposed by [74] as a relatively quicker replacement to DoG which was a slower operator. This method was called Speed Up Robust Feature (SURF). In SURF, key points were localized by geometric operators such corner, blob, or T-junction detectors. Second important aspect of the SURF that was improved over SIFT was related to a lower memory usage. They also proposed an "upright" version which was even quicker and more distinctive. A machine-learning based approach was proposed by [75] that employed a decision tree classifier. Their approach improved edge detection. This method was called Feature from Accelerated Test (FAST). Similar to its name, they demonstrated 30 times speed up of execution time in comparison to SIFT. In some cases, this method was demonstrated a better performance to the other methods. Calonder et al. [76] proposed a new method to enhance speed and memory requirements of SIFT. They proposed descriptor was based on binary strings which made it possible to quickly build and compared by a Hamming distance metric. This method was labeled as Binary Robust Independent Elementary Feature (BRIEF).

BRIEF and FAST was later combined by [77] as Oriented FAST and rotated BRIEF (ORB) method which was demonstrated to be two orders of magnitude quicker than SIFT. These methods was later bases on real experiments for many research works, e.g. [78] compared PCA-SIFT, SIFT, and SURF based on different scenarios. This research demonstrated that SIFT was slower than others while acting better in some cases. This led to a conclusion that choice of a suitable method depends on the application. In another article, performance of ORB, SURF, and SIFT was compared by

[79] against parameters such as shearing, fisheye lens distortion, and affine transformation. It was demonstrated that SIFT produced better matching, while ORB was the quickest approach.

Simultaneous localization and mapping (SLAM) concerns a problem that aim to simultaneously localize a sensor and generate a map of the surrounding environment of the sensor. For the image-based localization step, two general type of solutions are available: BBA, and direct-georeferencing methods. Many research works concentrated on image-based SLAM solutions. A review of recent state-of-art monocular SLAM methods was presented by [80]. Loop closure was employed in many image-based SLAM as the foundation, e.g. a loop-closure based monocular SLAM was proposed by [81] and compared to other techniques such as image-to-map and image-to-image. Challenges in real-time SLAM solution was discussed by [19], and consequently a method based on path initialization and loop detection was proposed. Employing a stack of key points based on BRIEF and FAST was proposed by [82] as a bag of words. A monocular SLAM method proposed by [30] based on feature tracking, loop-detection, and path optimization. This method was demonstrated as a noise-resistant method and called ORB-SLAM. Inertia data from an IMU [83] and GNSS data [18] was later fused with a monocular SLAM to robustly estimated the path. A sfm solution for oblique images proposed by [84]. They demonstrated a pair selection strategy to create a sparse structure of a network. They employed overlap analysis to optimally select stereo pairs. Accurate status of images was calculated by bundle adjustment.

Direct georeferencing is the process of determining the status of images by employing a) MMS status observations from GNSS and IMU, and b) calibration information regarding relative position and orientation of MMS components [85]. Direct georeferencing is especially important to estimate trajectory of a platform in real-time application such as UAV mapping. In direct georeferencing, positional data from a GNSS antenna and orientation data from an IMU sensor are converted into exterior orientation parameters of images in a global coordinate system. This process is in contrast to post processing determination of image status which usually involves employing ground control points. Direct georeferencing is usually less accurate than post-processing BBA with Ground Control Points (GCP). Two main requirements should be ensured prior to direct georeferencing: a) an accurate synchronization between camera shots, GNSS, and IMU should be established, and b) lever-arm vector and boresight angles should be accurately estimated. These parameters are called mounting parameters [17]. Direct georeferencing has been widely studied by many researchers, e.g. a

mobile mapping system with two cameras in horizontal and vertical positions were employed for a road survey task by [86]. They reported 20-40 cm accuracies of detected center of roads. Direct georeferencing in urban area was studied by [87] with a multi-camera mobile mapping system. They employed BBA with few check points to improve check-point residuals from 40 cm to approximately 4 cm.

2.5 Classification (Article IV)

Two deep-learning techniques have been employed in this thesis for a tree-species classification: convolutional neural network and multi-layered perceptron (MLP). Convolutional neural networks (CNN) are important classifier that consist of feature extraction and classification [88–90]. Unlike MLP, nodes of one layer of CNN are not forced to be connected to all nodes of a next layer, therefore the number of parameters could be significantly reduced in comparison to an MLP. Two operations of applying convolution filters and pooling operations [91] are included in the feature extraction phase. CNNs and MLPs are both similar in their application in estimating non-linear functions. CNNs are structurally more general than MLPs in a sense that they have the possibility to contain MLPs as their internal units. The major difference between the two classifiers comes from the convolutional and pooling layers that are designed to address parameter explosion. Unlike MLPs that contain dense structures, CNNs are supposed to have manageable sets of parameters. Convolution filters plays an important role here to gradually decrease the data dimension through a mechanism that produce features that best describe the objects of classes. Those weights are optimized during the training phase as the free network parameters.

Among several possible pooling operators that are usually considered for a CNN, the maximum pooling operator was employed in the architecture that is discussed in this thesis. This operator similar to the 3D convolutional operator works with a sliding window principle meaning that it moves with a window over the whole image [92]. The classification step in a CNN consists of applying an MLP or a convolutional later on the output of the last layer. Rectified linear unit function (ReLU) is a common type of activation function to avoid vanishing gradient problems [93] that happens in a gradient-based network optimization when gradients becomes too small to have a contribution on parameters [94]. A ReLU layer is usually installed before the last layer, since the non-linearity of the model is increased in this way. Batch normalization layers are used to speed up the

training phase by normalizing output of each internal working unit of a CNN. Average values and standard deviations are calculated in the batch normalization step to whiten the output signals [95].

Early successful attempts of employing CNNs were related to image classification problems, e.g. a CNN capable of detecting digits was proposed by [88] to classify individual letters (LeNet). This network was fed with 32×32 pixel normalized images of letters. A data preparation step was proposed by [90]. They applied elastic deformation to generate a better training set. They proposed a CNN to classify letters in MNIST dataset of handwriting digits [96]. Modern CNNs are considerably large, and able to classify millions of images to tens or even hundreds of classes [97], e.g. [98] proposed a CNN to classify 1.2 millions of images in 1000 different classes.

Tree-species classification is an important task in applications such as forest inventory. Layers of data such as hyper-spectral channels, point clouds, RGB images, as a single layer or in combinations are usually employed for the classification task. Usually, RGB cameras installed on a UAV provide the most affordable source of data for tree-species classification. On the contrast, hyper-spectral cameras provide more accurate data in a trade of with higher operational cost and more complex work-flow. Hyper spectral images have been successfully demonstrated in many classification problems [99–103].

A wide list of recent publication exists on this topic [21], e.g. two airborne lidar sensors was employed by [22] to automatically label Scot pine, Norway spruce, and birch. They employed intensity variables to achieve an accurate classification accuracy of 90%. Support Vector Machine (SVM) and Random Forest classifier were employed and compared by [104] to classify Norway spruce, Scots pine, scattered birch, and other broadleaves by two hyper spectral sensors. They achieved a high classification accuracy of 90% by employing one of the sensors. They reported no major difference between the classifiers. The birch class was demonstrated less distinguishable in comparison to other species with a classification accuracy of 61.5%. Full-waveform Lidar was employed by [105] for a tree-species classification task of 6 classes. Multiple additional data sources of lidar, HS, and color infrared (CIR) with an SVM classifier were later investigated by [106] to recognize beech, oak, pine, and spruce. Full-waveform was also investigated by [107] to classify coniferous and deciduous trees by a supervised and an unsupervised classification method. A high classification accuracy of 93%-95% was reported by them. A slight improvement of 2% in classification accuracy was reported for the supervised method. A crop/weed segmentation algorithm bases on a convolutional neural network were proposed

by [108] to classify multi-spectral orthomosaic. They reported 96% overall accuracy in pixel-wise segmentation of crop and weed. The locality of solutions is a problem in tree-species classification that was highlighted by [21]. Multi-layered perceptron, SVM, and RF was compared by [109] in a tree-species classification task by employing hyper-spectral image layers. The fully connected network was reported to produce better results (77% overall accuracy) in comparison to other classifiers. Scots pine, Norway spruce, and birch was classified by [110]. They employed an RF classifier on multispectral airborne laser scanning data. A high classification accuracy of 85.9% was reported in this work. A rotating multispectral sensor was employed by [111] in a UAV setting for a tree-species classification task. They employed RF classifier with overall accuracy of 78%. A 3D-CNN was employed by [99] to recognize tree-species in hyper-spectral and RGB data, An accurate classification accuracy (96.2%) was reported by them. A collection of 16 bands (shortwave infrared and visible - near infrared) from a satellite multispectral sensor was employed by [112] to recognize 8 tree species in tropical forest by SVM. On average, a classification accuracy between 60%-90% was reported for most classes.

Chapter 3

Materials and methods

In this chapter research objectives, materials and methods regarding Articles **I-IV** are presented. Table 1 lists research objectives of this thesis and the corresponding methods to address each objective. Figure 4 is a schematic view that demonstrates contributions made by each Article in the objectives on the thesis. Section 3.1 presents the calibration method for single and multi-cameras. Observational equations for GCP, GNSS and IMU readings, and scale bars are described, and a sparse BBA scheme is presented. The model is updated by an approach presented in Sections 3.2 and 3.3 for an MMS. Real-time aspects of the trajectory estimation problem are presented in Section 3.4. In Section 3.5 tree-species classification is presented that is followed by performance assessment method that is presented in Section 3.6. More details of the presented material could be found in Articles **I-IV**.

3.1 Calibrating a single or multi-camera (Articles I-III)

The calibration process that were discussed in Article **I** and **II** consists of two parts: a) calibration target and calibration room design and implementation (Article **I**), and b) calibration method design and implementation (Articles **I-III**).

3.1.1 Automatic coded-target detection (Article I)

Asymmetric coded-targets could be designed by at-least four strategies: 1) randomly assign a label, and try to find the correct correspondences, 2) design many coded-targets with different shapes, locations and sized of circles,

Table 1: Research objective and methods described in this thesis.

1. To propose a suitable coded-target with photogrammetric applications
Designing and investigating a code-target with photogrammetric applications with sub-pixel accuracy and ease of automatic detection. (Article **I**)
Developing a MATLAB software to read the coded target. (Article **I**)
Designing a camera calibration room with the proposed coded-target. (Article **I**)
2. To develop a multi-projective camera calibration model
Proposing a sensor model for multi-projective cameras with relative interior orientation constraints. (Article **I**)
Proposing a novel multi-projective camera calibration scheme based on employing the proposed camera calibration room. (Article **I**)
Improving the proposed model by adding GNSS and IMU calibration parameters. (Article **II**)
3. To propose a real-time trajectory computation scheme for an unmanned aerial vehicle
Proposing a multi-level pyramid matching scheme to decrease the matching time. (Article **III**)
Proposing a sub-window propagation and matching scheme to increase image-matching precision, and tie-point frequency. (Article **III**)
Overlap analysis and loop-closure. (Article **III**)
Observational equations for GNSS, IMU, initial values of cameras, GCPs. (Articles **II**, **III**)
Quality control and assessment. (Article **I-IV**)
4. To propose a new angular parametrization of BBA
Parametrization of BBA based on spherical angles. (Article **III**)
5.To propose a metric scheme for non-stitching panoramic generation
Adding metric value to a non-stitching panoramic image by proposing a corresponding and a contribution map. (Article **III**)
Fast panoramic compilation. (Article **III**)
Offering a pre-processing step for effective panoramic camera design. (Article **III**)
6. To investigate a tree-species classification problem
Proposing a convolutional neural network for the tree-species classification task. (Article **IV**)
Investigating the structure of the network, as-well-as a comparison with a multi-layered perceptron. (Article **IV**)
Investigating different aspects of feature selection. (Article **IV**)

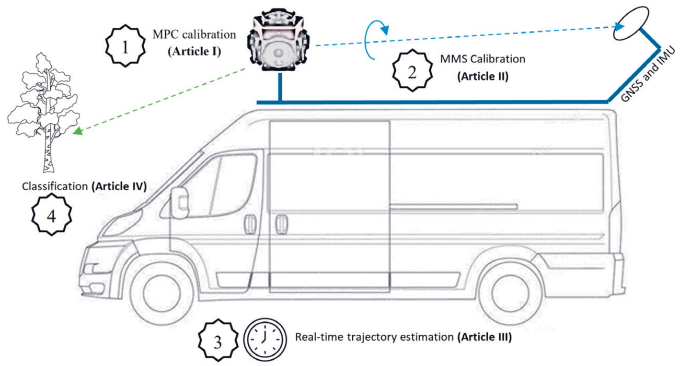

Figure 4: Schematic view of contributions of papers (I-IV) to the objectives of the thesis.

such that each shape will represent a label, 3) use different colors as the labeling feature, and 4) employing an embedded identifier inside a uniform coded-target body. The last solution is employed is this work. A vertically asymmetric code-target was designed that consists of 18 circles (Figure 5). A binary code was printed as 8 smaller circles. An on/off mechanism was considered by filling the id circles by a black color. Therefore, 256 distinguishable coded-targets was arranged. More labels were easily achievable by adding a second binary line. A more-precise target point was added for precise surveying instruments. The following criterions was considered for designing the coded target:

1. operational distance range of (30 cm - 4 m) with acceptable level of visibility in an image with resolution $>$ 1 mega pixel,

2. sub-pixel accuracy,

3. the relative ease of automatic detecting by considering a close cluster, and

4. embedded identifier and embedded scaling.

Figure 5 demonstrates an example of the coded target. Two algorithms were proposed to detect coded targets. The first algorithm was based on Maximally Stable Extremal Regions algorithm (MSER) [113] to initially estimate the location blobs, then refine those blobs into accurate blobs. MSER provides the location and diameter of all extreme blobs in an image. This algorithm was later replaced with a thresholding algorithm, since

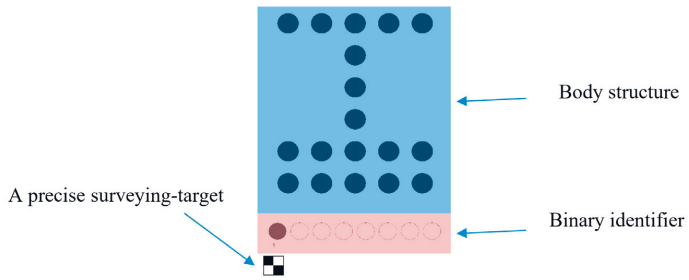

Figure 5: A vertically asymmetric coded-target that was designed and employed in this study.

thresholding was observed to be more effective in varying light condition. Simple thresholding was not suitable for this purpose since a single threshold was unable to appropriately act on a large rectangle in an image where light condition was affected by factors such as a shadow. A better thresholding algorithm was adaptive thresholding with first-order statistics that distinguishes black blobs from a background more efficiently with local thresholds; the following algorithm was finally employed to robustly detect coded targets:

1. employing adaptive thresholding with first-order statistics separates black blobs from background,
2. using connected-component labeling to find cluster of adjacent white pixels,
3. extracting boundary of each cluster as a list of image points, and fitting a rotated ellipse to each boundary,
4. combining ellipses to form local clusters, analyzing each cluster for a possible location of a coded target, and
5. localizing binary identifier of the extracted coded targets by a projective transformation and finally reading the id.

3.1.2 The calibration test field (Article I)

The calibration target that demonstrated in Subsection 3.1.1 was employed to cover a small space of $356 \times 519 \times 189$ cm (Figure 6). This space was labelled as "the FGI's camera calibration room". The coded targets were

printed on A4 papers and installed on all the surfaces of the calibration room, including walls, ceiling and floor. A minimum distance of 15-20 cm was maintained to separate coded targets from each other. Two projective cameras were employed to estimate the location of coded targets: 1) a Canon EOS 6D with 20 MPix sensor, Canon EF 24 mm f/2.8 IS USM lens, focal length = 20.65 mm ± 2 micron, and 2) a Samsung NX300, with 20 MPIX sensor, and ultra-wide-angle lens f/2.4. The coded targets printed on A4 papers were considered as rigid objects, however parameters such as change in humidity, temperature, and sun-light could deform the papers. Therefore, prior to each multi-camera calibration, the coded targets were measured by employing the Canon or the Samsung cameras. This ensured the best quality that was achievable by the proposed method.

The selected capturing strategy for constructing the room is such that the first pair is taken at a distance of approximately 2 meters with a base of 20-30 cm that is enough to have good overlap. Other images are consequently taken so that each photo could be connected to the previous image. A number of eleven datasets were captures by both cameras to estimate the calibration field. Approximately 100 images were taken in each dataset that cover the whole room.

The coplanarity method that was described in Sections 2.2 and 2.3 were then employed to initialize a network. Consequently, bundle adjustment (Section 2.7) was employed to optimize the network and propagate errors.

3.1.3 Cameras (Articles I,II)

Two projective cameras were employed to initiate the structure of the calibration room. The first projective camera was a Canon EOS 6D with Canon EF 24 mm f/2.8 IS USM lens. It has a 20 mpix sensor of size 5472×3648 pix. The lens had a focal length of 20.65 mm ± 2 μm. The second projective camera was a Samsung NX300 that was equipped with a Samsung ultra-wide-angle lens f/2.4. The sensor size was 20 mpix with image size of 5472×3648 pix. The lens had a focal length of $71.41^\circ \pm 1'$.

Two multi-projective cameras with 36 and 6 cameras were calibrated by the proposed scheme. The first camera was a Panono (Figure 7a) which is a panoramic ball of 36 cameras. Panono was introduced by a Berlin-based company, Professional360 GmbH, in 2017 mainly for high-resolution 360 imaging. Individual images are accessible by extracting raw image files. Individual projective sensors are 2064×1552 pixels. The final panoramic compilation is about 108 MPix.

The second camera was a Ladybug v.5 (Figure 7b) that is a multi-projective camera with 5 side-looking cameras and an upward-looking cam-

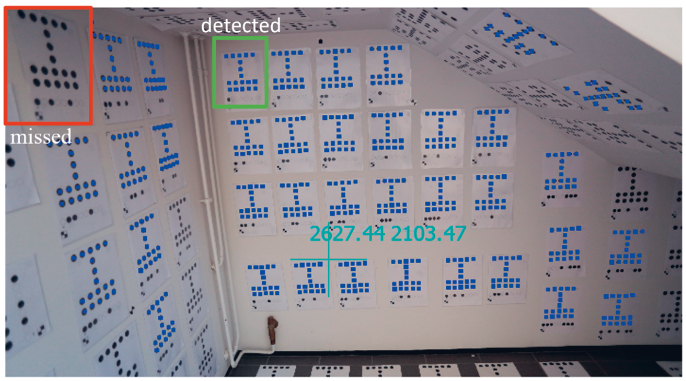

Figure 6: A sample photo of the FGI's camera calibration room. Detected targets are painted in blue. A sample missed target is demonstrated by a red rectangle.

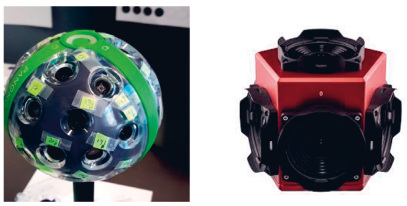

Figure 7: (a) A Panono camera with 3 sides. Each side contains 12 projective cameras. The cameras are individually labelled. (b) A Ladybug v.5 camera with 6 projective cameras.

era. The LadyBug camera was developed by FLIR. The area that was covered by the camera was 360° horizontally by 145° vertically. It had a sturdy design that made it suitable for mobile mapping applications. The camera was connected to a computer with a USB cable that carried the trigger signal. Individual projective sensors are 2464×2048 pixels. The camera had a focal length of 18 mm with a sensor size of $(35.8 \times 29.7$ mm). The approximated field of view (FOV) was 89.3°.

3.1.4 Initial estimation of the calibration room (Article I)

The projective cameras presented in Subsection 3.1.3 were employed to take several shots of the room. The images were captured such that an

overlap of at-least 50% were maintained. The image sequences were continuously covered all the coded targets. Image coordinates of coded targets were automatically detected by the proposed algorithm. A local coordinate system was selected by considering focal center of the first image as the center. The normalized 8-point algorithm [51] was employed to solve relative orientation between image pairs. The network then initialized by sequentially connecting all images. After adding any new image, a BBA step was performed by keeping the sensor parameters and location and orientation of previous network images fixed. This optimization step was efficient specially when initial parameters of the sensor were not accurate. Finally, BBA was solved in few iterations by keeping the sensor parameters fixed. Finally, a minimum-constraint self-calibrating sparse BBA was taken for few steps to estimate all the unknown parameters of the network and sensor parameters. The covariance of unknowns was finally analyzed to extract the confidence intervals of unknown parameters and positional uncertainties for camera locations and image points as 3D error ellipsoids.

There are many ways to initiate a network, each comes with certain cost and benefits. Depending on the application, a time-friendly low-accuracy, or a time-consuming high-accuracy method is selected. The following algorithm was used as a relatively fast network initialization:

Algorithm 1

1. Pick the first image
2. Put it as the network base.
3. Iterate over the image set to solve the relative orientation with the network base.
4. If no image is found, then select the next image as the network base and go to (3), otherwise continue.
5. Find all object points, and consider an arbitrary scale. Add the first pair to the network.
6. Pick the next image (I_i) in the list of unoriented images.
7. create a list of already oriented images with sufficient link to I_i and sort it based on the number of tie-points with I_i. Call it L_i.
8. Iterate over L_i and solve relative orientation until a successfully oriented pair is found.

9. Combine the solved pair with the current network by finding a 3D transformation.
10. Go to end if there is no more unoriented image, otherwise go to (6).
11. End

This algorithm is expected to work under the conditions that 1) no-outlier exists, and 2) the structure of a network is dense, meaning enough existence of high-frequency tie-points in images. This way, we can expect to solve 60-100 images per second. Two sources of errors could affect this process as labeling error, and image coordinate error. Labeling error occurs when two unrelated tie-points are falsely labeled to point to the same point. Image coordinate error occurs when image-location of tie-points are not accurate enough, for example, when the tie-points have been extracted from a top-level pyramid this error could occur. Random sample consensus (RANSAC) [114] could be employ to address the labeling error is in the relative orientation phase. Image coordinate errors could be detected by adding an optional bundle adjustment step when each pair is added. Then image points with relatively high standard deviation are filtered. These two steps could relatively increase the computational complexity of the original algorithm.

The simple network creation strategy is implemented as the following pseudocode:

Algorithm 1 **Pseudocode**

```
Given:
    Images - a list of unoriented images.
    Image points – a list of sufficient image tie-points.
Return:
    A network of oriented images.

i=0;
currentImage=Image{1}; currentImageIndex=1;
bFirstPair=false; List_of_oriented_images={};
List_of_unoriented_images=All_images;
List_of_failed_images={};
while !bFirstPair
    for i= currentImageIndex+1:numImages
      b :=SolveRelativeOrientation (currentImageIndex,i)
      if(b)
```

```
            bFirstPair := true
            Add this pair to the List_of_oriented_images
            Remove this pair from the List_of_unoriented_images
            break;
        end if
    end for
    currentImageIndex + = 1
end while
if(!bFirstPair)
    return "empty network"
end if
while List_of_unoriented_images is not empty
    BestImage=select the best image from List_of_unoriented_images
    If no BestImage exists
        break;
    end if
    if SolveRelative(BestImage,CurrentNetwork) is successful
        add it to CurrentNetwork
        remove it from List_of_unoriented_images
    end if
end while
```

3.1.5 Angular parametrization (Article III)

A rotation matrix could be expressed by different angular parameterization. In this context, the Euler angular framework is one of the most basic parametrizations. Euler angles are simple rotations around three main Cartesian axes. Each rotation is expressed by a 3×3 rotational matix. A complete 3D is possible by multiplying matrices, therefore, the order of applying rotations is important. A rotational matrix based on Euler angles is differentiable with respect to its parameters, therefore, it could be employed in a BBA. Euler angles have a significant drawback that occurs when the rotational planes become coplanar. This situation is called gimbal lock. One degree of freedom is lost when gimbal lock occurs, therefore the Jacobian matrix of the BBA becomes singular. One solution to avoid this singularity is to employ quaternions, or rotation axis-angle parametrization. In rotation axis-angle parametrization, every rotation is expressed as a right-hand-sided rotation around a unit rotation vector. An angular framework based on rotation angle-axis [115] was employed for parametrization of sparse BBA to address the gimbal lock singular situation. Rotation axis-angle parametrization was equivalently reformulated

by three rotations, two of which specify the rotation axis in a hypothetical unit sphere, and a right-hand-sided rotation around this vector. The benefit of the latter parametrization is that the length constraint is eliminated, therefore three independent parameters are remained. This parametrization typically inherits its parent property that is insensitivity to the gimbal lock situation.

The spherical rotation coordinate system is called spherical angles in this work, since a 3D rotation was expressed as spherical coordinated of a rotation vector and a right-hand sided momentum around it. The analytical Jacobian matrix of the new rotation matrix was straight-forward and easy to compute.

3.1.6 Observational equations (Articles I-III)

There are generally two sets of equations in a bundle adjustment: observational equations, and constraint equations. Observational equations connect a set of observable variables to unknown parameters. Observational equations act as a building block to propagate uncertainty from observations to unknowns. Constraint equations put a logical limitation on unknowns. Constraints could be simplified as observational equations by considering pseudo observations. In BBA, the collinearity equation (Equation 2) is employed to build image-point observational equations. These equations are rewritten for a multi-projective camera as

$$x_{ij}.(\mathbf{M}_3)_{t_1 j}(\mathbf{X}_i - (\mathbf{X}_0)_{t_1 j}) + (\mathbf{M}_1)_{t_1 j}(\mathbf{X}_i - (\mathbf{X}_0)_{t_1 j}) = 0, \tag{11}$$

and

$$y_{ij}.(\mathbf{M}_3)_{t_1 j}(\mathbf{X}_i - (\mathbf{X}_0)_{t_1 j}) + (\mathbf{M}_2)_{t_1 j}(\mathbf{X}_i - (\mathbf{X}_0)_{t_1 j}) = 0, \tag{12}$$

where x_{ij} and y_{ij} are scalar components of i^{th} image point in jth image, $(\mathbf{M}_i)_{t_1 j}$ is i^{th} row of image (j) at time (t_1), $(\mathbf{X}_0)_{t_1 j}$ is a 3×1 matrix of position of j^{th} image point at time (t_1), and $\mathbf{X}_i$ is a 3×1 matrix containing i^{th} object point.

GNSS and IMU observations are usable in a BBA if a hypothetical link between GNSS and IMU from one side and the camera from the other side is established. This link is expressed as a shift (lever-arm vector) and three spherical rotations (boresight angles) through the observational equations

$$\begin{pmatrix} \phi \\ \lambda \\ \kappa \end{pmatrix} - \mathbb{R}((\mathbf{R}_C)_I.(\mathbf{R}_{C_t})) = 0, \tag{13}$$

$$(\mathbf{X}_{0_{I_t}}) - (\mathbf{X}_{C_t}) + (\mathbf{R}_{C_t}).(\mathbf{R}_C)'_I.(\mathbf{X}_C)_I = 0, \tag{14}$$

where $\begin{pmatrix} \phi \\ \lambda \\ \kappa \end{pmatrix}$ are IMU spherical angles that at time (t) in the local coordinate system of IMU, $(\mathbf{R}_C)_I$ is the rotation of camera with respect to the IMU, and R_{C_t} is rotation of camera at time (t). $\mathbb{R}$ in Equation 13 is the function that maps a rotation matrix into spherical angles. Figure 8 demonstrates the relationship between GNSS, IMU and the camera. Equations 13 and 14 are designed in a simple manner that each equation contains only one observation. It therefore keeps the favorable simplicity in the Jacobian matrix of the observations.

A Ground Control Point (GCP) is a point with a measured location by a GNSS receiver, and good visibility in a network of images. Three observational equations are considered for a GCP:

$$\mathbf{X}_g(\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_m) - \mathbf{X}_{go} = 0, \tag{15}$$

where $\mathbf{X}_g$ is unknown position of the GCP that is a function of its corresponding image points $(\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_m)$, and $\mathbf{X}_{go}$ is the observed location of the GCP. It should be noted that $\mathbf{X}_{go}$ should be transferred to the local coordinate system of the BBA.

The observational equation of scale bars addresses one degree of uncertainty in the local-coordinate system of the BBA

$$\mathbf{L}_s(\mathbf{X}_i, \mathbf{X}_j) - \mathbf{L}_o = 0. \tag{16}$$

3.1.7 Initial values of parameters (Articles I-III)

Initial values observational equations are employed in two situations: a) a prior distribution of a parameter is given, and b) a network is not strong-enough to optimize a parameter. In the second case, an alternative solution is to lock (remove) the "weak" parameter from the optimization process, or add an observational equation with sufficient weight to solve the singularity. Initial values of the parameters and their standard deviations are either proved by system manufacturers, or estimated through a calibration process.

3.1.8 Bundle block adjustment (Articles I-III)

BBA is the phase when a cost function based on observation equations is considered and optimized. This cost function acts as a base to propagate uncertainties from observables to unknows. We may assume a system of

observational equations as

$$\begin{pmatrix} f_1(l_1, \{\mathbf{X}\}_1) \\ f_2(l_2, \{\mathbf{X}\}_2) \\ \vdots \\ f_m(l_m, \{\mathbf{X}\}_m) \end{pmatrix} = \mathbf{0}_{m\times 1}, \tag{17}$$

which could be rewritten in vector form as

$$\mathbf{f}_{m\times 1}(\mathbf{x}_{n\times 1}, \mathbf{l}_{m\times 1}) = \mathbf{0}_{m\times 1}. \tag{18}$$

The standard least-square frame-work involves optimizing a cost function based a weighted sum of square errors of observational Equation 16 as

$$argmin_{x,l}\, \mathbf{f}'.\Sigma_L^{-1}.\mathbf{f}. \tag{19}$$

A standard solution to Equation 19 is

$$\hat{\mathbf{X}} = -[A^{'}.(B.P^{-1}.B^{'})^{-1}.A]^{-1}.[A^{'}.(B.P^{-1}.B^{'})^{-1}.W], \tag{20}$$

$$A = \frac{d\mathbf{f}}{dx}, B = \frac{d\mathbf{f}}{dl}, P = \sigma_0^2.\Sigma_L^{+}, W = f(\mathbf{x}_0, \mathbf{l}), \tag{21}$$

$$C_X = -A^{'}.(B.P^{-1}.B^{'})^{-1}. \tag{22}$$

A selected ordering of unknown in Equation 18 has a considerable effect on structure, and sparsity of the A matrix. A good ordering usually helps to find a sparse structure for the A matrix. Usually an ordering of unknowns consists of sensor parameters, camera position and orientations, and object coordinates. Size of the A matrix depends on the number of observational equations which leads to allocating considerable amount of Random-Access Memory (RAM) in most real cases. To address this issue, indirect allocation of $N = A^{'}.(B.P^{-1}.B^{'})^{-1}.A$ was proposed which requires considerably less amounts of RAM, since N's dimensions are independent to the number of equations. Another important consideration is made by labeling object points as nuisance parameters by splitting N in a main/nuisance parts$N = \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}$. Since N_{22} is a large sub-matrix of N with 3×3 dense matrices on its main diagonal, an efficient sparse structure could be implemented for N (Figure 9).

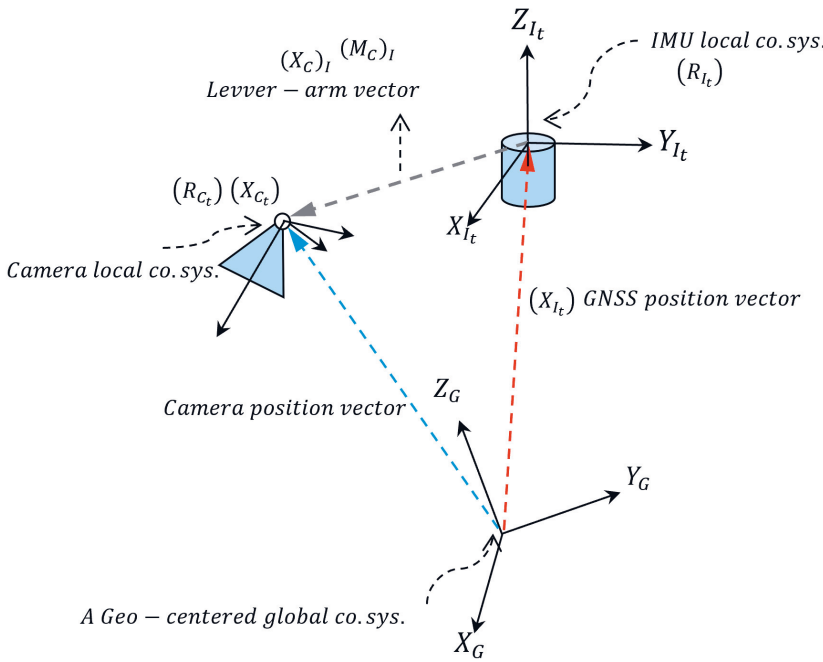

Figure 8: Relative position and orientation of a global navigation satellite system (GNSS) antena and IMU sensor with respect to the local coordinate system an image.

Figure 9: Structure of the N matrix.

3.1.9 Multi-camera calibration (Article I)

A multi projective-camera model is a model that collects individual came parameters and also relative orientation and position all individual cameras of a system with respect to an arbitrary local frame. This frame could be for example the local coordinate system of the first camera. Therefore, rest of the cameras are oriented and positioned with respect to this frame. If we assume a multi-projective camera with (n) cameras, then the MPC sensor model comprises of $10 \times n + 6 \times (n-1) + 1 = 16 \times n - 5$ parameres. The last parameter relates to a calibration scale.

Multi-projective camera calibration was enabled in this thesis by assuming a calibration body. The calibration room described in Subsection 3.1.2 was employed as the fixed object for the calibration. Initially, individual cameras of a selected multi-projective camera were resected to the point cloud. The output of this step was approximate location of image shots with respect to the calibration room. The image resection was possible by considering at-least 5 image points that their corresponding 3D location was known. In few occasions image resections failed due to weak distribution of image points. In next step, an approximated structure for the MPC was consequently retrieved by combining all individual projects. A minimum-constraint sparse BBA was finally performed by assuming the point cloud fixed. In this step, parameters such as location and orientation of image shots, and sensorial information of the multi-projective camera were set as free parameters. By combining all individual projects, image residuals considerably increased, since the relative orientation constraints were enforced through the approximated multi-projective sensor. The parameters that were optimized on this step were: 1) MPC sensor parameters, and 2) location and orientation of multi-images. Few steps of optimization process (an over-constraint self-calibrating sparse BBA) was enough for the problem to be converged into an acceptable solution. In the final solution, quality control parameters such as image residuals, or uncertainties regarding sensor parameters was considered to ensure the success of the optimization.

To calibrate a Panono camera, two datasets of 34 and 84 multi-image shots with 1224 and 3024 images respectively were captured. In the first dataset, three height-levels were maintained by a tripod for visual quality control purposes. In the second dataset, two height-levels were considered. In the lower high-level of the second dataset, the camera was installed on a linear slider such that every few images were taken almost on a line. This linearity of shots was later employed for visual quality control.

To calibrate the Ladybug camera, two datasets of 26 and 53 multi-image shots were captured. Physical constraints were applied to both datasets. In the first dataset, a horizontal linear slide was employed, therefore, image shots were almost on the same high level. Two height levels were enforced on the second dataset by employing a tripod.

3.1.10 Non-stitching panoramic generation (Article II)

A panorama could be generated by image stitching techniques [116–118], however, the final panoramic compilation will only have visualization and artistic values. The metric information is lost during the stitching operation.

A non-stitching panoramic compilation is possible by employing IOPs and ROPs of a calibrated multi-projective camera. In this scheme, footprints of individual projective cameras are calculated on the surface of the panorama. This footprint is called the contribution map of individual cameras. The contribution map helps to understand the amount of surface that is covered by each projective camera. Therefore, it is valuable tool to analyze structure of a multi-camera in design phase. Since cameras will have overlap on edges, a criterion should be considered to create a unique correspondence map for the final compilation. This criterion could be for example minimum incident angle or pixel size.

The correspondence map provides links between the final panorama and individual cameras at a pixel or sub pixel level. The correspondence map also facilitates the process of non-stitching panoramic compilation by providing a fast look up table. For any pixel on the compiled panorama, a vector of 3 numbers are saved as the components of the correspondence map. This vector consists of camera identifier, and pixel coordinates. Compiling a correspondence map of 7200×3600 pixel needs approximately 7 minutes of computation time on a single-thread procedure; then the map helps to quickly compile a new panorama in 2.1 seconds.

The compiled panorama by this method are geometrically accurate in comparison to stitched panoramas that are non-metric, meaning that each pixel could accurately be transformed to its corresponding image location. A discontinuity is expected over the edges of the correspondence map where the label is changing as a systematic error [119]. The discontinuity is larger for objects that are closer to the camera. This systematic error is addressed by the proposed correspondence map. A color blending should also be considered to soften the color steps on edges. This post-processing will enhance the final quality of the non-stitching panoramas.

3.2 Mobile mapping system (Articles II, III)

The FGI mobile mapping system (Figure 10) was employed to show the efficiency of the upgraded BBA (Article **II**). This system consisted of a Ladybug v.5 multi-projective camera that was firmly mounted on a truss structure. A GNSS receiver NovAtel PwrPak7, with A NovAtel IMU-ISA-100C, and a laser scanner were mounted on the same truss structure on top of a Skoda car. This system was also employed to show the efficiency of the calibrated parameters in direct georeferencing and sfm. More details of the mobile mapping system could be found in (Article **II**). A total number of 8424 multi images were taken by the LadyBug from a small region in Inkoo, Finland. The mobile mapping system calibration was done in an outdoor configuration.

3.3 System calibration for direct georeferencing (Article II

GNSS and IMU calibration relates to the problem of updating the sparse BBA solution for multi-projective cameras in order to accept six additional parameters that are related to relative location and orientation of the multi-projective camera with respect to the GNSS and IMU local coordinate systems. In this thesis we call this step MMS calibration.

The proposed solution was based on two separate steps. In the first step the multi-projective camera was calibrated by the proposed calibration scheme that was discussed in Subsection 3.1.5. In the second step (MMS calibration), the calibrated multi-projective sensor was assumed fix with only one degree of freedom (unknown scale). Considering the scale was due to the fact that scale bars in the calibration room could be considered as "inaccurate or missing". Therefore, the calibrated sensor resulted from "MPC calibration" step had one degree of uncertainty that was addressed by the scale parameter. Then, few consecutive multi-images (10-30) were selected in a suitable location with accurate GNSS and IMU readings.

The path was initially estimated by orienting multi-projective images. Few ground control points with good visibility were chosen. The global position of the GCPs were known. Local position of GCPs were calculated by image intersection from sfm. The two coordinate systems were consequently oriented with respect to each other by employing GCPs through a 7-parameter transformation (3 displacements, 3 orientations, and 1 scale). The estimated transformation was employed to estimate approximate global position and orientation of multi-projective images. Relative position and

orientation of the multi-projective camera was finally estimated by employing these global coordinates. Finally, an updated over-constraint sparse BBA by considering additional parameters of lever-arm vector and boresight angles, and additional observations of GNSS and IMU readings was employed to estimate MMS parameters. This process resulted to the calibrated sensor parameters of the MMS.

3.4 Real-time simultaneous localization and mapping (Article III)

In this section, hardware and algorithms involved in monocular SLAM are described.

3.4.1 Systems and datasets

The study regarding monocular SLAM was carried out using the FGI's Tarot 960 hexacopter UAV that contained a Samsung NX500 camera with Samsung 16 mm f/2.4 lens was employed to capture two datasets in Article **III**. This UAV was equipped with a Raspberry PI computer, GNSS-receiver NVS NV08C-CSM[1] and Vectornav VN-200 IMU.

A quadcopter UAV with Gryphon Dynamics was employed for to capture two calibration datasets in Article **III**. This UAV was equipped with a positioning system consisting a Trimble's APX-15 EI UAV GNSS-Inertial OEM System. The positioning system was comprised a multiband GNSS and Internal onboard IMU and a Harxon HX-CHX600A Antenna (Figure 11).

3.4.2 Multi-level matching

Multi-level matching was proposed in Article **III** as a pyramid-based image-matching scheme that decreased the computation time of image matching in a sequential trajectory-estimation problem. Multi-level matching was based on reducing the size of input images to find approximate location of sub-regions of a matched image on a reference image. A high-resolution sub-window matching was followed afterward to ensure the quality of matching. When a list of images was sequentially matched, a history of locations of matched sub-regions were kept in memory. Those regions were sequentially propagated along the camera path to ensure acquiring high-frequency tie points.

[1]NVS Navigation Technologies Ltd., Montlingen, Switzerland

Figure 10: The FGI's mobile-mapping system.

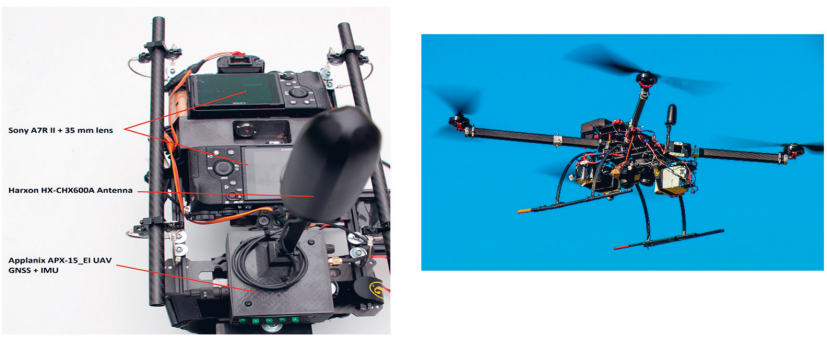

Figure 11: Unmanned aerial vehicle (UAV) used in Article **III** for aerial calibration and tests.

The scale of low-resolution pyramid of original images was determined by the time limit that was considered for this step. In our case, a scale of 0.25 was selected to ensure that the initial low-resolution matching was executed in an acceptable timing frame. Size of sub-windows was the second adjustable parameter of the proposed approach. In this case, a window of 200×200 pixels were selected. An important step to speed up the process was to employ Graphical Processing Unit (GPU) to down-scale input images by employing a suitable down-sampling approach (e.g. bicubic) to preserve geometric attributes of input images.

The matching kernel of the multi-level matching approach was optional. The cosine distance measure for descriptors were employed here to find correct correspondences. Distinctiveness of matches was improved by employing a ratio-check. In this filtering process, for any key point on a "left image" cosine distance ratios of the two closest counterparts on a "right image" were compared with a threshold. Only key points with a ratio less than the threshold were marked as "matched". By downscaling images for the matching phase, the execution time for pair orientation also decreased which led to faster network initialization. The estimated status of images from this step was later enhanced through BBA. Despite executing the ratio-check filtering process, many outliers were still remained in most images. To filter out the remaining false matches, a RANSAC model based on a projective transformation was executed. To propagate "sub-windows" from one image to another, a projective transformation kernel was used to localize the position on other images. These low-level rectangles were then used for a limited high-resolution matching. The whole process considerably decreased the matching time. In this process, accessing to a list of highly distinctive key points was an important factor that affect the frequency of resulted image tie-points. After this step, a new image was analyzed to locate parts of images that was uncovered. Then, new rectangular patches were added to uncovered placed. Loop detection was enabled by photogrammetric overlap analysis after 10 or 20 images. This step was necessary to strengthen the quality of the network. To enable loop detection, a search was performed for the tail image to find the furthest image that had an acceptable overlap; then image tie-points were found for this pair by the proposed multi-level matching. The structure of the sequential approach was improved by this modification.

3.4.3 Monocular SLAM

The proposed multi-level matching strategy described in Subsection 3.4.2 creates a foundation to automatically connect aerial images. The copla-

narity condition presented in Subsection 2.5.2 could be employed to form stereo pairs of captured images. These pairs could be combined together to estimate the camera trajectory. In general, at-least three strategies are possible to combine stereo pairs: 1) dense reconstruction, 2) sequential reconstruction, and 3) customized loop-based approach. In the dense approach, an initial network is formed from the first stereo pair. Then any new image is matched and oriented with respect to all images of the network. The network gradually grows as new images are combined into its structure. This method of network creation leads to a dense structure at the maximum computational cost. A sequential approach assumes the first pair as the initial network; any new image is matched and oriented only to the previous image of the network. The final network's structure will be weak; however, this method is efficient in terms of the computational resources that it needs. Since the positional errors are accumulating along the estimated path, large deformations are expected in a sequential network creation scheme. A modified approach tries to take advantages of benefits of a sequential approach, while addressing its shortcomings by proposing strategies such as loop-detection.

3.5 Tree-species classification (Article IV)

To enable tree classification application (Article **III**), a dataset of 11 plots from Vesijako area were captured by a hexacopter UAV that was equipped with a hyper-spectral camera based on tunable Fabry-Pérot interferometer [120–122]. The hyper-spectral images were 1024×768 pixels with a pixel size of 11 μm. The focal length of the FPI camera was 10.9 mm with a field-of-view (FOV) of $\pm$ 18° and $\pm$ 27° in straight and cross flight directions respectively. In total, 33 spectral bands of 11-31 nm were employed in the classification. An RGB camera was integrated inside the UAV to capture high-resolution images. The RGB camera was a Samsung NX1000 with a 20.3 megapixels sensor that was equipped with a 16 mm lens. The flight altitude was 87.5 m from the ground that resulted to an average ground sampling distance (GSD) of 8.6 cm for the FPI camera. The average GDS for the RGB camera was 2.3 cm. A dense point cloud was extracted from images with 5-cm point interval. The canopy height models (CHM) were considered from the point clouds by employing a digital terrain model (DTM). The variation in hyper-spectral images were normalized by radiometric processing resulted to uniform reflectance mosaics. More details about radiometric adjustment could be seen in Article **III**. RGB mosaics were generated with GSD of 5 cm.

A number of 3896 trees were selected for the labeling process. The labelled trees were among the most common tree types in Finland. Three major types of silver birth, Norway spruce, Scots pine were selected for the classification. Among all records, 80% were randomly selected for training and 20% selected for testing. The input dataset was imbalanced meaning the number of members of each class was different from others. In total, 2001 pines, 626 spruces, and 466 birches were selected for the classification.

A square of 25×25 pixel was considered around each tree in the RGB mosaic. The same rectangular area was considered for CHM, and hyper-spectral bands. The result turned into a hyper cube of data for each tree (3 RGB layers + 1 CHM layer + 33 hyper-spectral layers).

Two classification aspects went under investigation of the Article **IV**: 1) the most efficient set of features among RGB, CHM, and HS, and 2) the applicability and benefits of employing a 3D convolutional neural network in comparison to a multi-layered perceptron. To address the first aspect, different combinations of input features were individually labelled by a 3D-CNN classifier of similar structure. The blue channel of RGB was specially considered to investigate its importance in the classification. To address the second aspect, two separate classification were by a 3D-CNN and a multi-layered perceptron were compared.

MLPs that were employed to classify the data had a simple structure of one input layer of size $25 \times 25 \times n$ ($n : 1 = 37$), a hidden layer with 10 nodes and an output layer of 3 nodes. To avoid overfitting of networks in the training phase, 10% of its training set were randomly labeled as the cross-validation set. Scaled conjugate gradient descent was employed to optimize the networks. Four stopping creation were set to stop the training: 1) reaching maximum iteration, 2) reaching maximum training time, 3) reaching a small limit in the cost function, or 4) few consecutive decrease in classification accuracy of the cross-validation set.

A simple structure was considered to build a 3D-CNN for the classification of tree types (Figure 12). The proposed 3D-CNN was designed to gradually decrease the data dimension and ultimately perform the classification task. It consisted of convolutional layers, batch-normalization layers, max-pool layers, a ReLU layer, and a Softmax layer (Figure 12). More details about the sizes of inputs and outputs of each individual layer could be found in Article **IV**. The installed ReLu layer before the last convolutional layer was important to increase the non-linearity of the proposed model. The Soft max layer finally converts the result of process into a probability distribution that represents the probabilities that an object belongs to each class.

3.6 Performance assessment methods (Articles I-IV)

Elements of the proposed photogrammetric scheme were independently assessed by appropriate metrics which were inconsistent in nature. Several quality assessments methods were employed to express the success or fail of proposed approach, or to demonstrate the achieved accuracies, and precisions. Those metrics were not necessarily global, since their formulation was based on the situation of the specific problems.

Covariance matrices were analyzed and employed to assess quality of the output parameters in a BBA. Diagonal elements of covariance matrices were employed to find confidence intervals for individual parameters. To express uncertainties for image positions and 3D object points, 3×3 submatrices were extracted from covariance matrices and converted into 3D error ellipsoids. This interpretation was employed as spatial confidence regions. (Articles **I-III**)

Two different approaches were employed to assess geometric accuracy of a multi-projective cameras in localizing 3D points in a calibration room. The fist approach was based on excluding a set of 3D object points from the BBA as reference check points. Therefore, the result of BBA became independent of the check points. Next, the optimized exterior orientation and interior parameters orientation parameters along with the estimation location and orientation of multi-image shots were employed to estimate 3D location of check points. Finally, the computed values were compared with the estimated counterparts. The second approach was based on estimating scale bars from image datasets and comparing the estimations with reference values (Article **II**).

The accuracy of the approach described in Section 3.3 for MMS calibration and georeferencing was assessed by considering few check sites with 5-20 3D check points that their global positions were measured. For each check site, additional observations of GNSS and IMU and MMS calibration data were converted into status of multi-images. Then, check points were manually located inside multi-images, therefore, an independent global position of check points were calculated by image intersection. Those two sets of coordinates were compared for quality control (Article **II**).

To assess image matching algorithm that was described in Section 3.4 (Article **III**), time factor and inlier percentages were studied. Other image-matching methods such as ORB, SURF, and SIFT were compared with the proposed approach in various configurations. Quality of image-point intersections was assessed by image-point residuals. RMSE of image-point

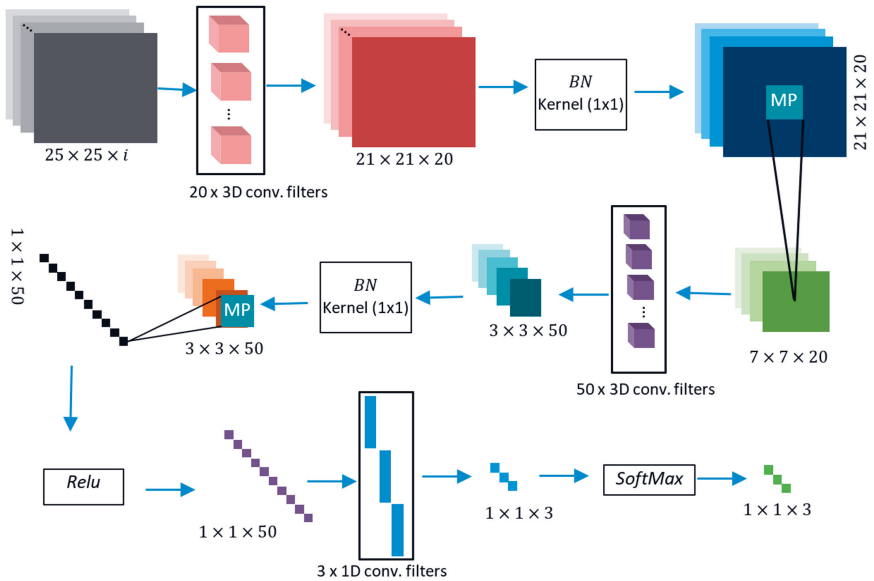

Figure 12: Structure of the proposed convolutional neural networks (CNN) for an input of i data layers of size 25×25.

residuals were employed to evaluate the correctness of connecting a new image a network. Diagonal elements of the covariance matrix of unknowns were employed for quality control of optimized parameters. 3D check points were employed to assess the correctness of the proposed sparse BBA. Differences between GNSS/IMU readings and projected output of the sparse BBA was employed to assess the correctness of the estimated lever-arm vector and boresight angles (Article **III**).

The optimization process was controlled by employing a cross-validation dataset in Article **IV**. Area under the curve (AUC) of receiver operator curves (ROC) were calculated for each proposed classifier in Article **IV**. Overall classification accuracy, user accuracy, and producer accuracy were also employed to assess and demonstrate classification quality of the proposed classifiers in Article **IV**. Confusion matrices were calculated and plotted for each classifier.

Chapter 4

Results

In this chapter, some experimental results of the proposed approached are presented. The order of this presentation is consistent with the order of the articles. More detailed numerical results could be found in the original articles.

4.1 Camera Calibration (Articles I-II)

Two single projective cameras (Canon EOS 6D and a Samsung NX300) were employed to take continuous shots of the calibration room. The algorithm described in Section 3.1 was employed to detect coded targets inside images with sub-pixel accuracy (0.1-0.2 pixels). Embedded identifiers were then automatically read by the algorithm.

In each image, approximately 80% of coded targets were successfully detected. Approximately, a number of 15-80 coded target were visible in most images of both camera datasets. Initial structure of the room was successfully estimated by the coplanarity equation for each dataset.

A self-calibrating BBA was then employed by setting all interior and exterior orientation parameters, camera location and orientations, and object point locations free. A minimum constraint BBA was enabled by considering the first image of a network as the local coordinate system, and a coordinate system with axis parallel to the first image's frame. To address the scale, largest component of the furthest image to the first images were set fixed.

Few optimization steps by the Newton method was enough for the convergence of the algorithm into a stable optimum solution. Covariance of unknowns were estimated as the covariance matrix. Most of the parameters were considered meaningful since their value was bigger than their

corresponding standard deviation. Among all the interior orientation parameters, scale and shear were less significant for both cameras, which implied that the cameras were not significantly affected by these two parameters. For correlated parameters such as components of image positions, sub matrices of the covariance matrix were extracted and analyzed. Positional uncertainties were demonstrated as rotated 3D error ellipsoid that were extracted from covariance matrices of 3×3. These error ellipses were employed for quality control and outlier filtering.

Image residuals were employed as an indicator that shows the goodness of fit of the model. On average 0.14 pixel image residual was obtained for Canon EOS 6D. This number was 0.15 pixel for Samsung NX300. 3D Error ellipsoids of object points indicated that higher accuracies were achieved whenever the intersection geometry was stronger (Figure 13). Error ellipsoids were larger for the cases where the number of intersections were low. The embedded scale inside the coded target was helpful in interpreting and understanding positional covariance values. Most of the 3D points had standard deviation values of approximately 60 μm in all directions, which demonstrated accurate positioning of the coded targets. Both cameras were successful in target accuracies, however, positional error ellipsoids of the Canon camera were smaller than the Samsung; Root mean square error (RMSE) of the differences between ideal values of scale bars and the computed values was 0.2 mm. The maximum difference was 0.6 mm.

In total 11 dataset were captured by two cameras. Different datasets were co-registered to investigate the variability of the room between measurement sessions. In X, Y, and Z directions 0.52, 0.51, ad 0.67 mm RMSE with 98% confidence interval was observed.

A Panono camera (Figure 14a) was calibrated by the methodology stated in Subsection 3.1.6 inside the camera calibration room. To enable this calibration, the Canon camera was employed to estimate the camera calibration room's structure prior to the Panono calibration. All coded targets in multi-images of the Panono datasets were automatically extracted by the proposed method stated in Subsection 3.1.1. Individual cameras of the Panono camera were resected to the point cloud. Initial structure of the Panono was retrieved by combining individual projects through averaging of relative camera status with respect to the first camera. This initial structure was noisy and not resembling to the physical reality of the camera. A BBA was then employed to enhance the initial camera structure by enforcing a multi-projective sensor model. The structure of the room was considered fixed during the BBA, since the quality of estimated room was high. Few steps of BBA were sufficient to converge into an ac-

ceptable solution. Approximately 8-12 gigabytes of RAM were allocated for the optimization. This number was later significantly reduced to less than 2 gigabytes by employing a sparse BBA model. The optimized sensor was very similar to visual appearance of the Panono (Figure 14b). All three surfaces of the multi-projective camera with 12 cameras were visible in the optimized sensor. The average RMSEs of the Panono was about 0.5 pixel which was larger than Canon or Samsung datasets. The physical constraints such as same height level or linearity of shots were visually controlled on the output of the BBA. The visual check confirmed the correctness of the proposed model. Significance of sensor parameters were analyzed by blocking effect of an underlying interior orientation parameter or a group of parameters.

Uneven distribution of coded targets inside the calibration room caused a variation in the standard deviation values of parameters of different cameras. The cameras that were located in the upper and lower part of the Panono had bigger error ellipses than the side cameras. This situation could be improved in future by considering a more suitable space such as a cubic space with suitable distances of walls, or by systematically taking more shots from different angles by employing a specialized tripod with arm. Positional standard deviation values of most cameras were approximately 1mm for X, Y, and Z coordinates.

With a similar methodology, the LadyBug camera (Figure 14c) was calibrated using the camera calibration room. Sub-pixel image residuals of 0.4 pixel with standard deviation value of 0.45 pixel were achieved for the LadyBug camera. The calibrated sensor was precisely resembling to the physical reality of the sensor (Figure 14d). The radial distortions of individual cameras were large, in comparison to the SCs and the Panono. Distribution of coded targets had similar effect on the LadyBug camera such as standard deviation values of interior orientation parameters of the side-looking cameras were lower than those of the top camera.

4.2 Mobile-mapping system calibration (Article II)

The method described in Section 3.2 was employed to calibrate FGI's mobile mapping system. Local coordinates of GCPs were calculated by intersection of at-least 6 rays. A mean residual of 9 cm for GCPs was resulted from MMS calibration in the outdoor calibration site. Average difference of 0.2° was obtained between the output of modified sparse BBA and IMU readings which was the expected accuracy based on the specification of the

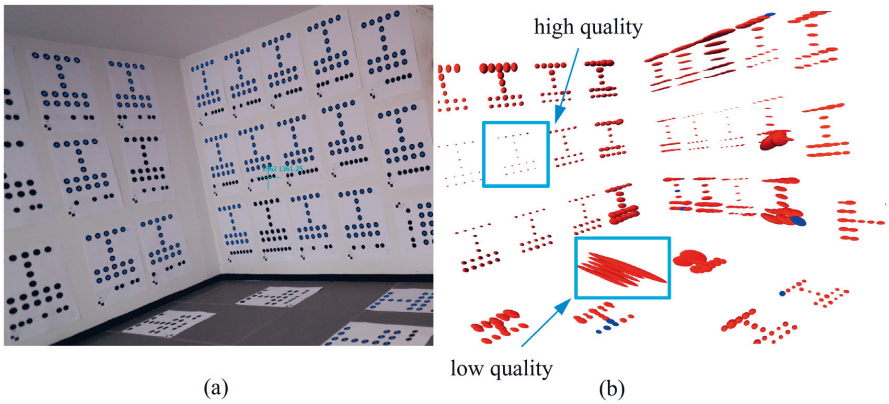

Figure 13: The FGI's camera field (a), and estimated 3D error ellipsoids with 100x magnification (b).

IMU. Observed difference between GNSS and BBA output was on average 11.3, 2.8, 0.6 cm in X, Y, and Z coordinates, respectively.

Direct georeferencing were employed to estimate global status of multi-images on check sites. To enable direct georeferencing, the calibrated MPC sensor parameters, the calibrated MMS parameters, and GNSS and IMU observations were contributed in the model. On cross-check sites 1 and 2, a mean difference of 5.6 cm and 1.4 cm between observed positions of check points and intersected positions with RMSE of 0.69 cm and 0.18 cm were obtained.

Relationship between quality of intersection geometry and image-based 3D object point positioning were also studied in Article **II**. As expected, object points with incident angles smaller than 20 degrees were positioned with a lower accuracy. This situation was significantly improved for points with incident angles over 20 degrees. The most stable situation was expected on $\frac{\pi}{2}$.

4.3 Non-stitching panoramic compilation (Article II)

The method described in Subsection 3.1.6 was employed to compile non-stitching panoramas from the images captured by the Panono and the LadyBug cameras. Footprints of projective cameras of the Panono camera

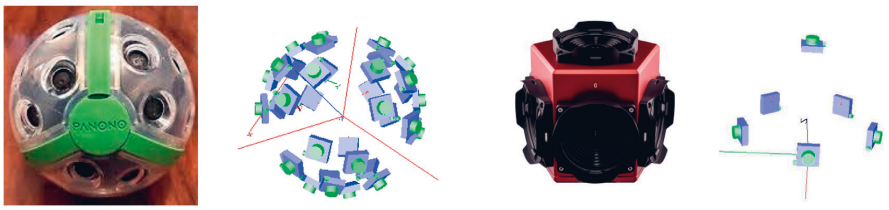

Figure 14: (a) Panono camera, (b) calibrated Panono sensor, (c) LadyBug v.5 camera, and (d) calibrated LadyBug sensor.

on a non-stitching panorama (Panono contribution map) is plotted in Figure 15. This figure highlights that a contribution map is essential in designing a multi-projective camera. Some of the cameras of the Panono were observed as redundant, since their footprint was already covered by other cameras; because of the overlaps between cameras, some pixels of the contribution map belong to more than one camera. This situation was converted into a map of one-to-one correspondence (correspondence map) by employing a selection criterion. Figure 16 demonstrates the correspondence map for the Panono camera by employing the minimum incident-angle criterion. It is obvious from this map that a better distribution of cameras could be considered in design phase of the Panono to better cover the top and bottom areas, while some side-looking cameras were redundant.

A sample of non-stitching panoramic compilation for the Ladybug camera is demonstrated in Figure 17. Original images are plotted on the first row of this figure. On the second row, undistorted images are plotted, and finally a non-stitching panorama is depicted in the fourth row of this figure.

4.4 Real-time SLAM (Article III)

The proposed multi-level image-matching described in Subsection 3.4.2 was implemented to find high-quality image tie-points for aerial images. The proposed matching strategy was compared to image matching methods e.g. ORB, SURF, and SIFT. In general, SIFT key points had the best distribution in images, while it needed considerable amount of RAM, and computational resources. ORB, on the other hand demonstrated the best time and accuracy in comparison to SIFT and SURF. The proposed approach demonstrated a much better performance in comparison to other stated methods. The proposed sub-window propagation scheme produced

Figure 15: Contribution map of the Panono camera.

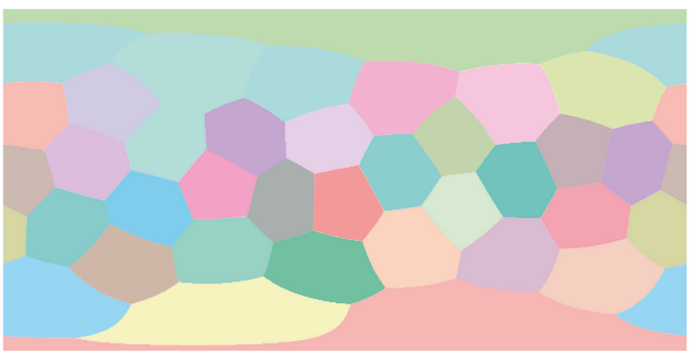

Figure 16: Correspondence map of the Panono multi-projective camera.

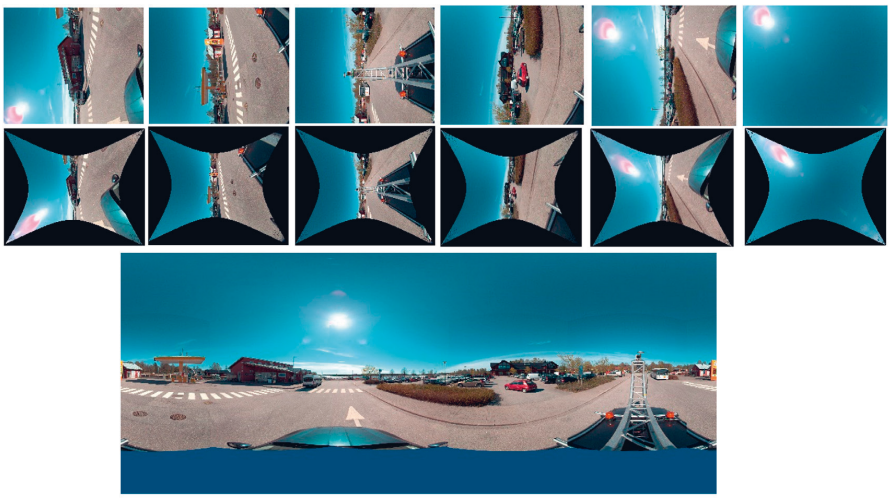

Figure 17: A sample of six images from side and top cameras of Lady-Bug (first row), undistorted images (second row), a metric non-stitching panoramic compilation of the first row (third row).

high-frequency tie-points that were extracted from different locations in an image, therefore, it had a considerable contribution in strengthening a network structure. The overall image-matching cost decreased to 3 seconds which was at the same level of image acquisition. Therefore, the proposed approach was a suitable image matching strategy for a real-time trajectory estimation.

The loop detection strategy was successfully implemented and tested. The overall structure of the network was improved by the proposed loop-detection algorithm. The overall cost of matching was not significantly increased, since only few more pairs were added to the network.

The gimbal lock singular case was addressed appropriately by employing the angular framework stated in Subsection 3.1.5. Derivatives of the rotational matrix with respect to the spherical angles were manually formulated. Two calibration datasets were individually analyzed. They successfully resulted the MMS calibration parameters. These two datasets were combined for a visual sanity check of the sensor configuration. The resulted multi-sensor configuration was similar to the reality of the stereo-camera sensor.

The algorithm was assessed in a post-processing step by check points. Few centime residuals were observed in the adjusted model. Lever-arm vec-

tor and boresight angles applied to GNSS and IMU readings to estimate status of images independent to the BBA output. Then, the differences between the calculated values and outputs of the sparse BBA were ploted. As expected, the differences were at the accuracy level of the GNSS and IMU observations. Approximately, 0.2 degrees differences for image rotations were observed. On average 2-4 cm positional differences were observed. In overall, the algorithm was successfully demonstrated as a monocular SLAM.

4.5 Tree type classification (Article IV)

The proposed fully-connected network and 3D-CNN were successful in classifying different combinations of input data. The 3D-CNN model with all data layers had a parameter file that was 169 megabytes. It took approximately 2.5 hours to train the proposed 3D-CNN with a core i7-6820HQ processor with 32 GB of RAM and a Nvidia Quadro M1000M graphical processing unit (GPU) with 2GB of DDR5 RAM. MLP training was considerably quicker. It took 16 seconds to train the MLP. The classification accuracy obtained from the MLP was overall 94.5% in recognizing tree species. Area under the curve (AUC) of receiver operator curve (ROC) for the MLP was 0.9961 for pine, 0.9590 for spruce, and 0.9685 for birch. Classification accuracies on the test dataset were 98.4% for pine, 82.2% for spruce and 95.6% for birch. The biggest drop in classification accuracy between the training and the test set was observed for spruce with 8.5% difference.

An improvement in classification accuracy was obtained for the proposed 3D-CNN with overall accuracy of 97.6% by employing all data layers (hyper spectral, RGB layer, and canopy-high model), therefore the 3D-CNN was superior to the MLP. Different 3D-CNN models were trained for all layers, as-well-as different combinations. Area under the curve (AUC) of receiver operator curve (ROC) for the 3D-CNN with all layers was 0.9999 for pine, 0.9941 for spruce, and 0.9956 for birch. These indicators demonstrate that the proposed 3D-CNN acted better that the MLP. The classification accuracies of the model that was trained with all layers and the model that was trained with hyper-spectral and RGB data was almost similar. Approximately 0.1% difference in overall classification accuracy was observed between these two models. This result confirmed that the canopy-height model was an irrelevant data layer in this classification task. Therefore, the best result was obtained when the hyper spectral and RGB layers were fed into the model. The model with RGB layers resulted to very good results

in comparison to the model with hyper spectral and RGB. A small difference of 0.4% was observed between these two models, which demonstrated the applicability of RGB cameras in a tree-species classification task. The biggest difference was related to the user accuracy in spruce which was 2.8% better in the model with hyper spectral and RGB data layers. The model with hyper spectral data layers was slightly comparable to the model with RGB layers. A small difference of 0.1% were observed between the two models. The best classification accuracy for pine was obtained with the model that employed hyper spectral and RGB data layers with 99.6% producer accuracy. Almost all the other models (except for the model with canopy height model) was able to accurately recognize pine. Spruce recognition signal was best demonstrated in RGB layer data. Classification accuracies dropped by at-least 4% in model that did not contain RGB layers for spruce. These results highlight that spruce was more distinguishable in RGB data. A hypothesis was tested to check if a combination of hyper spectral and blue channels were suitable to separate spruce and pine classes. The hypothesis was rejected since no meaningful difference was observed in comparison to a model that was based on hyper-spectral layer.

Chapter 5

Discussion

In this chapter we will answer to the research questions that were listed in the beginning of the article. The answers are based on the discussion of Articles **I-IV**.

• **RQ1: What are the important parameters to consider when designing a coded target?**

Important parameters in design phase of a coded target are

1. accuracy of localizing coded targets in images,
2. good visibility of coded targets (at-least 1 complete coded-targets) for a range of cameras with designed field of view (45-360 degrees), focal length (15-60 mm), sensor size (1-120 mpix) that were located in approximate distance of (30cm-4m),
3. simplicity of the proposed structure, rotation invariance,
4. embedded identifier,
5. embedded scale, and
6. ease of automatic detection.

The accuracy of the coded target element detection was fulfilled by fitting rotated ellipses to approximate locations of ellipses by employing the least-square fitting. The achieved accuracy was successfully demonstrated as image residuals in many calibration datasets. On best cases, 0.05-0.1 pixel image residual was observed for the coded targets. The detection accuracy was approximately 2-5 times better in comparison to a chess-board target [123] and square targets [64] however, its accuracy was comparable

or even higher than the coded targets with a similar detection mechanism such as [14].

The visibility was suitable for most cameras that were presented in this article. The best visibility was for SCs with wide lenses. The worse visibility was for individual cameras of multi-projective cameras specially when images were taken in close distance. The visibility is defined as a manageable parameter by upscaling and downscaling the coded target.

The simplicity of the structure enabled us to easily recognize the coded target among a complex stack of extracted ellipses. The embedded identifier enabled automatic labeling phase, therefore resulting to an easy network creation.

The embedded scale was helpful in realizing a correct scale for a calibrated multi-projective sensor. This scale was inaccurate; therefore, a scale parameter was later considered for an accurate MMS calibration.

• **RQ2: How to build a camera calibration room in an easy and efficient way for single and multi-cameras? How to accurately estimate the structure of the camera calibration room?**

Building a calibration room required special attention to hardware and software aspects. Hardware aspects was related to assigning an appropriate space for a specific range of cameras. A cubic space with suitable dimensions is advantageous in calibrating multi-cameras since coded targets are attached on all sides to create correlation between side-looking cameras. Lighting is also an important physical factor. A sufficient ambient light source helps to increase the quality of images and enhance the process of automatic target detection. Existence of patterns on surfaces of the calibration room is a disturbing factor which was considered by covering up unnecessary surfaces. A suitable space was selected from available options based on parameters such as GSDs, focal lengths, minimum and maximum desirable distances to camera, and camera FOVs. The considered space was suitable to calibrate most cameras that are stated in this thesis.

Software aspects concerns the coded target design, automatic coded-target detection and labeling, sensor modeling for single and multi-camera, fast initialization of the room, calibration strategy for single and multi-camera, and sparse BBA. The proposed coded-target was employed to build the calibration room. Some parameters such as sufficient number of coded targets on walls was addressed to make sure the suitability of the calibration room for multi-projective cameras. A minimum constraint self-calibrating sparse BBA was successfully implemented and employed to estimate the calibration-room's geometry, image positions and orientations, and camera parameters. All 10 interior orientation parameters of SCs were set free in

the optimization process. These parameters included focal length (1 parameter), principal-point location (2 parameters), radial distortion parameters (3 parameters) and tangential distortion parameters (2 parameters) , and scale and shear parameters (2 parameters).

The error propagation was enabled through the standard first-order non-linear least-square. The standard deviations for all interior orientation parameters indicated that they were meaningful. This conclusion was made by comparing the optimized values with the standard deviation values. The importance of an individual interior orientation parameter was assessed by blocking the effect of an individual or a group of correlated parameters from the overall correction that was made by the interior orientation parameters. Some parameters (such as radial distortion parameters) were assessed in groups since their high correlation caused false significance numbers. Based on this former assessment, scale and shear factors were not significant on the calibration process of SCs (Samsung NX300 and Canon EOS 6D). The standard deviations of those parameters confirmed this conclusion, since the optimized values were relatively close to standard deviations. Therefore, the SCs were not significantly affected by these parameters.

Image residuals were considered as an important indicator for assessing model fitness. In most calibration datasets average image residuals for SCs were 0.10-0.15 pixel which confirmed the suitability and accuracy of the designed coded-target and the calibration method.

The concept of a camera calibration room was implemented in similar works e.g. [11, 41, 61], however the proposed coded-target reach was simpler in structure, equal or better in accuracy, while it contained an embedded scale.

The calibration room could be improved by employing coded targets with different scales. In this proposed model, a scale number is embedded in the structure of the coded target that will be automatically recognized by the algorithm in a similar way to the coded-target identifier. Each corner of the calibration room will be filled with different scales which improves the usability of the room in calibrating different cameras.

• RQ3: How to calibrate a multi-projective camera in an efficient and easy way? What are the challenges in this regard?

Calibrating a multi-projective camera was enabled in this thesis through employing the proposed calibration room, calibration target, and sensorial model. The general overall of the proposed calibration scheme consists of estimating the calibration room's structure, creating separate projects for individual cameras of the underlying multi-projective camera by resecting its images to the point cloud, combining the individual projects and es-

timating the initial multi sensor, and final adjustment by considering the calibration room fixed.

The first challenge was related to the suitability of coded target for individual cameras of an underlying multi-projective camera. The design phase ensured that the targets were suitable for underlying cameras. The second challenge was related to automatic reading of the targets which was addressed in the coded-target implementation. The third challenge was related to the suitability of the calibration room for the calibration process. The calibration room should contain targets on parallel walls. This configuration ensures that side-looking cameras will have enough correlation in the optimization process. The fourth challenge was related to the singularity of the sparse BBA for multi-projective cameras, when structure of the sensor was enforced in the model. This challenge was addressed by setting the calibration room as a fixed structure. An efficient and easy calibration scheme for a general multi-projective camera was presented in Article **I**. Our proposed calibration method could be employed to calibrate simple or complex MPCs even with small or no overlaps that other calibration methods such as [59] and [61] are limited to address. A weakness in BBA is pointed out in Article **I** that is missing in simulated works such as [124].

• **RQ4: How to integrate a calibrated multi-projective camera in a mobile mapping system? What are the challenges and opportunities?**

A calibrated multi-projective camera could be integrated into a mobile mapping system for precise surveying and visualization purposes if 1) the MMS's sensors are synchronized through a central mechanism, 2) scale of the calibrated multi-camera will be the same as the other sensors, or a calibrated scale will be provided through a separate calibration process, 3) precise relative orientation and position of sensors with respect to a local coordinate system of the MMS will be known, and 4) GNSS and IMU observation will be accurate enough for precise surveying.

The first challenge that we focused in this regard was relates to the MMS calibration. An updated over-constrained sparse BBA was introduced that involved relative orientation and position of a multi-camera with respect to the GNSS and IMU. This model was employed to calibrate two MMSs (Article **II** and **III**). The results indicated that the calibrated system was able to localize 3D object points with high accuracy. high surveying accuracy.

The second challenge was related to the effect of weak intersection geometry on the accuracy of 3D object point localization. This challenge was studied by analyzing properties of intersections such as number of intersecting rays or incident angles on the quality of retrieved 3D point.

A calibrated MMS provides several opportunities related to precise surveying, geometrically accurate panoramic generation for application e.g. point-cloud painting or street-view. We demonstrated in Article **II** that Precise surveying is feasible to cm level accuracy, if a system is calibrated, and intersection geometry is strong. Our achieved accuracies are few times better than those reported by works such as [83] and [84].

• RQ5: How to compile metric panoramas from multi-projective images?

The proposed calibration sensor model for multi-projective cameras lays the foundation to compile non-stitching panoramas from multi-projective images. A non-stitching panorama is in contrast to the panoramas that are compiled through an image stitching mechanism. Stitched panoramas have visualization and artistic values, but metric information is lost due to the stitching mechanism. Non-stitching panoramas have steps on the edges where images meet, however, generating a precise map that corresponds each pixel of the compiled panorama to the parent image is feasible. This map was called the correspondence map. The correspondence map was employed in the panorama compilation process to generate a metric panorama. The correspondence map was assumed as the required metadata file for metric applications. The proposed scheme take the systematic error described by [119] into equations. Therefore, the calibration scheme could be employed to compile metric panoramas.

• RQ6: How to address the gimbal-lock singularity in Jacobian matrix of a BBA?

If the BBA is performed with Euler angles, then the gimbal-lock singularity could occur in the Jacobian matrix under the condition that rotations planes become approximately coplanar. To solve this singularity, a suitable rotational frame-works that is not susceptible to gimbal-lock singularity should be employed. Two well-known examples of such a system are quaternion and axis-angle rotational framework. Both of these systems are based on four parameters with 3 degrees of freedom. Therefore, a constraint should be applied on both systems to decrease the number of free parameters to three. In this thesis, we proposed a new angular parametrization of BBA based on spherical rotation coordinate system that has only three independent parameters, therefore, the constraint equation is no longer a requirement. The angular system was called spherical angles since spherical coordinates with two rotation parameters are employed to specify the rotation vector, and a momentum is considered as the rotation scalar value around the axis. The proposed sparse BBA was modified to work under this

rotational framework by employing analytical gradients that were manually calculated.

• **RQ7: What are the software challenges to build a real-time photogrammetry system? How to address real-time trajectory estimation challenge efficiently?**

Several important software challenges arise dealing with a real-time photogrammetric solution. These challenges are mainly related to sensor calibration, mobile mapping system calibration, real-time "trajectory and orientation" estimation, and technical difficulties that arises in real-time processing steps such as classification or object detection.

Geometric calibration of sensors is helpful and important prior to a real-time mission since self-calibration of a mapping system is not always feasible during a mission. Factors such as geometric weakness of a real-time situation, or computational limitations to address a self-calibration process are two key factors to consider with this regard.

Geometric calibration of single and multi-projective cameras was address in Article **I**. The mobile-mapping system calibration was addressed and discussed in Article **II** and **III**. Real-time trajectory and orientation estimation algorithm was proposed and discussed in Article **III**. A multi-level image matching scheme was proposed that significantly reduced the image matching time while preserved the quality of the network. A photogrammetric loop detection algorithm was proposed to further strengthen the network structure. A modified connect-to-next approach for network initialization was proposed that significantly decreased the execution of a connect-to-all approach. The proposed approach addressed the weaknesses of a sequential network creation approach. A monocular SLAM solution was therefore proposed by combining all small solutions. The proposed SLAM solution was updated to employ GNSS and IMU observations whenever possible. Calibrated MMS parameters were employed for high-quality direct georeferencing. The proposed multi-level matching approach was considerably faster than other approaches such as [72, 74, 77]. The proposed monocular-SLAM approach addressed the cases that other methods such as [30] were unable to address.

• **RQ8: How to integrate deep learning to a UAV-based multi-sensorial photogrammetric mapping system? What are the challenges and opportunities?**

The first step to employ a deep-learning method to classify UAV-based data is data preparation. Geometric and radiometric processing of RGB and hyperspectral images are essential preprocessing steps to produce suit-

able inputs for the classification phase. A supervised classification approach is achievable by labeling individual classes in data layers. The trained classifier could be then employed in a smart photogrammetric solution. To address this question a tree-species classification problem by employing RGB and hyperspectral images and deep-learning methods was investigated. Normalized hyper-spectral images, RGB images, and canopy height models were investigated as data layers to automatically recognize tree-species (Article **IV**). Two deep-learning methods were employed as a 3D-CNN and an MLP that were efficient in terms of archived accuracy and simplicity. The benefits of employing the proposed 3D-CNN over the conventional MLP were twofold. Firstly, it provided better accuracies than the conventional MLP. Secondly, it was more efficient in terms of the number of parameters, which implied as a simpler model. Different combination of features were studies to show the challenges and opportunities of employing each measurement technique. The best classification result achieved when all data layers were combined, which was expected. RGB images were successfully demonstrated as a powerful source of data where hyper-spectral images are not available. CHM data had the least discrimination power among all data layers in recognizing tree types.

The presented model based on all layers demonstrated higher classification accuracy than [125]. According to our knowledge, the classification accuracies on tree-level were the best published results in comparison to other works such as [22, 104–107, 109–111] up to the publication date of Article **IV**. The presented results are dependent to parameters such as characteristics of a forest, data normalization, and effect of noises and blunders, therefore, more research should be organized with higher amounts of data under varying conditions.

Chapter 6

Conclusions

Multi-projective cameras turned into important blocks of many smart solutions such as mobile mapping systems. Embedding multi-projective cameras in systems equipped with GNSS and IMU sensors provides great opportunities for direct georeferencing of multi images, and image-based mapping.

Coded targets and camera calibrations rooms are efficient tools to address challenges regarding geometric accuracy of cameras employed in a smart solution. When a camera model is under investigation, coded targets remove the extra and unnecessary burden of finding image tie-point. Employing coded targets to build a camera calibration room is an easy and efficient way to investigate geometric aspects of an optical system. Design phase is essential to build a camera calibration room since several parameters are involved to make the room efficient and applicable for a range of cameras.

The first hypothesis of this thesis was that a camera test field is required to geometrically calibrate a multi-projective camera. FGI's calibration room is a suitable space for camera calibration that is covered with coded targets. This room was designed during this study. The calibration room was constructed by proposing an easy-to-read and accurate coded-target. The proposed coded-target was based on an asymmetrical pattern of filled circles with embedded identifier and scale. A precise algorithm was presented to automatically recognize the coded target with high sub-pixel accuracy. A network creation method was proposed to initiate the structure of the room. Careful geometric assessments were performed based on image residuals and analysis of uncertainties. Covariance sub-matrices were analysed to present positional uncertainties as rotated 3D error ellipsoids. This presentation of positional uncertainty was the main tool to assess geometric strength of spatial intersections, as well as a test to assess model

goodness of fitness. Confidence intervals were analysed to ensure a robust estimation. The empirical study showed that the proposed test field was accurately estimated by the single projective cameras.

The first hypothesis was proven by employing the calibration field to calibrate two complex multi-projective cameras. A sensor model for multi-projective cameras was proposed based on physical constraints of the camera class. Two multi-projective cameras (Panono and Ladybug) with 36 and 6 projective cameras went under investigation. A weakness in internal structure of complex multi-projective cameras was observed as the main barrier in proposing an independent calibration scheme. This problem was efficiently addressed by estimating the structure of the calibration room by two SCs (Canon EOS 6D and Samsung NX300) prior to the multi-camera calibration and then treating the calibration room as a rigid object with known geometry. A calibration scheme was proposed to estimate initial structure of a multi-projective camera. A customized bundle adjustment was proposed by enforcing relative orientation constraints into the model that successful resulted in revealing the internal structure of the cameras under investigation. Embedded scale bars in coded targets were finally employed to address the scale uncertainty. The second hypothesis was that a system calibration is essential to acquire high geometric accuracies from a mobile mapping system equipped with a multi-projective camera, a GNSS and an IMU. The potentials of the proposed calibration model and scheme went under investigation in a high-accuracy mapping application. The proposed multi-projective camera calibration scheme was employed to calibrate a LadyBug camera prior to its usage in the mobile mapping system. The camera calibration phase was performed in FGI's camera calibration room. The sparse BBA model was updated to accept mounting parameters of the multi-projective camera with respect to GNSS and IMU. The mobile-mapping system mounting parameters were accurately estimated by the algorithm. The calibrated mounting parameters were employed for direct georeferencing which resulted to accurate image-based positioning. The geometric assessment was performed by employing check sites. The second hypothesis was proven by a demonstration of centimetre-level accuracy by employing the proposed MMS calibration scheme.

The problem of real-time trajectory estimation of a SC is an important part of this thesis that was addressed in the third article. Different challenges were considered in this part regarding camera calibration by employing an aerial calibration field. A new angular parameterization based on spherical rotation coordinate system was presented to address the gimbal lock singularity. The angular parametrization is named spherical angles

since it is based on spherical coordinates that locate a rotation vector, and an angular momentum around it. The sparse BBA was successfully modified to accept the spherical angles. A multi-level matching scheme was presented to address the performance issues in the image tie-point extraction phase. A window propagation scheme was presented to reduce the key-point localization and descriptor extraction domains on the first image scale. A loop detection algorithm was presented based on photogrammetric overlap analysis for a near-planar object. An image-based SLAM algorithm was presented by employing the proposed multi-level matching and the sparse BBA. The algorithm was successfully assessed by performance measures such as execution time. Geometric aspects were assessed by employing 3D error ellipsoids and 3D check points that were considered for each project.

To study different aspects of integrating classifiers into a futuristic vision-based photogrammetric system, a tree-species classification problem by employing a UAV is investigated in this thesis. The UAV was equipped with an RGB and a Hyper spectral camera. Different aspects of the problem such as the suitability of data layers in a classification task was under special attention. In total, 37 data layers were investigated in this problem; a number of 33 Hyper spectral layers, 3 colour RGB layers, and a canopy high model layer were the main data layers in the classification phase. Different combinations of the data layers were combined in separate classifications to investigate the most efficient set of features. It was demonstrated that a 3D convolutional neural network achieves better classification accuracies than the classic approach of multi-layered perceptron. The proposed structure for the 3D CNN was simple, while yielding high classification accuracies. Classification metrics as well as AUC of ROC curves were employed to assess the classification ability of the proposed method in comparison on an MLP. The results suggested that a combination of hyper spectral layers and RGB layers led to the best classification output. The classification with only RGB layers also resulted into high-accuracy outcomes which highlighted the significance of employing a regular camera for this problem. The hyperspectral layers were successful in detecting spruce and birch; however, pine was less visible than the state where RGB layers were also employed. Pine was almost detected equally well by all models. Spruce was detected well in hyper-spectral layer with 91% detection rate. Pine detection was also slightly improved to 94% by including RGB data. Birch was accurately classified by employing hyperspectral data with 96% detection rate. Birch accuracy was slightly dropped to 92% by employing only RGB layers. The overall accuracy of the proposed model was 98.3% which outperformed

previous studies. Different aspects of a futuristic photogrammetric solution are individually studied and presented in this thesis; however, a uniform solution is achievable in future by integrating the presented solutions into a real-time smart photogrammetric system.

References

[1] G. Fangi, The Multi-image spherical Panoramas as a tool for Architectural Survey, *Heritage Documentation*, vol. 21, pp. 311–316, 2011.

[2] K. Scheibe, H. Korsitzky, R. Reulke, M. Scheele, and M. Solbrig, Eyescan-a high resolution digital panoramic camera, *International Workshop on Robot Vision*, pp. 77–83, 2001.

[3] H. Kauhanen, P. Rönnholm, V. Lehtola, Motorized panoramic camera mount–calibration and image capture, *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 3, no. 5, 2016.

[4] M. B. Campos, A. M. G. Tommaselli, E. Honkavaara, F. D. S. Prol, H. Kaartinen, A. El Issaoui, and T. Hakala, A Backpack-Mounted Omnidirectional Camera with Off-the-Shelf Navigation Sensors for Mobile Terrestrial Mapping: Development and Forest Application, *Sensors*, vol. 18, no. 3, Art. no. 3, doi: 10.3390/s18030827, 2018.

[5] T. Luhmann, A historical review on panorama photogrammetry, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 34, no. 5/W16, p. 8, 2004.

[6] C. Badue, R. Guidolini, R. V. Carneiro, P. Azevedo, V. B. Cardoso, A. Forechi, and L. de Paula Veronese, Self-driving cars: A survey, *Expert Systems with Applications*, p. 113816, 2020.

[7] F. Nex and F. Remondino, UAV for 3D mapping applications: a review, *Applied geomatics*, vol. 6, no. 1, pp. 1–15, 2014.

[8] D. Anguelov, C. Dulong, D. Filip, C. Frueh, S. Lafon, R. Lyon, R., and J. Weaver, Google street view: Capturing the world at street level, *Computer*, vol. 43, no. 6, pp. 32–38, 2010.

[9] A. Habib and M. F. Morgan, Automatic calibration of low-cost digital cameras, *Optical Engineering*, vol. 42, no. 4, Art. no. 4, 2003.

[10] D. Schneider and H.-G. Maas, Geometric modelling and calibration of a high resolution panoramic camera, *Optical 3-D Measurement Techniques VI*, vol. 2, pp. 122–129, 2003.

[11] E. Schwalbe, Geometric modelling and calibration of fisheye lens camera systems, *Institute of Photogrammetry and Remote Sensing-Dresden University of Technology*, Dresden, 2005.

[12] S. Aghayari, M. Saadatseresht, M. Omidalizarandi, and I. Neumann, Geometric calibration of full spherical panoramic Ricoh-Theta camera, *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-1W1, vol. 4, pp. 237–245, 2017.

[13] H. Aasen, E. Honkavaara, A. Lucieer, and P. J. Zarco-Tejada, Quantitative remote sensing at ultra-high resolution with UAV spectroscopy: a review of sensor technology, measurement procedures, and data correction workflows, *Remote Sensing*, vol. 10, no. 7, p. 1091, 2018.

[14] D. A. Cucci, Accurate Optical Target Pose Determination For Applications In Aerial Photogrammetry, *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 3, pp. 257–262, doi: 10.5194/isprs-annals-III-3-257-2016, 2016.

[15] J. Amiri Parian and A. Gruen, Sensor modeling, self-calibration and accuracy testing of panoramic cameras and laser scanners, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 1, pp. 60–76, 2010.

[16] S. Harwin and A. Lucieer, Assessing the accuracy of georeferenced point clouds produced via multi-view stereopsis from unmanned aerial vehicle (UAV) imagery, *Remote Sensing*, vol. 4, no. 6, Art. no. 6, 2012.

[17] A. Habib, T. Zhou, A. Masjedi, Z. Zhang, J. E. Flatt, and M. Crawford, Boresight Calibration of GNSS/INS-Assisted Push-Broom Hyperspectral Scanners on UAV Platforms, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 5, Art. no. 5, 2018.

[18] X. Chen, W. Hu, L. Zhang, Z. Shi, and M. Li, Integration of low-cost gnss and monocular cameras for simultaneous localization and mapping, *Sensors*, vol. 18, no. 7, p. 2193, 2018.

[19] W. Forstner and R. Steffen, On visual real time mapping for Unmanned Aerial Vehicles, *Proceedings of XXI ISPRS Congress*, 2008.

[20] R. A. Oliveira, E. Khoramshahi, J. Suomalainen, T. Hakala, N. Viljanen, and E. Honkavaara, Real-time and post-processed georeferencing for hyperpspectral drone remote sensing, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2018.

[21] F. E. Fassnacht, H. Latifi, K. Stereńczak, A. Modzelewska, M. Lefsky, L. T. Waser, and A. Ghosh, Review of studies on tree species classification from remotely sensed data, *Remote Sensing of Environment*, vol. 186, pp. 64–87, 2016.

[22] I. Korpela, H. O. Ørka, M. Maltamo, T. Tokola, and J. Hyyppä, Tree species classification using airborne LiDAR–effects of stand and tree parameters, downsizing of training set, intensity normalization, and sensor type, *Silva Fennica*, vol. 44, no. 2, pp. 319–339, 2010.

[23] J. Amiri Parian and A. Gruen, A sensor model for panoramic cameras, *6th Optical 3D Measurement Techniques*, vol. 2, pp. 130–141, 2003.

[24] H.-G. Maas, Close range photogrammetry sensors, *Advances in Photogrammetry, Remote Sensing and Spatial Information Sciences: 2008 ISPRS Congress Book*, pp. 81–90, 2008.

[25] M. Laiacker, M. Schwarzbach, and K. Kondak, Automatic aerial retrieval of a mobile robot using optical target tracking and localization, in *2015 IEEE Aerospace Conference*, pp. 1–7, 2015.

[26] R. Gonzales, R. Woods, and S. Eddins, Digital Image Processing, *Prentice Hall, New Jersey*, 2002.

[27] W. Schnotz, An integrated model of text and picture comprehension, *The Cambridge Handbook of Multimedia Learning*, vol. 49, p. 69, 2005.

[28] A. Habib, I. Detchev, and E. Kwak, Stability analysis for a multicamera photogrammetric system, *Sensors*, vol. 14, no. 8, Art. no. 8, 2014.

[29] M. Irsigler, J. A. Avila-Rodriguez, and G. W. Hein, Criteria for GNSS multipath performance assessment, in *Proceedings of the 18th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2005)*, pp. 2166–2177, 2005.

[30] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, ORB-SLAM: a versatile and accurate monocular SLAM system, *IEEE transactions on robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[31] R. Hartley and A. Zisserman, Multiple view geometry in computer vision. *Cambridge university press*, 2003.

[32] M. Potmesil and I. Chakravarty, Synthetic image generation with a lens and aperture camera model, *ACM Transactions on Graphics (TOG)*, vol. 1, no. 2, pp. 85–108, 1982.

[33] P. Alho et al., Mobile laser scanning in fluvial geomorphology: Mapping and change detection of point bars, *Zeitschrift für Geomorphologie*, vol. 55, no. 2, pp. 31–50, 2011.

[34] D. Schneider and H. Maas, Application and accuracy potential of a strict geometric model for rotating line cameras, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 34, no. 5/W16, Art. no. 5/W16, 2004.

[35] D. Schneider and H.-G. Maas, A geometric model for linear-array-based terrestrial panoramic cameras, *The Photogrammetric Record*, vol. 21, no. 115, Art. no. 115, 2006.

[36] J. A. Parian and A. Gruen, An advanced sensor model for panoramic cameras, *ISPRS International Archives Photogrammetry Remote Sensing Spatial Information Sciences*, vol. 35, pp. 24–29, 2004.

[37] G. Fangi and C. Nardinocchi, Photogrammetric processing of spherical panoramas, *The Photogrammetric Record*, vol. 28, no. 143, Art. no. 143, 2013.

[38] M. B. Campos, A. M. G. Tommaselli, J. Marcato Junior, and E. Honkavaara, Geometric model and assessment of a dual-fisheye imaging system, *The photogrammetric record*, vol. 33, no. 162, Art. no. 162, doi: 10.1111/phor.12240, 2018.

[39] L. Barazzetti, M. Previtali, and F. Roncoroni, 3D Modeling with the Samsung Gear 360, *ISPRS International Archives Photogrammetry Remote Sensing Spatial Information Sciences*, vol. XLII-2/W3, pp. 85–90, Feb. 2017, doi: 10.5194/isprs-archives-XLII-2-W3-85-2017, 2017.

[40] S. Ray, The fisheye lens and immersed optics, *Applied Photographic Optics*, 327-3322002, pp. 326–332, 2002.

[41] D. Schneider, E. Schwalbe, and H.-G. Maas, Validation of geometric models for fisheye lenses, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 64, no. 3, Art. no. 3, 2009.

[42] S. Abraham and W. Förstner, Fish-eye-stereo calibration and epipolar rectification, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 59, no. 5, Art. no. 5, doi: 10.1016/j.isprsjprs.2005.03.001, 2005.

[43] S. Blaser, S. Cavegn, and S. Nebiker, Development of a Portable High Performance Mobile Mapping System Using the Robot Operating System., *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 4, no. 1, Art. no. 1, 2018.

[44] G. Fangi, R. Pierdicca, M. Sturari, and E. Malinverni, Improving Spherical Photogrammetry Using 360°Omni-Cameras: Use Cases and New Applications, *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII–2, pp. 331–337, doi: 10.5194/isprs-archives-XLII-2-331-2018, 2018.

[45] L. T. Losè, F. Chiabrando, and A. Spanò, Preliminary Evaluation of a Commercial 360 Multi-Camera Rig for Photogrammetric Purposes., *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, no. 2, Art. no. 2, 2018.

[46] D. Scaramuzza, Omnidirectional camera, *Computer Vision: A Reference Guide*, pp. 552–560, 2014.

[47] W. Song et al., Design and assessment of a 360°panoramic and high-performance capture system with two tiled catadioptric imaging channels, *Applied optics*, vol. 57, no. 13, Art. no. 13, 2018.

[48] A. Gruen and T. S. Huang, Calibration and orientation of cameras in computer vision, vol. 34. *Springer Science & Business Media*, 2013.

[49] H. C. Longuet-Higgins, A computer algorithm for reconstructing a scene from two projections, *Nature*, vol. 293, no. 5828, pp. 133–135, 1981.

[50] T. S. Huang and A. N. Netravali, Motion and structure from feature correspondences: A review, in Advances In Image Processing And Understanding: *A Festschrift for Thomas S Huang, World Scientific*, pp. 331–347, 2002.

[51] R. I. Hartley, In defence of the 8-point algorithm, in *Proceedings of IEEE international conference on computer vision*, pp. 1064–1070, 1995.

[52] D. Nistér, An efficient solution to the five-point relative pose problem, *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, no. 6, Art. no. 6, 2004.

[53] H. Stewénius, D. Nistér, F. Kahl, and F. Schaffalitzky, A minimal solution for relative pose with unknown focal length, *Image and Vision Computing*, vol. 26, no. 7, pp. 871–877, 2008.

[54] G. He, K. Novak, and W. Feng, Stereo camera system calibration with relative orientation constraints, in *Videometrics*, vol. 1820, pp. 2–8, 1993.

[55] B. King, Methods for the Photogrammetric Adjustment of Bundles of Constrained Stereopairs, *International Archives of Photogrammetry and Remote Sensing*, vol. 30, pp. 473–480, 1994.

[56] H. Zhuang, A self-calibration approach to extrinsic parameter estimation of stereo cameras, Robot. *Robotics and Autonomous Systems*, vol. 15, no. 3, Art. no. 3, 1995.

[57] T. Svoboda, H. Hug, and L. Van Gool, ViRoom—low cost synchronized multicamera system and its self-calibration, in *Joint Pattern Recognition Symposium*, pp. 515–522, 2002.

[58] I. Detchev, M. Mazaheri, S. Rondeel, and A. Habib, Calibration of multi-camera photogrammetric systems, *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL–1, pp. 101–108, doi: 10.5194/isprsarchives-XL-1-101-2014, 2014.

[59] J. L. Lerma, S. Navarro, M. Cabrelles, and A. E. Seguí, Camera calibration with baseline distance constraints, *The Photogrammetric Record*, vol. 25, no. 130, Art. no. 130, 2010.

[60] B. Li, L. Heng, K. Koser, and M. Pollefeys, A multiple-camera system calibration toolbox using a feature descriptor-based calibration pattern, in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1301–1307, 2013.

[61] A. Tommaselli, M. Galo, M. de Moraes, J. Marcato, C. Caldeira, and R. Lopes, Generating virtual images from oblique frames, *Remote Sensing*, vol. 5, no. 4, Art. no. 4, 2013.

[62] A. M. G. Tommaselli, L. D. Santos, R. A. de Oliveira, A. Berveglieri, N. N. Imai, and E. Honkavaara, Refining the Interior Orientation of a Hyperspectral Frame Camera With Preliminary Bands Co-Registration, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 7, pp. 2097–2106, 2019.

[63] J. Amiri Parian, Sensor modeling, calibration and point positioning with terrestrial panoramic cameras, *PhD Thesis*, ETH Zurich, 2007.

[64] S. Garrido-Jurado, R. Muñoz-Salinas, F. Madrid-Cuevas, and M. Marín-Jiménez, Automatic generation and detection of highly reliable fiducial markers under occlusion, *Pattern Recognition*, vol. 47, pp. 2280–2292, doi: 10.1016/j.patcog.2014.01.005, 2014.

[65] A. Tommaselli, M. Galo, M. de Moraes, J. Marcato, C. Caldeira, and R. Lopes, Generating virtual images from oblique frames, *Remote Sensing*, vol. 5, no. 4, Art. no. 4, 2013.

[66] G. An, S. Lee, M.-W. Seo, K. Yun, W.-S. Cheong, and S.-J. Kang, Charuco Board-Based Omnidirectional Camera Calibration Method, *Electronics*, vol. 7, no. 12, Art. no. 12, doi: 10.3390/electronics7120421, 2018.

[67] D. Jarron, D. D. Lichti, M. M. Shahbazi, and R. S. Radovanovic, Multi-camera panormamic imaging system calibration, 2019.

[68] J. M. Junior, A. Tommaselli, and M. Moraes, Calibration of a catadioptric omnidirectional vision system with conic mirror, *ISPRS J. Photogramm. Remote Sens.*, vol. 113, pp. 97–105, 2016.

[69] B. C. L. Lok, Interacting with dynamic real objects in virtual environments, *PhD Thesis*, University of North Carolina at Chapel Hill, 2002.

[70] D. G. Lowe, Distinctive image features from scale-invariant keypoints, *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[71] Y. Ke and R. Sukthankar, PCA-SIFT: A more distinctive representation for local image descriptors, in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2004 (CVPR 2004)*, vol. 2, p. II–II, 2004.

[72] Wu, C. A GPU Implementation of Scale Invariant Feature Transform (SIFT). 2007. *Available online: https://github.com/pitzer/SiftGPU* (*Accessed on 28 Sept. 2020*).

[73] J.-M. Morel and G. Yu, ASIFT: A new framework for fully affine invariant image comparison, *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 438–469, 2009.

[74] H. Bay, T. Tuytelaars, and L. Van Gool, Surf: Speeded up robust features, in *European Conference on Computer Vision*, pp. 404–417, 2006.

[75] E. Rosten and T. Drummond, Machine learning for high-speed corner detection, in *European Conference on Computer Vision*, pp. 430–443, 2006.

[76] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, Brief: Binary robust independent elementary features, in *European Conference on Computer Vision*, pp. 778–792, 2010.

[77] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, ORB: An efficient alternative to SIFT or SURF, in *2011 International Conference on Computer Vision*, pp. 2564–2571, 2011.

[78] L. Juan and L. Gwon, A comparison of sift, pca-sift and surf, *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 8, no. 3, pp. 169–176, 2007.

[79] E. Karami, S. Prasad, and M. Shehata, Image matching using SIFT, SURF, BRIEF and ORB: performance comparison for distorted images, *ArXiv Prepr.* ArXiv171002726, 2017.

[80] Y. Wu, F. Tang, and H. Li, Image-based camera localization: an overview, *Visual Computing for Industry, Biomedicine, and Art*, vol. 1, no. 1, pp. 1–13, 2018.

[81] B. Williams, M. Cummins, J. Neira, P. Newman, I. Reid, and J. Tardós, An image-to-map loop closing method for monocular SLAM, in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2053–2059, 2008.

[82] D. Gálvez-López and J. D. Tardos, Bags of Binary Words for Fast Place Recognition in Image Sequences, *IEEE Transactions on Robotics*, vol. 28, no. 5, pp. 1188–1197, 2012.

[83] G. Fink, M. Franke, A. F. Lynch, K. Röbenack, and B. Godbolt, Visual Inertial SLAM: Application to Unmanned Aerial Vehicles, *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 1965–1970, 2017.

[84] S. Jiang and W. Jiang, Efficient sfm for oblique uav images: From match pair selection to geometrical verification, *Remote Sensing*, vol. 10, no. 8, p. 1246, 2018.

[85] M. Cramer, D. Stallmann, and N. Haala, Direct Georeferencing Using GPS/Inertial Exterior Orientations for Photogrammetric Applications, *International Archives of Photogrammetry and Remote Sensing*, vol. 33, 2000.

[86] H. Gontran, J. Skaloud, and P.-Y. Gilliéron, A mobile mapping system for road data capture via a single camera, *Advances in Mobile Mapping Technology* Taylor Francis Group Lond. UK, pp. 43–50, 2007.

[87] S. Cavegn, S. Nebiker, and N. Haala, A Systematic Comparison of Direct and Image-Based Georeferencing in Challenging Urban Areas., *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 41, 2016.

[88] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, Object recognition with gradient-based learning, in *Shape, contour and grouping in computer vision*, Springer, pp. 319–345, 1999.

[89] S. Behnke, Hierarchical neural networks for image interpretation, vol. 2766. *Springer*, 2003.

[90] P. Y. Simard, D. Steinkraus, J. C. Platt, and others, Best practices for convolutional neural networks applied to visual document analysis., in *Icdar*, vol. 3, 2003.

[91] D. Scherer, A. Müller, and S. Behnke, Evaluation of pooling operations in convolutional architectures for object recognition, in *International Conference on Artificial Neural Networks*, pp. 92–101, 2010.

[92] V. Dumoulin and F. Visin, A guide to convolution arithmetic for deep learning, *ArXiv Prepr.* ArXiv160307285, 2016.

[93] B. Xu, N. Wang, T. Chen, and M. Li, Empirical Evaluation of Rectified Activations in Convolutional Network, *ArXiv Prepr.* ArXiv150500853, 2015.

[94] S. Hochreiter, The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 6, no. 02, pp. 107–116, 1998.

[95] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, The Adaptive Computation and Machine Learning Series, *Cambridge, MA: The MIT Press*, 2016.

[96] Y. LeCun. The MNIST database of handwritten digits. http://yann.lecun.com/exdb/mnist/index.html (*Accessed 06 Nov. 2020*), 1998.

[97] M. Z. Alom et al., The History Began from Alexnet: A Comprehensive Survey on Deep Learning Approaches, *ArXiv Prepr.* ArXiv180301164, 2018.

[98] A. Krizhevsky, I. Sutskever, and G. E. Hinton, Imagenet Classification with Deep Convolutional Neural Networks, in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[99] I. Pölönen et al., Tree Species Identification Using 3D Spectral Data and 3D Convolutional Neural Network, in *2018 9th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1–5, 2018.

[100] J. Peña, P. Gutiérrez, C. Hervás-Martínez, J. Six, R. Plant, and F. López-Granados, Object-based image classification of summer crops with machine learning methods, *Remote Sensing*, vol. 6, no. 6, pp. 5019–5041, 2014.

[101] Y. Li, H. Zhang, and Q. Shen, Spectral–spatial classification of hyperspectral imagery with 3D convolutional neural network, *Remote Sensing*, vol. 9, no. 1, p. 67, 2017.

[102] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, Deep learning-based classification of hyperspectral data, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2094–2107, 2014.

[103] Z. Xie, Y. Chen, D. Lu, G. Li, and E. Chen, Classification of Land Cover, Forest, and Tree Species Classes with ZiYuan-3 Multispectral and Stereo Data, *Remote Sensing*, vol. 11, no. 2, p. 164, 2019.

[104] M. Dalponte, H. O. Ørka, T. Gobakken, D. Gianelle, and E. Næsset, Tree Species Classification in Boreal Forests With Hyperspectral Data, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 5, pp. 2632–2645, doi: 10.1109/TGRS.2012.2216272, 2013.

[105] J. Heinzel and B. Koch, Exploring full-waveform LiDAR parameters for tree species classification, *International Journal of Applied Earth Observation and Geoinformation*, vol. 13, no. 1, pp. 152–160, 2011.

[106] J. Heinzel and B. Koch, Investigating Multiple Data Sources for Tree Species Classification in Temperate Forest and Use for Single Tree Delineation, *International Journal of Applied Earth Observation and Geoinformation*, vol. 18, pp. 101–110, 2012.

[107] W. Yao, P. Krzystek, and M. Heurich, Tree Species Classification and Estimation of Stem Volume and DBH Based on Single Tree Extraction by Exploiting Airborne Full-Waveform LiDAR Data, *Remote Sensing of Environment*, vol. 123, pp. 368–380, 2012.

[108] I. Sa et al., Weedmap: A Large-Scale Semantic Weed Mapping Framework Using Aerial Multispectral Imaging and Deep Neural Network for Precision Farming, *Remote Sensing*, vol. 10, no. 9, p. 1423, 2018.

[109] E. Raczko and B. Zagajewski, Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images, *European Journal of Remote Sensing*, vol. 50, no. 1, pp. 144–154, doi: 10.1080/22797254.2017.1299557, 2017.

[110] X. Yu, J. Hyyppä, P. Litkey, H. Kaartinen, M. Vastaranta, and M. Holopainen, Single-Sensor Solution to Tree Species Classification Using Multispectral Airborne Laser Scanning, *Remote Sensing*, vol. 9, no. 2, p. 108, doi: 10.3390/rs9020108, 2017.

[111] S. E. Franklin and O. S. Ahmed, Deciduous tree species classification using object-based analysis and machine learning with unmanned aerial vehicle multispectral data, *Int. J. Remote Sens.*, vol. 39, no. 15–16, pp. 5236–5245, doi: 10.1080/01431161.2017.1363442, 2018.

[112] M. P. Ferreira, F. H. Wagner, L. E. Aragão, Y. E. Shimabukuro, and C. R. de Souza Filho, Tree species classification in tropical forests using visible to shortwave infrared WorldView-3 images and texture analysis, *ISPRS journal of photogrammetry and remote sensing*, vol. 149, pp. 119–131, 2019.

[113] J. Matas, O. Chum, M. Urban, and T. Pajdla, Robust wide-baseline stereo from maximally stable extremal regions, *Image and Vision Computing*, vol. 22, no. 10, pp. 761–767, 2004.

[114] D. Nistér, Preemptive RANSAC for live structure and motion estimation, *Machine Vision and Applications*, vol. 16, no. 5, pp. 321–329, 2005.

[115] P. L. Cheng, A Spherical Rotation Coordinate System for the Description of Three-Dimensional Joint Rotations, *Annals of Biomedical Engineering*, vol. 28, no. 11, pp. 1381–1392, 2000.

[116] L. Juan and G. Oubong, SURF applied in panorama image stitching, in *2010 2nd international conference on image processing theory, tools and applications*, pp. 495–499, 2010.

[117] S. K. Pasupaleti, M. Uliyar, and P. Gupta, Panorama image stitching, *Google Patents*, 2010.

[118] Y. Xiong and K. Pulli, Fast Image Stitching and Editing for Panorama Painting on Mobile Phones, in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, pp. 47–52, 2010.

[119] J.-Y. Rau, B. Su, K. Hsiao, and J. Jhan, Systematic Calibration for a Backpacked Spherical Photogrammetry Imaging System., *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 41, 2016.

[120] E. Honkavaara et al., Processing and assessment of spectrometric, stereoscopic imagery collected using a lightweight UAV spectral camera for precision agriculture, *Remote Sensing*, vol. 5, no. 10, pp. 5006–5039, 2013.

[121] J. Mäkynen, C. Holmlund, H. Saari, K. Ojala, and T. Antila, Unmanned aerial vehicle (UAV) operated megapixel spectral camera, in *Electro-Optical Remote Sensing, Photonic Technologies, and Applications V*, vol. 8186, p. 81860Y, 2011.

[122] H. Saari et al., Miniaturized hyperspectral imager calibration and UAV flight campaigns, in *Sensors, systems, and next-generation satellites xvii*, vol. 8889, p. 88891O, 2013.

[123] A. De la Escalera and J. M. Armingol, Automatic chessboard detection for intrinsic and extrinsic camera parameter calibration, *Sensors*, vol. 10, no. 3, pp. 2027–2044, 2010.

[124] S. Urban, S. Wursthorn, J. Leitloff, and S. Hinz, MultiCol bundle adjustment: a generic method for pose estimation, simultaneous self-calibration and reconstruction for arbitrary multi-camera systems, *International journal of computer vision*, vol. 121, no. 2, pp. 234–252, 2017.

[125] O. Nevalainen et al., Individual tree detection and classification with UAV-based photogrammetric point clouds and hyperspectral imaging, *Remote Sensing*, vol. 9, no. 3, p. 185, 2017.

Appendix A

Acronyms and abbreviations used in this thesis are the following:

SC	Single camera
MCS	Multi-camera system
UAV	Unmanned aerial vehicle
SLAM	Simultaneous localization and mapping
3D-CNN	3D convolutional neural network
FGI	Finnish Geospatial Research Institute
NLS	National Land Survey of Finland
MLP	Multi-layered perceptron
GCP	Ground control point
SVM	Support vector machine
FOV	Field of view
BBA	Bundle block adjustment
RANSAC	Random sample consensus
GNSS	Global navigation satellite system
IMU	inertial measurement unit
MMS	Mobile mapping system
DoG	Difference of gaussian
PCA	Principle component analysis
RAM	Random-access memory
SIFT	Scale-invariant feature transform
SURF	Speed-up robust feature
FAST	Feature from accelerated test
BRIEF	Binary robust independent elementary feature
GSD	Ground sampling distance
ORB	Oriented FAST and rotated BRIEF
GPU	Graphical processing unit
CPU	Central processing unit