Martin Weinmann

Reconstruction and Analysis of 3D Scenes

From Irregularly Distributed 3D Points to Object Classes

Springer

Reconstruction and Analysis of 3D Scenes

Martin Weinmann

Reconstruction and Analysis of 3D Scenes

From Irregularly Distributed 3D Points to Object Classes

Springer

Martin Weinmann
Institute of Photogrammetry and Remote Sensing
Karlsruhe Institute of Technology
Karlsruhe
Germany

ISBN 978-3-319-29244-1 ISBN 978-3-319-29246-5 (eBook)
DOI 10.1007/978-3-319-29246-5

Library of Congress Control Number: 2016930271

Printed on acid-free paper

This Springer imprint is published by SpringerNature
The registered company is Springer International Publishing AG Switzerland

Reconstruction and Analysis of 3D Scenes

– From Irregularly Distributed 3D Points to Object Classes –

Zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

von der

Fakultät für Bauingenieur-, Geo- und Umweltwissenschaften

des

Karlsruher Instituts für Technologie (KIT)

genehmigte
Dissertation

von

Dipl.-Ing. Martin Weinmann

aus Karlsruhe

Tag der mündlichen Prüfung: 24.07.2015

Referent: Prof. Dr.-Ing. habil. Stefan Hinz
Korreferent: Prof. Dr.-Ing. Uwe Stilla

Karlsruhe 2015

To my family

Foreword

Over the last decades, 3D point clouds have become a standard data source for automated navigation, mapping and reconstruction tasks. While, in the past, the acquisition of such data was restricted to highly specialized photogrammetric cameras or laser scanners, this situation changed drastically during the last few years. Many common off-the-shelf sensors of modern photogrammetry, remote sensing and computer vision provide 3D point cloud data meanwhile. The available sensor suite includes, for instance, stereo- or multiview vision sensors, laser scanners, time-of-flight cameras and structured light sensors.

Although raw point cloud data delivered by these sensors may give a visually striking impression of the recorded scene, the 3D points are usually unstructured, irregularly distributed and by far not inhering any semantics of the objects. This is the starting point of this work. It develops a coherent concept of 3D point cloud processing with the ultimate goal of identifying the objects present in the scene, given only a few training samples of the sought-after object classes. The modules of the developed system comprise highly sophisticated data processing methods, such as co-registration of different point clouds, geometric pattern analysis including optimal neighborhood selection, feature relevance estimation and feature selection as well as contextual classification. In addition, also the integration of imaging sensors, in particular thermal cameras, in the acquisition system is touched in this work. All methods are fully automated and almost free of parameters set by the user, except those defining the objects to classify.

The implemented concept is not restricted to a specific kind of object class like many other approaches are. Furthermore, it can be applied to large real-world scenes implying millions and millions of 3D points, thereby producing reasonable and high quality results—especially in the light of not needing any user interaction. Both from a scientific and from an application-oriented view, this work is a milestone and a big step forward towards the long-term goal of a fully automated mapping system applicable to complex outdoor scenes. Last but not least, a thorough evaluation based on typical real-world examples elaborates open issues to be tackled in future work.

Karlsruhe
December 2015

Stefan Hinz

Preface

The fully automatic processing and analysis of 3D point clouds represents a topic of major interest in the fields of photogrammetry, remote sensing, computer vision, and robotics. Focusing on general applicability and considering a typical end-to-end processing workflow from raw 3D point cloud data to semantic objects in the scene, we introduce a novel and fully automated framework involving a variety of components which cover (i) the filtering of noisy data, (ii) the extraction of appropriate features, (iii) the adequate alignment of several 3D point clouds in a common coordinate frame, (iv) the enrichment of 3D point cloud data with other types of information, and (v) the semantic interpretation of 3D point clouds. For each of these components, we reflect the fundamentals and related work, and we further provide a comprehensive description of our novel approaches with significant advantages. Based on a detailed evaluation of our framework on various benchmark datasets, we provide objective results for different steps in the processing workflow and demonstrate the performance of the proposed approaches in comparison to state-of-the-art techniques.

Target Audience

Although this book has been written as research monograph, the target audience includes a broad community of people who are dealing with 3D point cloud processing, reaching from students at undergraduate or graduate level to lecturers, practitioners and researchers in photogrammetry, remote sensing, computer vision and robotics. Background knowledge on 3D point cloud processing will certainly improve the reader's appreciation but is not mandatory since reviews on fundamentals and related work are provided for the different components of an end-to-end processing workflow from raw 3D point cloud data to semantic objects in the scene.

Difficulty

Beginner to intermediate and expert. This book is intended to address a broad community by reflecting the fundamentals and related work on the different components of an end-to-end processing workflow from raw 3D point cloud data to semantic objects in the scene. Additionally, a comprehensive description of novel approaches with significant advantages is provided, whereby a detailed evaluation of the proposed approaches in comparison to state-of-the-art techniques is carried out on various benchmark datasets. In this regard, beginners may obtain an impression about the performance of different approaches with respect to different evaluation criteria, whereas experts in the related fields may appreciate the novel approaches and be inspired to conduct further investigations.

Organization of the Content

This book addresses the reconstruction and analysis of 3D scenes and thereby focuses on the presentation of an end-to-end processing workflow from raw 3D point cloud data to semantic objects in the scene. To a reasonable extent, all the material about one topic is presented together in a separate chapter, making the book suitable as a reference work as well as a tutorial on different subtopics.

In Chap. 1, we provide an introduction which focuses on explaining the goals of this work, the main challenges when addressing these goals and the scientific contributions developed in this context. We additionally provide a list of our publications, take a glance view on our novel framework for advanced 3D point cloud processing from raw data to semantic objects and finally give an overview on the topics presented in the whole book.

In Chap. 2, we consider fundamentals in the form of general definitions that will be important for a consideration of our work in the global context. Since, in this work, we focus on the use of 3D point cloud data as input for our methodology, we briefly explain how such 3D point cloud data may be acquired indirectly from 2D imagery via passive techniques and how it may be acquired directly by involving active techniques. Particularly the active techniques offer many advantages and we therefore consider respective systems for data acquisition in the scope of this book. An analysis of such acquisition systems reveals that the acquired 3D point cloud data may easily be represented in the form of range and intensity images which, in turn, might facilitate a variety of applications in terms of simplicity and efficiency. In this regard, we consider point quality assessment where the aim is to quantify the quality of the single range measurements captured with the involved acquisition system, and we present and discuss two novel approaches and their potential in comparison to related work.

In Chap. 3, we briefly explain fundamental concepts for extracting features from 2D imagery and 3D point cloud data. In this regard, we derive a definition for the general term of a *feature* and consider different types of features that might occur in 2D imagery or 3D point cloud data. These fundamentals have become the basis and core of a plethora of applications in photogrammetry, remote sensing, computer vision and robotics, and they also provide the basis for the subsequent chapters.

In Chap. 4, we focus on transforming 3D point clouds—where each 3D point cloud has been acquired with respect to the local coordinate frame of the acquisition system—into a common coordinate frame. We consider related work and thereby particularly focus on efficient approaches relying on the use of keypoints. Based on these considerations, we present a novel framework for point cloud registration which exploits the 2D representations of acquired 3D point cloud data in the form of range and intensity images in order to quantify the quality of respective range measurements and subsequently reliably extract corresponding information between different scans. In this regard, we may consider corresponding information in the respective image representations or in 3D space, and we may also introduce a weighting of derived correspondences with respect to different criteria. Thus, the correspondences may serve as input for the registration procedure for which we present two novel approaches. An evaluation on a benchmark dataset clearly reveals the performance of our framework, and we discuss the derived results with respect to different aspects including registration accuracy and computational efficiency.

In Chap. 5, we focus on the enrichment of acquired 3D point clouds by additional information acquired with a thermal camera. We reflect related work and present a novel framework for thermal 3D mapping. Our framework involves a radiometric correction, a geometric calibration, feature extraction and matching, and two approaches for the co-registration of 3D and 2D data. For the example of an indoor scene, we demonstrate the performance of our framework in terms of both accuracy and applicability, and we discuss the derived results with respect to potential use-cases.

In Chap. 6, we focus on 3D scene analysis where the aim is to uniquely assign each 3D point of a given 3D point cloud a respective (semantic) class label. We provide a survey on related work and present a novel framework addressing neighborhood selection, feature extraction, feature selection and classification. Furthermore, we consider extensions of our framework toward large-scale capability and toward the use of contextual information. In order to demonstrate the performance of our framework, we provide an extensive experimental evaluation on publicly available benchmark datasets and discuss the derived results with respect to classification accuracy, computational efficiency and applicability of involved methods.

In Chap. 7, we summarize the contents of the book and provide concluding remarks as well as suggestions for future work.

Abstract

The fully automatic processing and analysis of 3D point clouds represents a topic of major interest in the fields of photogrammetry, remote sensing, computer vision, and robotics. Focusing on general applicability and considering a typical end-to-end processing workflow from raw 3D point cloud data to semantic objects in the scene, we introduce a novel and fully automated framework involving a variety of components which cover (i) the filtering of noisy data, (ii) the extraction of appropriate features, (iii) the adequate alignment of several 3D point clouds in a common coordinate frame, (iv) the enrichment of 3D point cloud data with other types of information, and (v) the semantic interpretation of 3D point clouds. For each of these components, we reflect the fundamentals and related work, and we further provide a comprehensive description of our novel approaches with significant advantages. Based on a detailed evaluation of our framework on various benchmark datasets, we provide objective results for different steps in the processing workflow and demonstrate the performance of the proposed approaches in comparison to state-of-the-art techniques.

In particular, our derived results reveal that (i) our presented point quality measures allow an appropriate filtering of noisy data and have a positive impact on the automated alignment of several 3D point clouds in a common coordinate frame, (ii) the extraction of appropriate features improves the automated alignment of several 3D point clouds with respect to accuracy and efficiency, and even allows a co-registration of 3D and 2D data acquired with different sensor types, (iii) our presented strategies for keypoint-based point cloud registration in terms of either projective scan matching or omnidirectional scan matching allow a highly accurate alignment of several 3D point clouds in a common coordinate frame, (iv) our presented strategies in terms of a RANSAC-based homography estimation and a projective scan matching allow an appropriate co-registration of 3D and 2D data acquired with different sensor types, and (v) our strategy for increasing the distinctiveness of low-level geometric features via the consideration of an optimal neighborhood for each individual 3D point and our strategy for only selecting compact and robust subsets of relevant and informative features have a significantly beneficial impact on the results of 3D scene analysis. Thus, our novel approaches allow an efficient reconstruction and analysis of large 3D environments up to city scale, and they offer a great potential for future research.

Kurzfassung

Die automatische Verarbeitung und Analyse von 3D-Punktwolken stellt in den Bereichen der Photogrammetrie, Fernerkundung, Computer Vision und Robotik ein wichtiges Thema dar. Im Hinblick auf eine allgemeine Anwendbarkeit wird in der vorliegenden Arbeit eine neue und vollautomatisierte Methodik vorgestellt, welche

die wesentlichen Schritte von der Erfassung von 3D-Punktwolken bis hin zur Ableitung von semantischen Objekten in der Szene betrachtet. Diese Methodik umfasst verschiedene Komponenten, welche (i) die Filterung von verrauschten Daten, (ii) die Extraktion von geeigneten Merkmalen, (iii) die angemessene Ausrichtung von mehreren einzelnen 3D-Punktwolken in einem gemeinsamen Koordinatensystem, (iv) die Anreicherung von 3D-Punktwolken mit zusätzlicher Information und (v) die semantische Interpretation von 3D-Punktwolken umfassen. Für jede Komponente werden die Grundlagen sowie der aktuelle Stand der Forschung aufgezeigt. Ferner werden die im Rahmen dieser Arbeit entwickelten Verfahren mit deutlichen Vorteilen gegenüber den bisherigen Verfahren genauer beleuchtet. Basierend auf einer umfassenden Auswertung auf verschiedenen Standard-Datensätzen werden objektive Ergebnisse für verschiedene Schritte in der Datenverarbeitung präsentiert und die Leistungsfähigkeit der entwickelten Methodik im Vergleich zu Standard-Verfahren verdeutlicht.

Im Speziellen zeigen die im Rahmen der vorliegenden Arbeit erzielten Ergebnisse, (i) dass die entwickelten Qualitätsmaße eine angemessene Filterung von verrauschten Daten ermöglichen und sich positiv auf die automatische Ausrichtung von mehreren einzelnen 3D-Punktwolken in einem gemeinsamen Koordinatensystem auswirken, (ii) dass die Extraktion von geeigneten Merkmalen bei der automatischen Ausrichtung von mehreren 3D-Punktwolken sowohl die Genauigkeit als auch die Effizienz der getesteten Verfahren verbessert und sogar eine Ko-Registrierung von 3D- und 2D-Daten, welche mit verschiedenen Sensortypen erfasst wurden, ermöglicht, (iii) dass die vorgestellten Strategien zur Punkt-basierten Registrierung von 3D-Punktwolken über ein projektives Scan Matching und ein omnidirektionales Scan Matching zu einer sehr genauen automatischen Ausrichtung von einzelnen 3D-Punktwolken in einem gemeinsamen Koordinatensystem führen, (iv) dass die vorgestellten Strategien einer RANSAC-basierten Homographie-Schätzung und eines projektiven Scan Matchings für eine angemessene Ko-Registrierung von 3D- und 2D-Daten, welche mit verschiedenen Sensortypen erfasst wurden, geeignet sind und (v) dass die vorgestellte Strategie zur Erhöhung der Einzigartigkeit von einfachen geometrischen Merkmalen über die Betrachtung einer optimalen Nachbarschaft für jeden individuellen 3D-Punkt sowie die vorgestellte Strategie zur Selektion einer kompakten und robusten Untermenge von relevanten und informativen Merkmalen einen signifikanten, positiven Einfluss auf die Ergebnisse einer 3D-Szenenanalyse haben. Auf diese Weise ermöglichen die entwickelten Verfahren eine effiziente Rekonstruktion und Analyse von großen Bereichen bis auf Stadtgröße und bieten damit großes Potential für zukünftige Forschungsarbeiten.

Note

This book has been developed from a Ph.D. thesis written at the Karlsruhe Institute of Technology (KIT) from 2011 to 2015, further details of which are outlined in the German-language title page, and in the English and German abstracts in the front matter.

Karlsruhe Martin Weinmann
October 2015

Acknowledgments

First and foremost, I would like to thank Stefan Hinz who accepted to supervise this work from the beginning and provided guidance throughout my time as Ph.D. student. He gave me the great opportunity to carry out investigations in different research directions, to become part of the scientific community and to address my research interests. Under his supervision, I felt all the freedom to develop the core of this work.

Furthermore, I would like to thank Boris Jutzi for his constant advice, the numerous valuable discussions, and his availability throughout my time as Ph.D. student. I greatly appreciate his knowledge and I feel that our cooperation has been inspiring and fruitful.

I would also like to express my sincere gratitude to Uwe Stilla for his interest in my work and for acting as referee of this work despite his heavy workload. Furthermore, I am immensely grateful to Boris Jutzi, Clément Mallet, and Franz Rottensteiner for their helpful comments and relevant remarks on different parts of my work. I also deeply appreciate the valuable discussions with my twin brother Michael Weinmann and his constructive feedback throughout the last years.

I am also very grateful to Nicolas Paparoditis for letting me to be part of his research group at IGN during my stays abroad in France. Furthermore, I would like to thank Clément Mallet who supervised my work at IGN. I greatly appreciate his deep knowledge in numerous research domains, his valuable remarks on my work and the many discussions we had over the last years. I feel that I did learn a lot and hope that our cooperation will continue for a long time. I also want to thank the members of IGN-MATIS for their warm welcome, the interesting discussions, and the many events we have been together.

I would also like to thank my co-authors who contributed to the publications representing fundamental parts of this work. These co-authors are Stefan Hinz, Boris Jutzi, Sven Wursthorn, Jens Leitloff, André Dittrich, Steffen Urban, Uwe Stilla, Ludwig Hoegner, Clément Mallet, Franz Rottensteiner, and Alena Schmidt. Furthermore, I thank the anonymous reviewers of these publications for their effort and their constructive feedback with many relevant remarks.

Moreover, I would like to thank KIT-GRACE, the Graduate School for Climate and Environment at the Karlsruhe Institute of Technology (KIT) for funding my work, my travels to conferences, my participation in summer schools and my first two stays abroad in France in order to support the collaboration between KIT-IPF and IGN-MATIS in 2013. Furthermore, I deeply thank the Karlsruhe House of Young Scientists (KHYS) at KIT for funding my third stay abroad in France in order to support the collaboration between KIT-IPF and IGN-MATIS in 2014.

In particular, I also want to thank Sven Wursthorn for the technical support and Ilse Engelmann for the administrative support throughout the last years.

Of course, I am also very grateful to all my colleagues, ex-colleagues, and friends at KIT-IPF for the comradeship throughout the last years, the lively discussions, and the many events we have been together. Particular thanks in this regard go to Boris Jutzi, Sven Wursthorn, André Dittrich, Steffen Urban, Clémence Dubois, Jens Leitloff, Rosmarie Blomley, Simon Schuffert, Ana Djuricic, Ilse Engelmann, Christian Lucas, Werner Weisbrich, Thomas Vögtle, and Uwe Weidner.

Finally, I am gratefully indebted to my parents Ursula and Robert Weinmann, my brother Frank Weinmann and his wife Tatsiana, and my twin brother Michael Weinmann for their unlimited support during all this time.

Karlsruhe Martin Weinmann
October 2015

Contents

Chapter 1
Introduction

For us humans, the rich understanding of our world is mainly guided by visual perception which offers a touchless close-range and far-range gathering of information about objects in a scene, their color, their texture, their shape, or their spatial arrangement. This information, in turn, allows to conclude about (i) the number of objects in a scene, (ii) the properties of objects, e.g., with respect to shape, materials, fragility or value, (iii) the distance of objects and their size, (iv) how to interact with the scene, e.g., by grasping an object where thin surface elements might easily break if the object is grasped in a non-adequate way, or (v) how to move through the local environment. Consequently, the imitation of these human capabilities in theory, design, and implementation allows machines to "see" and it is therefore of utmost importance in the fields of photogrammetry, remote sensing, computer vision, and robotics. However, in order to transfer the capabilities of human visual perception to fully automated systems, we may involve several and different sensor types such as digital cameras, range cameras, or laser scanners, and we may even go beyond the human capabilities by for instance involving thermal cameras or multispectral cameras which perceive information in the non-visible spectrum.

Instead of imitating human visual perception, we focus on transferring specific capabilities of human visual perception to fully automated systems and thereby involving different data sources in order to capture different and complementary aspects of real-world objects or scenes. In the scope of this work, we consider those sensor types which provide 3D point cloud data representing the densely sampled and accurately described counterpart of physical object surfaces as most important data source. The nature of such 3D point cloud data requires data structures and processing methods which differ from those used for 2D imagery, yet a representation of a scene in the form of a 3D point cloud provides a sampling of physical object surfaces which merits attraction for a variety of applications. Regarding for instance the acquisition of indoor and outdoor scenes, 3D point cloud data acquired with terrestrial laser scanning systems (e.g., Leica HDS6000 or Riegl LMS-Z360i), mobile laser scanning systems (e.g., L3D2 [1] or Stereopolis II [3]) or range cameras (e.g., PMD[vision] CamCube 2.0, MESA Imaging SR4000 or Microsoft Kinect) are increasingly used for applications reaching from 3D scene reconstruction via geometry processing,

M. Weinmann, *Reconstruction and Analysis of 3D Scenes*,
DOI 10.1007/978-3-319-29246-5_1

shape and appearance modeling, object detection, object recognition, and 3D scene interpretation to large-scale issues. While a framework containing numerous state-of-the art algorithms including filtering, feature estimation, surface reconstruction, registration, model fitting, and segmentation has been released with the Point Cloud Library (PCL) [4] which represents a standalone, large-scale, open project for 2D/3D image and 3D point cloud processing, further improvements toward an advanced 3D point cloud processing are still desirable.

A respective improvement may generally address a variety of different issues amongst which three crucial tasks may be identified. First, it has to be taken into account that, depending on the involved acquisition system, occlusions resulting from objects in the scene may be expected as well as variations in point density which becomes visible in Fig. 1.1 for the visualization of a 3D point cloud acquired with a terrestrial laser scanner. Thus, an adequate digitization and 3D reconstruction of a scene often relies on a combination of several acquired 3D point clouds describing different parts of the considered scene, and appropriate approaches for such a combination of 3D point clouds are required. Second, it might be desirable to map complementary types of data acquired with different sensor types (which typically provide data in the form of images) onto already available 3D point clouds, e.g., in order to obtain a photo-realistic depiction of the considered scene or in order to visualize the heat distribution in the scene. In this regard, for the 3D point cloud depicted in Fig. 1.1, radiometric data could for instance be useful to differentiate between different construction types for buildings as shown in Fig. 1.2. Third, the 3D point cloud depicted in Fig. 1.1 is already sufficient for us humans in order to conclude about specific objects in the scene such as buildings, ground, road inventory, or cars and, hence, it seems desirable to realize such a semantic interpretation of the

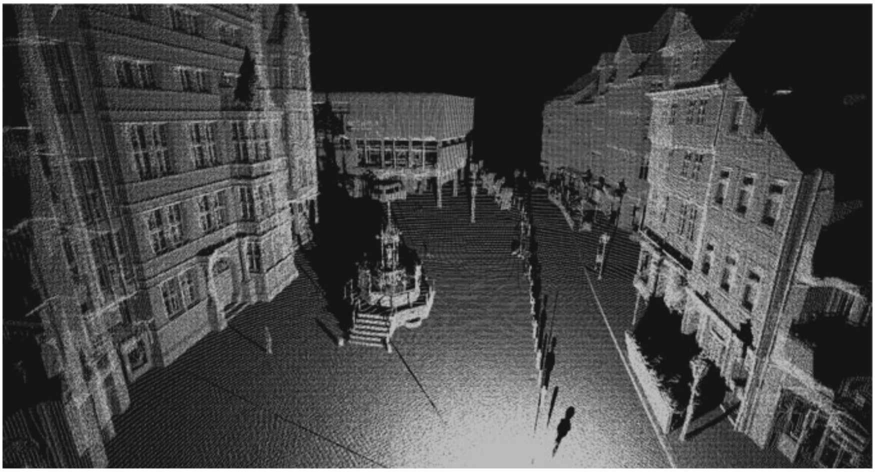

Fig. 1.1 Visualization of a 3D point cloud acquired with a terrestrial laser scanner: occlusions resulting from objects in the scene as well as a varying point density become clearly visible

Fig. 1.2 Visualization of a 3D point cloud acquired with a terrestrial laser scanner and colored with respect to the respective radiometric data

considered scene in a fully automated way. While these three tasks represent fundamental problems in photogrammetry, remote sensing, computer vision, and robotics—and have therefore been addressed in a large number of investigations—respective investigations have mainly focused on a selection of specific subtasks, whereas potential sources for further improvements have not adequately been addressed yet.

1.1 Goals

Generally, in order to understand a scene, we need to reason about the distance and spatial arrangement of objects in the world, gather various properties of these objects and capture the semantic meaning of the world. Accordingly, the understanding of our world via visual perception involves a mapping which, in turn, is composed of three different, but possibly interrelated mappings:

- a *geometric mapping* which maps the real world onto a geometric representation (e.g., a 3D point cloud or a 3D model),
- a *radiometric mapping* which maps the real world onto a radiometric representation (e.g., a color image or a gray-valued image), and
- a *semantic mapping* which maps the real world onto a semantic representation (e.g., a labeling with respect to different objects, different materials, accessible locations, abstract elements, etc.).

Since we focus on transferring capabilities of human visual perception to fully automated systems instead of imitating human visual perception, we may involve several and very different sensors as well as a variety of algorithms in order to obtain an appropriate mapping between the real world and abstract data. Thereby, we have to take into account both the limitations and the interrelations between respective mappings:

- There are sensor types which allow a direct geometric mapping, e.g., laser scanners or range cameras. Furthermore, a geometric mapping may also be derived indirectly by for instance reconstructing a scene from 2D imagery acquired with digital cameras.
- Many sensor types allow a radiometric mapping, e.g., digital cameras, thermal cameras, multispectral cameras, or even modern scanning devices such as laser scanners or range cameras.
- None of the available sensor types directly allows a semantic mapping.
- If we cannot directly obtain a semantic mapping, then we need to exploit the data obtained via a geometric mapping and/or a radiometric mapping in order to allow a semantic reasoning.
- If we cannot directly obtain a geometric mapping, then we need to exploit the data obtained via a radiometric mapping in order to allow a geometric reasoning.
- If we use different sensors for a geometric mapping and/or a radiometric mapping, then we need to correctly align the captured data which is commonly referred to as co-registration of data or data fusion.

The goal of our work consists of the development of novel concepts and methods for the digitization, reconstruction, interpretation, and understanding of static and dynamic indoor and outdoor scenes. Accordingly, we will first focus on the geometric mapping, where we may involve a variety of sensors in order to either directly or indirectly acquire 3D point cloud data representing the measured or derived counterpart of the real world. In this regard, we focus on a complete, dense, and accurate digitization/reconstruction of indoor and outdoor environments and, consequently, it might be advisable to exploit scanning devices in order to directly acquire 3D point cloud data from different viewpoints. As, for each viewpoint, the acquired 3D point cloud data is only measured with respect to the local coordinate frame of the sensor, the data acquired from different viewpoints has to be aligned in a common coordinate frame which is commonly referred to as point cloud registration and addressed in Chap. 4. In this regard, we take into account that most of the currently available scanning devices already provide radiometric information which may significantly facilitate an appropriate alignment. For scene interpretation, we follow the idea of involving further data sources in order to capture different and complementary aspects of real-world objects or scenes, where the additional information is typically represented in the form of 2D imagery. As a consequence, we will have to focus on an adequate co-registration of the acquired data in terms of a mapping of complementary information onto the already available 3D point cloud data which is addressed in Chap. 5. Based on all available data, a subsequent scene interpretation may be carried out in order to derive a semantic mapping and we will address such a semantic reasoning in Chap. 6. Thus, in a summary, the main goals of our work consist of

- a faithful reconstruction of 3D scenes (in both indoor and outdoor environments),
- data fusion in terms of an enrichment of existing 3D point cloud data by adding external data from multiple, but complementary data sources, and
- a semantic reasoning based on all available information.

As input data, we consider (partial) representations of a 3D scene in the form of raw 3D point cloud data consisting of a large number of 3D points as measured counterpart of physical surfaces in the local environment of the sensor, and raw data acquired via complementary data sources. All acquired data is provided to a "black box"—which we will explain in detail in this book—and the derived output data consists of a global reconstruction of the considered scene in the form of a 3D point cloud with correctly aligned radiometric information and respectively assigned point-wise semantic labels.

1.2 Challenges and Main Contributions

As outlined in the previous section, we address the tasks of (i) point cloud registration, (ii) co-registration of 3D point cloud data and complementary information represented in the form of 2D imagery and (iii) 3D scene analysis in the scope of our work. The scientific key contributions presented in this book are:

- **A geometric quantification of the quality of range measurements**:
  In order to assess the quality of scanned 3D points, we consider a 3D point cloud and the respectively derived 2D image representations in the form of range and intensity images. Based on the given range image, we propose two novel image-based measures represented by *range reliability* [8] and *local planarity* [11] in order to assess the quality of scanned 3D points. We explain the chances and limitations of both measures in comparison to other alternatives, and we demonstrate the performance of the proposed measures in the context of filtering raw 3D point cloud data. Furthermore, we demonstrate how to exploit these measures in order to define a consistency check which allows us to filter unreliable feature correspondences between the image representations of different scans and thus to facilitate an efficient and robust registration of 3D point clouds by retaining only those feature correspondences with reliable range information.
- **A shift of paradigms in point cloud registration**:
  Based on our investigations on feature extraction from 2D imagery [6] and 3D point clouds [7], we may conclude that local features are well-suited in order to allow an efficient and robust registration of 3D point clouds. The respective category of registration approaches—which we refer to as *keypoint-based point cloud registration approaches*—has indeed been addressed in literature for years. However, respective investigations mainly focused on the use of specific types of local features representing the most prominent approaches for deriving feature correspondences. Instead, our investigations also take into account that other approaches for extracting local features may even be more promising [5]. Furthermore, recent