

Indoor Mapping and Reconstruction with Mobile Augmented Reality Sensor Systems

Zur Erlangung des akademischen Grades eines

DOKTOR-INGENIEURS

von der KIT-Fakultät für
Bauingenieur-, Geo- und Umweltwissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von MSc.

Patrick Hübner

aus Karlsruhe

Tag der mündlichen Prüfung: 29.10.2021

Referent: Prof. Dr.-Ing. habil. Stefan Hinz
Institut für Photogrammetrie und Fernerkundung (IPF)
Karlsruher Institut für Technologie (KIT)

Korreferent: Prof. Dr. Ir. George Vosselman
Faculty of Geo-Information Science and Earth Observation (ITC)
Department of Earth Observation Science (EOS)
University of Twente

Karlsruhe 2021

Acknowledgment

My sincere gratitude and appreciation

To	For
Stefan Hinz	Providing the possibility to freely pursue my research interests Valuable supervision & advice
George Vosselman	Being second reviewer of this work
Martin Weinmann	Invaluable guidance & fruitful discussions
Sven Wursthorn	Our shared success, good times & lots of fun
The whole IPF	Making this institute what it is
Dennis Haitz Kristoffer Schneider	Persistent proofreading
My parents, Jürgen & Christiane	Selflessly providing me with a foundation in life, without which my achievements would not have been possible in the first place
LisMi	Supporting me in the daily struggle Making life worthwhile

This work is licensed under a Creative Commons “Attribution-NonCommercial-ShareAlike 4.0 International” license.



Abstract

Augmented reality (AR) is generally well-suited for the interactive visualization of all kinds of virtual, three-dimensional data directly within the physical environment surrounding the user. Beyond that, AR holds the potential of not only visualizing arbitrary virtual objects anywhere but to visualize geospatial data directly in-situ in the location that the data refer to. Thus it can be used to enrich a part of the real world surrounding the user with information about this environment and the physical objects within it. In the scope of this work, this usage mode is defined and discussed under the term of 'fused reality'.

An appropriate scenario to demonstrate and elaborate on the potential of fused reality is its application in the context of digital building models, where building specific information, e.g. about the course of pipelines and cables within the walls, can be visualized directly in the respective location. In order to realize the envisioned concept of indoor fused reality, some principal requirements must be fulfilled. Among these is the need for an appropriate digital model of a building environment at hand which is to be enriched with virtual content. While building projects are nowadays oftentimes designed and executed with the help of building information modeling techniques, appropriate digital representations of older stock buildings are usually hard to come by. If a corresponding model of a given building environment is available, the respective AR device needs to be able to determine its current position and orientation with respect to the model in order to realize a correct registration of the physical building environment and the virtual content from the model. In this work, different aspects about how to fulfill these requirements are investigated and discussed.

First, different ways to map indoor building environments are discussed in order to acquire raw data for constructing building models. In this context, an investigation is presented about whether a state-of-the-art AR device can be deployed to this task as well. In order to generate building models based on this indoor mapping data, a novel, fully-automated, voxel-based indoor reconstruction method is presented and evaluated on four datasets with corresponding ground truth data that were acquired to this aim. Furthermore, different possibilities to localize mobile AR devices within indoor environments are discussed and the evaluation of a straight-forward, marker-based approach is presented. Finally, a novel method for aligning indoor mapping data with the coordinate axes is presented and evaluated.

Zusammenfassung

Augmented Reality (AR) ermöglicht es, virtuelle, dreidimensionale Inhalte direkt innerhalb der realen Umgebung darzustellen. Anstatt jedoch beliebige virtuelle Objekte an einem willkürlichen Ort anzuzeigen, kann AR Technologie auch genutzt werden, um Geodaten in situ an jenem Ort darzustellen, auf den sich die Daten beziehen. Damit eröffnet AR die Möglichkeit, die reale Welt durch virtuelle, ortbezogene Informationen anzureichern. Im Rahmen der vorliegenden Arbeit wird diese Spielart von AR als "Fused Reality" definiert und eingehend diskutiert.

Der praktische Mehrwert, den dieses Konzept der Fused Reality bietet, lässt sich gut am Beispiel seiner Anwendung im Zusammenhang mit digitalen Gebäudemodellen demonstrieren, wo sich gebäudespezifische Informationen - beispielsweise der Verlauf von Leitungen und Kabeln innerhalb der Wände - lagegerecht am realen Objekt darstellen lassen. Um das skizzierte Konzept einer Indoor Fused Reality Anwendung realisieren zu können, müssen einige grundlegende Bedingungen erfüllt sein. So kann ein bestimmtes Gebäude nur dann mit ortsbezogenen Informationen augmentiert werden, wenn von diesem Gebäude ein digitales Modell verfügbar ist. Zwar werden größere Bauprojekte heutzutage oft unter Zuhilfenahme von Building Information Modelling (BIM) geplant und durchgeführt, sodass ein digitales Modell direkt zusammen mit dem realen Gebäude entsteht, jedoch sind im Falle älterer Bestandsgebäude digitale Modelle meist nicht verfügbar. Ein digitales Modell eines bestehenden Gebäudes manuell zu erstellen, ist zwar möglich, jedoch mit großem Aufwand verbunden. Ist ein passendes Gebäudemodell vorhanden, muss ein AR Gerät außerdem in der Lage sein, die eigene Position und Orientierung im Gebäude relativ zu diesem Modell bestimmen zu können, um Augmentierungen lagegerecht anzeigen zu können.

Im Rahmen dieser Arbeit werden diverse Aspekte der angesprochenen Problematik untersucht und diskutiert. Dabei werden zunächst verschiedene Möglichkeiten diskutiert, Indoor-Gebäudegeometrie mittels Sensorsystemen zu erfassen. Anschließend wird eine Untersuchung präsentiert, inwiefern moderne AR Geräte, die in der Regel ebenfalls über eine Vielzahl an Sensoren verfügen, ebenfalls geeignet sind, als Indoor-Mapping-Systeme eingesetzt zu werden. Die resultierenden Indoor Mapping Datensätze können daraufhin genutzt werden, um automatisiert Gebäudemodelle zu rekonstruieren. Zu diesem Zweck wird ein automatisiertes, voxel-basiertes Indoor-Rekonstruktionsverfahren vorgestellt. Dieses wird außerdem auf der Grundlage vierer zu diesem Zweck erfasster Datensätze mit zugehörigen Referenzdaten quantitativ evaluiert. Desweiteren werden verschiedene Möglichkeiten diskutiert, mobile AR Geräte innerhalb eines Gebäudes und des zugehörigen Gebäudemodells zu lokalisieren. In diesem Kontext wird außerdem auch die Evaluierung einer Marker-basierten Indoor-Lokalisierungsmethode präsentiert. Abschließend wird zudem ein neuer Ansatz, Indoor-Mapping Datensätze an den Achsen des Koordinatensystems auszurichten, vorgestellt.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Fundamental Concepts	1
1.2.1	Digital Building Models	2
1.2.2	Indoor Environments	4
1.2.3	Augmented Reality	4
1.2.4	Augmented Reality vs. Fused Reality	6
1.2.5	Indoor Fused Reality	8
1.3	Scope of this Work	9
2	Indoor Mapping	11
2.1	Fundamentals of Indoor Mapping	12
2.1.1	Indoor Mapping Systems	13
2.1.2	Evaluation of Indoor Mapping Systems	14
2.2	Methodology	15
2.2.1	Sensor Description	15
2.2.2	Evaluation Method	18
2.2.2.1	Depth Sensing	18
2.2.2.2	Tracking	20
2.2.2.3	Indoor Mapping	22
2.3	Results	23
2.3.1	Depth Sensing	23
2.3.2	Tracking	25
2.3.3	Indoor Mapping	29
2.4	Discussion	33
2.4.1	Depth Sensing	33
2.4.2	Tracking	34
2.4.3	Indoor Mapping	35
2.5	Conclusion and Outlook	37
3	Indoor Reconstruction	38
3.1	Fundamentals of Indoor Reconstruction	38
3.2	Methodology	42
3.2.1	Data Preparation	44
3.2.2	Ceiling and Floor Reconstruction	44
3.2.2.1	Ceiling Detection	44
3.2.2.2	Ceiling Refinement	46
3.2.2.3	Floor Detection	46
3.2.2.4	Ceiling and Floor Finalization	46
3.2.3	Voxel Classification	47
3.2.4	Voxel Model Completion	47
3.2.5	Voxel Model Refinement	48
3.2.5.1	Wall Geometry Refinement	48
3.2.5.2	Wall Opening Refinement	49

3.2.6	Room Segmentation	50
3.2.6.1	Indoor Spaces	51
3.2.6.2	Ramp Space Detection	52
3.2.6.3	Indoor Space Partitioning	52
3.2.7	Evaluation	55
3.3	Results	56
3.4	Discussion	69
3.5	Conclusion and Outlook	73
4	Indoor Localization	76
4.1	Fundamentals of Indoor Localization	77
4.2	Methodology	79
4.2.1	Localization Method	79
4.2.2	Implementation	81
4.2.3	Evaluation Scheme	83
4.3	Results	84
4.3.1	Observations on the HoloLens Camera	84
4.3.2	Evaluation of the Localization Method	86
4.4	Discussion	88
4.5	Conclusion and Outlook	89
5	Pose Normalization of Indoor Mapping Datasets	90
5.1	Related Work	92
5.2	Methodology	93
5.2.1	Rotation around the Vertical Axis	94
5.2.2	Orientation of the Vertical Axis	97
5.2.3	Unambiguousness of the Rotation around the Vertical Axis	102
5.2.4	Evaluation Method	105
5.3	Results	106
5.4	Discussion	111
5.5	Conclusion and Outlook	119
6	Synopsis	120
6.1	Summary	120
6.2	Discussion and Outlook	122
6.3	Conclusion	124
	List of Figures	125
	List of Tables	127
	Bibliography	128

Chapter 1

Introduction

1.1 Motivation

Currently, augmented reality (AR) and digital building models are both considerably active fields of research. Both promise an abundance of innovation and exciting perspectives for future applications. The topic of this work - the fusion of both these technologies - offers tremendous potential for both fields. On the one hand, the application within the context of digital building models provides the possibility to unfold the true potential of augmented reality by demonstrating its capabilities as an in-situ interface for the visualization of and interaction with geospatial data - directly in the location the respective geodata refer to. On the other hand, application domains concerned with digital building models such as architecture, construction, facility management, civil engineering and navigation can benefit significantly from what augmented reality technology has to offer them.

In this context, we can exemplarily envisage a mechanic, being guided by a navigation system via audio-visual hints on the way through a large facility complex to the place of commitment. There, a damage case needs repair work as automatically reported by the intelligent building management system. Navigation hints can be given by synthetic speech commands but also by the visualization of arrows and highlighting of relevant parts of the building environment directly within the field of view of the fictive repair person wearing AR glasses. Arriving at the destination, the faulty pipelines needing repair work are visualized within the walls, as if by x-ray vision. After executing the necessary repair work, the mechanic can update the properties of the replaced part of the pipeline in the digital building model directly in-place by way of gesture and voice commands.

Investigating the technological requirements necessary for realizing this illustrative example for the potential of applying augmented reality to digital models of building environments is the central topic of this work. To this aim, various relevant aspects to this theme are examined in the following chapters. Prior to this, Sec. 1.2 illustrates the fundamental concepts relied upon in this work: digital building models, augmented reality and the application of the latter in the context of the former. Subsequently, Sec. 1.3 gives an overview about the following chapters and the structure of this work.

1.2 Fundamental Concepts

This section introduces the fundamental concepts this work relies upon. First, Sec. 1.2.1 gives an overview of the different ways available to digitally represent building environments and possible applications thereof. Then, Sec. 1.2.2 presents a discussion on the peculiarities and characteristics of indoor environments. Afterwards, Sec. 1.2.3 defines the concept of augmented reality and presents the different technologies available for its realization. Subsequently, in Sec. 1.2.4, the concept of

fused reality as a theoretical contribution of this work is defined and discussed. Lastly, Sec. 1.2.5 exemplifies the vision of indoor fused reality by illustrating its numerous fields of application and its potential value therein.

1.2.1 Digital Building Models

In general, a multitude of possibilities exists to digitally represent building environments [102, 327]. One rather straight-forward way is to simply represent buildings directly by the raw data acquired in-situ when mapping a physically existing building - i.e. by point clouds, triangle meshes or depth images (see Chap. 2 on indoor mapping). However, this holds some disadvantages as indoor mapping data - especially point clouds - can quickly amount to large quantities, impeding their storage and processing. Furthermore, these large amounts of acquired raw data only represent the bare building geometry as captured by the respective sensors. Thus, they can be subject to a large amount of measurement noise and incompleteness due to fragmentary acquisition while lacking any immediately accessible semantic information.

Nonetheless, on some occasions, semantically enriched point clouds have been proposed as an adequate means of representing buildings or urban environments [465, 460, 463, 149]. This requires a process of semantic point cloud segmentation [620, 312, 464] which can be applied to structure the raw point cloud data in a way to make it accessible for semantic queries as e.g. in the case of the 'smart point cloud' infrastructure proposed by Poux and Billen [463]. As discussed by Döllner [149], this approach has the advantage that geometric modeling including simplification and abstraction is avoided. Thus, a large degree of detail and genericity is preserved.

Applying an analogous approach to depth images or RGBD images is in theory also conceivable. However, while images have been frequently applied to the task of reconstructing building models [246, 455, 335] (see Chap. 3), so far, pre-structured image data has rarely been used as queryable representation of building environments. In this context, urban models constituted of georeferenced depth images were proposed in the context of smart city applications [410].

Still today, 2D floor plans are a commonly used representation of building environments. Traditionally, they are used in analog, printed form. Due to their printable nature, semantic information is typically encoded in a way that makes it visually retrievable, i.e. by style conventions or text labels. Thus, automatically interpreting floor plans is not a straight-forward endeavor [9, 148, 344]. While analog floor plans are being more and more replaced by digital, three-dimensional representations of buildings, they are still an important means of architectural analysis of building layouts [9, 493].

Initially, three-dimensional, digital representations of buildings were mainly based on computer-aided design (CAD) techniques [625, 377, 134]. CAD models, like floor plans however, mainly focus on representing geometry while semantic information, again, has to be encoded by visualization styles or relying on the layer principle [234].

Later, CAD-based representations were increasingly superseded by building information modeling (BIM) [72] in the domain of architecture, engineering and construction (AEC). The importance of BIM-based building representations has since steadily increased [192, 497]. In the context of BIM, physical (such as walls, doors or building components like piping systems or ventilation ducts) as well as non-physical (like spaces partitioning the building interior into semantically defined sections) building components are modeled as parametrized objects and typically represented by voluminous constructive solid geometry (CSG) [488]. While in practice, proprietary data formats such as Autodesk Revit¹ or Graphisoft Archicad² are frequently used, from a scientific perspective, open BIM standards

¹<https://www.autodesk.de/products/revit> (Last visited on 28/05/2021)

²<https://graphisoft.com/de/archicad> (Last visited on 28/05/2021)

[254, 431] such as Industry Foundation Classes (IFC)³ [307] are certainly more interesting.

BIM models can be gainfully applied to the benefit of all stages of the building life cycle [549, 55, 385], i.e. in design and planning [374, 128], construction [167, 252], in the operational phase in the context of facility management [414, 390, 56, 184], in the case of modifications and renovation [146, 528] and eventually for dismantling and deconstruction [13, 11, 589]. More specifically, BIM models can e.g. be applied for clash detection among planned building elements [14, 90], for ensuring conformity with construction norms and regulations [191, 565] as well as for building performance analysis in the context of energy efficiency [202, 26, 169, 255]. Furthermore, BIM is frequently applied in the cultural heritage domain as historical BIM (HBIM) [352, 28, 531].

While BIM is widely used in the AEC domain, in the context of geoinformation Sciences (GIS), another way to represent buildings is more prevalent. Here, CityGML⁴ [295, 204] is a standard for the representation of digital city models for applications in the domain of urban planning [61, 543]. While the standard is originally based on Extensible Markup Language (XML)⁵, a more compact, JSON⁶-based variant called CityJSON is currently gaining in favor [316]. Contrary to BIM, where CSG-based geometry is typically used, CityGML is based on a boundary representation describing the geometry of objects by their surface instead of a voluminous representation. While originally focused more on the exterior of buildings, the latest version of CityGML [304] increases the capabilities for modeling indoor spaces over all levels of detail.

Another standard dealing with the representation of building environments is IndoorGML [264, 326, 139]. Contrary to BIM and CityGML, IndoorGML sets the focus less on representing building geometry itself but is more concerned with describing indoor spaces and their connecting topology in the context of indoor navigation applications [165, 301, 628, 629]. While indoor spaces may have associated geometries, this is not mandatory.

Currently, a topic of ongoing research is the interoperability between these different standards for representing building environments. Particularly, the conversion between the worlds of BIM and GIS (i.e. CityGML and IndoorGML) is still a challenging task [8, 225, 562].

Lastly, building environments can also be described by a voxel representation [691, 646] which is also used in this work for the indoor reconstruction approach presented in Chap. 3. Voxel grids may not be a particularly compact way to model buildings in terms of storage needs. However, they provide a straight-forward means for explicitly representing empty space and thus make it easily accessible in the context of building analysis applications aiming at the building interior. This makes them attractive for applications like visibility analysis [142, 16], navigation [537, 201, 200, 605] and all kinds of building performance simulation [197] like structural analysis [87] or fire simulation [534].

Nowadays building facilities are often-times planned and constructed relying on BIM techniques. In this case, detailed digital models of the respective buildings are available as they arise together with the buildings themselves. In the case of existing stock buildings however, corresponding digital models are often not available [595, 262]. While manually reconstructing digital models for existing buildings (as-built BIM) is indeed possible and frequently applied, it can be a tedious and time-consuming endeavor when relying on individual manual distance measurements taken in the building [444, 56]. In this context, indoor mapping technology as discussed in Chap. 2 can be a valuable means for the quick and efficient acquisition of building structures. The resulting indoor mapping data can subsequently be subjected to a process of indoor reconstruction as discussed in Chap. 3 in order to automatically generate digital building models for existing buildings.

³<https://www.buildingsmart.org/standards/bsi-standards/industry-foundation-classes/>
(Last visited on 28/05/2021)

⁴<http://www.citygml.org/> (Last visited on 28/05/2021)

⁵<https://www.w3.org/XML/> (Last visited on 06/06/2021)

⁶<https://www.json.org/> (Last visited on 28/05/2021)

1.2.2 Indoor Environments

Indoor environments are characterized by some peculiar properties that distinguish them from outdoor spaces. This, along with the fact that humans today typically spent the majority of their time within buildings [77, 278], has brought them to the focus of a range of theoretical studies. Within the context of indoor cartography, Chen and Clarke [103] provide an elaborate overview of different definitions of 'indoors'. Recently, an extensive survey of different concepts of space, explicitly including indoor spaces, in a multitude of different application domains was presented by Zlatanova et al. [693]. Sithole and Zlatanova [529] characterize concepts such as 'position', 'location' and 'place' in the context of indoor environments. Furthermore, the transition between indoor and outdoor environments is an active topic of research as well [297, 649].

Winter et al. [628, 629] discuss the particularities and challenges of indoor environments in the context of navigation and localization. They point out, that indoor spaces are highly structured environments decomposed in sub-spaces (i.e. rooms) and typically also structured vertically, encompassing different height levels (i.e. storeys). Despite typically having a complex structure, indoor spaces can also be characterized by a high degree of symmetry and structural repetitions on the level of different wings, floors or rooms. Besides this spatially complex situation, indoor environments can also be subject to temporal complexity as they are highly dynamic environments w.r.t. the placement of furniture, usage and access restrictions. Furthermore, global reference systems like global navigation satellite systems (GNSS) are typically not available within indoor environments.

Lehtola et al. [319, 320] point out, that indoor environments can have relevant spatial structures on a wide range of scales, depending on the respective focus of interest. These can encompass the course of cables with diameters of some millimeters as well as the course of large hallways potentially extending over tens or hundreds of meters. This, as well as the fact that indoor environments are often characterized by a high amount of clutter objects like furniture, makes them a particularly challenging environment w.r.t. the task of indoor mapping (see Chap. 2). In this context, Zlatanova et al. [692] also present an analysis of problems in indoor mapping and modeling. This is however more focused on the current state of technological development (at that time) and less on the particularities of indoor environments themselves. Some of the identified problems like a lack of mobility among the available indoor mapping systems are meanwhile solved to a large extent (see Chap. 2).

1.2.3 Augmented Reality

Augmented reality (AR) is a visualization technique that allows to display three-dimensional, virtual objects directly within a real, physical environment [287, 111, 338]. While the term 'augmented reality' was first introduced by Caudell and Mizell [88], Azuma [37] provided a definition that is today widely accepted. According to it, AR is characterized 1) by spatially combining virtual and real objects in a real environment, 2) by working in real-time and 3) by being interactive. AR is closely related to the concept of virtual reality (VR), which also provides interactivity and real-time capability but is limited to a purely virtual environment without any spatial relation to the real world [686, 663].

Due to recent marketing campaigns of AR and VR hardware vendors⁷, the term 'mixed reality (MR)' is currently in the focus of public attention. MR was originally introduced by Milgram and Kishino [386] in the context of a mixed reality continuum encompassing different intermediate stages between a purely virtual reality and the real world with AR being situated somewhere in-between. Due to the recent developments however, there is currently a considerable ambiguity in the usage of

⁷e.g. <https://docs.microsoft.com/en-us/windows/mixed-reality/discover/mixed-reality>
(Last visited on 29/05/2021)

the term 'mixed reality'. While some consider it to still refer to the concept of the MR continuum, others meanwhile regard it as a synonym to AR, as some kind of combination of AR and VR or as a somehow 'evolved' version of AR providing a higher degree of immersion [536]. Furthermore, Hönig et al. [230] provide an extended definition of MR by involving not just one but multiple virtual environments which can be useful in the context of human-robot and multi-robot interaction.

Generally, AR is frequently applied in the field of human-robot interaction [210, 271, 627]. Other applications of AR include tourism [135, 667, 125], game design [483, 310], medicine [209, 153, 303, 398] and education [12, 29, 156]. Furthermore, AR is also popular in museums and cultural heritage [570, 440, 89] as well as in the context of navigation applications [189, 259, 664, 558]. Besides, it can be benefiting to industry applications [17, 188, 302, 681], e.g. for assembly guidance tasks [611, 389, 99].

A range of different hardware platforms can be deployed to realize AR applications. Some AR systems are designed as stationary installations. In this context, half-transparent mirrors [62, 286] or projector systems [277, 321] are often used for visualization. While suchlike stationary systems are suited e.g. for application in museum exhibitions, mobile AR systems [97] are generally more flexible in their deployment. Among these, projector-based visualization is also feasible [513, 662]. However, hand-held systems like tablets or smartphones [220, 475] and head-worn systems [40] are by far more common.

The advantages of smartphones and tablets as AR devices are their prevalence in the general public, low cost and them being off-the-shelf equipped with a broad range of sensors such as GNSS receivers, inertial measurement units (IMU) and cameras. For these devices, video-based visualization on the screen is used to display augmentations. This requires the operator to point the device towards the scene where augmentations are to be displayed - a usage mode known as 'magic lens' [351]. Head-worn AR glasses on the other hand, can be used more intuitively as they directly augment the current field of view of the operator wearing the device. In this context, visualizations can also be provided via video screens [539, 451] or by transparent optical see-through displays [552, 162]. The Microsoft HoloLens⁸ for instance, a current state-of-the-art AR glasses that is also used in the scope of this work, uses the latter visualization technique [298].

In order to provide augmentations that are perceived as geometrically stable and convincing independently from the movement of the user, AR devices need to be able to determine their pose (i.e. position and orientation) with respect to the surrounding environment in real-time. This task is commonly referred to as 'tracking' [36, 368, 657].

To this aim, a range of different tracking methods can be deployed. Among these, marker-based tracking methods use planar, artificial markers as reference [272, 186, 70]. This is a simple and efficient approach, however it requires the placement of artificial markers within the environment to be augmented. Similarly, three-dimensional, rigid objects with known geometry can also be used as reference for tracking [86, 638, 2, 217].

The requirement of placing artificial references in an environment that is to be augmented can be alleviated by relying on natural features as reference for tracking [435]. In this context, visual odometry (VO) [459, 221] or simultaneous localization and mapping (SLAM) [180, 538, 608, 268] algorithms can be used. As in many fields of research, deep learning techniques are currently also frequently applied to the task of marker-less tracking [640, 648, 438].

Besides depending on a single tracking system, hybrid tracking systems rely on a fusion of multiple independent tracking systems using different sensors with complementary properties [164, 323, 613]. Furthermore, a tracking system without visualization capabilities can be complemented to a full AR system by coupling it with an independent visualization system by means of relative tracking [240].

⁸<https://www.microsoft.com/en-us/hololens> (Last visited on 29/05/2021)

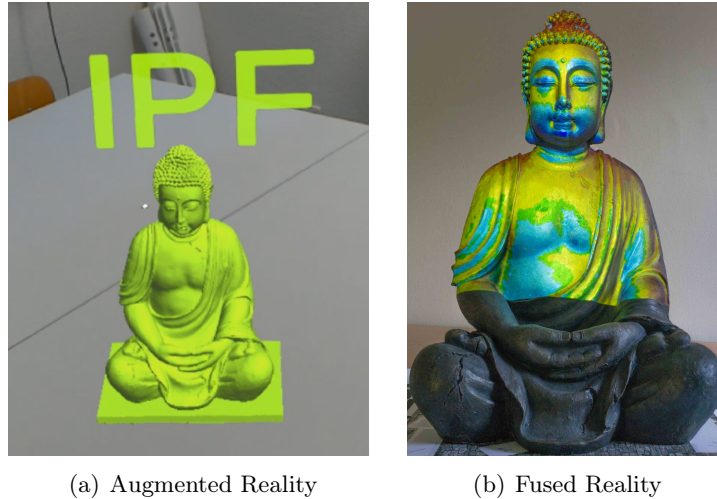


Figure 1.1: Example for the conceptual difference between (a) augmented reality and (b) fused reality. While AR simply encompasses the placement of arbitrary virtual objects like the stature depicted in (a) within the real world, FR implies the augmentation of a physically existing object with additional, virtual information. In (b) for instance, a physical statue is augmented with colors representing the difference between two scans of the statue. Source of (b): [277].

1.2.4 Augmented Reality vs. Fused Reality

The definition of AR as discussed in Sec. 1.2.3 leaves room to a broad range of applications that can be categorized as AR. These can however differ significantly from each other on a conceptual level. This section points out that there actually is a conceptual difference between two usage modes that are currently both subsumed under the term of augmented reality.

The Microsoft HoloLens for instance is frequently referred to as an 'indoor AR device'. This designation however only aims at the fact that the device is intended to be used within indoor environments. The concept of many HoloLens applications comprises placing virtual objects within the indoor environment around the user wearing the device. In this context, the placement can be realized by letting the user select the position where a virtual object should be visualized. Alternatively, the HoloLens is also able to analyze the surrounding environment by means of its depth sensor (see Chap. 2) and autonomously decide on a suitable location for placement. For example, a virtual statue can be placed on a real physical table surface as depicted in Fig. 1.1(a). This allows the user to get a good impression of the respective rendered object by walking around the table and observing it from different perspectives. Meanwhile, the virtual object itself stays in its apparent position, independent from the movement of the user.

This kind of three-dimensional data visualization certainly holds great potential, e.g. for the better comprehension of complex three-dimensional objects [229, 676] or for getting an impression of how a particular piece of furniture would look in a given place [436]. However, this usage mode is nonetheless qualitatively different from the indoor AR scenarios presented in Sec. 1.1 and in the following Sec. 1.2.5 where the indoor environment itself is to be augmented with spatially related information referring to said indoor environment itself. This concept is demonstrated in Fig. 1.1(b) by an example from [277]. Here, a physical statue is augmented by a projector-based AR system with virtual, color-coded information presenting the difference between scans of said statue acquired with two different scanning systems.

This section intends to highlight, that there is a distinct, qualitative difference between AR

applications that visualize arbitrary virtual objects and AR applications where geospatial data is to be visualized directly in-situ on the location the geodata refer to. While both mentioned usage scenarios fall within the definition of AR by Azuma [36], this section argues that the latter scenario should be clearly distinguished within the more general conception of AR. To this aim, the term 'fused reality (FR)' is proposed which emphasizes that in a suchlike AR scenario, a virtual, geospatial object is fused with its physical corresponding counterpart in-situ in the location that the respective geodata-object is referring to. By doing so, the user is informed about the surrounding environment or a part thereof. Furthermore, the physical environment can thus be made to be interactively accessible via its underlying digital counterpart.

The concept of fused reality as defined in this work has strong relations to the concept of digital twins [541, 542, 497, 597]. The concept of a digital twin emerged in the context of product design and manufacturing to denote a virtual, digital equivalent of a product [203]. Kritzinger et al. [300] renders this definition more precisely by distinguishing a digital twin from a mere digital model of a physical object by requiring it to be dynamically linked with its physical counterpart by bidirectional data flow. I.e. changes in the state of the physical object cause updates in the state of its digital twin and vice-versa.

The role of AR as a link between a digital twin and its physical counterpart has already been pointed out by Saracco [505]. In this context, Saracco [505] also emphasizes the importance of the digital model being interactively operable in the sense of Kritzinger et al. [300]. Moya et al. [395] for instance, demonstrate a suchlike AR application based on a digital twin in the form of a comparably simple scenario of a single hyperelastic beam under dynamic loads while Qiu et al. [477] explore the fusion of AR and digital twins in the context of complex product assembling. To use a digital twin which is bidirectionally linked with its physical counterpart in the sense of Kritzinger et al. [300] in the context of fused reality as it is discussed here, would comprise the most elaborate form of FR, enabling it to unleash its full potential. However it has to be noted that the definition of FR as provided here does only require a corresponding digital model for a given physical object or environment without insisting on a automatic, dynamic data-exchange between the digital model and its physical counterpart.

The term 'fused reality' was already used by Bachelder [41] for a patented system for helicopter aircrew training of this name. This system combines visual renderings of flight situations with the physical hardware of a real helicopter cockpit that the trainee has to learn to handle. In this context, a helmet-based video display visualizes the cockpit environment while dynamically exchanging the view through the cockpit windows with simulated renderings. In a sense, the physical helicopter is augmented with data relating to its digital counterpart performing flight operations in a virtual environment. This system could thus be considered as falling within the more general definition of FR provided here.

Furthermore, Strauss and Fleischmann [548] also mention 'fusing space with data'. This, however, relates to augmenting a stage environment in the art context of hypermedia storytelling. As the relation of the virtual content and the physical stage environment is purely of a creative or imaginative nature, this scenario rather falls within the context of traditional AR. Finally, the value of AR for presenting geodata has also been pointed out by Dalla Mura et al. [131]. In this context, the authors also discuss the potential of AR devices for acquiring new geodata in the field. This aspect is also explored in this work in the context of indoor mapping in Chap. 2.

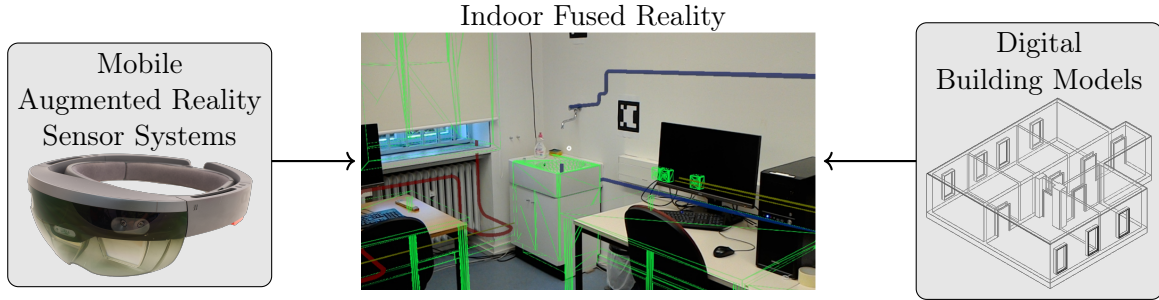


Figure 1.2: Schematic overview of the concept of indoor fused reality. By means of mobile augmented reality devices and digital building models, indoor environments can be augmented with location-dependent information improving the efficiency of in-situ work tasks and allowing direct, interactive access to the underlying building model via gestures and voice commands.

1.2.5 Indoor Fused Reality

As already indicated in Sec. 1.1, applying augmented reality techniques (see Sec. 1.2.3) in the context of digital building models (see Sec. 1.2.1) holds an enormous potential for all kinds of building-related applications. Thus, this combination of technologies has become the focus of intense research efforts [318, 486, 554, 19, 114, 525, 635]. Some proposed applications can be considered to be within the bounds of 'ordinary' AR (i.e. AR visualization of digital building models with arbitrary placement, e.g. in a kind of 'tabletop' mode) [121, 603, 84, 676]. Others rely on pure virtual reality [257, 592, 190, 526] instead of AR. However, many applications in this context aim at overlaying physical building structures with informative content from corresponding digital building models. These can be considered to fall within the definition of fused reality as presented in Sec. 1.2.4.

The concept of indoor fused reality is schematically depicted in Fig. 1.2. Mobile AR devices like the Microsoft HoloLens can be used to augment indoor environments with information from a corresponding digital building model. This not only provides the possibility for persons conducting on-site work tasks to instantly retrieve relevant information directly in-situ in the place of conduct. It furthermore also enables the possibility to directly interact with the building environment via its underlying model, e.g. by means of gestures or voice commands.

While some works in this context focus on conceptual frameworks for integrating BIM and AR technology [152, 610, 362], others aim at analyzing the potential benefits of AR techniques within the AEC domain [382, 428, 158] or investigate the challenges of deploying AR applications within the underground construction industry [170]. Other works focus on analyzing the requirements that potential users within the AEC domain put at AR-BIM systems [305, 194, 119].

The value of FR applications has been demonstrated for many tasks that relate to digital building models like e.g. design evaluation and communication [81, 92, 256]. During the construction phase, some works put the focus on dynamical aspects [216, 381] while others aim at visualization aspects [632, 662, 263], on-site process controlling [612, 669, 482, 171], progress monitoring [198, 342, 481] or at the detection of construction defects [442, 306] and discrepancies between planned and built structures [45, 93]. Furthermore, the potential of FR for construction safety [333, 524], on-site risk assessment [609, 133] and collaboration [515, 154] has been investigated.

Another important application domain for indoor FR is the domain of facility management [193, 633, 376, 385, 507, 279]. As the operation and maintenance phase of buildings is known to cause about 60 to 80 % of the total life-cycle costs of a building [35], improving efficiency by means of FR technology can be of significant benefit in this domain. In this context, special emphasis has also been put on on-site building asset management [409], plumbing facilities [140] and healthcare facilities

[248]. Furthermore, the value of FR for on-site inspection tasks has been investigated [689, 470, 112]. Finally, the deployment of sensors for on-site building monitoring holds potential in this context as well [445, 408, 101]. Besides, Indoor FR can also be applied in the domain of cultural heritage [115, 44, 177, 79].

1.3 Scope of this Work

The central topic of this work is that of indoor fused reality as proposed in Sec. 1.2.4 and elaborated on in Sec. 1.2.5, i.e. to use AR techniques (see Sec. 1.2.3) to visualize information from digital building models (see Sec. 1.2.1) directly in-situ within the indoor environment (see Sec. 1.2.2) of the corresponding real building. For the realization of this scenario, several components are needed that are successively presented in the following chapters. Fig. 1.3 gives a schematic overview of the structure of this work. Each chapter first presents an general overview of its respective topic. Afterwards, an own contribution is discussed respectively.

The own contributions to the topic of each presented chapter have already been published. In the following, paragraphs designated by vertical lines along the side margins are demarcated as originating from such a publication. They are adopted in the same form as originally published with the exception of small editorial changes to correct errors and improve the readability.

As a digital building model may not exist for a particular building where an indoor FR application is to be deployed, indoor mapping can be used to efficiently acquire geometric representations of existing building structures. Chap. 2 provides a comprehensive overview of the characteristic challenges of this task, the different systems that can be deployed to this aim along with their respective properties and the existing approaches to evaluate the performance of indoor mapping systems. As a contribution to the topic of indoor mapping, Chap. 2 furthermore presents a thorough quantitative evaluation of the mobile AR device Microsoft HoloLens assessing its adequacy for the task of indoor mapping. The results show, that state-of-the-art AR devices can indeed be used for indoor mapping, providing the basis for digital building models that can later be used in the context of indoor FR scenarios.

To generate actual, semantic and geometric building models from unstructured indoor mapping data, automatic indoor reconstruction methods can be used. After providing an extensive overview of existing approaches to this aim, Chap. 3 present a novel, fully-automatic, voxel-based indoor reconstruction and room partitioning procedure. Furthermore, adequate evaluation methodology is presented which is used to quantitatively assess the quality of the proposed indoor reconstruction approach. One of the advantages of the presented method is its genericity with respect to complex building structures which allows for non-planar room surfaces and complex vertical room layouts encompassing stair wells.

In order to use a digital building model in the context of an indoor FR application, the AR device used to this aim has to be localized within the respective building structure. Chap. 4 present an in-depth overview of the topic of indoor localization. Furthermore, a marker-based localization approach is presented, evaluated and discussed.

Finally, Chap. 5 presents a novel, automated method for pose normalization of indoor mapping data. Indoor mapping data is often-times acquired in arbitrary orientation with respect to the coordinate system. Often, the initial pose of an indoor mapping device determines the coordinate system. Automatically aligning indoor mapping data with the coordinate axes is a preprocessing step that can not only be used to facilitate automated indoor reconstruction tasks, but can potentially also be used for drift-reduction in indoor mapping systems and in the context of indoor localization.

These topics are discussed individually in their respective chapter. Afterwards, Chap. 6 presents

an overall discussion encompassing all the topics covered in this work, putting them in a common context while also discussing potentials for future research. Besides, Chap. 6 summarizes the work and closes with concluding remarks.

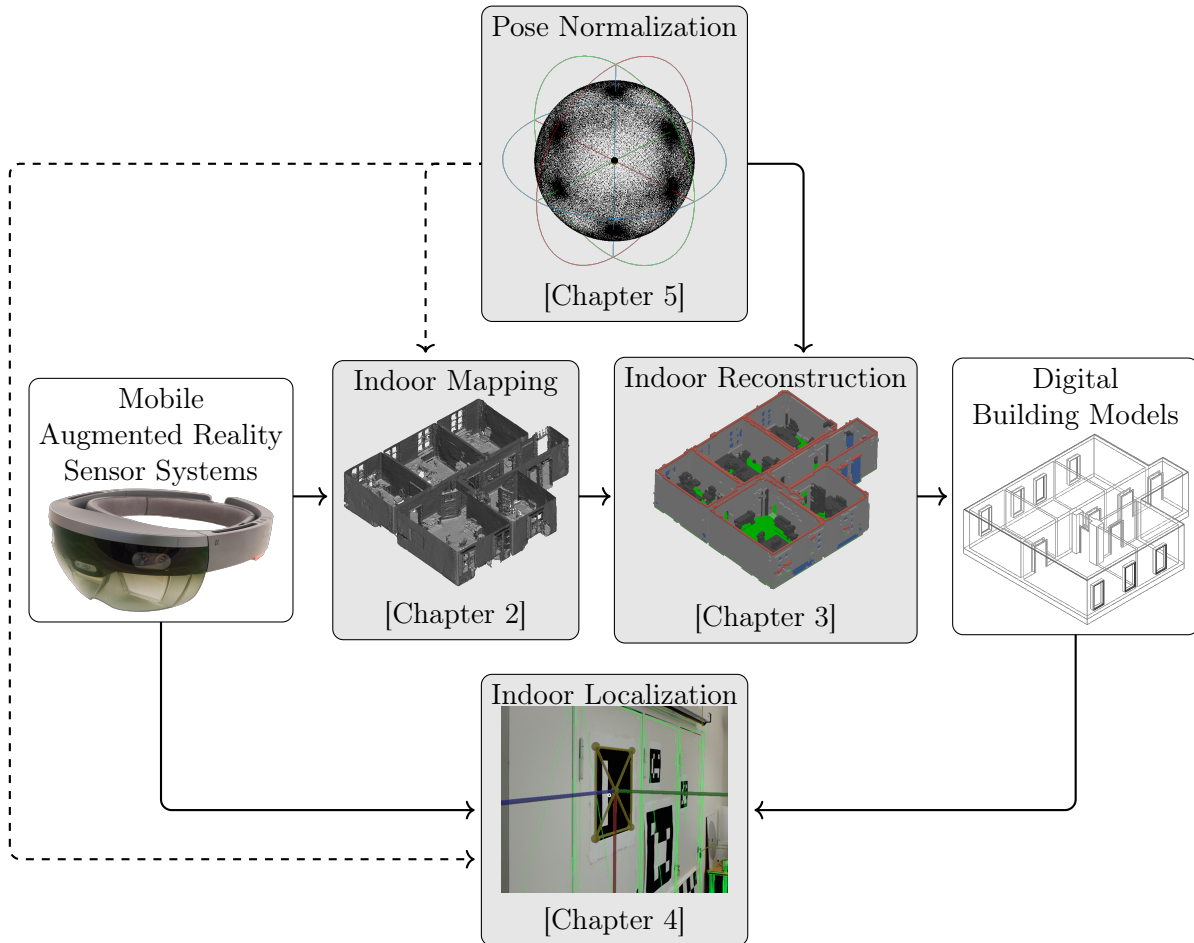


Figure 1.3: Schematic overview of the structure of this work: mobile AR sensor systems like the Microsoft HoloLens can be used to acquire point clouds or triangle meshes of indoor environments (Chap. 2, indoor mapping). The resulting indoor mapping data in turn, can be used to extract digital, geometric and semantic models of the mapped indoor environments (Chap. 3, indoor reconstruction). The resulting digital building models can again be applied in the context of indoor fused reality applications by means of mobile augmented reality devices. This requires a localization of the mobile AR device with respect to the building model (Chap. 4, indoor localization). In addition, an optional preprocessing step to automatically align indoor mapping data with the coordinate axes is presented (Chap. 5, pose normalization). This can facilitate the task of indoor reconstruction. Furthermore, the methodology presented for pose normalization can potentially also be of benefit for the tasks of indoor mapping and indoor localization (dashed lines, not realized within the scope of this work).

Chapter 2

Indoor Mapping

The topic of this chapter is indoor mapping, i.e. the acquisition of the visible geometry of indoor building structures by means of suitable sensor systems. Indoor mapping is one of the central concepts in the scope of this work as it provides input data for automatic indoor reconstruction tasks as detailed in Chap. 3. From raw indoor building geometries provided by indoor mapping systems, geometrically refined and semantically enriched indoor building models (see Sec. 1.2.1) can be generated which can serve as a data basis for indoor fused reality applications as presented in Sec. 1.2.5. To this aim, the task of indoor localization, i.e. determining the pose of a mobile AR device within the respective building environment has to be solved. In this context, the respective mobile AR device needs to have the capability of sensing its immediate building environment in order to localize itself within it. Thus, indoor mapping is also of importance in the context of indoor localization as detailed in Chap. 4.

The aim of this chapter is to evaluate, whether the sensor equipment of state-of-the-art mobile indoor AR devices is sufficient to the task of indoor mapping on building-scale level. In this case, a respective AR device could be deployed to the raw data acquisition of digital building models as well as to the realization of indoor fused reality scenarios making use of said building models. Thus, a comprehensive, quantitative evaluation study of a suchlike mobile AR device (the Microsoft HoloLens) is presented in this chapter, showing that its indoor mapping capabilities are indeed sufficient to this aim. In this context, the evaluation is approached by postulating that the quality of an indoor mapping system is mainly determined by the respective quality of its two essential sub-components, i.e. the tracking system and the range sensor. While the range sensor is used to acquire indoor building geometries from each respective pose of the indoor mapping system along its trajectory through the building, the poses provided by the tracking system are used to transform the locally measured geometry into a global coordinate system in which the whole building structure is mapped. In the course of the presented evaluation, both components (i.e. range sensors and tracking system) of the Microsoft HoloLens are evaluated individually against respective ground truth data. Additionally, an overall evaluation of the end result of the indoor mapping process against ground truth data is presented as well. This evaluation concept is depicted schematically in Fig. 2.1.

In the following, first, Sec. 2.1 provides a comprehensive overview on the topic of indoor mapping, the different types of sensor systems available to this aim and established evaluation methodology used to assess their quality. Afterwards, Sec. 2.2 presents the evaluation procedure deployed in the scope of this work. The respective results are presented in Sec. 2.3 and further discussed in Sec. 2.4. Finally, Sec. 2.5 concludes the chapter with a summary and suggestions for future work.

Within this chapter, material published in [239] and [238] is used.

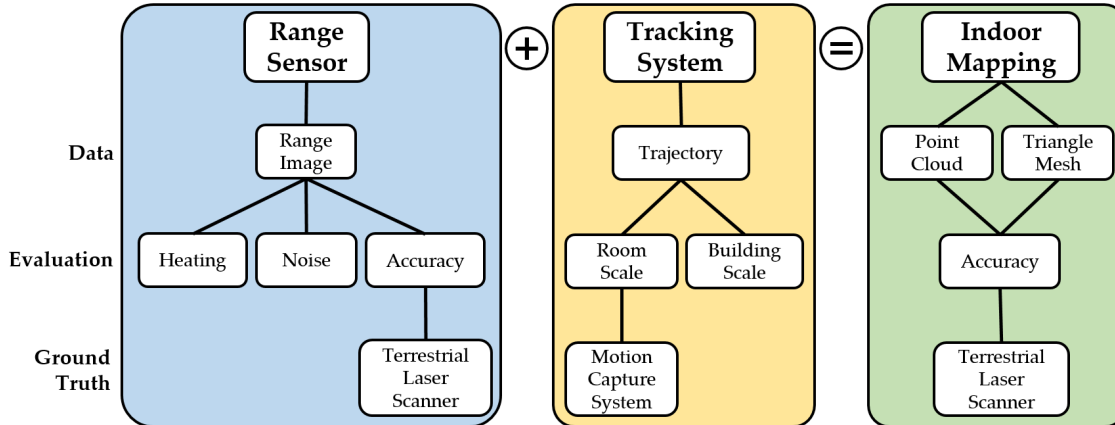


Figure 2.1: Schematic overview of the evaluation of the Microsoft HoloLens as indoor mapping system with range sensor and tracking system as the relevant subsystems constituting it. Source: [238].

2.1 Fundamentals of Indoor Mapping

This section provides an overview on the topic of indoor mapping. The geometric acquisition of indoor building structures by means of mobile sensor systems is an active field of research [692, 53, 63, 457, 424, 320, 434]. While the term ‘indoor mapping’ is widely used for the acquisition of geometric representations of indoor environments, e.g. in the form of point clouds or triangles meshes, Chen and Clarke [103] point out, that these are actually not maps but raw data for generating actual 3D indoor maps via cartographical techniques. In the same sense, Han et al. [218] designate point clouds of indoor environments as ‘spatial maps’ in contrast to semantic maps.

In consideration of the complexity and further peculiarities of indoor environments as already discussed in Sec. 1.2.2, the task of indoor mapping is a rather challenging one [320]. Indoor environments are characterized by complex arrangements of rooms typically containing lots of clutter objects occluding each other and room surfaces. Thus, to adequately map indoor environments, numerous different sensor poses have to be used during the acquisition process. Furthermore, indoor environments are particularly challenging for the tracking of mobile mapping systems. As GNSS signals cannot penetrate into building interiors, tracking systems need to rely on IMU, image-based methods and active range sensors. Image-based tracking is further impeded by the presence of texture-poor wall surfaces in some indoor environments such as hallways and by the spatial bottlenecks of doors between rooms. These characteristics distinguish indoor mapping from the more general task of mobile mapping in outdoor environments, where GNSS can be relied upon and the environment itself is often less challenging in terms of tracking [472, 530, 614, 65, 222].

Besides its already mentioned relevance to the tasks of indoor reconstruction (see Chap. 3) and indoor localization (see Chap. 4), indoor mapping also holds importance for application areas such as the documentation of cultural heritage buildings [83, 138] or as a source of realistic, digital environments for gaming applications [173, 25, 593]. In this context, the importance of TLS data from archaeological documentation and digital reconstructions originally intended for a gaming scenario for the reconstruction and digital preservation of the damaged Notre Dame cathedral in Paris is worth mentioning [275, 219].

In the following, Sec. 2.1.1 presents an overview of the different sensor systems and approaches available for the task of indoor mapping. Subsequently, Sec. 2.1.2 provides an overview of how the evaluation of indoor mapping systems has been approached in the literature.

2.1.1 Indoor Mapping Systems

Generally, a number of different sensor systems can be deployed to the task of indoor mapping. Traditional methods for the generation of floor plans such as the manual measurement with measuring tape or electronic range meter as well as tachymeter surveys are rather laborious, especially when applied to acquire building geometries in 3D [291].

For this reason, from early on, there was a large interest in deploying more efficient methods to the acquisition of indoor building structures. For instance, terrestrial laser scanning [619, 322] has been frequently used to this aim [387, 51, 495, 504]. This however poses the problem, that complex indoor environments necessitate a large number of TLS positions that need to be registered which is again quite labor intensive.

Alternatively, photogrammetric techniques have been used to three-dimensionally reconstruct building geometries from images [580, 373, 147, 449]. This method is more flexible with regard to complex indoor spaces than the TLS-based method. However, the 3D acquisition is comparably sparse and can fail in poorly-textured areas. Furthermore, selecting suitable image poses requires a trained operator.

With SLAM [290, 160, 399, 584], image-based mapping evolved from using unordered batches of images with well-planned acquisition strategy to video-sequences of cameras moving freely through the environment. This provides further flexibility for the acquisition of indoor environments. However, the sparseness of the resulting point clouds and the liability to poorly-textured environments persist.

These problems can be countered by using range cameras, e.g. operating via the structured light (SL) [282, 671] or the time of flight (ToF) principle [672]. Special variants of SLAM have been developed for the use of the RGBD images provided by these sensors [223, 606, 690, 679] that allow their application for the task of indoor mapping [594, 208, 655]. Furthermore, for the task of indoor mapping, multiple RGBD cameras have been combined for increased field of view [100, 144, 656], e.g. in the case of the Matterport system⁹.

As range cameras are subject to a rather high level of noise in the range measurement, points from surfaces captured from different poses during the acquisition process can vary significantly. Thus, a process that aggregates and refines captured surfaces over multiple frames is frequently applied when using RGBD cameras for indoor mapping [251, 413, 415, 624, 117]. The resulting refined surfaces are typically represented in the form of triangle meshes. This form of describing the measured indoor geometry is significantly more compact in relation to point clouds while still allowing for comparable results in segmentation and classification tasks [50, 621].

Besides directly using the mapping results of SLAM algorithms for indoor mapping, SLAM can also be used as tracking method in combination with other kinds of range sensors for this task, e.g. in combination with low-cost 2D laser scanners [60]. Indoor mapping systems based on this principle often incorporate an IMU for further improvement of the tracking. Suchlike systems have been proposed in different configurations, i.e. wheeled, trolley-mounted systems [98, 348, 658] (like e.g. the NavVis M6¹⁰ or the Applanix TIMMS¹¹), backpack-carried systems [315, 64, 309, 270, 250] (like e.g. the Leica Pegasus¹² or the Gexcel HERON AC-2¹³) or even hand-carried [58, 236, 582] or head-mounted systems (like e.g. the GeoSLAM ZEB Revo RT¹⁴ or the Kaarta Stencil¹⁵).

⁹<https://matterport.com/de/kameras/pro2> (Last visited on 05/06/2021)

¹⁰<https://www.navvis.com/m6> (Last visited on 05/06/2021)

¹¹<https://www.applanix.com/products/timms-indoor-mapping.htm> (Last visited on 05/06/2021)

¹²<https://leica-geosystems.com/de-de/products/mobile-sensor-platforms> (Last visited on 05/06/2021)

¹³<https://gexcel.it/en/solutions/heron-mobile-mapping/heron-ac> (Last visited on 05/06/2021)

¹⁴<https://geoslam.com/solutions/zeb-revo-rt/> (Last visited on 05/06/2021)

¹⁵<https://www.kaarta.com/products/stencil-2-for-rapid-long-range-mobile-mapping/>
(Last visited on 05/06/2021)

Finally, some works focus on mounting suchlike mapping systems on autonomously moving robots [346, 5]. In this context, active SLAM methods that do not only map the environment but also decide on appropriate steering commands in order to autonomously explore an unknown environment are of interest [74, 458, 677]. While mapping systems have also been mounted on unmanned aerial vehicles (UAV) [228, 276], they are rarely deployed in indoor environments as closed doors pose a serious problem for them. Ground-based robots that are able to open doors have already been proposed [66]. However, for flying platforms, this does not seem to be the case yet.

2.1.2 Evaluation of Indoor Mapping Systems

The quality of an indoor mapping system is mainly determined by the respective quality of its range sensor and its tracking system. Thus, the evaluation of an indoor mapping system can either be approached by directly focusing on the resulting indoor mapping geometries (i.e. point clouds or triangle meshes) or by analyzing those two fundamental components separately.

Concerning the evaluation of range sensors, the accuracy of terrestrial laser scanners as well as the effect of different influencing factors like scanning geometry and scene properties have been investigated [535, 82, 519]. Similarly, a broad range of studies has been presented that aims at evaluating different RGBD cameras. Some of them only evaluate the precision and accuracy of the respective sensors in relation to geometric scene parameters such as distance and inclination of surfaces w.r.t the sensor [622, 85]. Others, in addition, also investigate aspects such as sensor calibration [282, 308], heating behaviour [199, 308, 181, 617] or the influence of material properties and colors [308, 617]. Some works also provide detailed discussion on potential error sources in range sensing with RGBD cameras [506, 181].

Concerning the tracking system, Azuma [36] provides an analysis of tracking accuracy requirements regarding the application within the context of augmented reality. Many works that aim to evaluate tracking systems use ground truth trajectories as reference. These can be independently acquired by means of an external motion capture system [159, 299, 315, 325, 679] or estimated by comparing the resulting maps with reference floor plans [166]. For quantifying tracking accuracy based on ground truth trajectories, the absolute trajectory error (ATE) and the relative pose error (RPE) are frequently used metrics [550]. Recently, a new theoretical approach for trajectory evaluation based on a probabilistic, time-continuous formulation has been proposed [683]. Furthermore, some works use manually measured reference points [64] or completely rely on simulated environments for evaluation [369]. Besides evaluating tracking based on accuracy, other quality measures can be used as well, e.g. robustness against tracking loss in AR scenarios [325].

Regarding studies that aim at evaluating the tracking system of the Microsoft HoloLens, some works focus on the spatial stability of the displayed virtual objects as a measure for tracking quality [590, 511]. Others rely on user studies where tracking quality is assessed by interviewing test persons about their impression of the quality of different indoor AR devices [489]. Finally, motion capture systems have been deployed for providing reference trajectories in the context of quantitative evaluation of the tracking quality of the HoloLens, e.g. with a focus on head tracking in seated position [349] or in the context of human-robot interaction [289].

An often-applied methodology for quantitatively evaluating the accuracy indoor mapping point clouds is to use reference data, e.g. acquired by TLS [106]. The respective indoor mapping point cloud is registered onto the reference point cloud, e.g. via the ICP algorithm [59, 682]. The indoor mapping accuracy can be determined by applying error metrics based on point-to-point distances between both datasets. For instance, this evaluation procedure is often applied for comparing handheld and backpack-based or trolley-mounted indoor mapping systems [568, 361, 309, 577, 499]. In the cited works, all indoor mapping point clouds are evaluated against a common TLS ground truth point cloud

as reference. A similar procedure is also applied to the evaluation of triangle meshes acquired with the Microsoft HoloLens [283] as well as with the Matterport system [247, 454] or via KinectFusion [378]. In the latter case, not only point-wise distance deviations to the reference are evaluated, but angular deviations in the normal directions as well. Furthermore, evaluation against TLS ground truth via point-wise distance measures has been applied for the evaluation of photogrammetric indoor mapping approaches [130, 373] and in an extensive evaluation study considering a multitude of different indoor mapping systems based on laser scanning [319]. Besides evaluating point-wise distances for whole indoor mapping point clouds, manually selected reference points have also been used for evaluation against TLS ground truth data [110, 454].

In the absence of ground truth data, indoor mapping results can be evaluated by fitting geometric primitives such as planes for walls or cylinders for columns to the data and determining the deviations from these ideal geometries [580, 423]. Another evaluation study focuses more on evaluating an indoor mapping system via its trajectory, while the resulting point cloud itself is only assessed qualitatively [315]. Furthermore, an approach for the evaluation of indoor mapping point clouds in the absence of reference data has been proposed [269]. Here, assumptions about typical indoor building geometries such as planarity, parallelity and orthogonality are exploited. In addition, evaluation against weaker ground truth data in comparison to TLS point clouds is discussed, i.e. two-dimensional, possibly outdated floor plans.

2.2 Methodology

This section describes the experiments conducted for assessing the capabilities of the Microsoft HoloLens in regard to its aptitude for the task of indoor mapping. Sec. 2.2.1 gives an overview of the different camera sensors the device is equipped with and their respective characteristics. Afterwards, Sec. 2.2.2 comprehensively elaborates on the evaluation procedures of the different experiments conducted in the course of this study.

2.2.1 Sensor Description

The Microsoft HoloLens device is equipped with a range of various imaging sensors providing data necessary to accomplish the different tasks constituting its mobile indoor augmented reality system such as tracking, re-localization in known environments and capturing the geometric structure of its surroundings by means of depth sensing. Tab. 2.1 gives an overview of these camera sensors and their respective characteristics, while Fig. 2.2 shows an overlay of images recorded by those sensors to give an impression about their arrangement on the device.

All camera sensors of the HoloLens can be queried via the Microsoft Windows 10 SDK¹⁶. However, for all cameras except for the color camera, a so-called 'research mode' has to be activated. This mode is only meant for research. Applications making use of it cannot be used in apps published on the Microsoft Store for applications.

The color camera can be queried in different resolutions. It is not used for tracking, but only for allowing the user to record screenshot videos and pictures. Virtual renderings augmenting the physical environment of the user wearing the device can optionally be rendered into the images captured with this camera. The center of the color image in Fig. 2.2 roughly aligns with the line-of-sight of the user wearing the device.

Besides this color camera, the device also includes four grayscale tracking cameras, two of

¹⁶<https://developer.microsoft.com/en-us/windows/downloads/windows-10-sdk> (Last visited on 12/05/2021)

Table 2.1: HoloLens camera sensors and their characteristics as indicated by the Microsoft Windows 10 SDK for the device. Source: [238].

Camera	Type	Field of View [°]	Image Size [pixels]	Effective pixels [%]	Frame rate [fps]	Data type
Photo Video	Color	40×25	1408×792	100	30	BGRAS ^a
			1344×756			
			1280×720			
			896×504			
Long Throw	Depth	60×54	448×450	24 ^b	1 ^c	Gray16
Long Throw	Intensity	60×54	448×450	24 ^b	1 ^c	Gray8
Short Throw	Depth	78×77	448×450	71 ^b	15	Gray16
Short Throw	Intensity	78×77	448×450	71 ^b	15	Gray8
4 × Tracking	Grayscale ^d	60×50	160×480 ^d	400 ^d	30	BGRAS ^d

^aThe alpha channel contains a constant value of 1.0.

^bOnly a fraction of the image actually contains values (see Fig. 2.3).

^cThe SDK reports a frame rate of 3 fps, but an actual frame rate of 1 fps was observed.

^dThe system returns a 160×480 4-channel image, which actually represents a 640×480 grayscale image spread line-wise over all four channels.

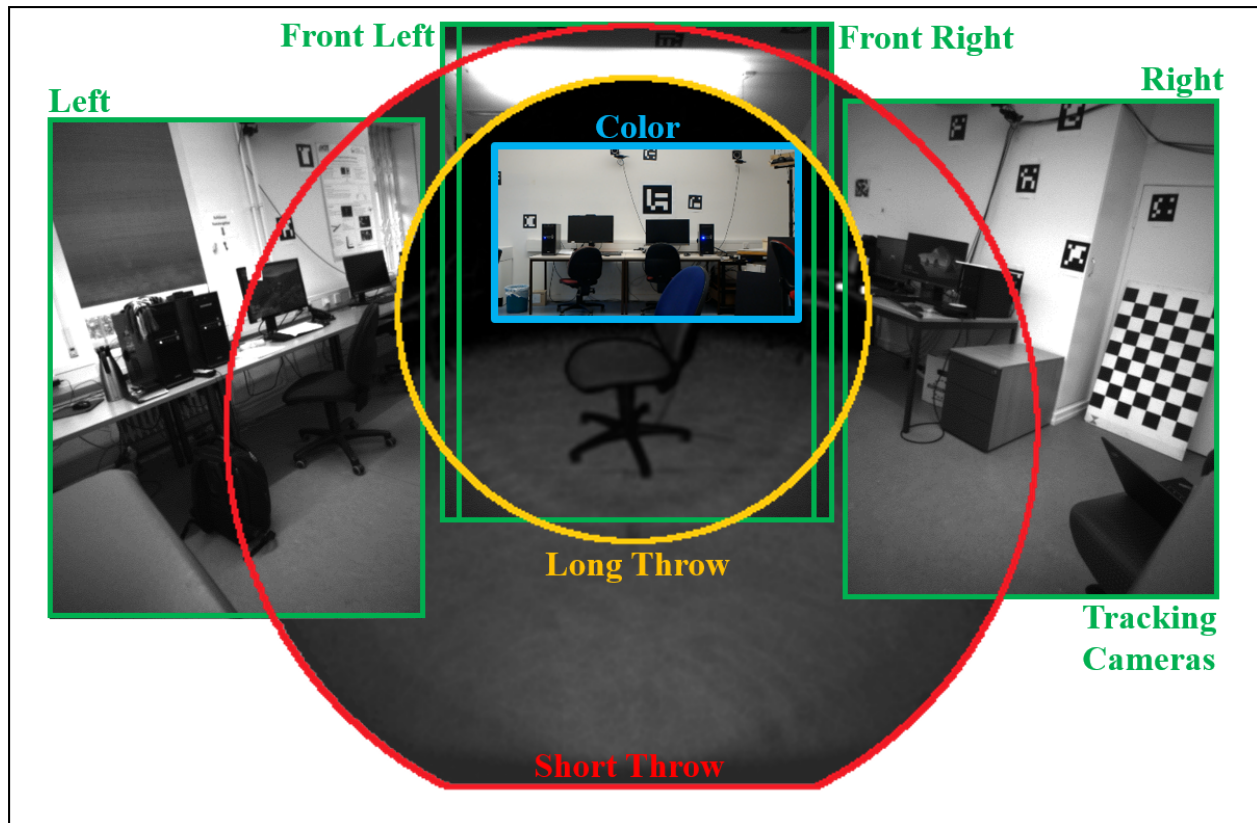


Figure 2.2: Overlay of images recorded by the different camera sensors of the Microsoft HoloLens as specified in Tab. 2.1. Source: [238].

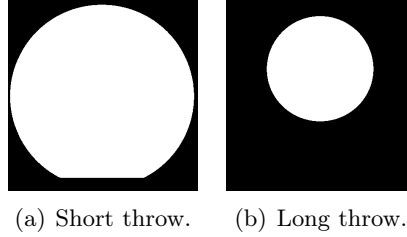


Figure 2.3: Pixels of the HoloLens range camera that actually contain values are shown in white for short throw and long throw mode. Source: [238].

which are oriented to the front in a stereo configuration with large overlap, while the other two are oriented to the right and left respectively with nearly no overlap to the center pair as depicted in Fig. 2.2. The images of these tracking cameras are provided by the SDK rotated by 90° , but their attached poses correct for this rotation. It is worth mentioning that the SDK returns 160×480 4-channel 8-bit images when querying the grayscale tracking cameras. These images actually represent 640×480 1-channel grayscale images where the intensity values are spread line-wise over all 4 channels. So the first pixel of the first line of the 160×480 image contains the first 4 pixels of the first line of the 640×480 image in its 4 channels.

The HoloLens device is furthermore equipped with a time-of-flight (ToF) range sensing camera, providing images with pixel-wise range measurements. These range images can be queried by the SDK in two different modes, termed 'long throw' and 'short throw'. Short throw data contain distance values in the range of 0 m to 0.8 m, while long throw data contain distance values from 0.8 m to about 3.5 m. For both modes, depth sensing data is delivered by the SDK in the form of 16-bit range images where the pixels contain integer values representing distance in millimeters. Furthermore, 8-bit grayscale images representing infrared reflectivity can be queried for both modes.

All images acquired by the ToF sensor have a size of 448×450 pixels; however different parts of the images actually contain values as depicted in Fig. 2.3. The part of the image actually containing values is circular for both modes. In the case of the long throw mode, this circular area containing range measurements is bigger and slightly clipped on the lower side.

The range sensing camera is oriented slightly downwards relative to the line-of-sight of the user as can be seen in Fig. 2.2. In typical usage scenarios, the short throw mode mainly observes the hands of the user for gesture recognition, while long throw range data are used for environment mapping. The field of view of the ToF camera overlaps with the one of the color camera; however the color image covers only a fraction of the range images.

The inner orientations of all camera sensors can be queried from the SDK in the form of a matrix, mapping from pixel coordinates $(x \ y)$ to metric 2D coordinates $(U \ V)$ on a plane in 1 m distance from the respective camera. An inverse mapping is also provided.

The range images delivered by the ToF sensor contain distance values along rays through the respective point $(U \ V)$ on the unit plane for each pixel $(x \ y)$. To transform these range values R to depth values D along an axis parallel to the image plane, the following equation can be used:

$$D = \frac{R}{\sqrt{U^2 + V^2 + 1}} \quad (2.1)$$

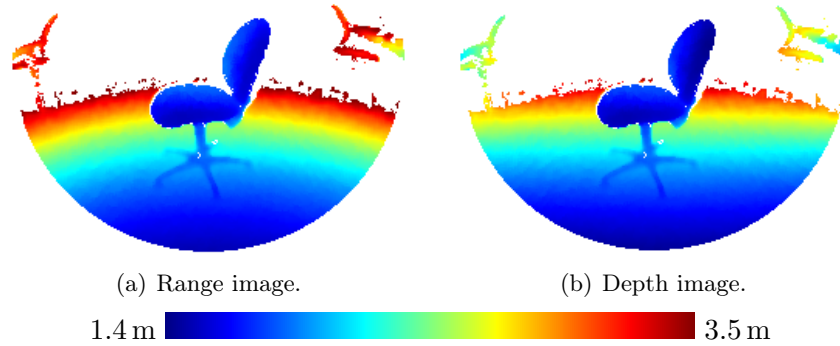


Figure 2.4: Range image and depth image corresponding to the long throw reflectivity image shown in Fig. 2.2. Source: [238].

Thus the Cartesian coordinates of a 3D point $(X Y Z)$ can be derived:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = D \begin{pmatrix} U \\ V \\ 1 \end{pmatrix} \quad (2.2)$$

Fig. 2.4 shows the range and depth image respectively corresponding to the reflectivity image for the long throw mode of the ToF camera shown in Fig. 2.2.

For all images, the corresponding camera pose in a coordinate frame defined by the initial pose when starting the respective HoloLens application can be obtained from the SDK. However, the camera pose is provided as split in two relative poses. One of these poses ($\mathbf{T}_{Device}^{Origin}$) consists only of a translational component and describes the position of the HoloLens device with constant orientation. The other pose is the inverse of the pose $\mathbf{T}_{Camera}^{Device}$ of the respective camera relative to this first pose. The current orientation of the device is encompassed in this second pose that furthermore contains a translational component for the offset of the respective camera from the point of reference of the device pose. The absolute pose of the respective camera then results to

$$\mathbf{T}_{Camera}^{Origin} = \mathbf{T}_{Device}^{Origin} \mathbf{T}_{Camera}^{Device} \quad (2.3)$$

Besides raw range images captured by the ToF sensor, the HoloLens SDK also provides preprocessed triangle meshes derived from the range data. Usage of these triangle meshes is not restricted to the research mode.

2.2.2 Evaluation Method

To assess the adequacy of the Microsoft HoloLens for the usage as a mobile indoor mapping device, a range of experiments were conducted which are detailed in this section. First, Sec. 2.2.2.1 describes the evaluation of the depth sensor of the HoloLens device. This is followed by the evaluation of its tracking system in Sec. 2.2.2.2. Finally, Sec. 2.2.2.3 describes the evaluation of the combined system for the use-case of indoor mapping.

2.2.2.1 Depth Sensing

The presented evaluation of the HoloLens depth sensing capability focuses on the long throw mode mentioned in Sec. 2.2.1. The short throw mode is only used for gesture recognition and



Figure 2.5: Ground truth data captured for a three-dimensional scene using a terrestrial laser scanner (Leica HDS6000). Source: [238].

thus not of relevance regarding the use-case of indoor mapping. For the conducted experiments, a plain, white, planar wall was used as reference object. The HoloLens device was fixed on a stand, facing the wall and recording long throw range images.

First, the influence of the heating process on range measurements was investigated by capturing a static scene and analyzing the temporal variation of the resulting range data. The device was positioned in a distance of ca. 1 m, approximately perpendicular to the wall surface with the wall filling the whole field of view of the long throw range images. The recording of the data was started with a completely cooled-down device, several hours after its last usage. Range images were recorded for a duration of 100 minutes with a frame rate of 1 fps in four consecutive recordings with 25 minutes length each and pauses of only a few seconds in between. The reason for splitting the measurement is that the device switches to sleep mode after 30 minutes without movement. Subsequently, the change in the mean depth value resulting from the respective range images was analyzed to characterize the influence of the warm-up process of the device over time.

Furthermore, to assess the influence of the measured distance on sensor noise, the distance of the sensor to the wall was varied while keeping it approximately perpendicular to the wall surface. In this manner, recordings of several minutes length each were made at varying distances over the whole working range of the sensor (0.8 m to about 3.5 m).

In the subsequent analysis, the standard deviation of the measured distance values per pixel was determined for each probed sensor position. Mean standard deviations over all pixels were then determined per sensor position, once for all pixels considered for evaluation and once only for those pixels that have recorded range values in at least 75 % of the images of the respective recording.

In doing so, only pixels representing wall surface were considered for evaluation. In some cases, parts of floor, ceiling, lateral margins of the wall surface, etc. are visible. This happens with growing distance of the sensor from the wall. For these, binary masks were created manually based on the reflectivity images for excluding those pixels not belonging to the wall surface from the evaluation.

Furthermore, the influence of the inclination of the wall surface on sensor noise was also investigated for inclinations between 0° and 80° . Here, the same analysis as in the above-mentioned evaluation of the influence of distance was conducted.



Figure 2.6: Rigid body of reflective sphere markers affixed to the HoloLens device. Source: [238].

Besides those experiments where a flat wall surface was used as reference object, sensor noise was also investigated on a three-dimensional scene comprising simple, geometric bodies (boxes, cylinders and spheres) as depicted in Fig. 2.5. This scene was captured by the HoloLens from three different distances.

Furthermore, the three-dimensional scene was also used for an evaluation of the accuracy of the captured distances using ground truth data acquired by a terrestrial laser scanner (Leica HDS 6000). To this aim, the point clouds derived from the range images were manually registered on the ground truth point cloud with a subsequent refinement of the registration via the Iterative Closest Point (ICP) algorithm [59, 682]. Then, Euclidean distances to the nearest ground truth point were determined for each HoloLens point. Afterwards, the HoloLens points were transformed back to the pixel grid of the original range images for better visual interpretation. For this analysis, the software Cloud Compare [195] was used. Furthermore, the software HoloLensForCV Microsoft [384] providing access to the device sensors was applied for accessing range images with accompanying poses from the ToF sensor.

2.2.2.2 Tracking

For assessing the tracking capacity of the HoloLens, the optical motion capture system OptiTrack Prime 17W¹⁷ with eight tracking cameras in a laboratory room with a size of approximately 8 m × 5 m × 3 m was used for the acquisition of ground truth data. For this purpose, the HoloLens device was equipped with a rigid body consisting of five reflecting sphere markers trackable by the motion capture system as depicted in Fig. 2.6.

Generally, in this work, the pose (i.e. the position and orientation) of an object A with respect to a coordinate system B is denoted by a 4×4 matrix

$$\mathbf{T}_A^B = \begin{pmatrix} \mathbf{R}_A^B & \mathbf{t}_A^B \\ \mathbf{0}^\top & 1 \end{pmatrix} \in \mathbb{R}^{4 \times 4} \quad (2.4)$$

consisting of an orthogonal rotation matrix $\mathbf{R}_A^B \in \mathbb{R}^{3 \times 3}$ with three degrees of freedom and a three-dimensional translation vector $\mathbf{t}_A^B \in \mathbb{R}^3$. Furthermore, \mathbf{T}_A^B can be understood as a three-dimensional Euclidean transformation from the local coordinate system of the respective object A to the given coordinate system B .

The spatial offset $\mathbf{T}_{RigidBody}^{Device}$ between the local coordinate system constituted by those rigid body markers and the local HoloLens device coordinate system, whose poses are recorded by

¹⁷<https://www.optitrack.com/products/prime-17w/> (Last visited on 12/05/2021)

the HoloLens tracking, had to be determined by a calibration procedure. For this purpose, a checkerboard pattern was observed by the HoloLens color camera in a static setting, while the device was equipped with the rigid body. The pose $\mathbf{T}_{Camera}^{Checkerboard}$ of the camera relative to the local coordinate system of the checkerboard was determined via the Perspective-n-Point (PnP) algorithm [185], while the relative pose $\mathbf{T}_{Camera}^{Device}$ of the camera with respect to the local coordinate system of the HoloLens itself was acquired from the Windows 10 SDK.

By manually measuring the positions of the sphere markers of the rigid body and the corners of the checkerboard pattern with a tachymeter (Leica TS06), the poses of the checkerboard ($\mathbf{T}_{Checkerboard}^{Tachymeter}$) and the rigid body ($\mathbf{T}_{RigidBody}^{Tachymeter}$) in the local coordinate system of the tachymeter were determined. The pose $\mathbf{T}_{RigidBody}^{Device}$ of the rigid body in the local coordinate system of the HoloLens device could thus be determined as:

$$\begin{aligned}\mathbf{T}_{RigidBody}^{Device} &= \mathbf{T}_{Camera}^{Device} \mathbf{T}_{Checkerboard}^{Camera} \mathbf{T}_{Tachymeter}^{Checkerboard} \mathbf{T}_{RigidBody}^{Tachymeter} \\ &= \mathbf{T}_{Camera}^{Device} \mathbf{T}_{Camera}^{Checkerboard}^{-1} \mathbf{T}_{Checkerboard}^{Tachymeter}^{-1} \mathbf{T}_{RigidBody}^{Tachymeter}\end{aligned}\quad (2.5)$$

Device trajectories of poses $\mathbf{T}_{Device}^{HoloLensOrigin}$ acquired via the HoloLens tracking system could now be transformed to trajectories of poses $\mathbf{T}_{RigidBody}^{HoloLensOrigin}$ of the rigid body attached to the device which could thus be compared to the ground truth trajectories of poses $\mathbf{T}_{RigidBody}^{MotionCaptureSystem}$ observed by the motion capture system:

$$\mathbf{T}_{RigidBody}^{HoloLensOrigin} = \mathbf{T}_{Device}^{HoloLensOrigin} \mathbf{T}_{RigidBody}^{Device}\quad (2.6)$$

A prevalent metric for the evaluation of estimated trajectories against ground truth trajectories is represented by the absolute trajectory error (ATE) and the relative pose error (RPE) [550].

For determining the ATE, it is essential to spatially align the trajectory with its corresponding ground truth trajectory when they are given in distinct coordinate systems as it is the case here. Furthermore, a temporal alignment by timestamps t_i between corresponding poses of both trajectories is required, that allows to assign each pose \mathbf{P}_i of the trajectory its temporarily closest ground truth pose \mathbf{P}_i^{GT} .

As the poses acquired by the motion capture system only have timestamps relative to the start time of the measurement, a temporal alignment between the HoloLens trajectory and the trajectory of the motion capture system had to be conducted. This was achieved by manually extracting timestamps at trajectory positions on the apex of distinct peaks in the trajectories.

The thus temporarily assigned pose pairs \mathbf{P}_i and \mathbf{P}_i^{GT} could then be used to spatially align both trajectories by the method of Horn [232] as proposed by Sturm et al. [550] while keeping the scale fixed. With the trajectories registered in a common coordinate frame, the ATE could be calculated by the root mean square error

$$ATE = \sqrt{\frac{\sum_{i=0}^N ||trans(\mathbf{D}_i)||^2}{N}}\quad (2.7)$$

of the translational components of the pose differences \mathbf{D}_i between corresponding HoloLens and ground truth poses:

$$\mathbf{D}_i = \mathbf{P}_i^{GT^{-1}} \mathbf{P}_i\quad (2.8)$$

The ATE is only meaningful as an aggregated value like the root mean square error over a complete trajectory as the quantity of translational differences of particular pose pairs results from the alignment process between both trajectories and not from the tracking quality in the respective poses themselves. Thus, the ATE can only be regarded as a measure for tracking quality over the whole trajectory. To eliminate the subjective influence of manually selecting a pose pair for temporal alignment between both trajectories, an optimization procedure is applied that determines the temporal alignment in millisecond-resolution by minimizing the ATE.

The RPE on the other hand is a metric for relative drift between an estimated trajectory and its ground truth trajectory. Like the ATE, it is calculated as root mean square error of the translational (or rotational) components of pose differences (see Eq. 2.7). Here, however, the pose differences \mathbf{D}_i are relative differences based on an offset Δ in the pose index:

$$\mathbf{D}_i = (\mathbf{P}_i^{GT^{-1}} \mathbf{P}_{i+\Delta}^{GT})^{-1} (\mathbf{P}_i^{-1} \mathbf{P}_{i+\Delta}) \quad (2.9)$$

As offset Δ , the number of poses corresponding to the time difference of one second was applied in order to get the RPE as a value for drift per second.

The ATE and RPE metrics were evaluated for trajectories recorded while walking around in the laboratory space covered by the cameras of the motion capture system. In doing so, the same pattern of movement was followed for each trajectory. The conditions were varied by masking the depth camera for some of the recorded trajectories.

Furthermore, to assess the influence of drift on large-scale trajectories through long corridors in large building complexes, a trajectory with accompanying triangle meshes was recorded along a long closed loop of a total length of 287 m on two floors of a building. The trajectory ended in the same room it started in, while the room was re-entered through a different door than it was left through. The course of the trajectory was planned and executed in a way that ensures that the relocalization system of the HoloLens is only able to detect the drift-induced failure in its position and correct for it, when the device has already re-entered the room.

2.2.2.3 Indoor Mapping

After the evaluation of the individual components relevant for indoor mapping (depth sensor and tracking system), the performance of the overall HoloLens system in regard to the task of indoor mapping was evaluated as well. For this purpose, the indoor space of an office environment comprised of six rooms with furniture was mapped which resulted in triangle meshes provided by the spatial mapping system of the HoloLens. For a subset of four of these rooms, range images in the long throw mode were also recorded. While the acquisition of the triangle meshes was conducted by walking through the rooms leisurely in typical walking speed, the acquisition of the range images was done by walking deliberately slow, as range images can currently only be acquired with a rate of one frame per second.

The range images were subsequently transformed to a global point cloud making use of their poses as provided by the tracking system. The resulting point cloud as well as the triangle meshes were manually registered on a ground truth point cloud of the mapped indoor environment acquired by a terrestrial laser scanner (Leica HDS 6000) with subsequent ICP-based refinement [59, 682]. As the ground truth point cloud was acquired in a furniture-less state with completely empty rooms, all objects that are not represented in the ground truth data were manually removed from the point cloud and the triangle meshes acquired with the HoloLens. The floor was also removed, as it was hard to manually separate it from furniture objects. The evaluation against the TLS ground truth data was then conducted by assigning each point (respectively

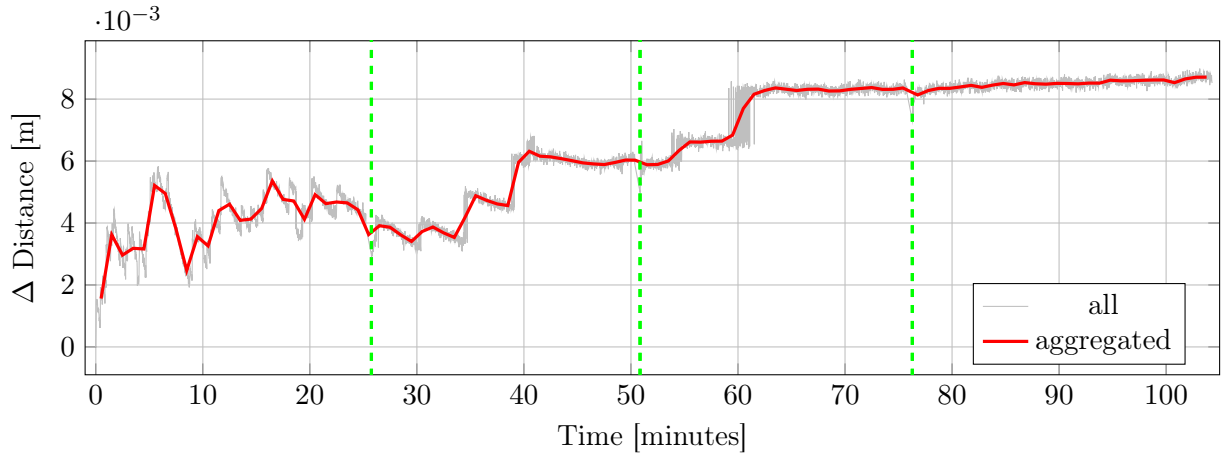


Figure 2.7: Change in distance measurement of the HoloLens range sensor relative to the start value over time during the warm-up process of the device. Vertical green lines indicate when the recording of range measurements had to be stopped and immediately restarted to prevent the device from switching to sleep mode. Source: [238].

vertex in a triangle mesh) the Euclidean distance value to the nearest point of the ground truth data. Again, the software Cloud Compare [195] was used for this analysis.

2.3 Results

This section presents the results of the different experiments detailed in Sec. 2.2. First, Sec. 2.3.1 presents the results of the evaluation of the HoloLens depth sensing capabilities. Afterwards, the results of the tracking evaluation are presented in Sec. 2.3.2, while Sec. 2.3.3 concludes with the results of the evaluation of the overall system for the use-case of indoor mapping.

2.3.1 Depth Sensing

This section presents the results of the evaluation of the HoloLens ToF range sensor as described in Sec. 2.2.2.1. Fig. 2.7 shows the variation of the measured range values over time during the warm-up process of a completely cooled-down device relative to the first range measurement. In the first 40 minutes after the beginning of the measurements, the range value is subject to strong fluctuations in the range of few millimeters. Afterwards, from about 40 to 60 minutes, the range value remains more or less constant at a value of 6 mm above the initial value. Then, at a measurement time of about one hour, the measured value rises again by two to three millimeters accompanied by strong fluctuations. Afterwards, it remains stable with a slightly increasing trend for the rest of the measurement.

Tab. 2.2 on the other hand, shows the results of the investigation of sensor noise against distance and inclination of a captured plane. The given values for noise are calculated as mean standard deviations of the measured distance values over all pixels. The sensor noise is calculated once for all pixels on the captured wall surface and once only for those pixels that contain range values in at least 75 % of images of a respective recording. The results are further visualized in Fig. 2.8.

As Fig. 2.8(a) shows, the sensor noise stays below 5 mm for measured distances smaller than about 2.5 m. From 2.5 m upwards, a rapid increase in noise is observable. This increase is mainly

Table 2.2: Noise of the distance measurements of the HoloLens ToF sensor against distance and inclination of a captured plane. Stable pixels are pixels that have range measurements in at least 75% of images. Part a): variation of distance; part b): variation of inclination. Source: [238]

	Horizontal Angle [°]	Vertical Angle [°]	Depth [m]	Noise [m]	Noise in Stable Pixels [m]	Fraction of Stable Pixels [%]
a)	0.5	0.6	0.82	0.0018	0.0018	100
	0.2	5.9	0.93	0.0025	0.0025	100
	-1.9	4.3	1.08	0.0023	0.0023	100
	-3.1	4.7	1.24	0.0028	0.0028	100
	2.5	3.8	1.41	0.0030	0.0030	100
	1.8	3.9	1.52	0.0032	0.0032	100
	0.1	4.8	1.72	0.0036	0.0036	100
	-2.4	5.3	1.88	0.0038	0.0038	100
	-0.9	4.7	2.07	0.0039	0.0039	100
	1.8	5.0	2.28	0.0042	0.0041	100
	3.9	5.0	2.56	0.0054	0.0044	95
	3.8	4.6	2.82	0.0070	0.0047	84
	3.8	4.5	3.05	0.0096	0.0054	74
	0.4	4.5	3.17	0.0113	0.0059	69
	5.6	5.2	3.33	0.0162	0.0072	51
	5.5	5.1	3.45	0.0201	0.0076	18
b)	0.1	4.8	1.72	0.0036	0.0036	100
	13.1	5.8	1.84	0.0040	0.0040	100
	19.5	8.6	1.94	0.0049	0.0044	98
	37.8	-2.0	1.89	0.0065	0.0053	96
	48.6	-2.0	1.84	0.0067	0.0061	97
	58.6	1.1	1.69	0.0081	0.0076	97
	73.8	-8.4	1.55	0.0103	0.0100	98
	80.8	-29.6	1.38	0.0124	0.0121	99

Table 2.3: Evaluation of a three-dimensional scene as depicted in Fig. 2.5 captured by the HoloLens ToF sensor from three different distances regarding noise and accuracy against TLS ground truth. Additionally, a HoloLens triangle mesh of the scene is also evaluated against TLS ground truth. For visual representation of the results, see Figures 2.9, 2.10, 2.11 and 2.12. Source: [238].

Recording	Mean Depth [m]	Mean Noise [m]	Mean Noise in Stable Pixels [m]	Mean Distance to Ground Truth [m]
Near	1.30	0.0043	0.0042	0.0188
Midrange	2.02	0.0062	0.0060	0.0138
Far	2.64	0.0190	0.0104	0.0939
Mesh	—	—	—	0.0458

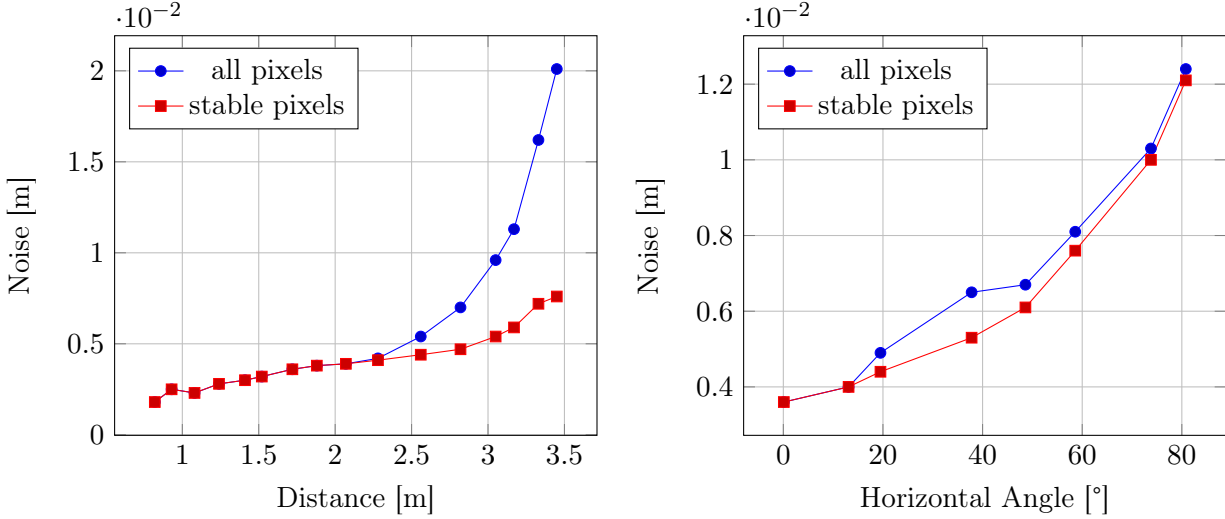


Figure 2.8: Noise of the distance measurements of the HoloLens ToF sensor against distance and inclination of a captured plane. Stable pixels are pixels that have range measurements in at least 75 % of images. Source: [238]

caused by pixels, that only sporadically return range measurements. When only considering stable pixels having range measurements in at least 75 % of the recorded images, the increase in noise with distance proves less steep and stays below 1 cm for the whole measurable distance range.

In the case of the influence of surface inclination as visualized in Fig. 2.8(b) however, sensor noise increases by approximately the same rate for only the stable pixels as in the case of considering all pixels. In both cases, noise remains below 5 mm for inclinations below 20°.

Tab. 2.3 presents the results of the evaluation of the three-dimensional scene depicted in Fig. 2.5. The results are also visualized in Fig. 2.9 as depth images, while Fig. 2.10 visualizes the noise and Fig. 2.11 the accuracy of the range measurements evaluated against TLS ground truth data for all three distances the scene was captured from by the HoloLens range sensor (near, midrange and far). Furthermore, Fig. 2.12 visualizes the accuracy of the HoloLens triangle mesh of the same scene evaluated against the TLS ground truth data.

2.3.2 Tracking

In this section, results of the evaluation of the HoloLens tracking system as detailed in Sec. 2.2.2.2 are presented.

The evaluation of eight trajectories against ground truth data determined by the motion capture system results in a mean ATE value of 1.9 ± 0.4 cm and a mean RPE value quantifying drift per second of 1.6 ± 0.2 cm and 2.2 ± 0.3 °. Seven similar trajectories recorded with covered range sensor resulted in a mean ATE of 1.3 ± 0.1 cm and a mean RPE of 1.6 ± 0.1 cm and 1.5 ± 0.3 °.

One of the evaluated trajectories of the rigid body on the device as tracked by the HoloLens tracking system is depicted in Fig. 2.13. This trajectory was recorded with non-covered range sensor. Fig. 2.14 shows the associated velocity and RPE values over the course of the trajectory. The color range in both figures symbolizes time, going from blue to red.

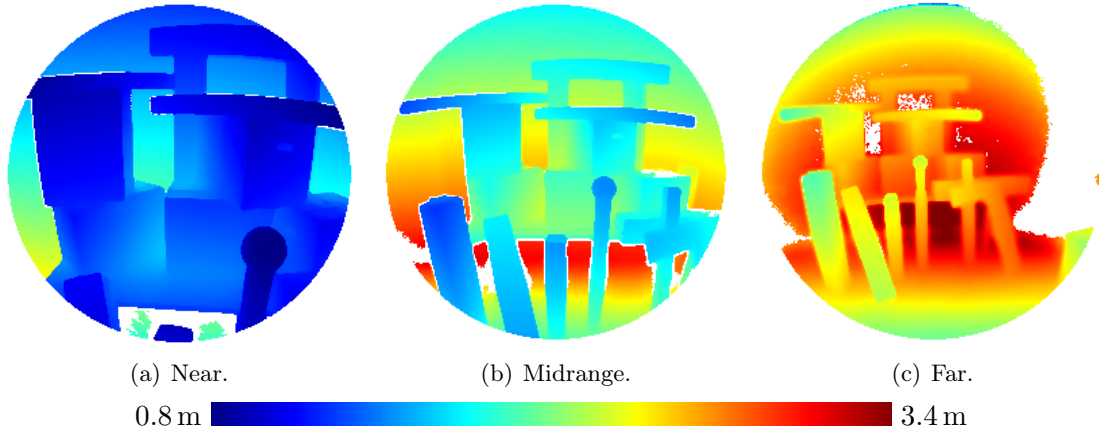


Figure 2.9: Depth images derived from the HoloLens ToF sensor for the three-dimensional scene depicted in Fig. 2.5 captured from three different distances. Source: [238].

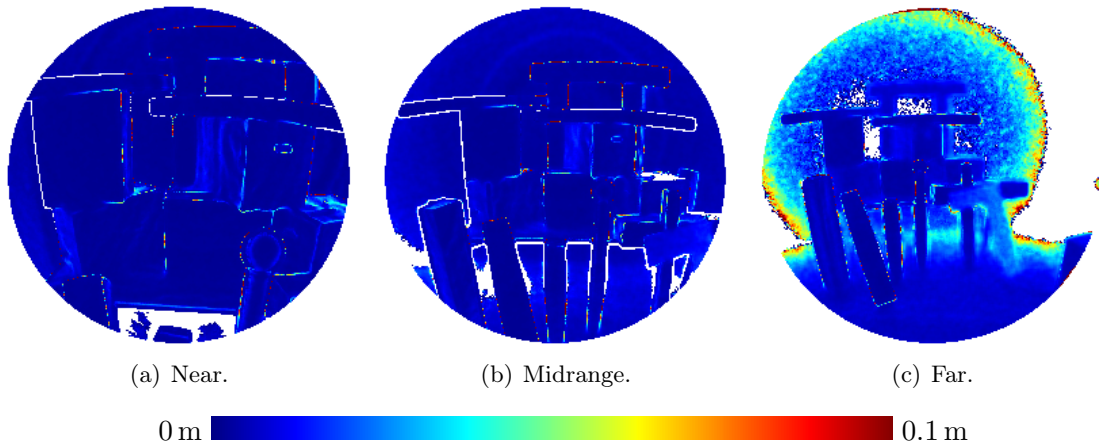


Figure 2.10: Noise of the range measurements of the HoloLens ToF sensor for the three-dimensional scene depicted in Fig. 2.5 captured from three different distances. Source: [238].

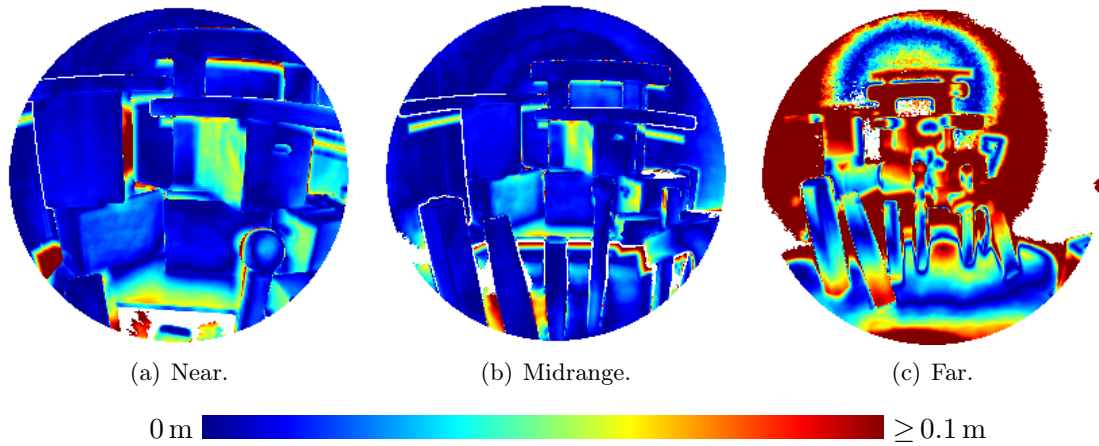


Figure 2.11: Accuracy of HoloLens range measurements evaluated against TLS ground truth data for the three-dimensional scene depicted in Fig. 2.5 captured from three different distances. Source: [238].

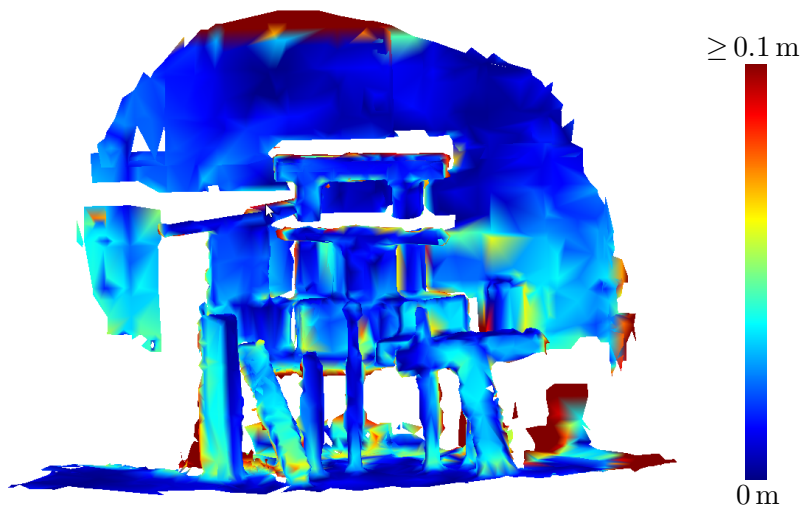


Figure 2.12: Accuracy of the HoloLens triangle mesh of the three-dimensional scene depicted in Fig. 2.5 evaluated against TLS ground truth data. Source: [238]

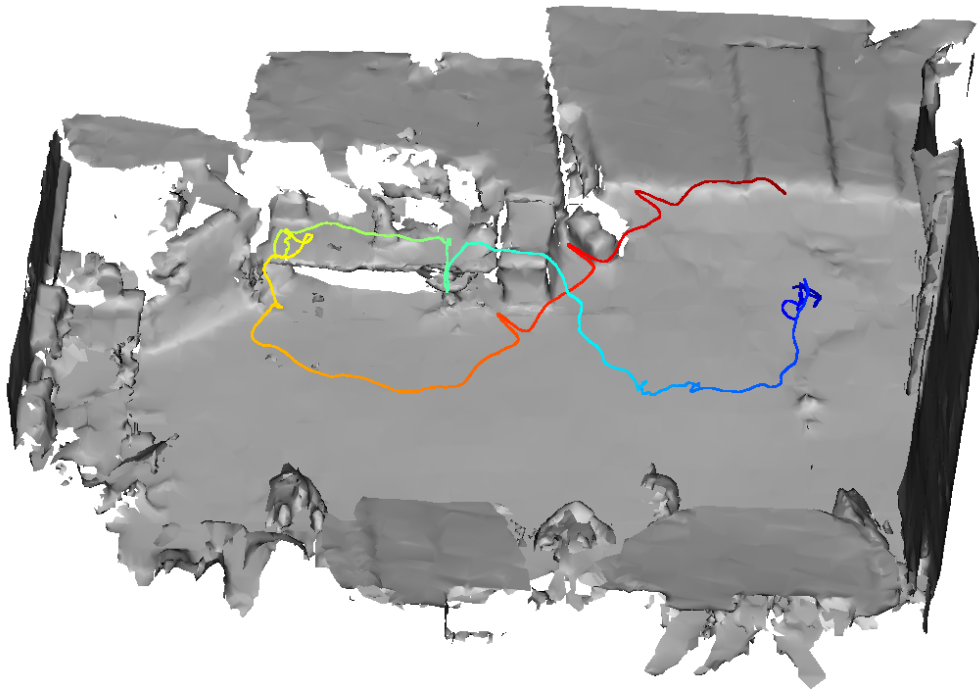


Figure 2.13: HoloLens trajectory in a room equipped with a motion capture system. The color represents time. The trajectory starts at blue and ends at red. The evaluation results in Fig. 2.14 also refer to this trajectory. They are visualized in the same color range representing time. Source: [238].

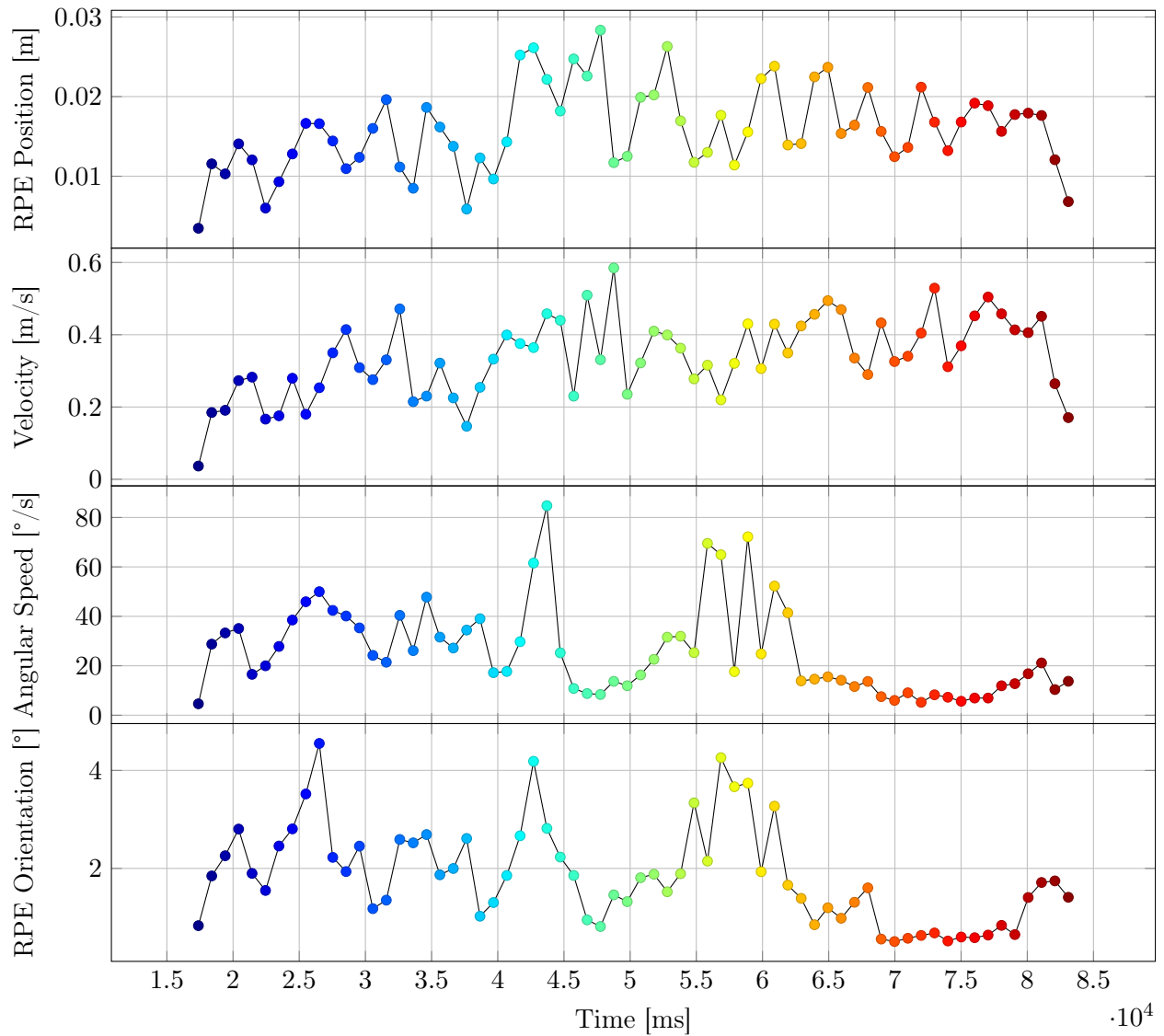


Figure 2.14: Velocity and RPE values over the course of the trajectory depicted in Fig. 2.13. The color range is the same as the one from Fig. 2.13, symbolizing the time going from blue to red. Source: [238].

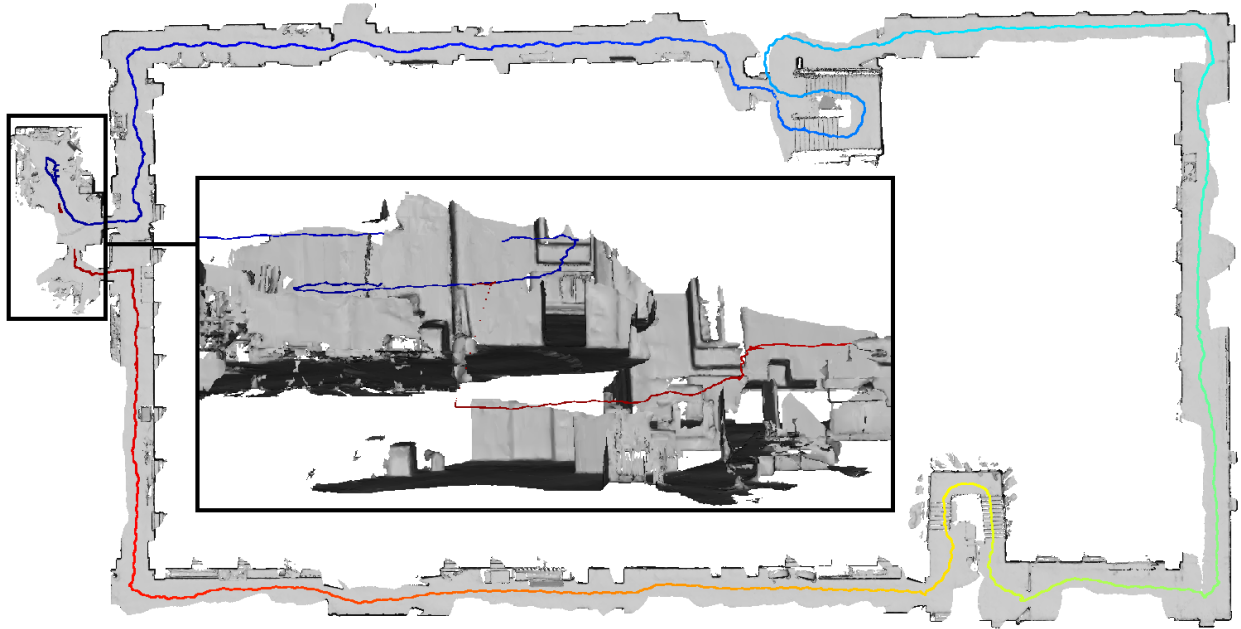


Figure 2.15: Closed trajectory of a total length of 287 m. The color represents time. The trajectory starts at blue and ends at red. The positional error caused by drift amounted to 2.39 m upon reentering the room. Source: [238].

Finally, Fig. 2.15 shows the result of the experiment to assess drift on large-scale trajectories described in Sec. 2.2.2.2. The traveled distance of the depicted trajectory totals to 287 m (including drift). The offset caused by drift upon re-entering the room amounts to 2.39 m.

2.3.3 Indoor Mapping

In this section, the results of the evaluation of the overall HoloLens system for the use-case of indoor mapping as described in Sec. 2.2.2.3 are presented.

Fig. 2.16 depicts the captured triangle mesh of an indoor office environment consisting of five rooms and a small hallway. In Fig. 2.17, the accuracy of the triangle mesh evaluated against TLS ground truth data is visualized. In this case, the mesh was registered on the ground truth data while keeping the scale fixed. The average accuracy of the complete mesh evaluated amounts to 2.3 cm. Fig. 2.18 on the other hand depicts the accuracy evaluated against the same ground truth data resulting from a registration which also adapts the scale of the mesh. In this way, a scale factor of 0.9938 was determined, while the mean accuracy amounts to 1.7 cm.

The same indoor environment was also mapped with the HoloLens in the empty, furniture-less state the TLS ground truth was acquired in. In this case, five independently-acquired triangle meshes were recorded. These were registered at fixed scale with each other to determine the precision of repeated acquisitions of the same environment with the results being depicted in Fig. 2.19. Furthermore, the same triangle meshes were evaluated against the TLS ground truth data. The resulting accuracy averaged over all five triangle meshes is visualized in Fig. 2.20. Finally, all five triangle meshes were manually split in the individual rooms. The resulting meshes for each room were then individually registered with fixed scale onto the respective room of the ground truth data. Fig. 2.21 shows the resulting averaged accuracy evaluated against the TLS ground truth.



Figure 2.16: Triangle mesh of an indoor office environment captured by the HoloLens. The ceiling is removed for better visibility. Source: [238].

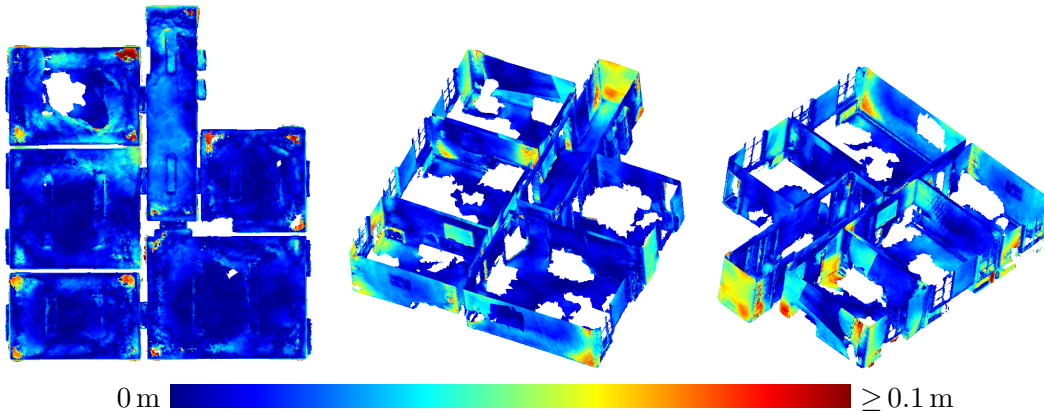


Figure 2.17: Accuracy of the HoloLens triangle mesh evaluated against the TLS ground truth. The registration was done with fixed scale. The mean distance to the ground truth amounts to 2.3 cm. Source: [238].

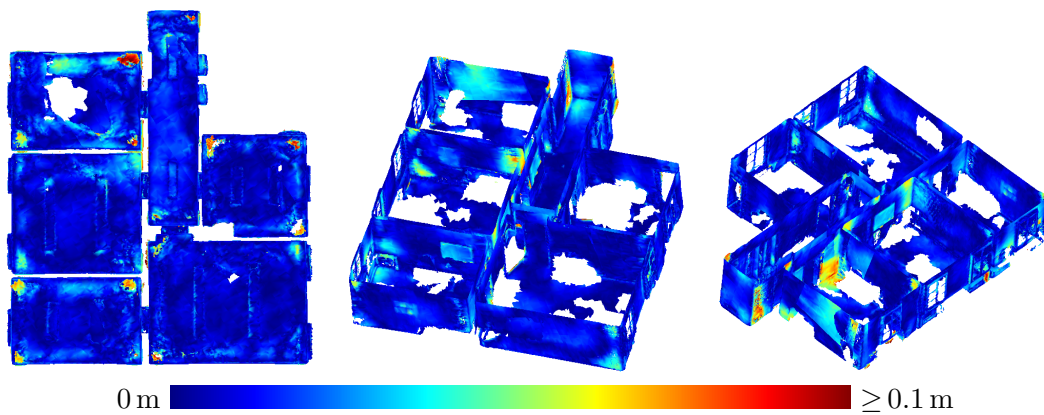


Figure 2.18: Accuracy of the HoloLens triangle mesh evaluated against the TLS ground truth. The scale factor was determined to 0.9938 by registration. The mean distance to the ground truth amounts to 1.7 cm. Source: [238].

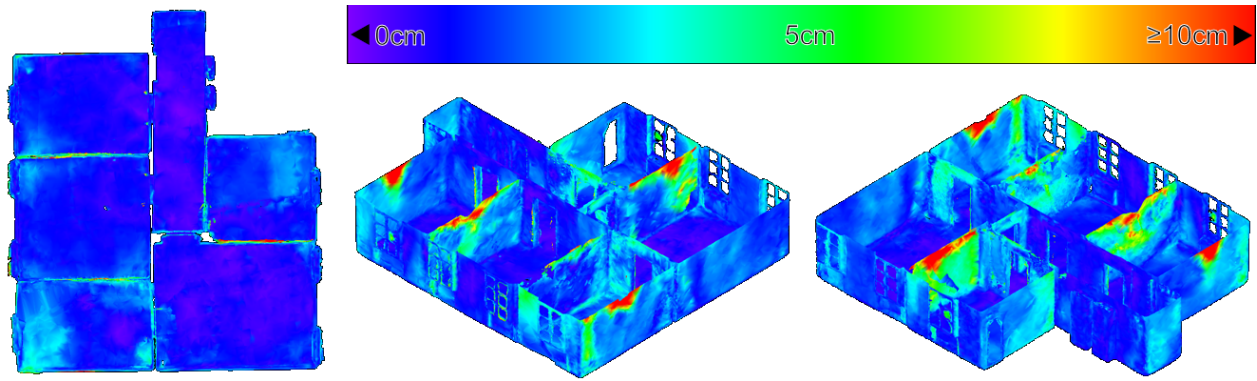


Figure 2.19: Precision of five triangle meshes of the same, furniture-less indoor environment, acquired independently with the Microsoft HoloLens. The registration was done with fixed scale. Source: [239], modified.

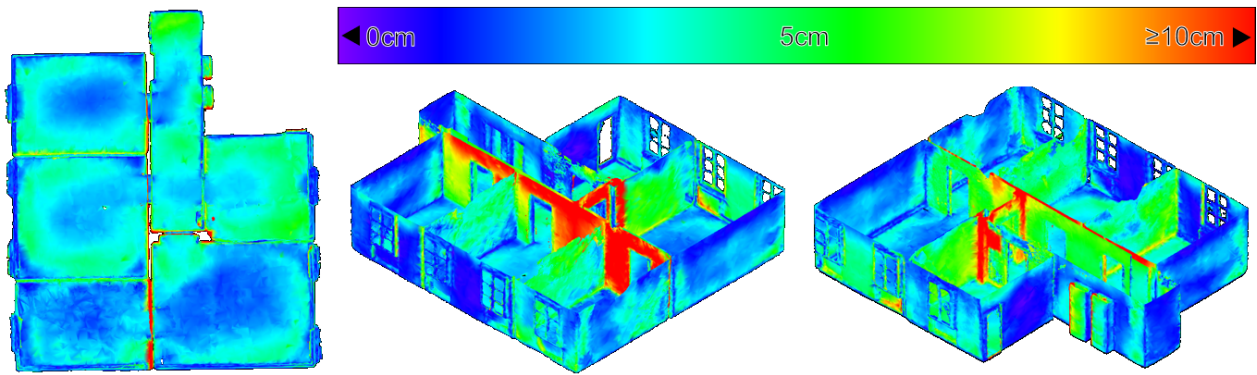


Figure 2.20: Averaged accuracy of five triangle meshes of the same, furniture-less indoor environment, acquired independently with the Microsoft HoloLens and evaluated against TLS ground truth. The registration was done with fixed scale. Source: [239], modified.

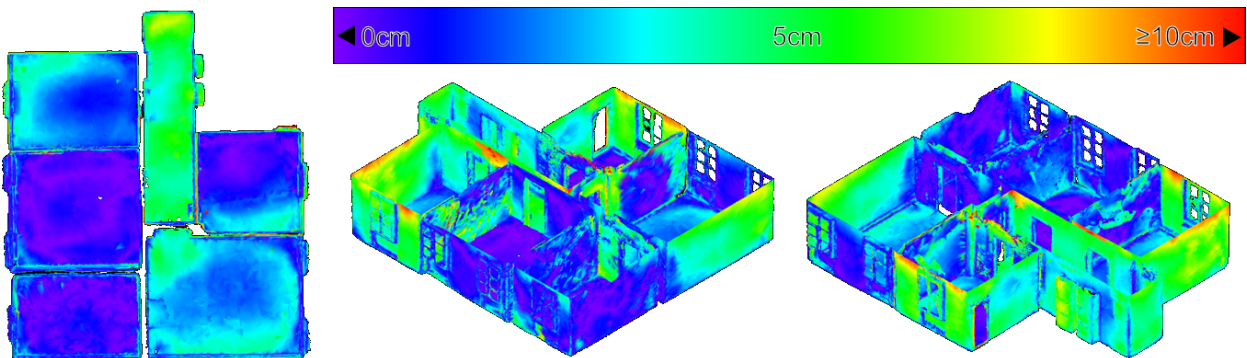


Figure 2.21: Averaged accuracy of five triangle meshes of the same, furniture-less indoor environment, acquired independently with the Microsoft HoloLens and evaluated against TLS ground truth. The registration was done with fixed scale. In this case, each triangle mesh was manually split in different rooms. The rooms were individually registered on the corresponding room from the TLS ground truth dataset. Source: [239], modified.

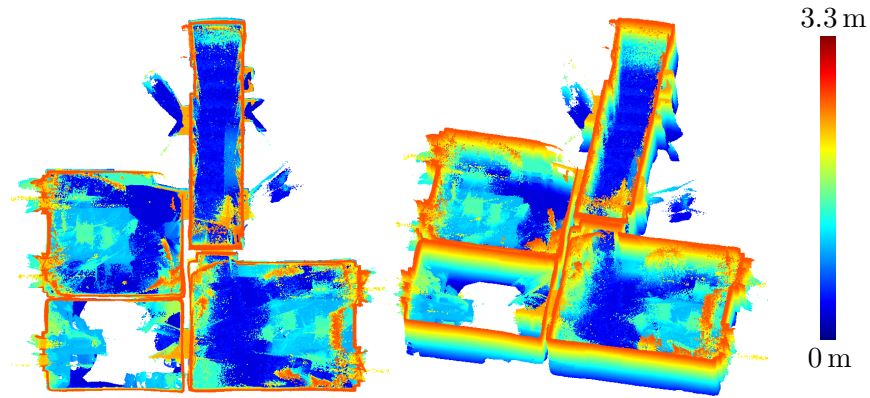


Figure 2.22: Point cloud of an indoor office environment captured by the HoloLens range sensor. The colors visualize point height with blue for the lowest and red for the highest points. The ceiling is removed for better visibility. Source: [238].

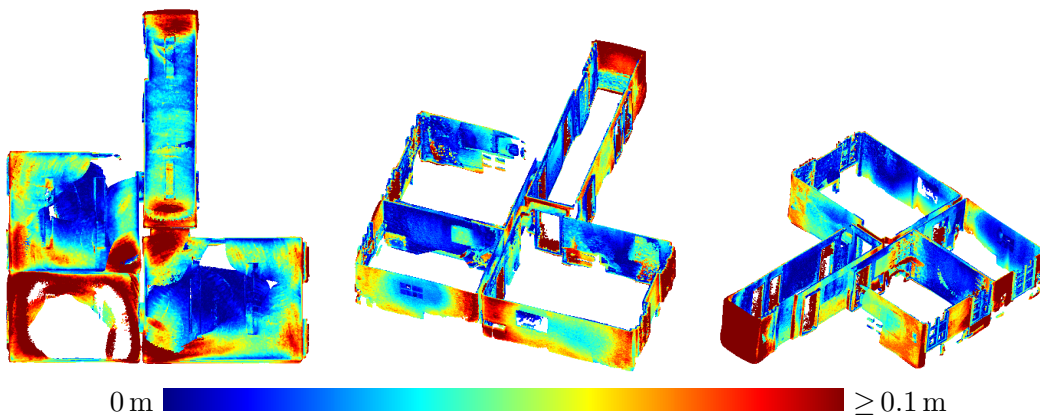


Figure 2.23: Accuracy of the HoloLens point cloud evaluated against the TLS ground truth. The registration was done with fixed scale. The mean distance to the ground truth amounts to 4.0 cm. Source: [238].

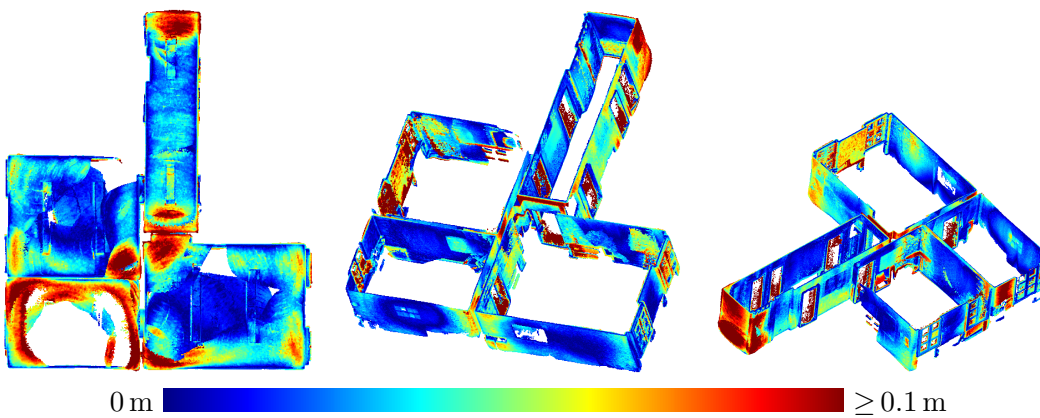


Figure 2.24: Accuracy of the HoloLens point cloud evaluated against the TLS ground truth. The scale factor was determined to 0.9887 by registration. The mean distance to the ground truth amounts to 2.4 cm. Source: [238].

Fig. 2.22 shows a point cloud of a subset of three of the rooms and the hallway that was derived from range images captured by the HoloLens range camera and registered via the camera poses provided by the tracking system. The evaluation results for this point cloud are depicted in Fig. 2.23 for fixed scale with a resulting mean accuracy of 4.0 cm and in Fig. 2.24 for a scale factor of 0.9887 determined by registration on the TLS ground truth point cloud and a resulting mean accuracy of 2.4 cm.

2.4 Discussion

In the following sections, the results of the experiments presented in Sec. 2.3 are discussed. Again, the section starts with the evaluation of depth sensing in Sec. 2.4.1, continues with discussing the results of the evaluation of the HoloLens tracking system in Sec. 2.4.2 and concludes with Sec. 2.4.3, where the results of the experiments dedicated to indoor mapping are discussed.

2.4.1 Depth Sensing

Regarding the influence of the warm-up process of the device on the accuracy of range measurements as presented in Fig. 2.7, it can generally be recommended to let the device warm-up for at least one and a half hours before starting measurements with the HoloLens range camera, when precision is of importance. When using the device for indoor mapping tasks, this warm-up-induced drift in range measurements can potentially further increase drift effects caused by drift in tracking as reported in Sec. 2.3.2.

In the findings presented in Fig. 2.8, noise in range measurements of up to 2 cm under unfavorable conditions (long distances, high inclination) was ascertained. However, in the context of indoor mapping, the influence of such effects cannot be easily assessed, as indoor mapping is generally a dynamic process affected by the movement of the user wearing the device through the environment to be mapped. In contrast, the findings presented in Sec. 2.3.1 apply to static situations, where a scene is captured from one fixed sensor position over a certain range of time. In the context of indoor mapping, it will rarely happen that a part of the scene is observed from only one position.

A user mapping his environment should take care to observe all surfaces of interest from a distance of at most about 2.2 m and from a not too steep angle. However, even if all relevant parts of the scene are captured by favorable sensor positions, there will always also arise range measurements beset by high noise caused by large distances or oblique angles. Raw indoor point clouds derived from HoloLens range images will thus always contain a high amount of noise as is apparent in Fig. 2.22. This is further contributed to by errors in the sensor poses obtained from the tracking system.

Thus, HoloLens range measurements have to be further processed e.g. by removing single points affected by high noise or whole range images affected by tracking errors to yield reasonable results. The triangle meshes provided by the HoloLens system, although produced in a black-box manner, can be regarded as the result of such a processing. As shown in Tab. 2.3, the accuracy of the triangle mesh falls in the range between the range images captured under favorable conditions and the one suffering from a too large distance of the sensor to the scene. The mesh represents the accumulated knowledge the HoloLens system has of its environment after capturing range images from the three positions detailed in the table. However, the accuracy of the mesh as specified in Tab. 2.3 is lower than the accuracy of the range images captured from shorter distances. It can be suspected, that this is caused by the reduction of spatial resolution due to

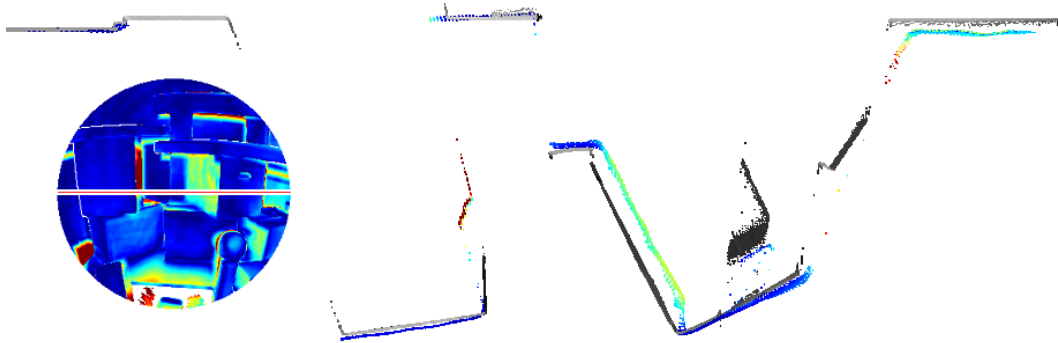


Figure 2.25: Cross section from Fig. 2.11(a) corresponding to the red line on white background shown in the small range image in this figure. The large deviation from the ground truth (visualized in gray-scale) of the range measurements on the left side of the right box are possibly caused by multi-path effects. Source: [238].

the triangulation process.

Besides inaccuracy caused by sensor noise, there are also systematic effects degrading the accuracy of HoloLens depth sensing in some parts of the data. In Fig. 2.11(a), e.g., the left side of the box on the right is indicated by turquoise to yellow coloring to deviate quite strongly from the ground truth data. A horizontal cross section of this part is shown in Fig. 2.25. This deviation could possibly be caused by multi-path effects. Other deviations possibly caused by multi-path effects include the upward bulging of triangle meshes occurring in corner situations on ceilings as indicated by red color in the top view visualizations on the left-hand side of Fig. 2.17 and 2.18.

2.4.2 Tracking

The evaluation results presented in Sec. 2.3.2 show that the HoloLens tracking system is capable of marker-less inside-out tracking in indoor environments with an accuracy of two centimeters or better. This is also supported by the apparent spatial stability of virtual objects as perceived by the user wearing the device. Fig. 2.14 seems to imply a correlation of positional RPE values with velocity and rotational RPE with angular velocity over the course of the trajectory.

It is noteworthy, however, that the presented results seem to indicate a higher tracking accuracy when covering the range sensor. It can be assumed, that this is caused by the ToF range sensor of the HoloLens interfering with the motion capture system. In this case, both conducted experiments would not adequately assess the true HoloLens tracking accuracy as i) in the case of the uncovered range sensor, the ground truth values would be distorted and ii) in the case of the covered range camera, the system would not be evaluated in its usual working mode. In any case, the presented results can be regarded as a lower bound for the quality of the HoloLens tracking system, as it is not to be assumed that using the range sensor degrades tracking quality.

Anyhow, even quite small drift effects as those observed in room-scale trajectories accumulate with traveled distance as shown in Fig. 2.15. In situations as the one depicted, loop closure is detected by the HoloLens system after re-entering the room and the position of the device is corrected accordingly. The triangle meshes however are only corrected locally in the direct surrounding of the place of the detected loop closure by removing falsely positioned meshes that do not correspond to physical structures when they get in the field-of-view of the range sensor.

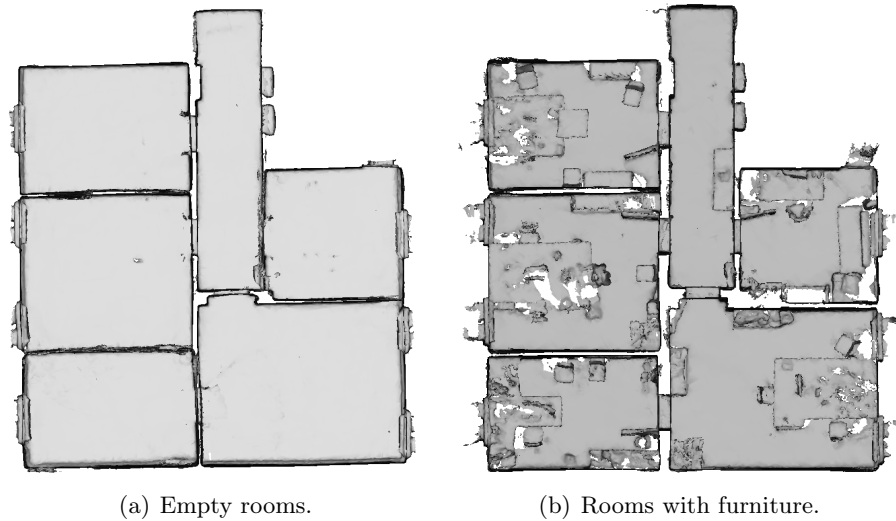


Figure 2.26: HoloLens triangle meshes of the same indoor environment acquired a) in an empty state without furniture b) with furniture.

2.4.3 Indoor Mapping

The precision of the HoloLens triangle meshes is overall quite high as indicated by Fig. 2.19. However, the accuracy evaluated against TLS ground truth data of the triangle meshes acquired in an empty, furniture-free indoor environment (see Fig. 2.20) is significantly lower as the accuracy of triangle meshes of the same indoor environment when furniture is included (see Fig. 2.17). As Fig. 2.26 suggests, the difference between the acquisition of an empty indoor environment to the acquisition of the same rooms including furniture is given by the fact that some of the walls in the indoor mapping results with empty rooms seem to have a significantly underestimated thickness. While the geometry and dimensions of the individual rooms seem to be quite consistent in both datasets, the rooms in the dataset without furniture seem to be somewhat displaced to one another in comparison to the dataset including furniture.

A schematic overview of intensity and direction of the deviations of a HoloLens triangle mesh in relation to the TLS ground truth data is presented in Fig. 2.27. The detail view on the right-hand side of this figure indicates, that the accuracy is actually worse than Fig. 2.20 suggests. The distance to the nearest point of the TLS reference point cloud is used to quantify the accuracy. Thus, mesh vertices of walls that are deviated to a degree so that they are nearer to the reference wall surface of an adjacent room than to the actual wall surface can show a misleadingly low distance to the ground truth data.

Furthermore, it is apparent that the major deviations occur on walls in which doors have been crossed during the indoor mapping process. In most cases, the direction of the respective deviation is the opposite of the direction into which the rooms have been entered coming from the central floor. This gives the impression that the rooms themselves are captured quite successfully taken individually but that the individual rooms are shifted towards one another along the directions of the doors connecting them. This impression is confirmed by Fig. 2.21, showing that the accuracy significantly improves when the rooms are registered onto the ground truth point cloud on a per-room basis.

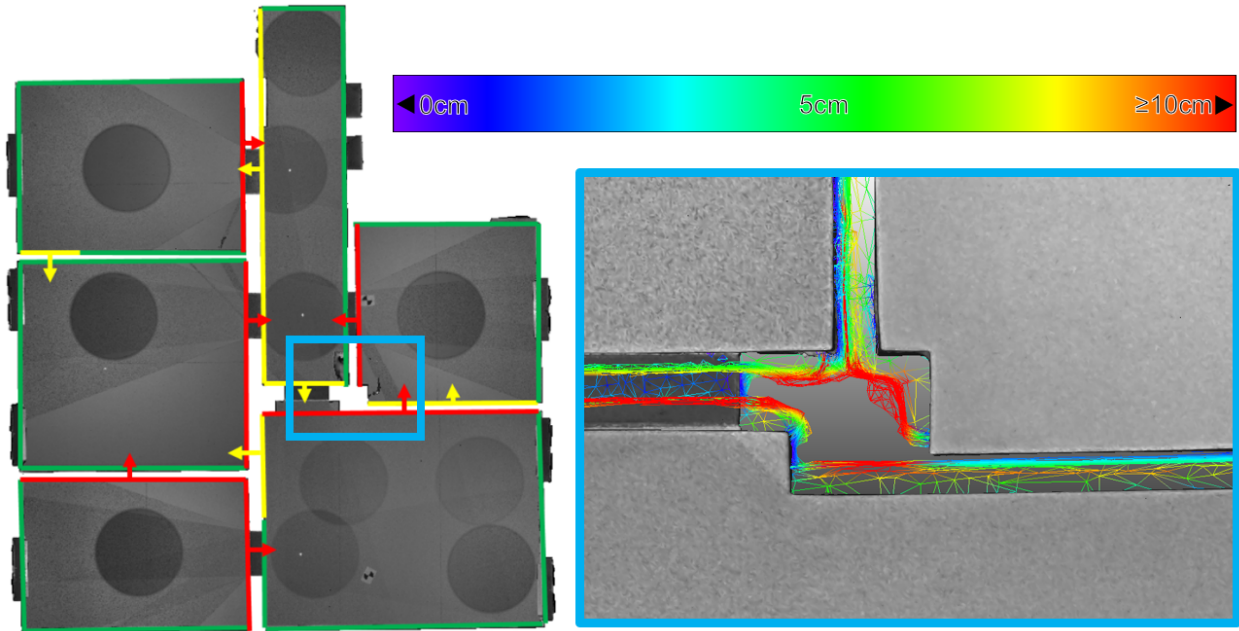


Figure 2.27: Deviations of a triangle mesh of a furniture-less indoor environment acquired with the Microsoft HoloLens compared to TLS ground truth data. On the left, low deviations are demarcated in green, medium deviations in yellow and large deviations in red. The arrows indicate the direction into which the HoloLens triangle mesh deviates from the ground truth data. In the detail view on the right of the section demarcated by the blue frame, the HoloLens triangle mesh is colored according to distance to the TLS ground truth with the color scale being depicted above. Source: [239], modified.

Contrary to the large displacements of adjacent rooms in the case of the furniture-free environment, no displacements of such a magnitude were noticed in the indoor triangle mesh presented in Sec. 2.3.3. Only the wall between the upper and middle room on the left-hand side in Fig. 2.16 shows a slight narrowing towards the room corners. This could be caused by multi-path effects as it is probably also the case with the outward bulging of ceilings in room corners visible as red spots in some room corners in the top view visualizations in Figures 2.17 and 2.18.

It can be suspected that the large deviations between the individual rooms in Fig. 2.20 were maybe caused by this furniture-less state, the white texture-less walls causing a deterioration in tracking performance. As, in this case, the ground truth data set also did not contain furniture, all parts of the triangle mesh representing objects not present in the ground truth data had to be removed manually.

Although the evaluation still establishes a significant scale factor, its impact on the accuracy of the triangle mesh is by no means as strong as in the case of the furniture-free environment. With a mean accuracy of 1.7 cm for corrected scale and 2.3 cm for the original scale of the triangle mesh, these results demonstrate the high potential of the HoloLens for the use-case of indoor mapping.

Large-scale drifts in tracking as discussed in Sec. 2.4.2 however still prove an obstacle. In these cases, it would be necessary to distribute a detected offset over the whole trajectory and its attached meshes in the event of loop closure detection. Corrections like this are not taken into consideration for the HoloLens as it is not needed for its actual use-case as an augmented

reality device where only the correctness of the triangle mesh in the direct vicinity of the user is of importance.

The evaluation of the point cloud of a subset of four rooms of the same indoor environment derived from range images of the HoloLens ToF camera resulted in an accuracy of 2.4 cm for corrected scale and 4.0 cm for the original scale. This accuracy is lower than the resulting accuracy of the triangle meshes of the same environment whereas the evaluation of the scene presented in Sec. 2.3.1 resulted in the triangle meshes showing a lower accuracy than range measurements captured under suitable conditions. In this case, however, the sensor was capturing the scene in a static setting for a certain duration whereas, here, it was constantly moving through the environment with the user. Thus, as already discussed in Sec. 2.4.1, every part of the mapped indoor environment can be expected to not only have been captured in favorable constellations, but also from high distances or steep angles. Furthermore, inaccuracies in the tracking of the device pose propagate to the global position of points resulting from range images. The resulting point clouds are thus characterized by a huge amount of noise, as can be seen in Fig. 2.22. Besides, some parts of the indoor environment were only ever captured under unfavorable conditions by the range sensor. For example, in the case of the lower left room depicted in the top view visualization in Figure 2.23 and 2.24, the operator mapping the environment did not look upwards to the ceiling. The ceiling surface in this room was thus only captured partially and only under oblique angles, which results in low accuracy in the respective part of the point cloud.

2.5 Conclusion and Outlook

In this chapter, a thorough evaluation of the Microsoft HoloLens regarding its adequacy for the use-case of indoor-mapping was presented. After a brief survey of the different camera sensors the device is equipped with, the performance of its depth sensing and tracking system were investigated independently from each other. Subsequently, the complete system w.r.t. the task of mapping indoor environments was evaluated.

While the potential of the HoloLens as an off-the-shelf tool for indoor mapping could thus be demonstrated, its shortcomings also need to be highlighted. It however has to be remembered that the HoloLens was not primarily designed as an indoor mapping device. It rather is a mobile augmented reality headset. Thus, its capabilities in capturing the geometry of its surrounding are geared towards the needs of an AR device, where typically only the direct vicinity of the user that is to be augmented with virtual content needs to be consistently known. Large-scale drift in tracking and the deviations in the captured meshes caused by it are not a problem from the viewpoint of augmented reality, as the user only ever perceives his current vicinity which is captured sufficiently consistent to allow for virtual content to realistically interact with the physical environment.

Nevertheless, the HoloLens as an off-the-shelf, rather low-cost device that is easy to use still holds great promise for effortlessly capturing the geometric structure of large indoor environments.

Regarding potential future work on the evaluation of the HoloLens or similar sensor systems, the investigations presented in this chapter can certainly be further extended and deepened. The evaluation of the range sensor should particularly be extended by a wider variety of test objects and scenarios. For instance, examining further test geometries and constellations could enable further insight on the behaviour of multi-path effects. In addition, investigating the influence of different surface materials holds potential for further research. Furthermore, the second version of the HoloLens should also be comparatively examined regarding its potential for the use-case of indoor mapping.

Chapter 3

Indoor Reconstruction

Indoor mapping systems as discussed in the previous Chap. 2 can be used as a basis for realizing indoor fused reality scenarios as envisioned in Sec. 1.2.5. In this regard, primary indoor mapping data like point clouds, triangle meshes or image data are not really usable in the context of suchlike applications as they are typically characterized by a large quantity of data with a rather low level of semantic information. In order to be efficiently usable in the context of indoor fused reality applications, indoor mapping data thus needs to be processed to a more accessible form of representation, i.e. building models as described in Sec. 1.2.1.

While building models can be derived from indoor mapping data via a process of manual modeling, this is a rather laborious, time-consuming endeavor. Thus, methods that can be applied towards the automated reconstruction of semantically enriched and geometrically completed and generalized building models hold a significant importance for the realization of indoor fused reality scenarios. This task, often termed as automated indoor reconstruction, is the topic of this chapter. First, Sec. 3.1 presents an overview of the topic of indoor reconstruction and the approaches typically applied to this aim. Afterwards, Sec. 3.2 presents, as an own contribution to this field of research, a voxel-based indoor reconstruction approach that can be used to derive semantically enriched and geometrically refined voxel models of indoor environments from triangle meshes acquired for instance with the Microsoft HoloLens whose indoor mapping capacities were evaluated in the preceding Chap. 2. Furthermore, appropriate methodology for the quantitative evaluation of indoor reconstruction results in voxel representation is presented as well. Subsequently, Sec. 3.3 presents six datasets acquired with the Microsoft HoloLens and the respective quantitative evaluation results of the proposed indoor reconstruction approach. Lastly, Sec. 3.4 provides further discussions before Sec. 3.5 closes this chapter with a conclusion and an outlook on future work.

Within this chapter, material published in [243] is used. An early version of the proposed approach was also published in [242].

3.1 Fundamentals of Indoor Reconstruction

Automatically reconstructing building models from indoor mapping data is a broad and active field of research with a large number of different approaches proposed over the last decade [24, 444, 359, 353, 267, 320, 456]. While indoor mapping approaches are sometimes denoted as 'reconstruction' or 'modelling' of indoor environments [117, 145, 675, 690], this chapter focuses on approaches that do not merely aim to geometrically acquire indoor building environments by means of sensor systems. Rather, the discussed indoor reconstruction approaches aim to automatically derive compact, abstracted building models in the sense of Sec. 1.2.1 from the given indoor mapping geometries.

This implies the necessity to exceed a mere semantic segmentation task, where the represented indoor mapping geometries are labeled as belonging to different semantic classes [113, 466, 684, 108]. Indoor reconstruction rather comprises the geometric abstraction of the given input data to refined geometries. For instance, planar walls should be represented e.g. by planes or voluminous geometries and not just by noisy points labeled as 'wall'. Furthermore, also the parts of wall surfaces that are occluded by furniture or incompletely acquired by the respective indoor mapping system need to be geometrically reconstructed. In this regard, the task of automated indoor reconstruction is related to the fields of three-dimensional, geometric inpainting [498, 587] and semantic scene completion [129, 347, 124].

Besides, indoor reconstruction is also related to the task of automatically reconstructing building models from outdoor mapping data such as point clouds acquired by terrestrial or airborne laser scanning systems [618, 471, 213, 323, 601, 233, 645]. Here however, the typical characteristics of indoor building environments as discussed in Sec. 1.2.2 pose specific additional challenges in comparison to outdoor building reconstruction. In this context, a potentially interesting object of research is the fusion of indoor and outdoor building reconstruction. For instance, a reconstruction approach applicable to interior as well as exterior building structures has been proposed [469]. Other approaches focus on registering indoor and outdoor building representations by detecting and registering window openings in both datasets [120, 293]. In addition, the field of indoor reconstruction is also related to automatic reconstruction approaches targeting all kinds of other man-made objects with clearly defined geometry characterized by a high degree of regularities and symmetries. In this context, approaches for automatically reconstructing structures such as pipes [265, 598, 521], columns [532, 360, 491], briges [15, 317, 476] and power pylons [678, 602] or other steel structures [654] from point clouds have been proposed.

Indoor reconstruction approaches have been utilized in a range of different application scenarios such as change detection in building structures [557, 419, 294, 573], facility management [541] and progress tracking in construction projects [91, 52, 360]. Furthermore, indoor reconstruction has been applied in the context of cultural heritage [27, 490, 623], the simulation of 5G signal distribution in indoor environments [126, 540], building energy analysis [116, 187, 433, 249] and indoor navigation [174, 537, 407, 141, 176, 418].

While the reconstruction of building models from data such as point clouds is frequently approached manually or semi-automatically [665, 43, 448, 651], automating this modeling process has been in the focus of research efforts for a long time [559, 533, 355, 597]. In this context, a range of studies have been conducted analyzing the challenges and specific requirements posed on the task of automated indoor reconstruction [406, 215, 478, 604].

While this work focuses on deriving three-dimensional building models from three-dimensional input data such as point clouds or triangle meshes, there are also approaches that aim at deriving building models (2D and 3D) from scanned floor plans [148, 344, 253, 564, 636]. Furthermore, a range of approaches have been proposed that aim at deriving floor plans from three-dimensional input data such as point clouds [430, 579, 343, 105, 182, 168]. As the resulting floor plans can be converted to three-dimensional building models by vertical extrusion, these approaches can be considered indoor reconstruction methods. Worth of mention is furthermore a range of approaches aiming to derive floor plans from trajectories [21, 452, 404]. Lastly, two-dimensional, possibly outdated floor plans have been used as an additional source of information in the context of indoor reconstruction from point clouds [46].

Besides indoor reconstruction approaches relying on three-dimensional input data such as point clouds or triangle meshes, some approaches also aim at reconstructing 3D room layouts from single (panorama) images [235, 659, 553, 694]. A similar approach is followed in [501], where TLS point cloud data is represented as a panorama image from which room models are reconstructed.

Two fundamental components in the context of indoor reconstruction applications are the reconstruction of wall structures from potentially cluttered and occluded indoor mapping data [4, 401, 427, 467, 340] and the segmentation of indoor building structures in individual subspaces, i.e. room partitioning [71]. While room partitioning has also been applied to floor plans [366, 10, 132] or building models [642, 439], it is frequently used in the context of indoor reconstruction from indoor mapping data as well [31, 405, 607, 572, 653].

Room partitioning is however not necessarily used in all indoor reconstruction approaches. Some methods are applicable only to single room scenarios [80, 643, 631, 123, 501, 335]. These approaches are only applicable to indoor mapping data representing a single room. Thus, indoor mapping data of more complex indoor environments need to be segmented into rooms beforehand, either manually or by means of a separate room segmentation approach. Furthermore, some indoor reconstruction approaches are applicable to more complex indoor environments comprised of multiple rooms but do not partition the resulting building model into subspaces such as rooms [578, 427, 641, 340]. In these cases, space is only partitioned into indoor space and outside space.

In general, a range of different strategies have been put forward when approaching the task of indoor reconstruction. A commonly applied approach is to detect planes representing walls and to assemble them into rooms which solves the mentioned problems of wall reconstruction and room segmentation. In doing so, two principal strategies can be followed. Some approaches rely on detecting local plane patches (e.g. by region growing approaches) [596, 341, 466] and aim to assemble these plane patches into room surfaces [416, 422, 641, 518, 429]. In contrast to this bottom-up strategy, the problem of indoor reconstruction can be approached in a more top-down manner as well. This encompasses firstly the detection of major global planes (e.g. by RANSAC [175, 512] or Hough Transform [73]). Then, the detected planes are intersected with one another to create a cell complex whose cells are subsequently assigned to the interior space and the outside or partitioned into rooms. In most cases, a two-dimensional cell complex is created from detected wall surfaces projected to the horizontal plane as lines [402, 427, 607, 328, 574]. In multi-storey scenarios, one such 2D cell complex is created per storey. In some cases however, planes resulting from floors and ceilings are taken into account as well to create a 3D cell complex [400, 123]. Furthermore, a 2D cell complex created by intersecting lines has been used for the reconstruction of occluded walls with openings from orthoprojected raster representations of planar walls [383]. Lastly, some approaches combine elements of both, top-down and bottom-up strategy [22, 403, 425].

Other indoor reconstruction approaches rely on trajectories of the respective indoor mapping system [420, 157, 468, 687, 339]. Trajectory information can for instance be utilized for facilitating room partitioning and door detection. However, it limits the applicability of an indoor reconstruction application to indoor mapping datasets containing trajectory information.

Furthermore, some indoor reconstruction approaches rely on representing the input data in discrete rasters, i.e. voxel grids [67, 174, 421, 176, 178]. Besides three-dimensional voxel grids, some approaches also use horizontal 2D raster representations for room partitioning [246, 363, 261, 105] or vertical 2D rasters resulting from orthoprojection of wall surfaces for wall reconstruction [4, 151, 143, 599, 588]. The latter approaches achieve good result in the reconstruction of occluded wall surfaces and the detection of wall openings like doors and windows. However, they assume wall surfaces to be planar. Furthermore, raster representations are well-suited for deep learning approaches which are recently gaining in prevalence in the context of indoor reconstruction [104, 124, 288, 183, 653].

Besides ceilings, floors and wall surfaces, some approaches also aim at reconstructing other building elements such as furniture [588, 647, 492], elements affixed to room surfaces such as fire alarms or power plugs [6, 556] or door openings [143, 38, 176, 653]. The detection of door openings as transition spaces between rooms is also of importance in the context of applications aiming to reconstruct the room topology of indoor environments [575, 492, 653]. Besides the topology of rooms,

some approaches also focus on extracting the topology of wall structures [49, 673].










In order to tackle the challenging problem of indoor reconstruction, facilitating assumptions about the structure of indoor environments to be reconstructed are often applied. One such assumption concerns the already discussed indoor reconstruction approaches that assume that the given input data represents exactly one room. Another commonly applied assumption is the Manhattan world assumption, i.e. that all surfaces of a building are perpendicular to one of the three coordinate axes [80, 502, 372, 288, 496]. Other indoor reconstruction approaches weaken this assumption about the structure of indoor environments by allowing for horizontally diagonal [402, 426, 607, 520, 574] or even curved walls [427, 328, 652, 340, 634] or non-horizontal, slanted ceilings [67, 123, 421, 339]. Another important aspect regarding the flexibility of indoor reconstruction approaches is the handling of multi-storey building structures. In this context, some approaches are able to deal with complex vertical building structures such as stairwells [502, 174, 328, 421, 339].

While many indoor reconstruction approaches only output surface models of reconstructed indoor environments [502, 401, 427, 57, 340, 634], e.g. in the form of triangle meshes, some approaches aim at reconstruction walls as actual, volumetric objects instead of just their surfaces [569, 260, 425, 418, 574]. Some approaches even aim at directly generating actual BIM data formats such as IFC [23, 364, 47, 354].

Concerning the evaluation of indoor reconstruction approaches, quantitative metrics such as precision, recall or Intersection-over-Union (IoU) are frequently used [328, 48, 574, 418]. Besides, there are also several approaches which evaluate the quality of reconstruction by quantifying the metric distances of points of the input data to the reconstructed room surfaces [126, 641] or measure the distances between certain corresponding walls in the ground truth model and achieved reconstruction results or room areas, respectively [335]. For some other approaches, results are only visually presented and qualitatively discussed [176, 425]. Results for room partitioning are mostly evaluated on the basis of comparing the number of rooms (and doors where applicable) to the respective ground truth value [328, 126, 652, 418].

Besides the respective evaluations of proposed indoor reconstruction approaches, there are also some studies explicitly targeting evaluation methodology for indoor reconstruction tasks. For instance, a framework for the unsupervised evaluation of indoor reconstruction results has lately been proposed which checks the internal consistency of the derived indoor model against a formal grammar containing semantic, geometric and topological consistency rules [417]. Others, in turn, deal with methodology for the geometric validation of indoor reconstruction results against point clouds as a kind of unlabeled, solely geometric ground truth [69]. The proposed procedure is however based on manual interaction of the user. Another framework for evaluating the quality of indoor reconstruction results and also relying on manual user interaction targets reference indoor models provided by expert manual modelers [285]. Other works dealing with evaluation methodology concern themselves not with indoor reconstruction results, but with the indoor mapping point clouds that are the input data for indoor reconstruction procedures [106, 34]. In this context, evaluation methodology is presented to assess accuracy and consistency of these data as a prerequisite for a successful indoor reconstruction.

Table 3.1: Color scheme for voxel classes. Source: [243].

	Ceiling		Wall		Wall Opening
	Floor		Interior Object		Empty Interior
	Not Classified		Wall & Ceiling		Wall & Floor

3.2 Methodology

A novel method for the reconstruction of voxel models of indoor environments from unstructured 3D data with oriented normals is presented in the following. Fig. 3.1 visually summarizes the proposed approach, while the colors used for semantic voxel classes are detailed in Tab. 3.1. The given input data representing indoor environments are voxelized to a three-dimensional voxel grid. In this voxel representation, a model of the indoor environment is reconstructed by assigning voxels to rooms and semantic classes.

The proposed method is generally applicable for unstructured 3D data with oriented normals. It could thus also be applied to point clouds given that the normals of the points can be consistently oriented to differentiate between inside and outside of the indoor environment. However, this chapter focuses on triangle meshes as captured with the Microsoft HoloLens or the Matterport system. Especially in the case of the Microsoft HoloLens, the input triangle meshes can be expected to be affected by noise and incomplete scene acquisition (see Chap. 2). Furthermore, it is taken into consideration, that the represented indoor environments can contain a large amount of clutter like furniture objects that can occlude room surfaces.

With respect to the represented building structure, the presented approach is intended to be as generic as possible regarding the actual shape of room surface geometries in indoor environments. In this context, room surfaces such as ceilings, floors and walls are not necessarily assumed to be planar or to adhere to the Manhattan World assumption. Only floor and ceiling surfaces are presupposed to be approximately horizontal within a range of $\pm 60^\circ$ for ceilings and $\pm 45^\circ$ for floors. Walls are expected to vertically connect the borders of ceiling and floor surfaces. They can however have protrusions and recesses in their vertical course. Furthermore, the upward direction is assumed to be known and the input data to be vertically aligned accordingly. The proposed approach is predominantly based on assumptions about the size and proportions of indoor spaces, derived from the typical dimensions of their human inhabitants. These assumed parameters are formulated in terms of metric values to be independent of the applied voxel resolution.

The general aim of the proposed approach is to automatically generate basic semantic and geometric models of challenging indoor environments with arbitrary building structures from noisy, cluttered and incomplete indoor mapping data as attainable with easy-to-use, off-the-shelf devices like the Microsoft HoloLens. The resulting voxel models are partitioned into rooms and connecting transition spaces. Furthermore, they provide a semantic distinction between the building structure itself (i.e. ceilings, floors and walls) and furniture objects and other clutter contained in the building within the geometric bounds of the respectively selected voxel resolution. The resulting voxel models can serve as a basis for a multitude of potential application scenarios, i.e. automated BIM model reconstruction, indoor navigation or indoor simulation/analysis scenarios e.g. with regard to temperature, noise or substance dispersion, air quality or emergency scenarios.

In the following, the chapter proceeds with outlining necessary steps of data preprocessing in Sec. 3.2.1. Subsequently, the various steps of the proposed reconstruction algorithm are detailed

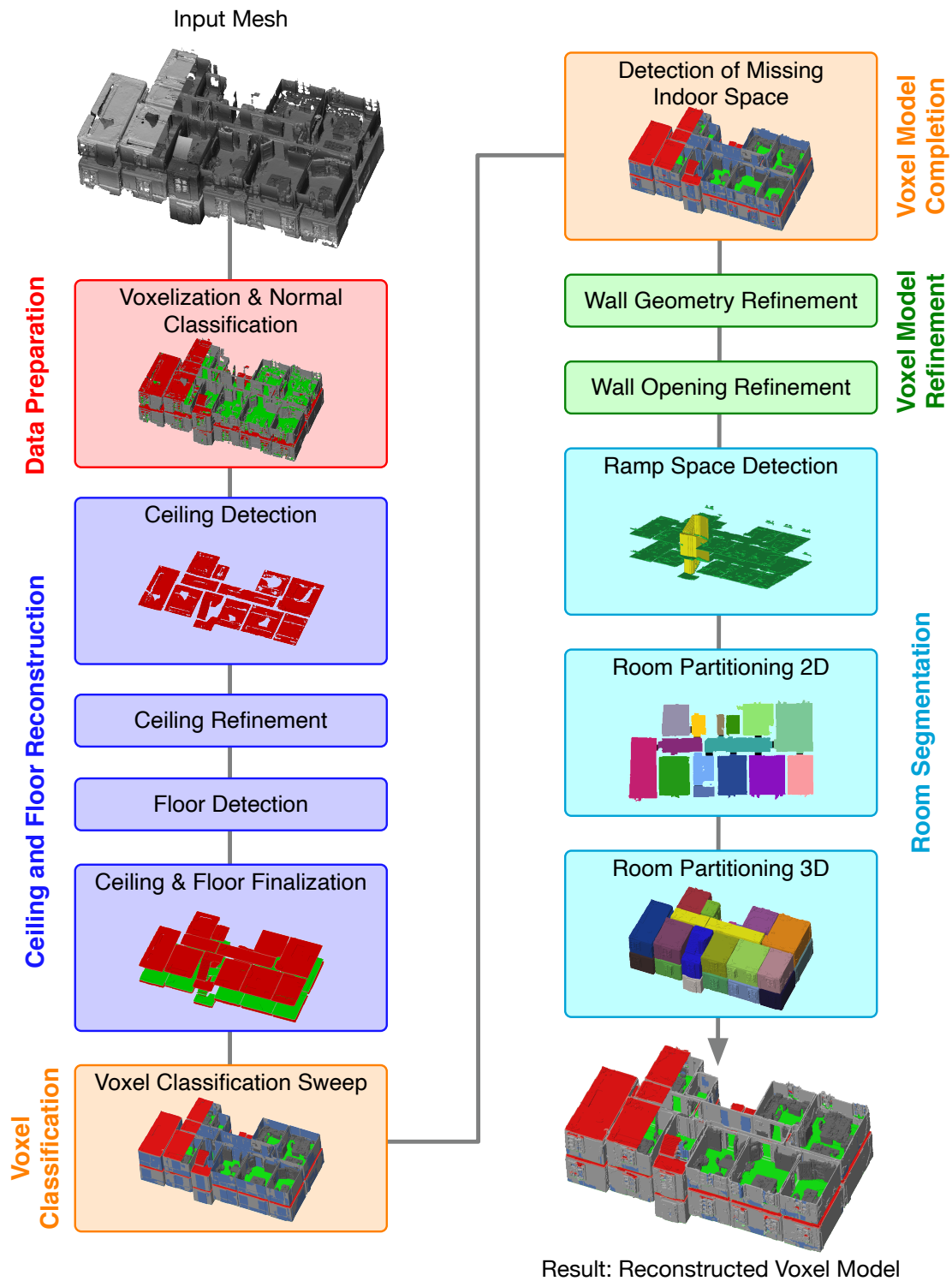


Figure 3.1: Overview of the processing workflow of the proposed reconstruction method: the input data are represented by a triangle mesh, while the output is given by the reconstructed voxel model. The individual sections are given in the vertical text. Source: [243].

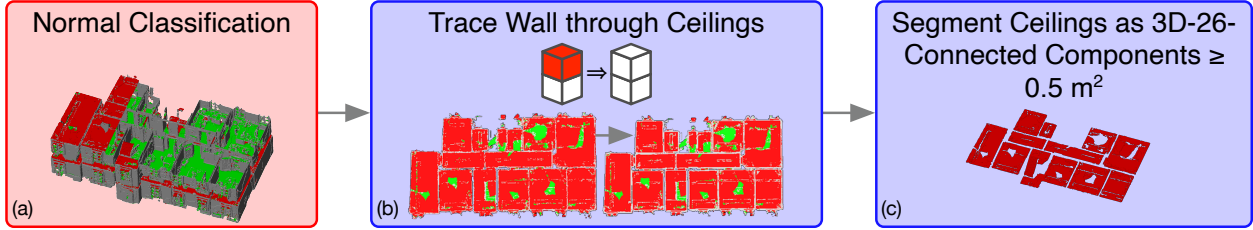


Figure 3.2: Overview of the ceiling detection process. In (a) and (b), voxels are colored in red for 'Normal Down', green for 'Normal Up' and white for 'Normal Horizontal'. In (c), the detected ceiling segments are visualized in red (only the upper floor is shown for better visibility). Source: [243].

in Sec. 3.2.2 to 3.2.6. In doing so, the workflow of the semantic classification and geometric reconstruction in voxel space is summarily detailed in Sec. 3.2.1 to 3.2.5. Furthermore, Sec. 3.2.6 presents a novel indoor space partitioning approach based on the resulting voxel model. Finally, Sec. 3.2.7 describes the methodology applied for performance evaluation.

3.2.1 Data Preparation

As a preliminary processing step, the input triangle mesh is voxelized into a regular voxel grid of fix but arbitrarily choosable resolution. In doing so, all voxels that do not intersect any triangles are assigned the value 'Empty'. The non-empty voxels that do intersect triangles get assigned the values 'Normal Up', 'Normal Down' or 'Normal Horizontal' based on the normal vector directions of the contained triangles. If the majority of the normal vectors of the contained geometries is pointing downwards within the range of $\pm 60^\circ$, a respective voxel is classified as 'Normal Down'. If the majority of the contained normal vectors is pointing upwards within the range of $\pm 45^\circ$, it is classified as 'NormalUp'. Else, it is classified as 'Normal Horizontal'.

The resulting voxel grid serves as input for the reconstruction algorithm specified in the following sections. The main aim of the proposed reconstruction algorithm consists of a segmentation of this voxel grid into single rooms and a classification of the voxels belonging to a room as 'Ceiling', 'Floor', 'Wall', 'Wall Opening', 'Empty Interior' or 'Interior Object'.

3.2.2 Ceiling and Floor Reconstruction

The reconstruction process starts with the initial detection of ceiling segments in the voxel grid derived from the input data. These ceiling segments are subsequently refined and corresponding floor surfaces are detected underneath. The following sections describe this process in detail. Furthermore, the workflow is visualized in Figures 3.2 and 3.3.

3.2.2.1 Ceiling Detection

In a first step visualized in Fig. 3.2(b), the grid of voxels classified by normal direction is traversed from bottom to top. In doing so, every 'Normal Down' voxel, that has a 'Normal Horizontal' voxel directly underneath is turned to 'Normal Horizontal' itself. This traces the walls through the ceiling surfaces and causes ceiling surfaces of neighboring rooms to be separated by the walls. Subsequently, in Fig. 3.2(c), ceiling segments are segmented as 26-connected 3D components among the 'Normal Down' voxels. Ceiling segments with a horizontal extent of less than 0.5 m^2 are discarded at this step. The corresponding parts of the building are however incorporated later on in Sec. 3.2.4.

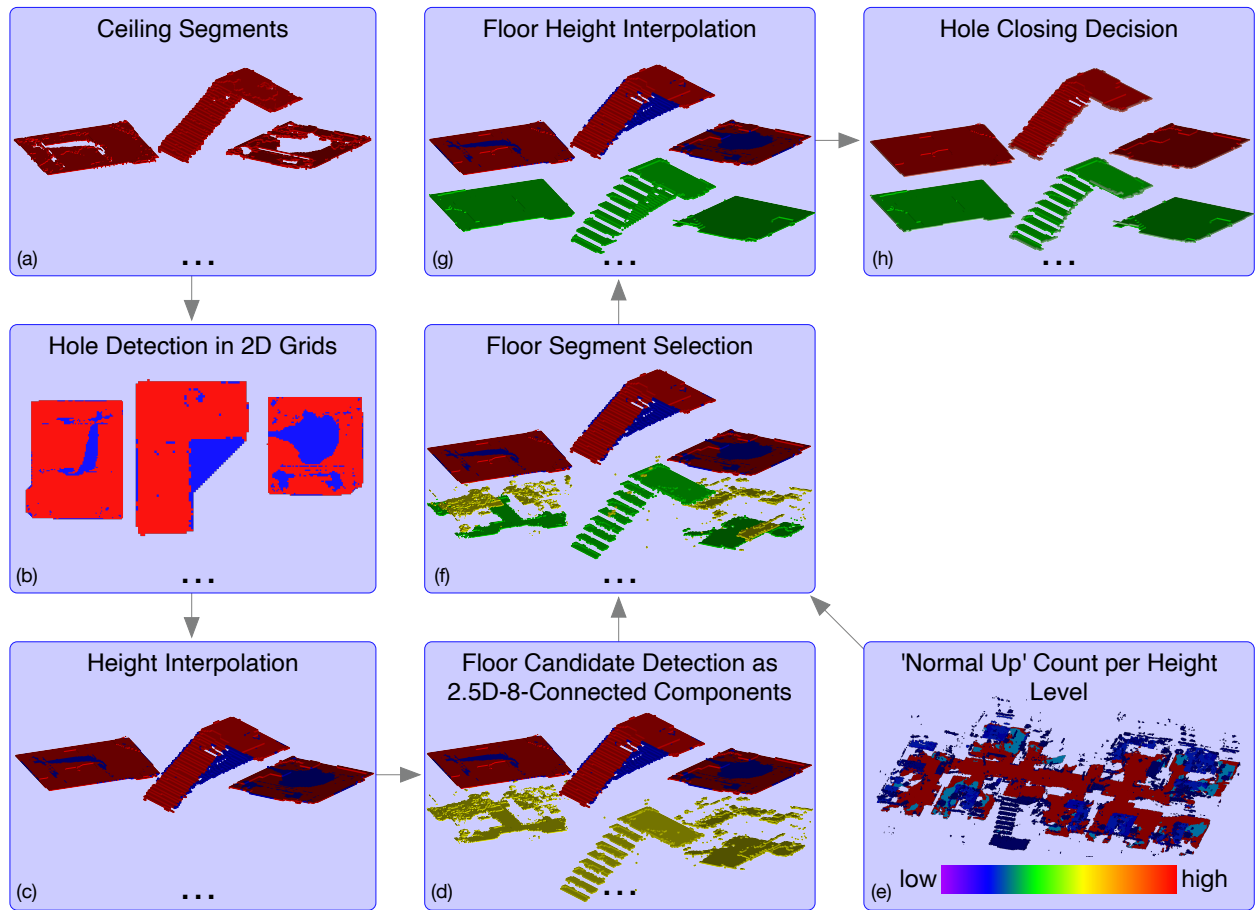


Figure 3.3: Overview of the ceiling and floor reconstruction process. All subfigures except (e) are colored according to Tab. 3.1. Furthermore, holes detected within the ceiling segments are depicted in blue. In (d) and (f), floor candidate segments that are not identified as the actual floor are depicted in yellow. Source: [243].

3.2.2.2 Ceiling Refinement

The detected ceiling segments are used as a starting point for reconstructing the rooms belonging to the respective ceilings. This is done by vertically traversing the voxel grid downwards from the ceiling voxels until the floor is encountered. This process described in Sec. 3.2.3 on the one hand presupposes that for each ceiling voxel, a corresponding floor voxel exists underneath to vertically terminate the voxel classification process. On the other hand, the ceiling segments detected so far can have holes distorting their actual geometry and accordingly the geometry of the rooms to be reconstructed underneath. These holes can be caused by incomplete acquisition during the indoor mapping process as well as due to occlusion by objects like lamps attached to the ceiling.

These holes have to be detected and eventually closed in order to allow for a correct reconstruction of the respective room. However, ceiling surfaces can also have openings like the interior of columns, inner yards or corners of walls pointing convexly inside the room. These need to be differentiated from the aforementioned holes in order to achieve a correct reconstruction of ceiling geometries. To this end, a refinement process as visualized in Fig. 3.3 is conducted for each detected ceiling segment.

The process commences in Fig. 3.3(b) with orthogonally projecting the individual ceiling segments in respective 2D pixel grids. In these 2D grids, the detection of potential holes is conducted by searching for empty pixels in-between ceiling pixels along the four directions of a 2D-8-neighborhood. The detected hole pixels are subsequently segmented as 4-connected 2D components.

Next, in Fig. 3.3(c), the height of all height pixels is interpolated based on the height of the ceiling pixels bordering the respective hole segment. Finally, a smoothing of the height values across the whole ceiling segment and its holes is conducted.

3.2.2.3 Floor Detection

For each resulting voxel position, the voxel grid is vertically traversed downwards until a 'Normal Up' voxel or a 'Normal Down' voxel belonging to another ceiling segment is encountered. The resulting floor candidate voxels as depicted in Fig. 3.3(d) are segmented to floor candidate segments as 8-connected 2.5D components with a threshold of 18 cm between neighboring floor voxels. This is a common value for the height of stair steps and enables floor segments to extend over stairs and ramps while mostly avoiding to spread over to surfaces on top of furniture like tables or chairs.

In the algorithm as originally published in [242], the respectively largest floor candidate segment was chosen to initialize the respective floor corresponding to each ceiling segment. This however can lead to large table surfaces being erroneously identified as floor in rooms where large parts of the actual floor surface are occluded. Thus, floor candidate segment voxels are weighted with the total amount of 'Normal Up' voxels at the height of each respective voxel as visualized in Fig. 3.3(e). For each room, the floor candidate segment with the largest summarized weight is selected as floor as depicted in Fig. 3.3(f).

3.2.2.4 Ceiling and Floor Finalization

The missing parts of the floor are subsequently filled as depicted in Fig. 3.3(g). Again, the height is interpolated based on the height of the borders of the existing floor segment. In the end, each ceiling or hole voxel has a corresponding floor voxel beneath it.

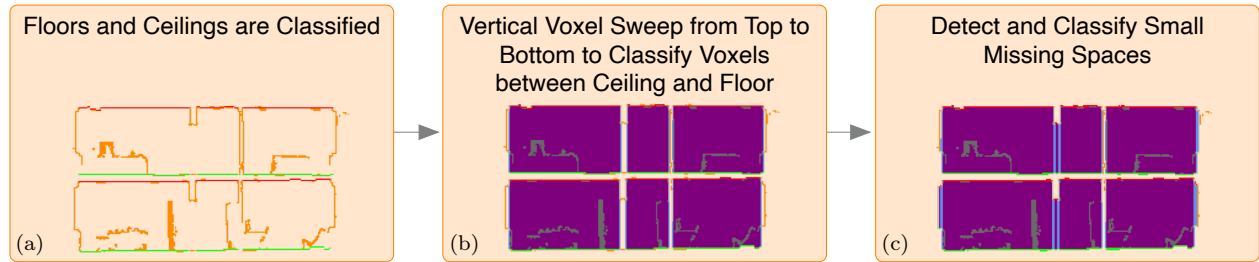


Figure 3.4: Overview of the voxel classification process. All subfigures are colored according to Tab. 3.1. Source: [243].

On this basis, ceiling holes are either closed by incorporating them into the ceiling surface or discarded along with the corresponding floor voxels as depicted in Fig. 3.3(h). To make this decision, the border and interior of each hole is checked along the whole height range between ceiling and floor. If intersections with other rooms in the interior or substantial evidence for the presence of walls along the border of a hole are found, it is discarded. Else, it is closed by incorporating it into the respective room. This approach might lead to unwanted results in case of inner yards delimited by large panorama windows. In this situation, a ceiling hole above an inner yard would be closed and incorporated into the surrounding room. Finally, walls are initialized along the borders of ceilings and floors.

3.2.3 Voxel Classification

At this point, ceilings and corresponding floors are identified. Classifying the voxels between ceiling and floor is now quite straightforward as visualized in Fig. 3.4. The classification is conducted just by checking if the respective voxels between a ceiling voxel and the corresponding floor voxel are empty or not. Empty voxels are assigned the class label 'Empty Interior'. Non-empty voxels on the other hand are classified as 'Interior Object'. Along the borders where wall voxels have been initialized, the class labels 'Wall' and 'Wall Opening' are assigned based on the same condition.

Note, that all non-empty voxels within the room interior are labeled as 'Interior Object' to demarcate them as not belonging to the building structure itself but to some object inside the respective room. Currently however, the proposed approach does not distinguish between different types of furniture objects. Nor does it segment 'Interior Object' voxels to actual object instances. Rather, it provides the basis for doing so in the context of applications making use of the created voxel models. In this context, it is also conceivable to go back to the actual indoor mapping geometries intersecting the respective voxels for segmentation or fine-grained classification tasks.

3.2.4 Voxel Model Completion

So far, only rooms covering a horizontal area of at least 0.5 m^2 with a continuously connected ceiling surface have been reconstructed. However, indoor environments typically also contain smaller sections with own smaller ceiling surfaces that are distinct from the main ceiling surface of larger rooms like alcoves, window recesses or the spaces inside door frames between rooms that should also be considered as part of the indoor space to be reconstructed.

These missing sections are detected and reconstructed in a manner similar to the algorithm

described so far. In doing so, however, no threshold on the ceiling area is applied this time. Larger areas of 'Normal Up' voxels that so far do not belong to any room are also considered. If no traces of suitable, corresponding ceiling surfaces can be found, a virtual, hypothetical ceiling surface on a fixed height of 3 m above the respective floor voxels is assumed.

This allows for the consideration of areas without a visible ceiling within the scope of the voxel-based indoor reconstruction framework. This can be the case for areas without any ceiling like inner courtyards that are nonetheless included in indoor mapping datasets or areas whose actual ceiling is not represented in the dataset. This can be the case for e.g. glass surfaces that cannot be captured with the ToF sensor of the HoloLens device or because the ceiling is too high to be captured given a respective sensor working range.

After detecting these additional indoor spaces, the aforementioned classification process for the voxels between ceiling and floor is applied for these spaces as well to complete the indoor voxel model. This is depicted in Fig. 3.4(c).

3.2.5 Voxel Model Refinement

With the previously described voxel sweep algorithm, voxels are assigned both class labels and room affiliations. Based on the resulting indoor voxel grid, subsequent refinement sweeps are applied to further improve the indoor voxel model. The respective refinement steps are visualized in Fig. 3.5.

3.2.5.1 Wall Geometry Refinement

Up to now, the reconstructed wall surfaces have a width of only one voxel owing to their initialization along the edges of ceiling segments that were propagated downwards during the voxel classification sweep described in Sec. 3.2.3. These were subsequently extruded vertically downwards as wall surface until encountering the floor. This causes protrusions and recesses on walls to either be classified as 'Interior Object' when protruding towards the room interior like window sills for instance or to be neglected when protruding towards the outside like window frames. However, these elements should rather be considered as belonging to the wall surface if they do not exceed a certain size. To resolve this, a wall geometry refinement is applied as depicted in Fig. 3.5(a).

In a first step of wall geometry refinement, non-empty voxels are added to room surfaces from the outside up to a distance threshold of 15 cm. Similarly, non-empty voxels are also added to the wall surface from the room interior. This again is applied to 'Interior Object' voxels within a distance of 15 cm from the wall surface. If a continuous succession of 'Interior Object' voxels beginning from the wall surface continues beyond this distance into the room, it is assumed that it belongs to a piece of furniture standing in front of the wall. Only smaller objects that protrude less than 15 cm are made part of the wall surface. However, note that pieces of furniture covering the complete vertical space of a room w.r.t. the respective voxel resolution cannot be distinguished from wall surfaces, i.e. a cupboard would be reconstructed as a protrusion of the wall if the spacing to the ceiling is negligible.

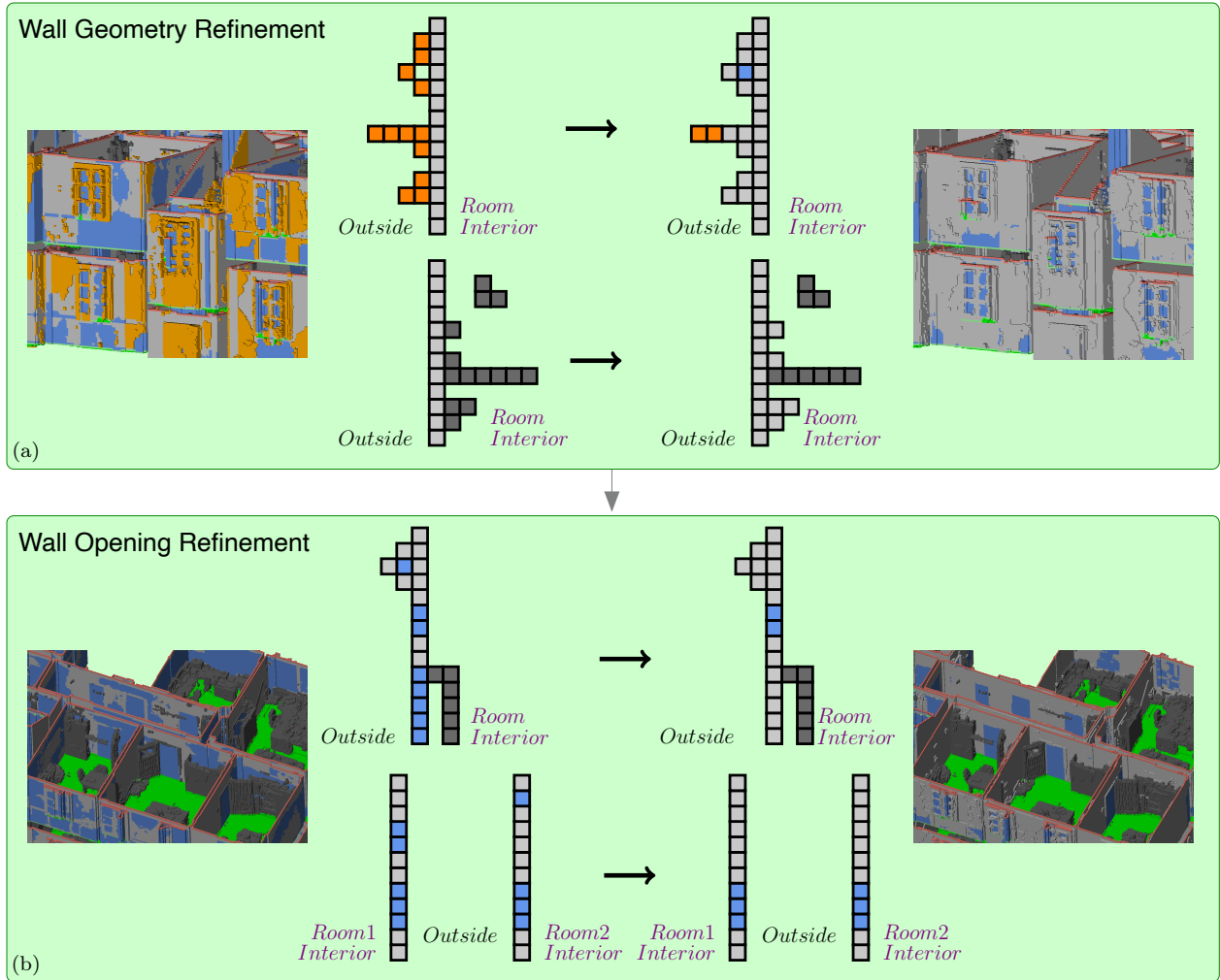


Figure 3.5: Overview of the wall surface refinement process. All images are colored according to Tab. 3.1. Source: [243].

3.2.5.2 Wall Opening Refinement

In a second wall surface refinement step visualized in Fig. 3.5(b), the occurrence of the class label 'Wall Opening' is adjusted. Currently, all voxels belonging to a reconstructed wall surface that are empty are labeled as 'Wall Opening'. Besides actual wall openings such as windows or doors, this however also encompasses false wall openings e.g. caused by occlusion. These parts of the wall surface are detected and eventually reconstructed by switching the respective voxel class labels to 'Wall'.

To this aim, each 'Wall Opening' voxel is checked, if it is caused by an occluding furniture object within a search distance of 70 cm towards the room interior. Furthermore, wall openings are checked if they are also open on the side of adjacent rooms (with an assumed maximum wall thickness of 0.5 m). If the adjacent room has a closed wall surface at the corresponding position, wall openings are closed as well.

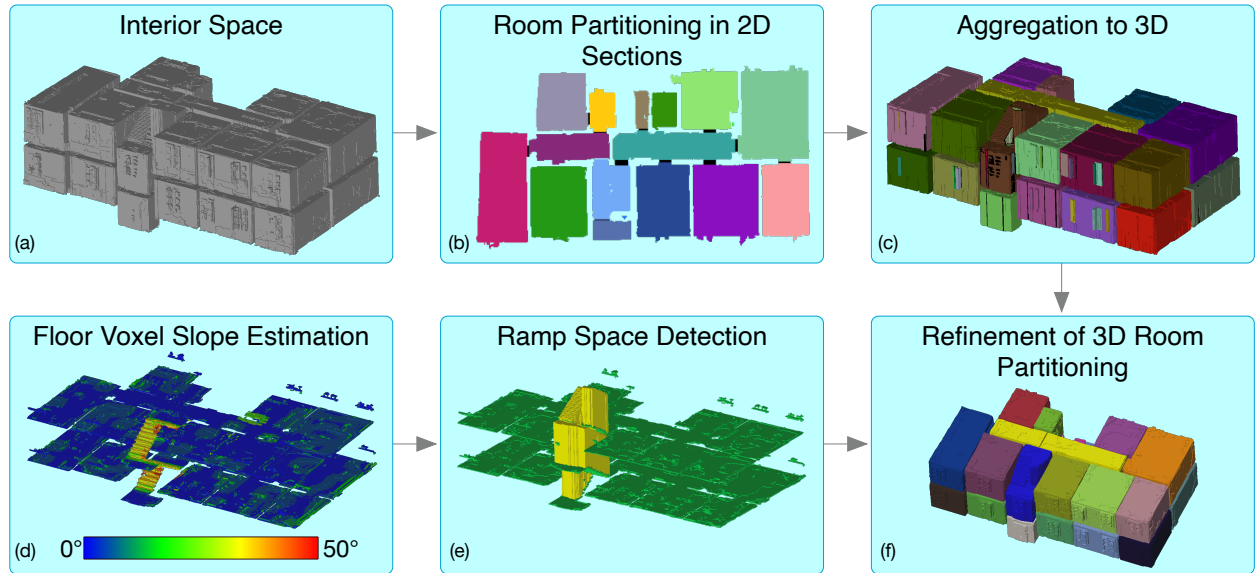


Figure 3.6: Overview of the room segmentation process. In (e), the floor surface is depicted in green while the detected ramp space is depicted in yellow. In (b), (c) and (f), different rooms are visualized in different colors while the connecting transition spaces are visualized in black. Source: [243].

3.2.6 Room Segmentation

The room segmentation resulting from the algorithm described so far is far from being ideal. On the one hand, over-segmentation of rooms occurs in case of height discontinuities in ceiling surfaces. Furthermore, structures like pipelines or cable ducts mounted on ceiling surfaces can also lead to the presence of 'Normal Horizontal' voxels within actual ceiling surfaces which, again, causes over-segmentation. On the other hand, under-segmentation can also occur in cases where multiple rooms that should still be considered as distinct are connected by a continuous ceiling surface. This can e.g. be caused by little pronounced door frames above the door openings connecting the rooms (with respect to the voxel resolution).

To resolve these problems, this section presents an algorithm for completely repartitioning the voxel-based interior space (classes 'Empty Interior', 'Interior Object' and 'Wall Opening') resulting from the presented reconstruction algorithm. Thus, the initial room segmentation used in the course of previous sections is discarded and the voxel model is repartitioned into rooms. An overview of the proposed procedure is presented in Fig. 3.6.

In the following, this section starts with elaborating on the theoretical concept of indoor spaces underlying the proposed room partitioning approach in Sec. 3.2.6.1. As stairwells are particularly challenging parts of indoor environments for room partitioning, these ramp spaces are detected beforehand to deal with them separately as described in Sec. 3.2.6.2. Finally, the proposed algorithm for voxel-based room partitioning of indoor environments is presented in Sec. 3.2.6.3.

3.2.6.1 Indoor Spaces

While many works in indoor reconstruction concern themselves with partitioning the indoor environment into rooms [246, 328, 421, 126], frequently, little is written about the concrete criteria by which this partitioning is conducted – e.g. the theoretical basis on the grounds of which two parts of an indoor space are considered as two distinct rooms or as one room.

In the scope of the indoor space partitioning approach presented here, rooms are considered as disjointly subdividing interior space, so that every part of the interior of an indoor model belongs to exactly one room. Generally, rooms are assumed to have a certain minimum size as can be expected in consideration of the fact that indoor environments are normally tailored to be used and inhabited by human beings with their own typical scale. Thus, rooms are expected to have a minimum horizontal extent of 0.5 m^2 and at least partially a vertical extent of at least 1.5 m.

Here, the partitioning of rooms is defined on the basis of geometric characteristics, as only these are available in indoor mapping datasets without further semantic knowledge about the represented indoor environment (like the intended usage function of parts of the indoor space that often underlies the definition of rooms in architecture). Concretely, the interior spaces of two distinct rooms are expected to be either completely disjoint or to be connected at most by transitional zones of interior space. These are geometrically distinct from both connected rooms by being considerably smaller in height and/or width.

These transition spaces – typically door frames – can thus be considered as a special kind of room, that can be smaller than the specified minimum dimensions for general rooms. They are connecting two large rooms via a continuous passage of interior space. The interior openings of transition spaces connecting the adjacent rooms are expected to be within a certain size range. So, transition spaces should have a certain minimum width to allow for the passage of human beings while still being of considerably smaller width than the adjacent rooms. On the other hand, in relation to a human observer, transition spaces of a certain large maximum width are not perceived as dividing two distinct rooms, even if their width is considerably smaller than the widths of both adjacent rooms. In consideration of these arguments and with respect to established standards for door frame widths, transition spaces between rooms are expected to have a width between 0.5 m and 2 m and like rooms to have a minimal height of 1.5 m. Transition spaces of smaller dimensions are considered as not passable by human subjects. They are thus discarded leaving the adjacent rooms disjoint. Transition spaces of a large width (e.g. exceeding 2 m) on the other hand, would not be perceived as dividing two distinct rooms and are thus merged together with the adjacent rooms to constitute one large room.

Besides transition spaces, there is an additional special kind of room considered in this work: the ramp space. As mentioned previously, stairwells represent particularly challenging environments for the proposed room partitioning algorithm and thus necessitate a special treatment. These critical parts of the indoor space thus need to be detected before applying the room partitioning. In this context, parts of the indoor space that have a considerable continuous inclination of its floor surface and a certain minimal vertical and horizontal extent are considered as ramp spaces. Here, the inclination of the floor surface is based on a horizontal distance (0.5 m) large enough to cause typical stair flights to appear as having a continuous inclination and thus be regarded as ramps. The indoor spaces above ramps as well as above stair flights are thus both treated as ramp spaces because stair flights might not be distinguishable from ramp surfaces anyway, depending on the voxel resolution. The detection of ramp spaces in the voxel model is described in the following Sec. 3.2.6.2.

Here, ramp spaces are always considered as distinct rooms. Depending on their size, adjacent

rooms can be merged with the ramp space or remain as distinct rooms. It can be argued for it being appropriate to count small horizontal platforms connecting the stair flights in stairwells as being part of the ramp space – i.e. the stairwell being itself one room of type ramp space. On the other hand, in the case of two rooms being vertically connected by a stair flight passing through the horizontal slab between the rooms, it seems inappropriate to merge both rooms together with the stair flight connecting them. In this case, the stair flight is regarded as an own room (of type ramp space), distinct from the two independent rooms that it connects. The stair flight is then delimited from the adjacent rooms by transition spaces consisting of the contact surface voxels to the respective rooms neighboring the ramp space. In this case, suchlike transition spaces are composed of the complete contact surface regardless of their spatial extents disregarding the size criteria for transition spaces formulated above. Currently, a size threshold of a horizontal extent of 3 m^2 is used as criterion for whether rooms adjacent to a ramp space should be merged to the ramp space or remain independent.

3.2.6.2 Ramp Space Detection

For the detection of ramp spaces defined in the previous Sec. 3.2.6.1, ramps and stair flights are assumed to have an inclination in the range of 20° to 50° in accordance with established building standards. For all floor voxels of the reconstructed indoor model, a local floor slope value is determined. This is done by determining the difference in height of the respective floor voxel to floor voxels in 0.5 m distance in all eight horizontal main directions. In those eight positions, a search is conducted in the vertical range corresponding to inclinations of -50° to 50° for the occurrence of floor voxels while disregarding directions where a vertical wall surface is crossed above the floor level. Thus, only inclination values for neighboring floor voxels in 0.5 m distance are considered, if they are part of the same room without the occurrence of a wall in-between. From the up to eight resulting inclination values, the largest absolute value is considered as local slope value for the respective floor voxel. Results of this local slope determination procedure are visualized in Fig. 3.6(d).

Ramp segments with slope values in the range of 20° to 50° are segmented based on a 2.5D neighborhood as already used in Sec. 3.2.2.3. The resulting ramp segments are then filtered by discarding segments with a horizontal extent below 0.5 m^2 or a vertical extent below 1.2 m . The remaining ramp segments are geometrically refined by closing holes. The interior space (classes 'Empty Interior', 'Interior Object' and 'Wall Opening') above the detected ramp surface then composes the detected ramp space as defined in Sec. 3.2.6.1 and visualized in Fig. 3.6(e).

3.2.6.3 Indoor Space Partitioning

As already stated in the introduction to Sec. 3.2.6, the initial room segmentation resulting from the reconstruction process as described up to Sec. 3.2.5 is discarded. Instead, the interior space composed of the voxel classes 'Empty Interior', 'Interior Object' and 'Wall Opening' is completely repartitioned from scratch. To this aim, transition spaces like door openings in-between rooms are detected in a binary voxel grid, where voxels of the aforementioned interior classes are distinguished from the rest of the voxel grid. The proposed approach first performs an initial room partitioning independently in horizontal 2D slices throughout the whole height range of the binary voxel grid as depicted in Fig. 3.6(b). The slice-wise results are aggregated over the height of the voxel grid in a 3D refinement step as depicted in Fig. 3.6(c).

Initially, transition spaces are assumed to have widths in the range of 0.5 m to 1.1 m . The detection of rooms and their connecting transition spaces is approached by applying morphological

erosion with a circular structuring element of 1.1 m diameter to the respective horizontal binary 2D slice of the interior space. This can lead to narrow corridors or small rooms being completely eroded. These narrow spaces are detected by dilating the erosion result with the same structuring element and subtracting the result from the initial binary image. The result contains the narrow rooms that have been completely eroded. They can now be eroded by using a smaller structuring element (e.g. 0.2 m diameter) with the result being added to the result of the first erosion step. The binary image resulting from this erosion process is then segmented with the segments initializing the rooms resulting from the 2D partitioning step visualized in Fig. 3.7.

Next, the transition spaces connecting these rooms are detected. To this aim, the initial room segments are grown layer-wise while stopping the growing process locally when a pixel belonging to one room neighbors a pixel of another room, until all pixels either belong to one of the initial room segments or these newly grown room domains around the initial room segments. For each contact zone, where pixels of different room domains are in direct neighborhood to one another, the presence of a transition space connecting the respective rooms can be assumed. The procedure described so far is illustrated in Fig. 3.7(a).

However, the location and shape of the transition spaces do not necessarily coincide with the detected contact zones between room domains and thus need to be refined. First, the direction of transit through the transition space from one adjacent room to the other is determined as illustrated in Fig. 3.7(b). This is done by determining the nearest pixel to the contact zone of the respective core room segments of both adjacent rooms. The direction between these pixels among the eight main directions of the pixel grid is determined. Next, the interior space between the two pixels is split layer-wise into sections perpendicular to this direction of transit and the sequence of the widths of these sections is analyzed to localize the actual transition space area as visualized in Fig. 3.7(c).

In this way, all pixels in the sections between both rooms are assigned to either one of the two rooms, to the transition space connecting them or are removed from the interior space because the transition space is found to be too narrow in respect to the assumed dimensions of transition spaces as formulated in Sec. 3.2.6.1. The resulting rooms can subsequently be checked for the presence of transition spaces of a larger width (e.g. in the range of 1.1 m to 2 m).

In a second step of the partitioning algorithm, the results of this 2D room partitioning approach applied independently to the respective horizontal slices of the voxel grid are then further refined by aggregating them along the vertical dimension of the voxel grid. In doing so, transition spaces must be present with sufficient frequency of occurrence of a sufficiently large vertical extent to be accepted as a 3D transition space. Here, transition spaces that are situated within the ramp spaces detected beforehand in Sec. 3.2.6.2 are discarded, as stairwells often suffer from frequent false positive transition spaces. They are expected to be uninterrupted indoor spaces in accordance with the considerations presented in Sec. 3.2.6.1.

As false positive transition spaces do also occur in corridors and other narrow room structures, the 3D transition spaces are subjected to another refinement step. Here, they are checked if they satisfy the criterion for transition spaces to be geometrically distinct from the adjacent rooms as formulated in Sec. 3.2.6.1 by either having a sufficiently lower height or width. Transition spaces that are not sufficiently geometrically distinguishable from the adjacent rooms are discarded and the respective rooms are merged. Finally, the resulting room partitioning of the interior space of the reconstructed voxel model is propagated to the surface voxels of classes 'Floor', 'Ceiling' and 'Wall'. The 3D room partitioning results are visualized in Fig. 3.6(f).

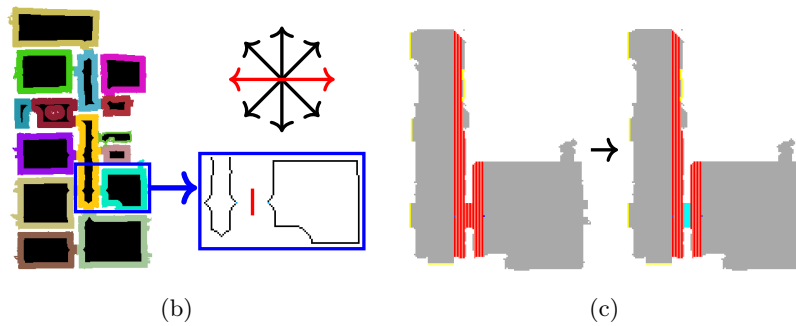
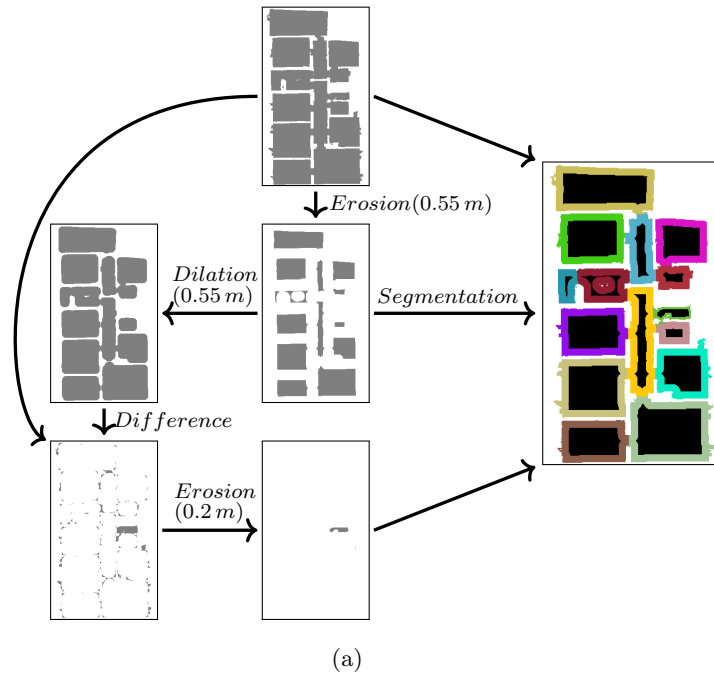


Figure 3.7: Overview of the 2D room segmentation procedure. (a) Detection of room segments and their respective domains by means of morphological operations. (b) Determination of the direction of transit between two rooms via the angle of the room segment pixels (blue) nearest to the contact zone of the two respective room domains (red). (c) The interior space (grey) between both blue pixels is split into sections (shades of red) perpendicular to the direction of transit between both rooms. The transition space (cyan) is detected by analyzing the width of the interior sections. Source: [243].

3.2.7 Evaluation

For evaluation purposes, this section proposes to manually generate ground truth data by labeling triangle meshes regarding class and room affiliation and geometrically completing them where necessary. In this context, room surfaces occluded by furniture objects or missing due to incomplete acquisition need to be manually modeled. Furthermore, the manual construction of geometries is also necessary in the case of the class 'Wall Opening'.

The resulting ground truth triangle meshes can be voxelized to generate ground truth data in voxel space that are directly comparable to the results of the presented voxel-based indoor reconstruction approach. This procedure allows to create ground truth voxel grids in arbitrary resolution and rotation around the vertical axis for given ground truth triangle meshes.

While voxelizing ground truth triangle meshes, all class labels and room affiliations of intersecting triangles are transferred to the voxels. An exception is made for the class label 'Wall Opening'. It is only passed to an intersecting voxel in the absence of other class labels. The class labels of the test data are then evaluated by checking if a corresponding ground truth voxel contains the required class label (it can however contain other class labels besides).

Manually constructing ground truth geometries for the class 'Empty Interior' is not feasible. Therefore, this class label is automatically assigned to all empty voxels between ceiling and floor of the same room. Transition spaces connecting ramp spaces to adjacent rooms are derived automatically as well in the ground truth voxel grid. To this end, ramp space interior voxels neighboring interior voxels of other rooms are detected.

The results of the room segmentation as presented in Sec. 3.2.6 and the voxel classification, i.e. the semantic classes assigned to voxels are evaluated independently. Furthermore, the effect of different rotation angles of the input data around the up-axis is investigated to demonstrate that the approach is not limited to the Manhattan World assumption as well as the effect of different resolutions of the voxel grid.

For evaluating the room segmentation results, first, a mapping between the detected rooms of the test dataset and the manually partitioned rooms of the ground truth dataset needs to be determined. To this end, a one-to-many mapping between rooms from both datasets is derived by comparing the room affiliations of voxels between ground truth and test voxel grid. Finally, a bijective one-to-one room mapping is derived by assigning a room of one dataset the corresponding room of the same type (transition space or general room) of the other dataset with the largest overlap. The resulting room mappings are discarded if they do not prove to be bijective (i.e. if room X in the ground truth data maps to room Y in the test dataset, room Y in the testdata set must also map to room X in the ground truth data). The amount of rooms in ground truth and test data for which a suchlike bijective mapping cannot be derived is part of the evaluation results and gets reported as the number of affected rooms in ground truth and test data as well as the corresponding fraction of affected voxels.

For the remaining rooms for which a bijective room mapping could be derived, an evaluation of the quality of the room segmentation is conducted. In doing so, the well-known metrics of precision and recall [480] are used where the precision

$$\text{Precision}_{Room_i} = \frac{\text{True Positives}_{Room_i}}{\text{Test Voxels}_{Room_i}} \quad (3.1)$$

of $Room_i$ denotes the fraction of voxels of the test data labeled as belonging to $Room_i$ that are

correctly labeled. The recall

$$\text{Recall}_{Room_i} = \frac{\text{True Positives}_{Room_i}}{\text{Ground Truth Voxels}_{Room_i}} \quad (3.2)$$

of $Room_i$ on the other hand denotes the fraction of voxels of the ground truth data belonging to $Room_i$ that are correctly detected in the test data. Furthermore, an accuracy measure like the F1-score

$$\text{F1-score}_{Room_i} = \frac{2 \cdot \text{Precision}_{Room_i} \cdot \text{Recall}_{Room_i}}{\text{Precision}_{Room_i} + \text{Recall}_{Room_i}} \quad (3.3)$$

can be derived from the recall and precision metrics.

Besides this evaluation of the room partitioning, the quality of the assigned voxel class labels 'Ceiling', 'Floor', 'Wall', 'Wall Opening', 'Interior Object' and 'Empty Interior' is evaluated as well. For this, again, the metrics of precision, recall and F1-score are used. However, in addition, an adjusted version of the precision/recall metrics is introduced. Here, a voxel labeling in the test data is considered as a true positive not only if the corresponding ground truth voxel has the respective class label, but also if the respective class label does occur in ground truth voxels within a certain neighborhood around the corresponding ground truth voxel. In the context of this work, a 3D-6-neighborhood is used for these 'neighborhood precision/recall' metrics. Note, that in this case, the recall should not be calculated by Eq. 3.2 but by applying Eq. 3.1 for precision with ground truth and test data interchanged respectively. This is due to the fact that when using Eq. 3.2 for calculating the neighborhood recall, the resulting value for recall could amount to values larger than 100% depending on the spatial distribution of the data.

3.3 Results

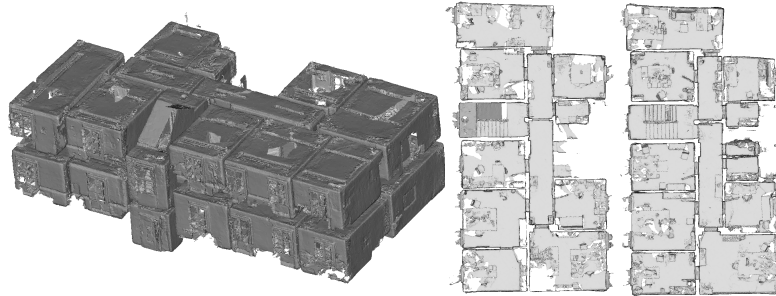
This section presents quantitative evaluation results for four datasets of triangle meshes of indoor environments acquired via the Microsoft HoloLens (version 1) depicted in Fig. 3.8. The dataset 'Office' as depicted in Fig. 3.8(a) represents a two-storey office environment comprised of 24 rooms including a stairwell that connects the two levels vertically. The dataset 'Attic' is depicted in Fig. 3.8(b). It represents an attic environment with slanted ceilings comprised of five rooms. Fig. 3.8(c) shows the dataset 'Basement' which represents a basement area including laboratories, offices and storage rooms connected by a long hallways with barrel-vaulted ceilings. It consists of 11 rooms and has with $48\text{ m} \times 22\text{ m} \times 6\text{ m}$ the largest spatial extent of the four presented datasets. The last dataset 'Residential House' as depicted in Fig. 3.8(d) represents a part of the indoor space of a residential house extending over three storeys connected by a stairwell which is connected openly towards the living space. It consists of 12 rooms. Further information about the datasets and their respective class distributions is given in Tab. 3.2.

The four presented datasets along with the manually modeled ground truth are made available to the public¹⁸. Furthermore, the code of the implementation of the voxel-based indoor reconstruction procedure presented in this chapter as well as the code for the automated evaluation against the ground truth data is released as well.

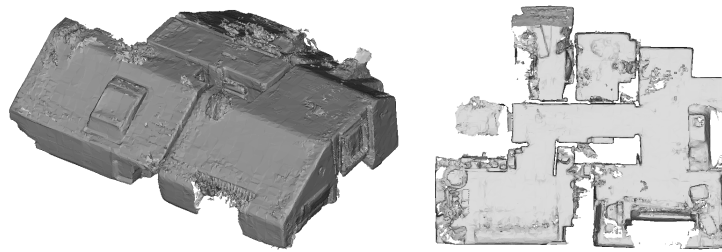
For each of these four datasets, ground truth triangle meshes were generated manually using the software Blender v2.91¹⁹ and subsequently voxelized according to the general procedure outlined in Sec. 3.2.7. The voxel-based indoor reconstruction approach presented in Sec. 3.2 was then applied to the datasets and the results were evaluated against the respective ground truth

¹⁸<https://github.com/huepat/voxir> (Last visited on 15/07/2021)

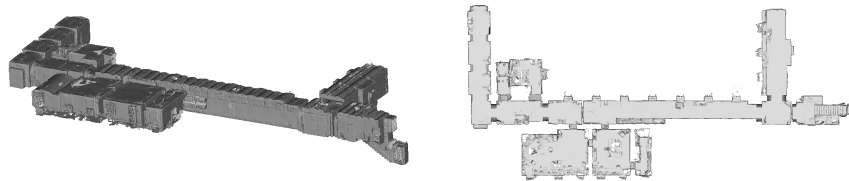
¹⁹<https://www.blender.org/> (Last visited on 13/07/2021)



(a) The dataset 'Office'.



(b) The dataset 'Attic'.



(c) The dataset 'Basement'.



(d) The dataset 'Residential House'.

Figure 3.8: The Microsoft HoloLens triangle meshes used for quantitative evaluation. Source: [243].

Table 3.2: Size and class distribution of the four provided benchmark datasets. Source: [243].

			Office	Attic	Basement	Residential House
Spatial Extent (m)			13×21×8	8×9×3	48×22×6	11×11×8
Mesh Faces			958,820	147,146	695,041	450,260
Number of Rooms (acc. to Ground Truth)			24	5	11	12
Fraction of Voxels (%) (for 5 cm Voxel Resolution)	Test Data	Ceiling	2.29	4.31	2.39	2.74
		Floor	2.10	3.46	1.95	2.44
		Wall	7.35	9.33	6.96	6.97
		Interior Object	2.32	3.11	1.76	4.26
		Empty Interior	85.71	79.62	86.63	83.1
		Wall Opening	0.22	0.17	0.31	0.49
	Ground Truth Data	Ceiling	2.24	4.39	2.47	2.78
		Floor	2.07	3.28	1.98	2.39
		Wall	6.1	7.27	5.53	5.64
		Interior Object	2.78	3.78	1.84	5.09
		Empty Interior	86.63	81.25	88.13	84.02
		Wall Opening	0.18	0.03	0.06	0.08

voxel grid by determining the evaluation metrics presented in Sec. 3.2.7.

The quantitative evaluation also aims to investigate the impact of two parameters on the results: horizontal alignment of the input data with the coordinate axes and voxel resolution. Tab. 3.3 to 3.6 show the quantitative results for each respective dataset at varying angles of rotation around the up-axis at a fixed voxel grid resolution of 5 cm. Furthermore, Fig. 3.9 exemplarily visualizes the results achieved for the dataset 'Residential House' for one horizontal slice of the voxel grid generated by our approach and the corresponding slice derived from the ground truth data. Similarly, another experiment was conducted to examine the effect of varying voxel grid resolutions at a fixed rotation angle representing optimal horizontal alignment of the input data along the coordinate axes. The quantitative results of this investigation are presented in Tab. 3.7 to 3.10 and an exemplary visualization for one horizontal slice of the dataset 'Residential House' is presented in Fig. 3.10.

To give an impression about the performance of the implementation of the proposed procedure in terms of processing time and memory consumption, these values are provided for varying voxel grid resolutions for the 'Office' dataset in Fig. 3.11. The values refer to a system with a i7-8550U CPU with 24 GB RAM and covers the whole procedure including voxelization of the triangle mesh but excluding data import and export. Large parts of the implementation are CPU-parallelized.

Besides this quantitative evaluation, qualitative results of the voxel-based indoor reconstruction approach were also assessed for two particularly challenging indoor environments that are part of the Matterport3D dataset [94]. The datasets as provided by Chang et al. [94] were captured with the Matterport system, a semi-mobile indoor mapping device comprised of three RGBD cameras mounted on a rotatable rig on a tripod. The resulting RGBD panoramas from multiple viewpoints are further processed to generate textured triangle meshes of the complete indoor environment covered from the viewpoints.

Fig. 3.12 shows triangle meshes and the corresponding results derived by the procedure presented in this work for a half-sphere-shaped building with two storeys. Triangle meshes and

Table 3.3: Evaluation results for the dataset 'Office' for a fixed voxel resolution of 5 cm and varying rotation angles around the up-axis (with 0° being aligned on the coordinate axes). Source: [243].

bad  good		Rotation angle around up-axis					
		0°	10°	20°	30°	40°	45°
Room mapping error (%) (fraction of affected voxels)		1.03	0.09	8.78	0.16	2.47	0.09
Room segmentation precision (%)		98.01	97.98	90.04	95.73	98.13	97.92
Room segmentation recall (%)		97.06	97.77	95.98	95.29	95.34	97.73
Room segmentation F1-score (%)		97.53	97.80	92.86	95.45	96.68	97.82
Precision (%)	Ceiling	85.14	85.57	85.69	85.59	85.27	84.37
	Floor	82.51	82.30	82.37	83.09	82.34	81.95
	Wall	68.58	67.64	67.74	68.38	69.84	70.19
	Interior Object	91.51	91.11	88.94	92.27	92.07	91.87
	Empty Interior	99.05	99.10	99.03	99.14	99.07	99.02
	Wall Opening	6.37	8.67	8.65	9.05	7.94	7.86
Recall (%)	Ceiling	86.92	86.68	86.28	85.50	85.06	85.04
	Floor	83.64	82.69	82.50	82.69	81.80	81.75
	Wall	82.47	84.06	83.77	85.63	85.65	85.26
	Interior Object	76.37	75.25	74.19	73.69	73.88	74.62
	Empty Interior	97.88	97.44	97.18	97.01	97.18	97.52
	Wall Opening	8.10	10.13	8.80	8.65	6.96	6.08
F1-score (%)	Ceiling	86.02	86.05	85.90	85.50	85.10	84.65
	Floor	83.07	82.45	82.40	82.80	82.05	81.80
	Wall	74.89	74.94	74.88	76.03	76.92	76.97
	Interior Object	83.26	82.39	80.84	81.88	81.93	82.34
	Empty Interior	98.46	98.24	98.06	98.06	98.08	98.25
	Wall Opening	7.09	9.33	8.72	8.85	7.42	6.78
Neighborhood Precision (%) (based on 3D-6- neighborhood)	Ceiling	96.52	96.53	96.49	96.42	96.39	95.95
	Floor	96.55	96.87	96.73	96.88	96.85	96.21
	Wall	92.40	91.64	90.43	90.15	90.26	90.38
	Interior Object	92.70	92.25	90.03	93.30	93.07	92.96
	Empty Interior	99.62	99.63	99.54	99.61	99.57	99.52
	Wall Opening	14.38	14.37	13.76	12.65	10.31	9.81
Neighborhood Recall (%) (based on 3D-6- neighborhood)	Ceiling	94.82	94.65	94.18	93.87	93.66	93.50
	Floor	96.20	96.36	96.01	95.93	95.86	95.41
	Wall	94.35	95.38	94.75	95.77	95.47	95.12
	Interior Object	81.64	81.01	80.46	79.99	79.92	80.24
	Empty Interior	99.44	99.37	99.23	99.11	99.07	99.11
	Wall Opening	25.59	24.94	22.59	21.38	17.53	15.61


Table 3.4: Evaluation results for the dataset 'Attic' for a fixed voxel resolution of 5 cm and varying rotation angles around the up-axis (with 0° being aligned on the coordinate axes). Source: [243].

bad  good		Rotation angle around up-axis					
		0°	10°	20°	30°	40°	45°
Room mapping error (%) (fraction of affected voxels)		0.00	0.00	17.36	0.00	0.00	0.48
Room segmentation precision (%)		97.46	97.17	88.67	97.51	97.02	97.24
Room segmentation recall (%)		95.49	94.38	92.29	94.87	94.31	95.34
Room segmentation F1-score (%)		96.46	95.75	90.44	96.17	95.65	96.28
Precision (%)	Ceiling	86.95	87.04	86.89	87.21	87.29	87.13
	Floor	68.28	63.47	65.13	66.05	65.17	65.95
	Wall	59.01	59.41	60.71	63.65	61.16	61.11
	Interior Object	84.58	84.44	85.73	84.07	83.30	85.32
	Empty Interior	98.64	98.75	98.87	98.85	98.73	98.80
	Wall Opening	0.00	2.41	2.37	6.26	6.35	1.58
Recall (%)	Ceiling	83.23	82.17	81.95	81.81	80.79	81.28
	Floor	70.39	66.03	66.97	67.08	66.04	66.41
	Wall	73.81	75.48	76.93	76.99	76.74	76.64
	Interior Object	67.83	62.48	61.60	63.14	62.68	65.67
	Empty Interior	94.25	93.30	93.09	93.99	93.32	93.59
	Wall Opening	0.00	10.67	18.75	20.76	18.09	4.88
F1-score (%)	Ceiling	85.05	84.53	84.35	84.42	83.91	84.10
	Floor	69.32	64.72	66.04	66.56	65.60	66.18
	Wall	65.59	66.49	67.86	69.69	68.07	68.00
	Interior Object	75.28	71.82	71.69	72.12	71.53	74.22
	Empty Interior	96.40	95.95	95.89	96.36	95.95	96.12
	Wall Opening	0.00	3.93	4.21	9.62	9.40	2.39
Neighborhood Precision (%) (based on 3D-6- neighborhood)	Ceiling	96.40	96.26	95.98	96.29	96.01	96.31
	Floor	86.67	82.10	83.35	85.11	83.59	84.60
	Wall	80.98	81.06	80.73	82.45	79.14	78.76
	Interior Object	86.53	86.18	87.43	86.17	85.18	87.48
	Empty Interior	99.64	99.64	99.71	99.72	99.61	99.67
	Wall Opening	0.00	2.41	2.37	6.26	6.35	1.58
Neighborhood Recall (%) (based on 3D-6- neighborhood)	Ceiling	91.78	91.21	91.25	91.15	90.20	90.68
	Floor	87.21	84.11	84.76	85.79	84.23	84.75
	Wall	87.83	87.88	88.69	87.54	87.21	86.91
	Interior Object	75.28	70.89	70.49	71.85	71.40	73.61
	Empty Interior	96.43	95.86	95.85	96.22	95.63	95.85
	Wall Opening	81.44	62.06	69.10	70.24	58.88	40.85

Table 3.5: Evaluation results for the dataset 'Basement' for a fixed voxel resolution of 5 cm and varying rotation angles around the up-axis (with 0° being aligned on the coordinate axes). Source: [243].

		Rotation angle around up-axis					
		0°	10°	20°	30°	40°	45°
bad  good							
Room mapping error (%) (fraction of affected voxels)		1.74	1.63	0.52	0.53	0.60	0.53
Room segmentation precision (%)		96.04	96.58	95.57	95.79	95.90	97.13
Room segmentation recall (%)		96.22	96.02	96.03	96.30	96.33	97.17
Room segmentation F1-score (%)		96.12	96.25	95.75	96.00	96.10	97.10
Precision (%)	Ceiling	84.06	85.06	84.74	85.39	85.42	85.71
	Floor	85.15	85.78	85.61	86.07	85.21	85.75
	Wall	62.51	64.94	66.28	67.21	66.75	67.12
	Interior Object	81.08	83.66	83.26	82.75	82.96	85.25
	Empty Interior	98.45	98.44	98.62	98.63	98.56	99.04
	Wall Opening	1.98	1.89	1.83	2.00	2.16	1.88
Recall (%)	Ceiling	81.98	81.46	80.72	80.36	80.50	80.07
	Floor	84.43	85.42	84.14	83.49	84.22	83.83
	Wall	79.17	81.41	81.21	81.98	82.82	81.41
	Interior Object	78.22	78.47	76.60	76.12	76.69	76.78
	Empty Interior	97.43	97.47	96.77	96.91	97.18	96.87
	Wall Opening	10.86	8.79	8.56	7.64	6.94	5.36
F1-score (%)	Ceiling	82.98	83.22	82.68	82.80	82.89	82.79
	Floor	84.79	85.60	84.87	84.76	84.71	84.78
	Wall	69.86	72.25	72.99	73.86	73.92	73.58
	Interior Object	79.62	80.98	79.79	79.30	79.70	80.79
	Empty Interior	97.94	97.95	97.69	97.76	97.87	97.94
	Wall Opening	3.35	3.11	3.02	3.17	3.29	2.78
Neighborhood Precision (%) (based on 3D-6- neighborhood)	Ceiling	95.14	95.65	95.58	96.50	95.97	96.54
	Floor	93.78	94.18	94.04	94.34	93.64	94.83
	Wall	87.43	87.46	87.38	87.09	85.79	86.22
	Interior Object	82.84	85.30	84.78	84.28	84.54	86.74
	Empty Interior	98.86	98.84	98.98	98.99	98.91	99.41
	Wall Opening	2.58	2.58	2.63	3.07	2.85	2.57
Neighborhood Recall (%) (based on 3D-6- neighborhood)	Ceiling	90.38	89.76	89.34	89.02	89.02	88.48
	Floor	92.64	93.76	92.79	92.01	93.16	92.58
	Wall	90.14	91.95	91.34	91.89	92.57	90.81
	Interior Object	83.92	84.94	82.91	82.29	83.02	82.57
	Empty Interior	98.89	99.13	98.47	98.53	98.74	98.25
	Wall Opening	25.47	21.33	23.96	23.27	21.61	17.61

Table 3.6: Evaluation results for the dataset 'Residential House' for a fixed voxel resolution of 5 cm and varying rotation angles around the up-axis (with 0° being aligned on the coordinate axes). Source: [243].

		Rotation angle around up-axis					
		0°	10°	20°	30°	40°	45°
bad  good							
Room mapping error (%) (fraction of affected voxels)		0.11	12.09	6.90	7.39	6.81	7.15
Room segmentation precision (%)		96.31	84.92	89.26	89.51	89.63	89.43
Room segmentation recall (%)		96.79	96.99	96.02	96.23	96.26	95.86
Room segmentation F1-score (%)		96.55	90.55	92.52	92.75	92.83	92.53
Precision (%)	Ceiling	90.89	90.13	91.00	90.97	90.76	89.74
	Floor	77.51	75.90	75.80	77.41	76.80	74.82
	Wall	60.90	61.06	61.75	64.01	64.15	65.86
	Interior Object	92.84	91.47	91.98	91.69	91.55	90.91
	Empty Interior	97.43	96.94	97.11	97.09	97.06	96.85
	Wall Opening	0.57	1.02	1.27	1.45	1.53	0.85
Recall (%)	Ceiling	90.33	90.07	89.48	89.26	88.82	88.84
	Floor	79.75	78.77	78.36	79.02	78.33	78.01
	Wall	75.81	75.67	77.28	78.30	77.76	77.93
	Interior Object	78.44	77.59	75.95	75.92	76.70	76.59
	Empty Interior	97.16	96.95	96.66	96.90	96.80	96.87
	Wall Opening	3.78	6.80	7.80	7.75	7.70	4.48
F1-score (%)	Ceiling	90.61	90.10	90.23	90.11	89.78	89.29
	Floor	78.61	77.31	77.06	78.21	77.56	76.38
	Wall	67.54	67.58	68.65	70.44	70.30	71.39
	Interior Object	85.03	83.96	83.20	83.06	83.47	83.14
	Empty Interior	97.29	96.94	96.88	96.99	96.93	96.86
	Wall Opening	0.99	1.77	2.18	2.44	2.55	1.43
Neighborhood Precision (%) (based on 3D-6- neighborhood)	Ceiling	96.74	96.09	96.42	96.66	96.15	95.56
	Floor	93.71	92.39	92.42	93.25	92.32	90.12
	Wall	83.70	84.12	83.30	84.78	83.67	84.31
	Interior Object	94.46	92.94	93.43	93.06	92.94	92.37
	Empty Interior	98.29	97.77	97.89	97.88	97.84	97.65
	Wall Opening	1.33	1.46	1.65	1.64	1.88	1.28
Neighborhood Recall (%) (based on 3D-6- neighborhood)	Ceiling	95.14	95.16	94.74	94.72	94.33	94.28
	Floor	94.81	93.92	93.97	94.34	93.66	92.55
	Wall	91.72	90.45	91.61	92.09	90.96	90.71
	Interior Object	85.26	85.54	84.39	84.21	84.29	84.15
	Empty Interior	98.96	99.07	98.89	98.99	98.88	98.76
	Wall Opening	17.49	20.06	22.69	21.02	20.98	15.75

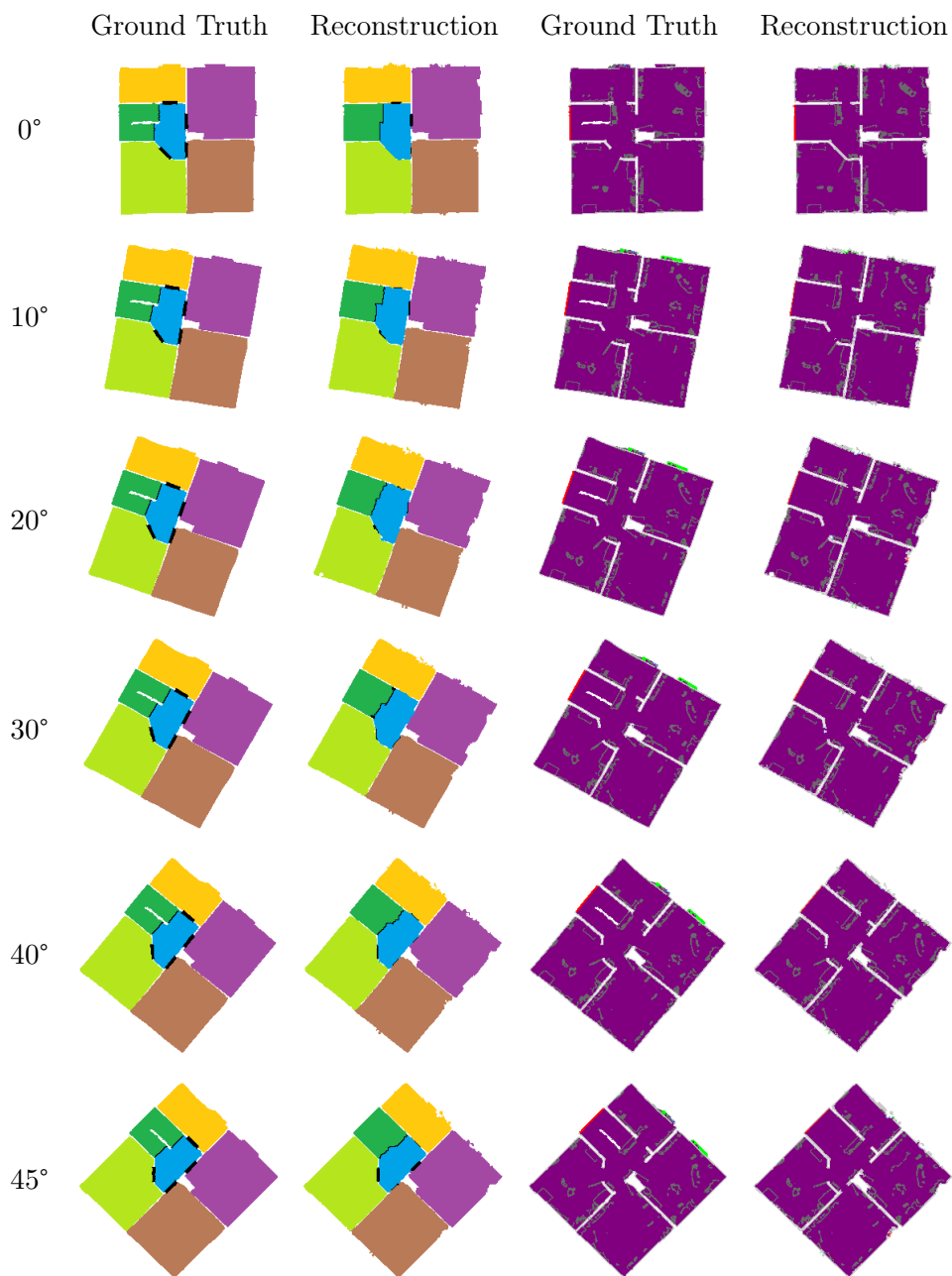


Figure 3.9: Exemplary results for one horizontal slice of the dataset 'Residential House' (cf. Figure 3.8(d), for numerical results see Tab. 3.6) for different rotations around the up-axis at 5 cm voxel resolution. Left: room partitioning; different colors indicate different rooms; black indicates transition spaces. Right: voxel classification, colored according to Tab. 3.1. Source: [243].

Table 3.7: Evaluation results for the dataset 'Office' for varying voxel resolutions. Source: [243].

bad  good		Voxel resolution						
		3cm	5cm	7cm	9cm	11cm	13cm	15cm
Room mapping error (%) (fraction of affected voxels)		1.73	1.03	0.01	2.56	0.00	3.46	4.12
Room segmentation precision (%)		97.24	98.01	97.68	97.16	97.52	97.58	97.30
Room segmentation recall (%)		97.35	97.06	97.83	95.66	97.25	93.99	92.96
Room segmentation F1-score (%)		97.25	97.50	97.70	96.34	97.38	95.75	95.08
Precision (%)	Ceiling	80.00	85.14	85.44	87.88	88.78	89.26	88.24
	Floor	76.08	82.51	82.98	84.43	84.32	86.35	79.80
	Wall	57.71	68.58	72.30	76.51	79.85	81.10	79.84
	Interior Object	90.64	91.51	93.44	91.83	77.48	81.11	80.59
	Empty Interior	98.92	99.05	98.78	98.34	98.61	98.90	98.85
	Wall Opening	5.69	6.37	8.72	11.92	13.39	19.36	16.82
Recall (%)	Ceiling	85.62	86.92	86.31	85.98	86.54	85.71	84.58
	Floor	80.10	83.64	83.96	83.36	82.36	86.36	81.01
	Wall	79.28	82.47	84.11	85.83	78.86	82.06	82.98
	Interior Object	77.16	76.37	73.83	69.54	71.13	66.38	60.18
	Empty Interior	97.83	97.88	97.64	97.90	97.71	97.42	96.75
	Wall Opening	7.36	8.10	11.30	15.82	19.37	24.79	22.39
F1-score (%)	Ceiling	82.71	85.99	85.85	86.84	87.65	87.45	86.31
	Floor	78.00	83.05	83.40	83.85	83.33	86.35	80.40
	Wall	66.77	74.86	77.75	80.89	79.35	81.58	81.32
	Interior Object	83.32	83.22	82.47	79.12	74.17	73.01	68.82
	Empty Interior	98.36	98.42	98.19	98.12	98.16	98.15	97.74
	Wall Opening	6.42	7.13	9.84	13.58	15.83	21.74	19.16
Neighborhood Precision (%) (based on 3D-6- neighborhood)	Ceiling	95.08	96.52	97.22	97.53	97.25	97.34	93.34
	Floor	93.92	96.55	96.93	97.56	96.03	94.45	87.48
	Wall	85.57	92.40	94.06	94.72	95.48	95.82	92.60
	Interior Object	91.45	92.70	94.92	93.10	79.99	83.40	83.23
	Empty Interior	99.43	99.62	99.49	99.13	99.44	99.59	99.56
	Wall Opening	11.03	14.38	13.29	21.80	22.76	33.83	22.81
Neighborhood Recall (%) (based on 3D-6- neighborhood)	Ceiling	92.94	94.82	95.43	95.78	95.97	96.05	96.91
	Floor	94.93	96.20	96.91	97.70	96.76	96.45	93.71
	Wall	91.01	94.35	95.99	96.69	97.62	98.21	98.07
	Interior Object	81.47	81.64	80.44	77.54	81.49	78.61	75.98
	Empty Interior	99.29	99.44	99.39	99.62	99.49	99.43	98.82
	Wall Opening	20.67	25.59	25.74	33.23	43.93	56.75	47.03

Table 3.8: Evaluation results for the dataset 'Attic' for varying voxel resolutions. Source: [243].


bad  good		Voxel resolution						
		3cm	5cm	7cm	9cm	11cm	13cm	15cm
Room mapping error (%) (fraction of affected voxels)		0.00	0.00	0.00	0.00	0.00	0.00	0.00
Room segmentation precision (%)		97.53	97.46	96.79	96.10	96.63	96.91	96.40
Room segmentation recall (%)		95.05	95.49	94.87	94.78	95.04	94.41	92.19
Room segmentation F1-score (%)		96.27	96.46	95.82	95.44	95.83	95.64	94.25
Precision (%)	Ceiling	81.50	86.95	86.68	90.81	90.34	90.85	84.77
	Floor	58.22	68.28	68.78	75.13	75.30	82.35	70.89
	Wall	51.95	59.01	65.56	67.93	70.07	71.62	65.39
	Interior Object	83.41	84.58	85.75	86.57	74.47	81.99	76.15
	Empty Interior	98.63	98.64	98.86	98.82	98.55	99.11	98.69
	Wall Opening	0.00	0.00	0.00	4.37	0.00	21.25	44.44
Recall (%)	Ceiling	82.28	83.23	84.10	82.60	80.10	81.78	77.14
	Floor	66.46	70.39	71.58	75.30	72.92	80.79	67.49
	Wall	70.48	73.81	77.57	80.54	72.69	78.88	77.33
	Interior Object	64.31	67.83	58.19	57.64	65.79	59.39	48.68
	Empty Interior	93.92	94.25	94.04	94.79	94.92	94.03	93.32
	Wall Opening	0.00	0.00	0.00	15.69	0.00	100.00	80.00
F1-score (%)	Ceiling	81.89	85.05	85.37	86.51	84.91	86.08	80.78
	Floor	62.07	69.32	70.15	75.21	74.09	81.56	69.15
	Wall	59.81	65.59	71.06	73.70	71.36	75.07	70.86
	Interior Object	72.63	75.28	69.33	69.20	69.86	68.88	59.39
	Empty Interior	96.22	96.40	96.39	96.76	96.70	96.50	95.93
	Wall Opening	0.00	0.00	0.00	6.84	0.00	35.05	57.14
Neighborhood Precision (%) (based on 3D-6-neighborhood)	Ceiling	93.29	96.40	96.77	98.73	98.27	98.36	92.54
	Floor	78.97	86.67	85.82	90.97	91.59	92.70	85.61
	Wall	76.01	80.98	83.06	83.77	88.09	85.90	83.04
	Interior Object	84.80	86.53	87.59	88.18	76.73	84.38	80.23
	Empty Interior	99.44	99.64	99.80	99.80	99.75	99.85	99.76
	Wall Opening	0.00	0.00	0.00	4.37	0.00	21.25	44.44
Neighborhood Recall (%) (based on 3D-6-neighborhood)	Ceiling	89.63	91.78	93.42	93.07	93.49	94.32	94.79
	Floor	82.56	87.21	88.80	92.16	92.79	93.76	91.40
	Wall	82.02	87.83	89.07	91.15	90.73	93.16	92.92
	Interior Object	70.39	75.28	68.19	69.53	82.24	78.45	69.29
	Empty Interior	95.72	96.43	96.01	97.02	97.45	96.38	96.44
	Wall Opening	2.14	81.44	65.31	88.24	87.10	100.00	100.00

Table 3.9: Evaluation results for the dataset 'Basement' for varying voxel resolutions. Source: [243].

bad  good		Voxel resolution						
		3cm	5cm	7cm	9cm	11cm	13cm	15cm
Room mapping error (%) (fraction of affected voxels)		0.65	1.74	3.71	4.37	18.61	24.14	5.07
Room segmentation precision (%)		95.72	96.04	95.34	94.15	95.13	94.82	94.93
Room segmentation recall (%)		96.94	96.22	96.27	96.19	78.97	75.13	92.48
Room segmentation F1-score (%)		96.33	96.13	95.80	95.16	86.30	83.83	93.69
Precision (%)	Ceiling	77.37	84.06	86.28	89.06	88.82	88.01	87.97
	Floor	79.35	85.15	85.19	86.45	88.09	85.04	87.27
	Wall	56.51	62.51	65.04	69.07	69.72	71.04	73.22
	Interior Object	76.48	81.08	84.72	87.97	74.02	69.16	73.67
	Empty Interior	98.31	98.45	98.19	98.47	98.51	98.56	98.64
	Wall Opening	1.25	1.98	2.10	2.13	1.12	1.85	2.95
Recall (%)	Ceiling	78.89	81.98	83.67	83.86	82.34	81.43	80.80
	Floor	82.81	84.43	85.52	86.29	87.59	86.22	86.17
	Wall	75.76	79.17	80.45	83.32	77.79	76.48	81.18
	Interior Object	78.77	78.22	76.81	72.04	71.36	64.81	63.51
	Empty Interior	97.69	97.43	97.32	97.04	96.32	96.14	96.28
	Wall Opening	6.17	10.86	12.05	11.49	7.74	10.03	16.67
F1-score (%)	Ceiling	78.12	83.01	84.95	86.38	85.46	84.59	84.23
	Floor	81.04	84.79	85.35	86.37	87.84	85.63	86.72
	Wall	64.73	69.86	71.93	75.53	73.53	73.66	76.99
	Interior Object	77.61	79.62	80.57	79.21	72.67	66.91	68.21
	Empty Interior	98.00	97.94	97.75	97.75	97.40	97.33	97.45
	Wall Opening	2.08	3.35	3.58	3.59	1.96	3.12	5.01
Neighborhood Precision (%) (based on 3D-6- neighborhood)	Ceiling	92.34	95.14	96.05	96.14	94.46	93.80	93.06
	Floor	92.83	93.78	93.78	93.94	94.03	90.81	92.55
	Wall	84.00	87.43	89.25	90.56	90.06	90.67	90.90
	Interior Object	77.52	82.84	87.08	89.89	77.21	72.95	77.14
	Empty Interior	98.72	98.86	98.74	98.98	99.07	99.12	99.09
	Wall Opening	1.57	2.58	2.48	2.26	1.31	2.67	2.95
Neighborhood Recall (%) (based on 3D-6- neighborhood)	Ceiling	85.97	90.38	91.68	92.74	92.69	91.96	92.47
	Floor	91.66	92.64	94.23	94.98	95.80	95.49	94.84
	Wall	85.90	90.14	92.93	94.46	95.21	94.50	95.92
	Interior Object	83.34	83.92	83.94	81.02	83.08	79.41	78.67
	Empty Interior	98.92	98.89	99.14	99.15	99.14	98.97	99.29
	Wall Opening	16.39	25.47	24.21	21.10	13.05	25.63	35.09

Table 3.10: Evaluation results for the dataset 'Residential House' for varying voxel resolutions. Source: [243].

bad  good		Voxel resolution						
		3cm	5cm	7cm	9cm	11cm	13cm	15cm
Room mapping error (%) (fraction of affected voxels)		3.93	0.11	0.08	3.13	13.70	8.59	17.10
Room segmentation precision (%)		92.17	96.31	96.23	93.25	91.03	92.41	88.35
Room segmentation recall (%)		96.84	96.79	96.17	95.95	90.60	90.55	84.84
Room segmentation F1-score (%)		94.45	96.55	96.2	94.58	90.81	91.47	86.56
Precision (%)	Ceiling	86.43	90.89	90.81	91.22	90.48	90.12	81.64
	Floor	68.80	77.51	78.61	79.94	81.29	80.63	68.35
	Wall	52.45	60.90	63.51	63.12	67.76	69.76	64.17
	Interior Object	92.58	92.84	94.00	92.53	89.65	91.04	89.65
	Empty Interior	97.01	97.43	97.57	97.36	97.77	98.42	99.08
	Wall Opening	1.07	0.57	0.61	0.22	0.31	0.51	0.57
Recall (%)	Ceiling	89.17	90.33	90.53	88.93	84.89	84.92	81.60
	Floor	74.68	79.75	80.32	81.64	81.33	83.18	68.32
	Wall	71.38	75.81	77.82	79.86	78.43	80.78	81.43
	Interior Object	78.43	78.44	74.74	66.78	69.86	64.92	54.45
	Empty Interior	97.27	97.16	96.62	95.79	94.75	94.81	91.03
	Wall Opening	7.29	3.78	3.94	1.55	2.82	4.23	1.94
F1-score (%)	Ceiling	87.78	90.61	90.67	90.06	87.60	87.44	81.62
	Floor	71.62	78.61	79.46	80.78	81.31	81.89	68.33
	Wall	60.47	67.54	69.94	70.51	72.71	74.87	71.78
	Interior Object	84.92	85.03	83.27	77.57	78.53	75.79	67.75
	Empty Interior	97.14	97.29	97.09	96.57	96.24	96.58	94.88
	Wall Opening	1.87	0.99	1.06	0.39	0.56	0.91	0.88
Neighborhood Precision (%) (based on 3D-6-neighborhood)	Ceiling	96.15	96.74	96.67	96.65	96.32	95.89	86.98
	Floor	89.76	93.71	92.22	94.00	93.71	90.78	77.86
	Wall	80.25	83.70	84.86	84.24	83.73	85.64	80.64
	Interior Object	93.64	94.46	95.81	94.08	92.06	93.42	92.71
	Empty Interior	97.81	98.29	98.53	98.27	98.51	99.01	99.83
	Wall Opening	1.74	1.33	1.07	0.22	0.55	0.58	2.56
Neighborhood Recall (%) (based on 3D-6-neighborhood)	Ceiling	93.47	95.14	95.85	95.74	94.78	94.86	94.08
	Floor	92.17	94.81	94.59	95.49	95.31	94.15	88.74
	Wall	88.47	91.72	93.23	93.00	93.12	95.19	95.39
	Interior Object	84.03	85.26	83.23	76.93	84.50	81.92	75.38
	Empty Interior	98.93	98.96	98.84	98.45	98.10	98.23	95.69
	Wall Opening	20.19	17.49	14.56	5.23	10.56	14.81	16.50

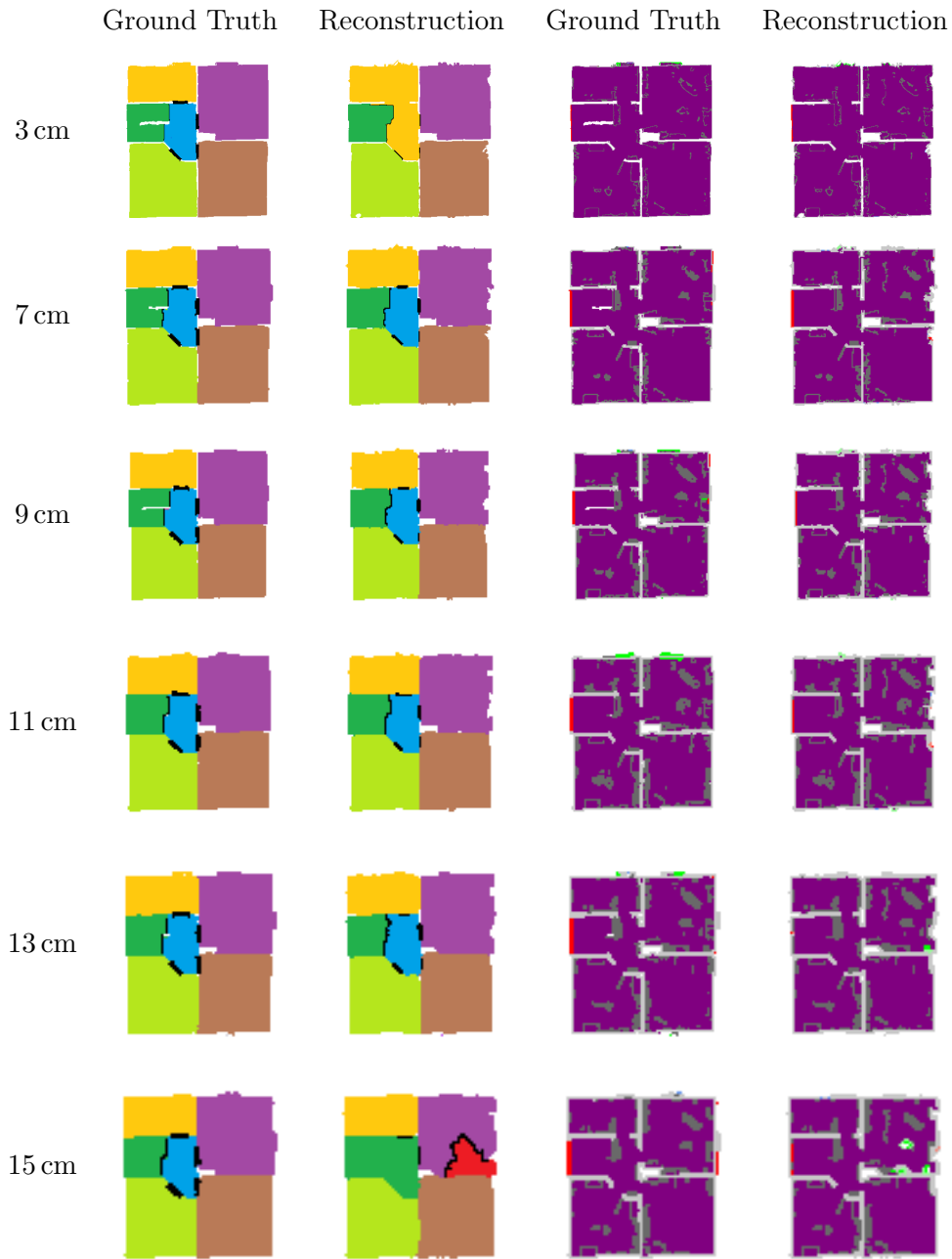


Figure 3.10: Exemplary results for one horizontal slice of the dataset 'Residential House' (cf. 3.8(d), for numerical results see Tab. 3.10) for different voxel resolutions (for 5 cm resolution, see first line of Fig. 3.9). Left: room partitioning; different colors indicate different rooms; black indicates transition spaces. Right: voxel classification, colored according to Tab. 3.1. Source: [243].

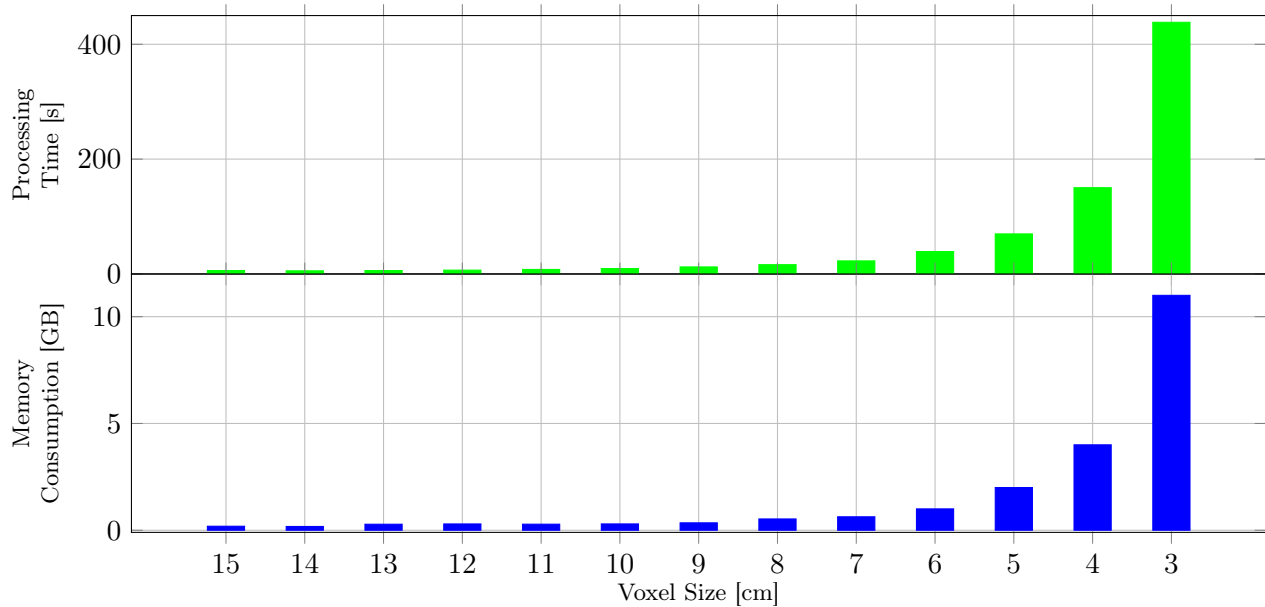


Figure 3.11: Dependency of processing time and memory consumption on voxel size for the 'Office' dataset. Source: [243].

corresponding results for another non-standard building are presented in Fig. 3.13.

3.4 Discussion

The quantitative evaluation results presented in Sec. 3.3 for the metrics introduced in Sec. 3.2.7 show in general low room mapping errors over the different datasets, resolutions and rotation angles around the up-axis. This implies an overall successful mapping of segmented room entities between ground truth and test data. Higher room mapping errors of up to 25 % occur occasionally, e.g. for the dataset 'Attic' at 20° rotation around the up-axis in Tab. 3.4 or for the dataset 'Basement' at 11 and 13 cm voxel grid resolution in Tab. 3.9, respectively. Furthermore, heightened room mapping errors occur frequently in the dataset 'Residential House' (cf. Tab. 3.6 and 3.10) as can also be seen in the visualization of one horizontal voxel slice of this dataset in Fig. 3.10. Here, under-segmentation of rooms occurs in the case of 3 cm and 15 cm voxel resolution while, in the latter case, over-segmentation does also occur.

The values for room segmentation precision and recall only refer to those rooms for which an unambiguous mapping between test and ground truth data could be derived. They are thus determined only for a fraction of 100 % – *Room Mapping Error* of non-empty voxels. The values for room segmentation precision and recall are mostly above 90 % with outliers down to about 75 %, e.g. in the case of the dataset 'Basement' at 11 cm and 13 cm voxel resolution.

The evaluation of the voxel classification, however, is performed completely independent of the room affiliation of the respective voxels and thus refers to all non-empty voxels, not only those with unambiguous room mapping. The values of precision and recall are for most voxel classes predominantly in the range of 70 % to 90 %. In most cases, these values rise significantly when using the metric of neighborhood precision/recall as defined in Sec. 3.2.7. By doing so, directly neighboring voxels of a ground truth voxel are considered as well for evaluating the class label of a given voxel. The difference between precision/recall and neighborhood precision/recall

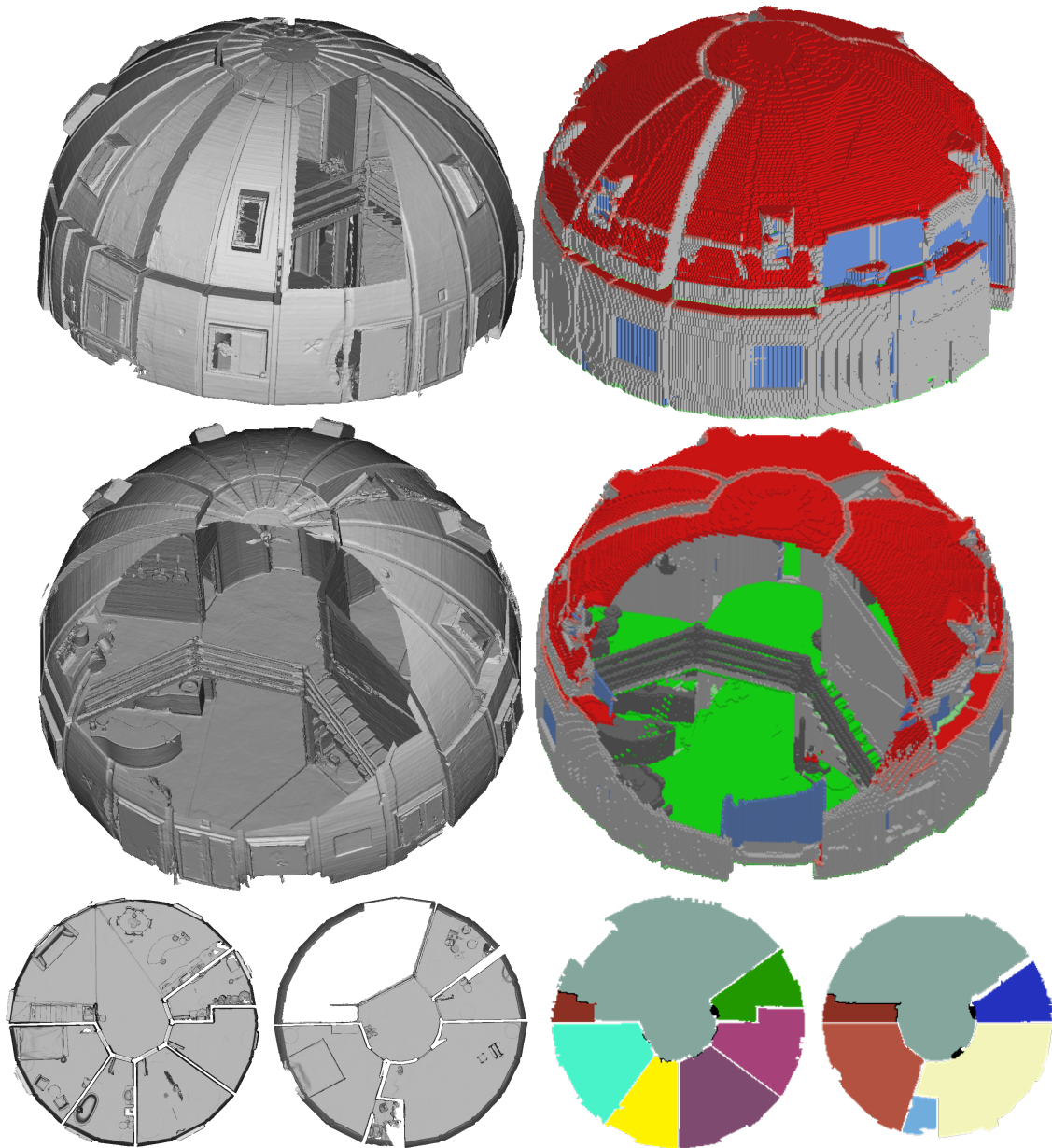


Figure 3.12: Left: Triangle mesh from the Matterport3D dataset [94]. Right: Results from our indoor reconstruction procedure applied on this mesh. In the first two rows, the results are colored according to Tab. 3.1. In the third row, different rooms are indicated by different colors while transition spaces are depicted in black. Source: [243].

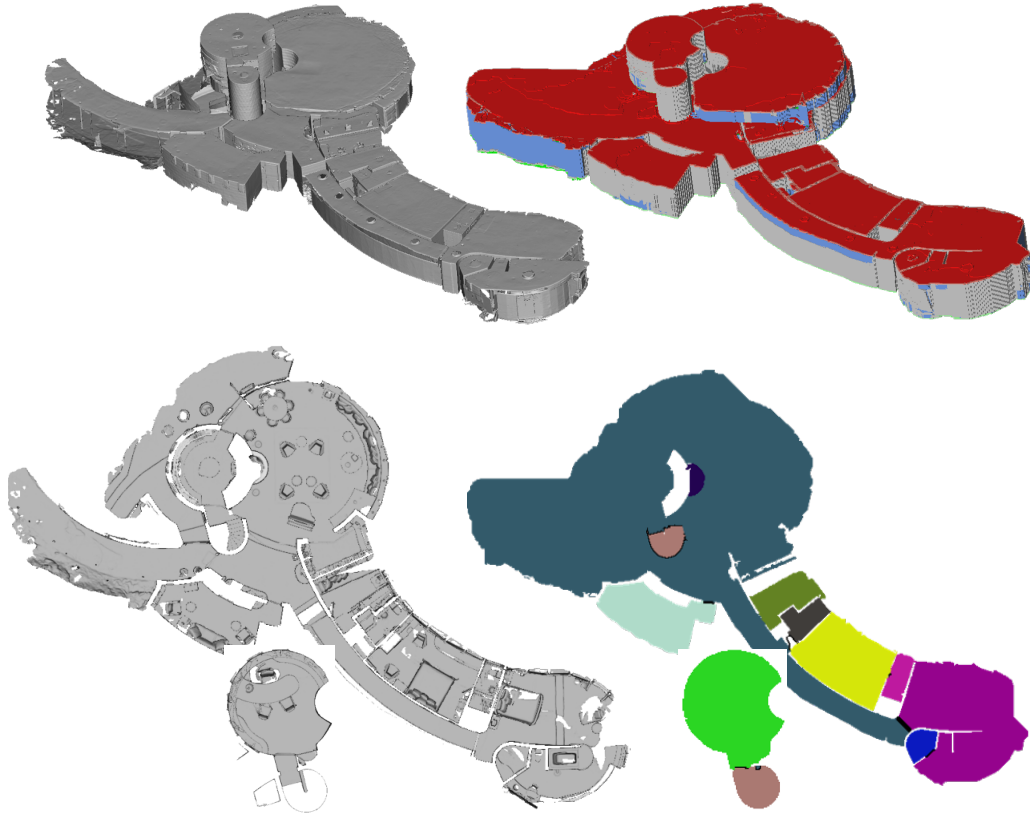


Figure 3.13: Left: Triangle mesh from the Matterport3D dataset [94]. Right: Results from our indoor reconstruction procedure applied on this mesh. In the first row, the results are colored according to Tab. 3.1. In the second row, different rooms are indicated by different colors while transition spaces are depicted in black. Source: [243].

implies that the majority of wrongly classified voxels are only one voxel away from a ground truth voxel having the correctly assigned value.

Generally, it seems less suitable for indoor reconstruction tasks to use classical metrics for pure classification tasks. It would rather be more appropriate to consider the metric distance of reconstructed geometric entities like wall surfaces to their respective counterpart from ground truth data. In the case at hand, however, quantifying the metric displacement between voxels proves to be unpractical due to the discretization caused by the rather coarse voxel resolution. Also, the difference between precision/recall and neighborhood precision/recall does not seem to change considerably with varying voxel resolution. For indoor reconstruction algorithms that output room surfaces in vector format or for very high voxel resolutions, however, a metric quantifying the distance between test and ground truth geometries like for instance the one proposed by Schuster and Weidner [514] would be more appropriate.

Overall, rather poor results – in regard to precision/recall as well as neighborhood precision/recall – are achieved in case of 'Wall Opening'. In this case, ground truth geometries had to be manually constructed. One of the reasons for the rather poor precision results for this class is the inaccuracy inherent in modelling the respective ground truth geometries into the noisy triangle meshes captured by the Microsoft HoloLens. Furthermore, 'Wall Opening' is in fact a challenging class for our reconstruction approach. In particular, the low results for recall

are caused by the abundance of false positive wall opening voxels caused by occlusion or missing input geometries due to incomplete mapping. We deal with this by checking for the presence of furniture objects causing wall occlusions and by enforcing consistency between adjacent room surfaces as detailed in Sec. 3.2.5.2. However, especially in the case of walls adjacent to the outside, it is not easy to distinguish between actual wall openings and missing room geometries, if no assumptions on the shape of wall openings are made. The achieved results could be further improved by considering a narrower definition of wall openings, e.g. by enforcing them to be rectangular.

The proposed approach is based on rather generic assumptions about the layout of indoor environments and thus its versatility is demonstrated by the results achieved on two datasets that are particularly challenging in this regard as presented in Fig. 3.12 and 3.13. Common assumptions made for indoor modelling like Manhattan world or planar room surfaces do not hold in these cases. The dataset presented in Fig. 3.12 is characterized by a spherical outer surface and a complex vertical room layout, where the room depicted in greyish color in our depicted room partitioning results extends over both storeys. Its two floor levels are connected via a staircase which is reconstructed as an own space (ramp space). This ramp space is separated from the main room by a transition space along their boundary surface in accordance with the general thoughts on the partitioning of indoor spaces presented in Sec. 3.2.6.1. Also in the case of the other dataset depicted in Fig. 3.13, the staircase (spiral stairs in this case) is reconstructed as a separate room, vertically connecting two distinct rooms. In the visualization of the room partitioning results, it is depicted in light brown. In its vicinity, it is apparent that the surrounding lower-level room is extended in the reconstruction result over the uncanopied outdoor space outside the circular main room. While the presented approach is generally able to reconstruct uncanopied space by assuming a fixed virtual ceiling height when no trace of a ceiling is found, the problems at hand result from the treatment of holes in ceiling surfaces as detailed in Sec. 3.2.2.2. Here, the decision if a hole is to be closed depends on the presence of walls along its borders. A large window surface adjacent to the outside like in the case at hand can thus cause inner yards and other parts of outdoor space to be wrongly attributed to a room. Generally, complex vertical room layouts and transitions between indoor and outdoor space are still a challenging and worthwhile subject to future research in indoor reconstruction and modelling.

A general drawback of voxel-based methods like the one presented here is scalability. As Fig. 3.11 demonstrates, processing time and memory consumption increase rapidly with the voxel grid resolution becoming increasingly finer. However, all presented examples could be processed within reasonable time on commodity hardware. Furthermore, the experiments investigating the effect of varying voxel resolution e.g. presented in Fig. 3.10 demonstrate that the proposed method still performs well with coarse voxel grid resolutions that are efficiently processable in terms of time and memory consumption. Thus, pyramid approaches are feasible, where the general room layout could be determined in coarse voxel resolution. The input data could then be processed in a finer resolution more efficiently by room-wise processing and loading into working memory from a database. Also the application of less redundant data structures such as octrees is conceivable.

The resulting voxel representation is certainly quite redundant in comparison to more prevalent formats for storing indoor building models. However, deriving vector surface representations from the resulting voxel indoor models is not a trivial matter, especially when maintaining the generic nature of this indoor reconstruction approach, and is thus outside the scope of this work. However, the proposed voxel representation holds some advantages that make it interesting for

future research taken by itself. For instance, the explicit representation of empty space and the straightforward neighborhood relations between voxels make it suited for tasks such as indoor navigation as demonstrated by Gorte et al. [201]. Besides, application in the field of indoor simulation and analysis such as heat dispersion or air quality modeling are conceivable.

3.5 Conclusion and Outlook

In this chapter, a novel fully-automatic voxel-based approach for the geometric and semantic reconstruction of indoor environments from triangle meshes was presented. First, the input triangle mesh is converted to a voxel representation, where the voxel values are based on the dominant normal directions of the contained triangles. Here, a voxel resolution of 5 cm is used, however this value is freely adjustable as demonstrated within the scope of the presented evaluation. Ceiling segments are detected in this voxel grid by region growing of voxels with main normal direction pointing downwards. For each ceiling segment, a corresponding floor segment is determined and holes in ceiling and floor caused by occlusion or gaps in the input data are closed. The voxels in-between ceiling and floor are classified in 'Empty Interior' or 'Interior Object' within the room interior and in 'Wall' or 'Wall Opening' along the lateral room boundary surfaces based on the voxels being empty or not. The geometry of the walls and especially the occurrence of wall openings behind occluding furniture are subsequently refined. Finally, the resulting interior space is repartitioned into rooms and transition spaces with special consideration of ramp spaces as vertically slanted floor surfaces like stair flights.

The proposed approach is applicable to a range of challenging scenarios encompassing indoor environments with curved walls and complex room layouts extending vertically over multiple storeys. The performance of the proposed approach is investigated by means of a thorough quantitative evaluation encompassing the variation of the key parameter represented by the voxel resolution. To this aim, a method for the automated quantitative evaluation of indoor reconstruction results against ground truth by means of appropriate metrics is presented. Furthermore, the datasets (including ground truth) used in this evaluation as well as the code of the implementation are provided to the community.

Further research should be directed to converting the reconstructed digital representations of indoor environments as voxel grid in less redundant surface-based or voluminous models that would allow conversion to prevalent data formats for the representation of digital building models such as IFC or CityGML. Another interesting direction of future research is focusing on supervised methods such as contextual classification approaches and deep learning approaches in contrast to the presented unsupervised, rule-based approach. While the proposed approach is based on a quite generic view on indoor building structures, there will always be specific buildings and datasets that are not suited for a given rule-based indoor reconstruction algorithm. Such cases could be identified to adapt the respective algorithm accordingly. More favorable in this regard however, would be a supervised, data-driven approach, that could be improved by simply adding the identified critical examples to the pool of training data. While mature machine learning based approaches already exist for the semantic segmentation of three-dimensional geometries [214, 311, 473, 312, 567], supervised methods are still rarely applied to the more complex task of indoor reconstruction which encompasses the creation of non-existent (occluded) geometries and assigning semantic meaning to empty space [129, 124].

Another promising subject of future research is further engagement with ground truth data as manually creating ground truth triangle meshes proved to be cumbersome and time-consuming. Available benchmark datasets like Stanford 2D-3D-S [30] or Matterport3D [94] provide an

abundance of labeled indoor geometries (as point clouds as well as triangle meshes) but are however missing non-visible geometries like occluded parts of walls that should also be considered when evaluating indoor reconstruction methods. Also geometrically not present entities such as wall openings are typically not considered and, while the data is usually partitioned into distinct indoor spaces, it is often not clearly stated on what basis this room partitioning is conducted. Enriching these datasets with the aforementioned missing information – preferably in an automated, unsupervised manner – would be a worthwhile endeavour. This would not only enable an automated large-scale evaluation on an abundance of challenging indoor environments but also provide the basis to apply supervised machine learning techniques such as deep neural networks to the task of three-dimensional indoor reconstruction.

Other benchmark datasets such as the ISPRS benchmark on indoor modelling [281] already provide ground truth data that exceed sole labelling of input geometries. However, here, input geometries are only given as point clouds instead of triangle meshes. While the method proposed in this work is generally also applicable to indoor mapping point clouds, it currently presupposes the presence of oriented normal vectors. While normal directions can easily be determined for point clouds, determining their orientation – i.e. the distinction between inside and outside – can be more challenging. Overcoming the need for oriented normals (or efficiently determining them for indoor mapping point clouds) would make the proposed method also applicable for these datasets.

Chapter 4

Indoor Localization

Digital building models that can for instance be created by acquiring data by means of indoor mapping techniques as discussed in Chap. 2 and a subsequent step of automatic indoor reconstruction as discussed in Chap. 3 can subsequently be used as data basis for indoor fused reality applications as presented in Sec. 1.2.5. To this aim, it is necessary to realize the spatial correspondence between the real indoor environment and the building model representing it by means of indoor localization techniques. This is the topic of this chapter.

While the term localization is sometimes used to describe the tracking of the movement of mobile devices within their environment (i.e. the 'L' in SLAM), in the context of this work, it is used to describe the determination of an absolute, initial pose of a mobile device with respect to its environment. In the context of indoor localization, this initial pose is used to localize a device within the coordinate system defined by a given digital building model in order to achieve a spatial coincidence between the building model and the physical building environment it represents. Thus, virtual content from the building model can be visualized in the correct position throughout the whole building environment when the mobile augmented reality device used to this aim can track its pose dynamically in real-time (see Sec. 1.2.3) relative to this initial, absolute pose determined by the indoor localization.

Consequently, indoor localization approaches do not necessarily need to operate continuously or in real-time, as they only need to be applied once or at least sporadically to set an absolute pose from which efficient tracking approaches with real-time capability can take over. Only when tracking loss happens, the indoor localization process has to be repeated in order to restore the alignment of the virtual building model with its physical counterpart. In this case, the last reliable pose before the tracking loss occurred could be potentially utilized to facilitate this task.

In the following, first, Sec. 4.1 gives a general overview on the topic of indoor localization, detailing the different approaches that have been proposed to this aim. Afterwards, in Sec. 4.2, a straight-forward, marker-based approach for localizing the Microsoft HoloLens within building environments is presented and evaluated. The evaluation results presented in Sec. 4.3 show, that once an absolute localization of the device within the respective building environment has been achieved, the tracking system of the HoloLens is indeed capable to visualize virtual content from a building model in its correct position with respect to the physical building environment in real time. Finally, Sec. 4.4 presents further discussions on this topic before Sec. 4.5 closes this chapter with a summary and an outlook on potential future research aims.

This chapter contains material published in [241].

4.1 Fundamentals of Indoor Localization

The topic of localizing mobile devices within pre-known indoor environments has been the focus of intensive research efforts [296, 660, 20, 205, 380, 668, 522]. Besides its fundamental role in the realization of indoor fused reality applications, indoor localization also holds importance for other fields of research such as on-site emergency response operations [331, 137], assistive systems for visually impaired persons [226, 527] and the deployment of autonomous robots in indoor environments [446, 437]. It differs from the more general task of localization in arbitrary environments as, particularly in open outdoor environments, one can rely on the availability of GNSS and extensive publicly available datasets such as Open Street Map²⁰ or Google Street View²¹ [485, 670, 453, 544]. As the microwave signals used for GNSS localization cannot penetrate sufficiently into indoor environments to allow for an adequate position estimation and as data about the geometric structure of building interiors is typically not publicly available in any centralized and standardized form due to the legally private nature of interiors, indoor environments are a particularly challenging environment for automated, absolute pose estimation.

Thus, some approaches for indoor localization in the context of indoor fused reality rely on user interaction for conducting the initial localization of the respective AR device [662, 122, 198, 296]. In this context, manual localization can e.g. be achieved by panning and rotating the virtual building structure in the AR visualization until an adequate overlay with the physical building structure is achieved and model based tracking methods can take over or via the manual selection of control points in both domains.

Other indoor localization approaches make use of indoor positioning systems that rely on a similar principal as GNSS in outdoor environments. In this case, fixed references deployed in an indoor environment are sending signals that allow mobile receivers within the building to determine their position. These systems can e.g. rely on ultrasound [245], Bluetooth²² technology [616] or the encoding of signals in electric light sources [392]. Suchlike indoor positioning system however rely on a dedicated infrastructure that has to be installed and maintained in every building where localization is required. Thus, it is more favorable, to rely on indoor localization approaches that do not require a dedicated infrastructure [629].

One possibility to so is to use the infrastructure already present in most buildings. In this context, WiFi signals can be used as well for indoor localization [585, 314]. This however requires a process of preparatory acquisition of the spatial patterns of the WiFi signals known as 'fingerprinting' [206]. This fingerprinting has to be updated at times in order to adapt to changing WiFi pattern in the respective building complex.

Due to their ubiquitous availability and their diverse sensor equipment [313], smartphones are frequently used devices for indoor localization tasks. Particularly their inertial measurement unit (IMU) is often used for a trajectory-based tracking approach called pedestrian dead reckoning (PDR) [54, 266]. In the context of indoor localization, floor plans are frequently used as reference, while other kinds of reference signals such as WiFi can also be integrated when available [33, 484, 639, 207]. These approaches often rely on particle filtering techniques [211], particularly Monte Carlo localization [136]. However, as in many fields of research, approaches based on machine learning have gained in favor recently [412, 494].

Suchlike approaches relying purely on non-visual sensors are however not particularly accurate in comparison to the required accuracies for augmented reality applications [36]. Furthermore, only positions and at most a bearing angle can be determined with this kind of indoor positioning in

²⁰<https://www.openstreetmap.de/> (Last visited on 12/06/2021)

²¹<https://www.google.de/intl/de/streetview/> (Last visited on 12/06/2021)

²²<https://www.bluetooth.com/> (Last visited on 12/06/2021)

contrast to the required poses with fully three-dimensional orientation. Due to this, camera-based approaches are of fundamental importance for indoor localization in the context of augmented reality applications [394]. However, the presented non-visual positioning methods can none-the-less be valuable for supporting camera-based approaches or for the determination of coarse, initial localization priors to ease visual localization.

A special case in this context are visual approaches where the cameras are mounted statically in the building environment (e.g. surveillance cameras), observing mobile units and thus able to support them in localization and tracking tasks [576, 227, 650]. More prevalent however are visual indoor localization approaches where the camera moves through the indoor environment and the task is the determination of an absolute camera pose.

An easy, straight-forward means for determining camera poses is to rely on artificial, planar markers [39, 450]. This approach has been frequently applied in the context of indoor localization [306, 612, 445, 390, 554]. A disadvantage of this method however is the necessity to equip the respective indoor environment with markers. Though, it needs to be mentioned, that many indoor environments are already equipped with some kind of planar markers that can readily be utilized for indoor localization if their respective poses in the building model are known, i.e. door plates or other kinds of unambiguously identifiable information signs [292, 432, 411].

More flexible however are marker-less indoor localization approaches relying on natural features. In this context, different possibilities are available to determine the absolute pose of a query image or video sequence within a known building environment. Many localization procedures rely on an image retrievable approach that aims to find the most similar one to a query image within a large collection of images and corresponding poses covering the respective indoor environment as completely as possible [332, 109, 555, 68, 644, 680]. The found image candidates with the highest similarity can subsequently be used to determine the pose of the target image.

Alternatively, other approaches aim to match the query image against a 3D representation of the respective indoor environment, either reconstructed from a similar collection of images or directly acquired three-dimensionally [443, 508, 324, 330, 329, 358]. These approaches based on 3D structures typically achieve a higher localization accuracy than approaches based on image retrieval. Recent research however has shown, that the image retrieval strategy can indeed achieve competitive results to structure-based approaches if the image coverage in the database is sufficiently dense with respect to the possible poses covering the respective indoor environment [571].

Lastly, image poses can be directly regressed by deep neural networks trained in an end-to-end manner [274, 397, 76, 42, 155]. However, pose regression based on deep learning has yet to achieve competitive results in comparison to structure-based localization and does not even clearly outperform approaches based on image retrieval [509]. Particularly, deep learning based approaches show difficulties in generalizing to poses that differ from the trajectories of image sequences used for training. It has been proposed to improve the generalizability via data augmentation approaches by rendering additional views from 3D reconstructions of the respective environment to increase coverage and amount of training data [396]. Furthermore, recent research hints at essential matrices being more favourable than directly regressing pose matrices [688].

All approaches for image-based indoor localization presented so far suffer from the drawback of requiring sufficiently up-to-date image data covering the whole indoor environment where localization should be possible. As indoor environments are often quite dynamic in their appearance due to changes in furniture and usage, frequent updates of the reference data used for localization are necessary. This is a major disadvantage as it implies repeated laborious on-site data acquisition in a manner similar to the WiFi-fingerprinting (visual fingerprinting [591]) and, in the case of approaches based on deep learning - time-consuming retraining of the networks. Thus, it seems favorable to use more generalized representations of indoor environments as reference for localization tasks. This

strategy has for instance been applied in the context of localization based on image retrieval where it has been proposed to generate the image database via virtual rendering from a BIM model instead of image acquisition in the real building [212].

Furthermore, model-based approaches for determining image poses on the basis of texture-less 3D CAD models as reference can be applied to this aim [150, 581, 566, 638]. In the context of augmented reality, model-based localization has e.g. been used with city models as reference [485, 32]. This being an outdoor scenario, GNSS can be relied upon for acquiring a coarse, initial positioning. However, also in the context of indoor localization where suchlike priors are not available, model-based localization approaches using untextured building models as reference have successfully been applied [583, 1].

An even more abstract form of reference data that has been used for image-based indoor localization are two-dimensional floor plans [118, 379, 586]. The presented approaches aim at mimicking the way humans read floor plans and work out their current location in the respective building. To this end, image-sequences instead of single frames are used to cope with ambiguous situations. This idea has recently also been applied to pose regression with untextured building models using deep learning [3].

Finally, instead of using single images or image sequences to query for absolute poses in indoor environments, three-dimensional data as acquired by indoor mapping systems (see Chap. 2) can be used as input. In this case, the reference data can be comprised of other indoor mapping results from earlier mapping campaigns [196, 388], untextured building models [224, 685] or floor plans [630, 367, 18].

Furthermore, the evaluation of indoor localization approaches is also subject of research [7, 462]. Evaluation efforts in this context have been driven by on-site competitions [357, 461, 356, 487], public benchmark datasets [393, 479, 503, 600] and simulation frameworks [172, 561].

4.2 Methodology

To achieve a spatially correct augmentation of an indoor scene with virtual building model content on the HoloLens, the spatial relationship between the environment in which the device is operating and its corresponding model has to be known. This comes down to determining the pose \mathbf{T}_{Model}^{App} of the building model in the coordinate frame of the HoloLens application.

In the following, Sec. 4.2.1 presents a simple, marker-based method to the aim of determining this building model pose \mathbf{T}_{Model}^{App} . Subsequently, the practical realization of this localization method with room-scale model data of a laboratory room is described in Sec. 4.2.2. Finally, in Sec. 4.2.3, an evaluation method suitable for quantifying the placement accuracy of virtual building model content that can be achieved with the proposed localization method is presented.

4.2.1 Localization Method

To the aim of overlaying indoor environments with virtual building model data, the pose of the HoloLens device with respect to the building model has to be determined. This equates to determining the pose \mathbf{T}_{Model}^{App} of the building model in the HoloLens application coordinate frame that leads to a correct alignment between the virtual and physical building geometry.

In the HoloLens application coordinate system *App*, the pose of the HoloLens device is described by its tracking system. The definition of the *App* coordinate system is derived from the pose of the device in the moment of starting the respective HoloLens application.

The current pose $\mathbf{T}_{HoloLens}^{App}$ of the HoloLens device in this *App* coordinate system can be queried by HoloLens applications via the HoloLens SDK²³. Furthermore, an arbitrary virtual object – a so-called ‘hologram’ – can be positioned by setting its pose $\mathbf{T}_{Hologram}^{App}$ in this *App*

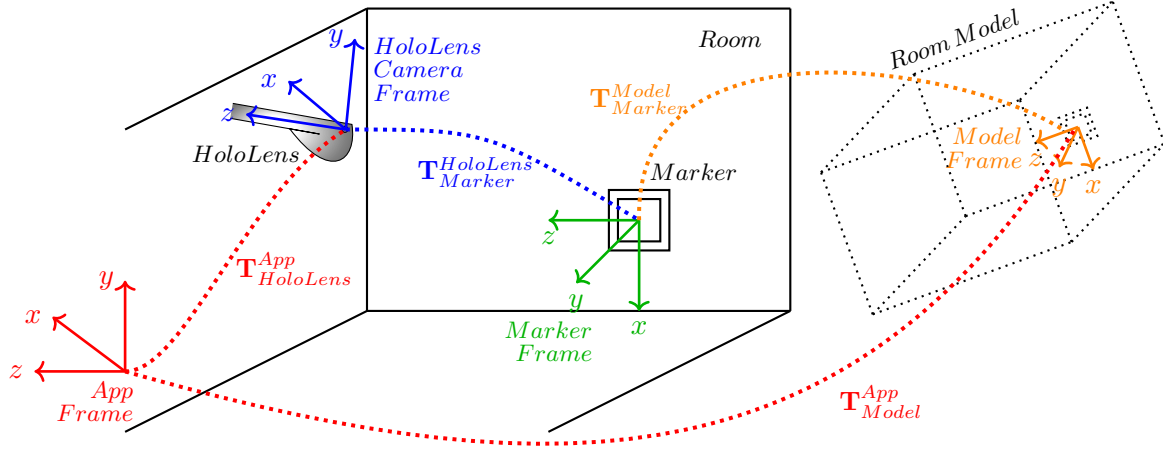


Figure 4.1: A room can be augmented with corresponding model data using a Microsoft HoloLens by determining the pose \mathbf{T}_{Model}^{App} of the virtual room model inside the HoloLens *App* coordinate frame. This pose can be determined via marker-based localization of the HoloLens camera. The pose $\mathbf{T}_{HoloLens}^{App}$ of the HoloLens device itself with respect to the *App* frame can be queried via the HoloLens SDK. Source: [241].

coordinate system via the SDK. In this manner, building model data can be positioned in the *App* coordinate system as holograms if the correct pose \mathbf{T}_{Model}^{App} of the building model can be determined.

An easy way to achieve this is by placing a marker in the building environment that is to be augmented with model data. The pose $\mathbf{T}_{Marker}^{Model}$ of this marker in the coordinate system of the corresponding building model has to be known. If the pose of this marker can also be determined in the HoloLens *App* coordinate system as ($\mathbf{T}_{Marker}^{App}$), the model pose \mathbf{T}_{Model}^{App} results in

$$\mathbf{T}_{Model}^{App} = \mathbf{T}_{Marker}^{App} \mathbf{T}_{Marker}^{Model} = \mathbf{T}_{Marker}^{App} \mathbf{T}_{Marker}^{Model^{-1}} \quad (4.1)$$

The marker pose $\mathbf{T}_{Marker}^{App}$ in the *App* coordinate system can be determined by observing the marker with the HoloLens color camera (the 'Photo/Video' camera from Tab. 2.1) and by determining its pose $\mathbf{T}_{Marker}^{HoloLens}$ in the local coordinate system of this camera. The camera pose in the *App* coordinate system in the moment of capturing the image of the marker can in turn be queried via the HoloLens SDK. The marker pose $\mathbf{T}_{Marker}^{App}$ in the *App* coordinate system then results to

$$\mathbf{T}_{Marker}^{App} = \mathbf{T}_{HoloLens}^{App} \mathbf{T}_{Marker}^{HoloLens} \quad (4.2)$$

A schematic overview of this marker-based method for localizing the HoloLens device inside building models corresponding to its surrounding indoor environment is depicted in Fig. 4.1.

²³<https://docs.microsoft.com/en-us/windows/mixed-reality/> (Last visited on 14/06/2021)

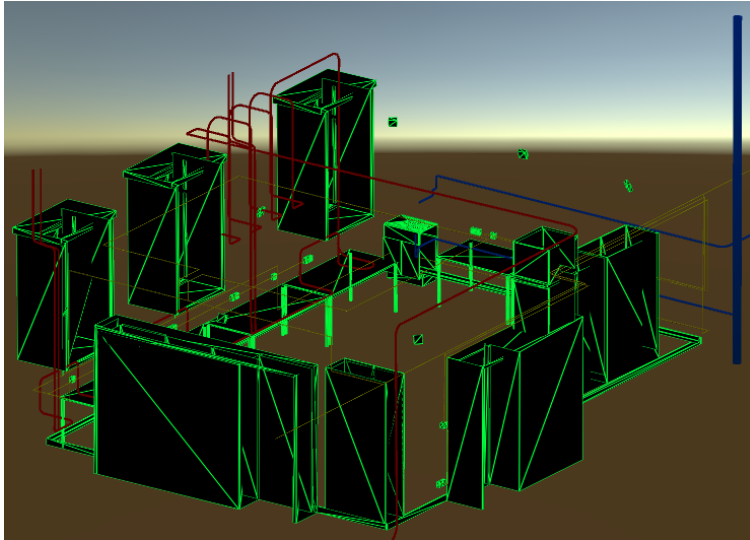


Figure 4.2: Room model: components like tables, cabinets, plug sockets and wall-mounted cameras are depicted in black and green wire frame, while infrastructure pipelines with fictive course inside the walls are depicted in red for heating pipes, blue for water pipelines and yellow for power supply lines. Source: [241].

4.2.2 Implementation

To demonstrate that the Microsoft HoloLens is capable of visualizing large room-scale virtual data with good spatial accuracy and stability, a simple marker-based localization method as described in Sec. 4.2.2 has been implemented.

For this purpose, a building model of a laboratory room of the dimensions of about $8\text{m} \times 5\text{m} \times 3\text{m}$ was created. This room model includes the room geometry like walls, windows and doors as well as furniture and various infrastructure pipelines with fictive courses inside the walls. An overview of the room model is depicted in Fig. 4.2. For reasons of clarity, the wall geometry has been omitted in this depiction.

The room which this model represents is equipped with various ArUco markers [186] of different sizes, as can be seen in Figures 4.3, 4.4 and 4.5. The local coordinate system used for the creation of the room model equates to the local coordinate system of one of those ArUco markers with corresponding points of origin.

Thus, the localization of the HoloLens device in the local coordinate frame of the room model can be achieved by capturing an image of the respective ArUco marker in the room with the HoloLens color camera and determining the pose of this camera relative to the marker. The current pose of the HoloLens camera in the coordinate frame in which the HoloLens tracks itself can be queried via the HoloLens SDK. In this way, the pose of the ArUco marker corresponding to the pose of the room model can be determined in the *App* coordinate system which in turn allows for the correct placement of virtual objects representing room components.

Fig. 4.3 for example shows the ArUco marker used for localization augmented with a virtual representation of this marker and its coordinate frame which corresponds to the coordinate frame of the room model. Fig. 4.4 shows the room augmented with the model depicted in Fig. 4.2 from the perspective of a camera placed directly in front of this localization marker looking into the room.

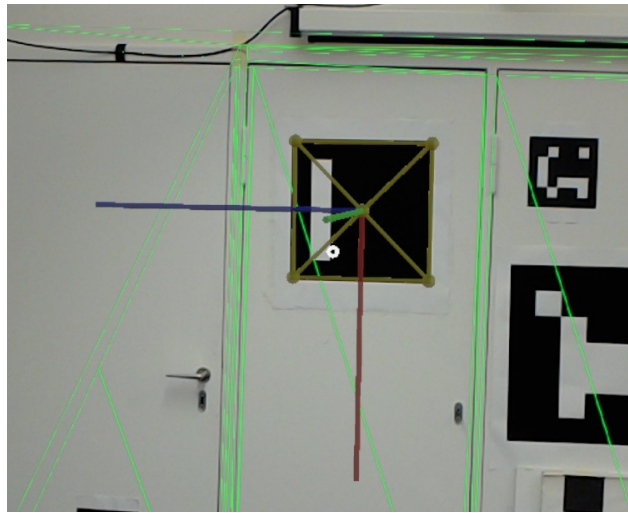


Figure 4.3: The ArUco marker used for localization augmented with a virtual representation of the marker and the coordinate system of the room model. Source: [241].



Figure 4.4: The real room augmented by the room model from Fig. 4.2 after localizing the HoloLens via the ArUco marker shown in Fig. 4.3. Source: [241].

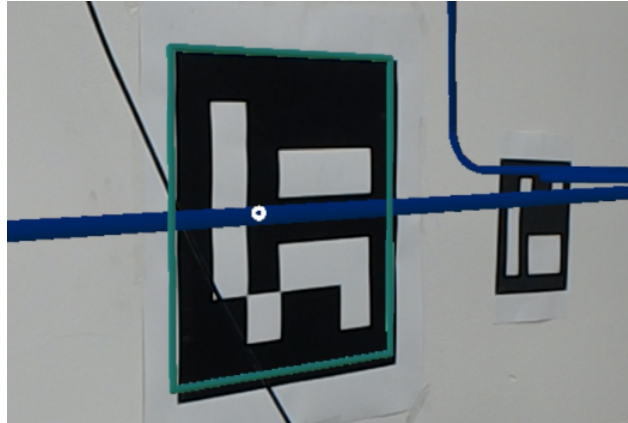


Figure 4.5: The ArUco marker used for evaluating the spatial accuracy of its overlay with the room model (blue rectangle). Source: [241].

4.2.3 Evaluation Scheme

For evaluating the spatial accuracy of the overlay between the real room and its virtual representation as visualized via the HoloLens device, another ArUco marker placed diametrically opposed to the one used for localization was augmented by a blue rectangle as part of the room model. This evaluation marker and its augmenting blue frame can be seen in Fig. 4.5. The spatial accuracy of the placement of the virtual room model can be quantified by determining the spatial deviation between the corners of the virtual rectangle and the respective corners of the physical evaluation marker.

However, for evaluating the apparent accuracy of the overlay of this evaluation marker, it is not advisable to use the 'Photo/Video' camera of the HoloLens. As shown in Sec. 4.3.1, the apparent position of virtual content with respect to the real objects in the background as depicted via the camera can differ considerably from its apparent position as perceived by the user wearing the HoloLens.

Consequently, an evaluation procedure has to be applied, that directly evaluates the impression as perceived by the user of the HoloLens device. Vassallo et al. [590] who evaluate the spatial stability of virtual content over time observed through the HoloLens achieve this by measuring the perceived position of the corners of a rectangular virtual object by contacting them with a stylus tracked with a motion capture system. This of course limits the possible evaluation distance of the virtual content to the range in which manual contacting can be performed by the user. In our case, the spatial accuracy of the placement of room-scale virtual content had to be evaluated. Thus, a contact-free measuring method was applied.

To this end, the corners of the evaluation marker and its virtual pendant were targeted with the laser pointer of a tachymeter and their Cartesian coordinates in the local coordinate system of the tachymeter were sequentially determined. A value for the spatial accuracy of the overlay between the physical and the virtual evaluation marker can be derived by averaging the Euclidean distances between the coordinates of their four corresponding corners.

In this way, deviations between the virtual and physical marker rectangle in the direction perpendicular to the wall which the physical marker is affixed to cannot be measured directly because the distance meter of the tachymeter always measures the distances to the wall surface regardless of the virtual marker being placed before or inside the wall. Thus, the projections of the corners of the virtual rectangle on the wall surface are actually measured and compared

with the physical marker corners.

Large deviations of the placement of the virtual marker rectangle perpendicular to the wall surface contribute to the determined value of the placement accuracy, in so far as they cause deviations of the position of the virtual corners projected on the wall surface. Actually, this apparent position deviation on the wall surface causes the impression of the virtual rectangle being positioned in front or behind the wall surface in the perception of the user wearing the HoloLens device. Thus, measuring the position of the marker corners projected on the wall surface can be considered as justified for quantifying the placement accuracy of virtual objects as perceived by the user.

4.3 Results

When evaluating the apparent spatial placement accuracy of virtual objects that can be achieved with a head-worn AR device like the Microsoft HoloLens, it is of importance to take into account that the three-dimensional impression of a virtual object can essentially only be experienced by an operator using the respective device and its near-eye display.

Head-worn AR devices may provide means of capturing images which themselves include renderings of the virtual objects augmenting the scene as is the case with the HoloLens and its camera. However, it is by no means certain that the apparent position of virtual objects depicted in suchlike images is in accordance with the augmented scene the user is experiencing while looking through the near-eye display of the respective head-mounted AR device.

In the case of the Microsoft HoloLens, Sec. 4.3.1 demonstrates, that the apparent position of virtual objects in augmented images captured with the 'Photo/Video' camera can differ noticeably from their apparent position as experienced by the user wearing the device. Therefore, it is imperative for an adequate and objective evaluation of the positioning accuracy of virtual objects to always measure the apparent position of a hologram directly as the user is experiencing it and not by means of an additional capturing mechanism like the HoloLens 'Photo/Video' camera.

The applied evaluation procedure as presented in Sec. 4.2.3 takes this into account. The evaluation results derived by this evaluation scheme are presented in Sec. 4.3.2.

4.3.1 Observations on the HoloLens Camera

Besides its tracking cameras, the Microsoft HoloLens is equipped with an additional camera, that allows for capturing images and videos that include visualizations of the holograms (i.e. virtual objects) the user wearing the device is seeing. The position of those virtual objects relative to the physical surroundings in the images of this camera however is not in every case identical to their position as it is perceived by the user looking through the see-through display of the device. The position of holograms overlaying physical objects can thus appear deviated in images captured with the HoloLens camera while the user wearing the device perceives them as correctly overlaying the surrounding environment.

This effect appears when the object to be overlaid with virtual content is not in the center of the image, i.e. when the cursor symbolizing the gaze direction of the user is not placed roughly on the object of interest. In those cases, when the user does not look directly at a hologram, but it is still within the field-of-view that can be augmented with virtual content by the HoloLens display, the apparent position of the respective virtual object in the image deviates noticeably from its apparent position as observed by the user.

This offset between the content of the images captured with the HoloLens camera and the

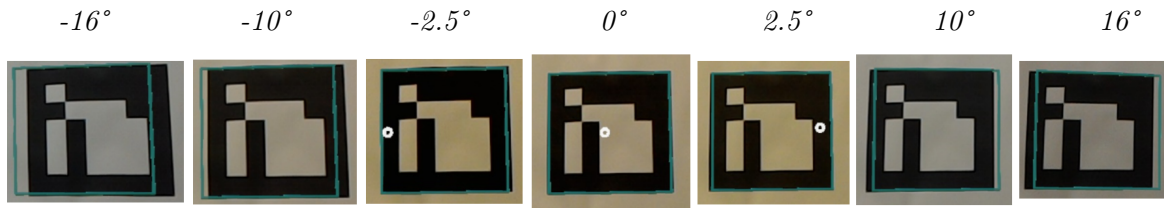


Figure 4.6: An ArUco marker is augmented with a manually positioned, blue virtual border. The images have been recorded with the HoloLens camera positioned in front of the marker while the camera is rotated subsequently to the left and to the right respectively. With increasing angle, an offset between the virtual rectangle and the physical marker is apparent that is not perceived by the user wearing the HoloLens device. Source: [241].

perception of the user wearing the device is demonstrated in Fig. 4.6. In this experiment, a blue rectangle was placed as a hologram manually by the user in such a way, that it exactly overlays the ArUco marker depicted in the images. Then, the user was standing in front of the marker looking directly at it. While the user gradually turned his head to the left and respectively to the right, images of the marker and the virtual rectangle augmenting it were captured with the camera. Some of these images labeled with the respective angle of horizontal rotation of the device are depicted in the figure.

While it is clearly visible in Fig. 4.6, that the virtual rectangle as depicted in the images increasingly shifts from the physical marker while the user is turning the gaze direction sideways, no such offset was perceived by the user through the see-through display of the HoloLens device. The user rather perceived an unvaryingly correct augmentation of the marker over the whole rotation range allowed by the field-of-view of the HoloLens display.

This aberration between the images captured with the HoloLens camera and the perception of the user has implications on the adequacy of this camera for various possible usage scenarios. Firstly, the spatial accuracy of the placement of virtual objects can appear worse in images and videos captured with the HoloLens camera than it actually is. This can cause problems in the context of demonstration and documentation of HoloLens applications where the precise placement of virtual content is of importance. Furthermore, the spatial accuracy of the placement of virtual objects cannot be evaluated automatically via those images under these circumstances (e.g. by placing virtual markers next to physical ones and comparing their offset as determined from the images with the actual reference offset).

This shift of holograms in the images of the HoloLens camera normally is not that conspicuous because in most use-cases, a hologram of interest is focused by the user (i.e. placed roughly in the center of the image) so that the offset between what the user sees through the HoloLens display and what the camera image shows is minor. Furthermore, holograms normally do not represent and overlay physically existing objects. Positional offsets of few centimeters are thus not that noticeable in prevalent use-cases.

This work however, intends to overlay physical building geometry with a corresponding virtual building model. In this scenario of overlaying physical objects with virtual content, even small deviations in the apparent position of virtual objects stand out noticeably. Furthermore, here, large-scale virtual objects are used which, even when directly focused, extend over the bounds of the region around the cursor where no offsets of the position of virtual content are

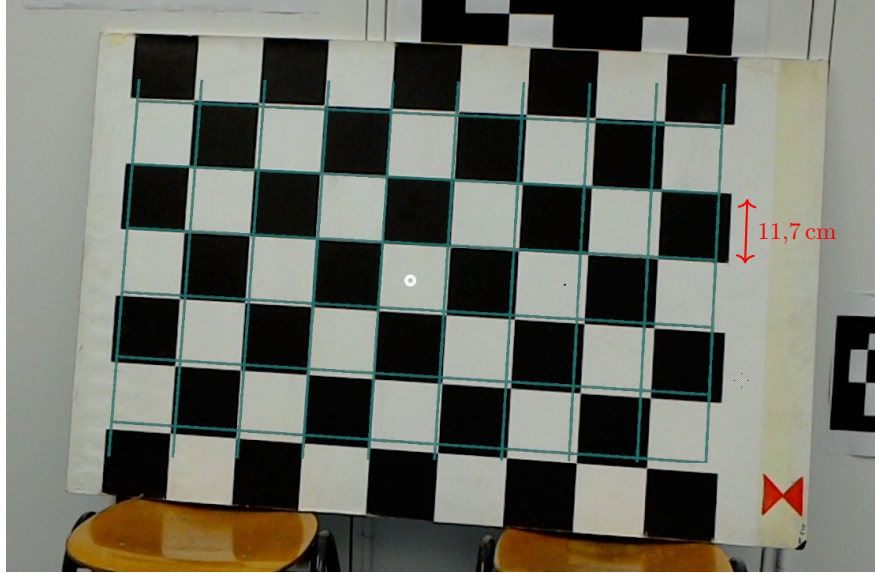


Figure 4.7: A checkerboard pattern augmented by its virtual counterpart in blue. The increasing offset between the virtual and real pattern with distance from the position of the HoloLens cursor only appears in the images captured with the HoloLens camera. The user wearing the device perceives an overall correct overlay. Source: [241].

visible in the images.

The inadequacy of spatially extended virtual objects for correct representation in the images captured with the HoloLens camera is demonstrated in Fig. 4.7. Here, a blue virtual grid was placed manually over a spatially extended checkerboard pattern to achieve a precise overlay in the perception of the user wearing the HoloLens device. While offsets between the checkerboard pattern and its virtual pendant are clearly visible in this image, the user perceived no such deviations.

4.3.2 Evaluation of the Localization Method

A quantitative evaluation of the augmentation of a room with building model data as described in Sec. 4.2.2 achieved by localizing the HoloLens device in the coordinate system of the respective model with the marker-based method proposed in Sec. 4.2.1 was performed according to the evaluation procedure presented in Sec. 4.2.3. In the course of doing so, the mean overlay error of the evaluation marker was measured from seven different points of view as depicted in Fig. 4.8 after localizing the HoloLens device via the opposing localization marker. The points of view differ in the angle under which the evaluation marker is observed and their distance to the evaluation marker. This procedure of localizing the HoloLens and the subsequent measurement of the overlay accuracy from the seven respective points of view was repeated 15 times.

The results of this evaluation (see Tab. 4.1) show that the mean overlay accuracy achievable by placing the virtual room model automatically via marker-based localization of the HoloLens device amounts to 2.3 cm. While the mean overlay errors per localization averaged over all points of view cover a wide range of values from 1.4 to 4.1 cm, the overlay errors per view point averaged over all 15 localizations all fall in the narrow range between 2.0 and 2.5 cm.

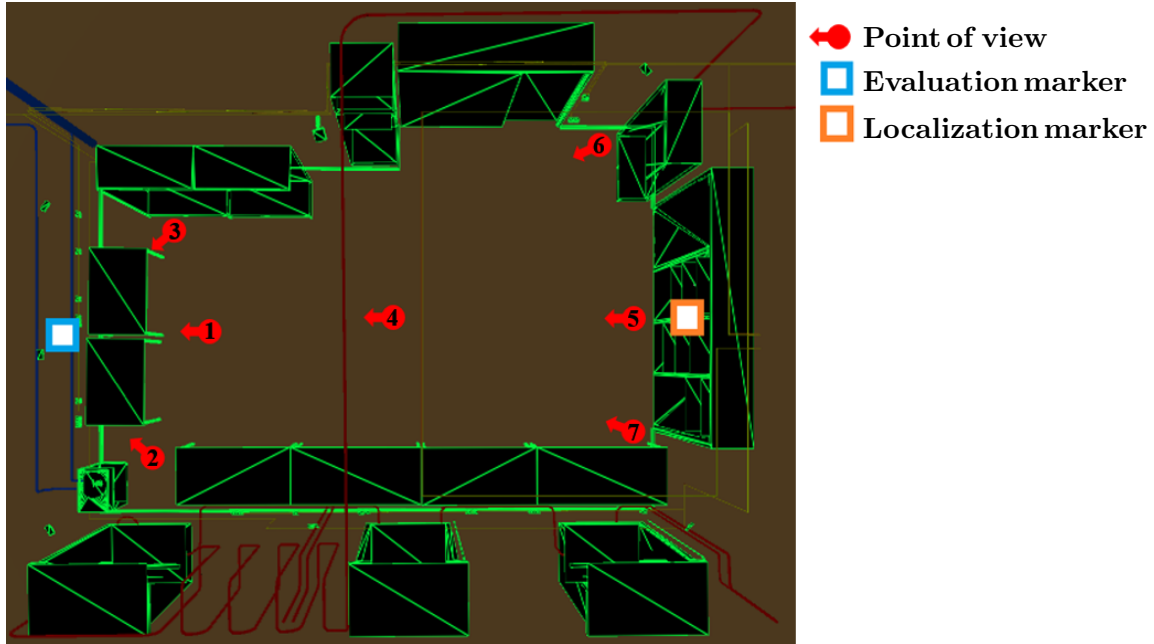


Figure 4.8: Placement of the seven different points of view that were used during the evaluation in relation to both ArUco markers. Source: [241].

Table 4.1: Overlay error [cm] between the real and the virtual evaluation marker depicted in Fig. 4.5 for all localizations and points of view (see Fig. 4.8 for positions of the points of view). Source: [241].

		Point of View							\emptyset
		1	2	3	4	5	6	7	
Localization	1	2.8	2.0	2.9	2.3	2.2	2.1	2.1	2.3
	2	1.9	1.7	1.9	1.3	1.2	1.3	1.4	1.5
	3	1.7	0.8	1.7	1.4	2.1	2.0	1.9	1.7
	4	1.1	0.6	0.9	1.8	3.2	3.2	3.0	2.0
	5	3.5	3.3	3.7	2.6	1.2	0.8	1.0	2.3
	6	3.3	2.7	3.4	3.9	4.8	4.9	5.1	4.0
	7	2.5	2.3	2.7	1.7	0.8	0.5	1.1	1.7
	8	0.2	0.3	0.4	1.1	2.6	3.0	2.4	1.4
	9	0.4	1.4	0.6	1.1	2.0	2.3	1.8	1.4
	10	3.2	3.1	3.2	2.7	2.9	2.2	2.5	3.5
	11	0.8	1.0	0.5	1.7	3.2	3.3	3.4	2.0
	12	3.7	4.0	4.0	3.7	4.2	4.3	4.8	4.1
	13	2.5	2.3	3.0	2.3	3.5	3.2	3.4	2.9
	14	1.4	1.8	1.7	0.7	1.9	2.0	1.5	1.6
	15	2.6	2.8	2.9	2.1	2.6	2.3	2.6	2.5
\emptyset	2.1	2.0	2.2	2.0	2.5	2.5	2.5	2.3	

4.4 Discussion

The experiment presented in Sec. 4.3.2 demonstrates the general applicability of the Microsoft HoloLens device for the on-site visualization of building model data. First of all, it could be shown that large room-scale model data can be visualized as holograms without experiencing adverse effects like jittering or noticeably reduced rendering frame rate. Building model data in an on-site indoor usage scenario is mainly perceived on a per-room basis, i.e. in most cases only the data in the direct surrounding of the user, e.g. just for the currently visited room, has to be visualized. So, the usage of large building-scale BIM data in this context should also be feasible, as it does not need to be visualized completely but mainly in room-scale parts around the current position of the user inside the building.

Furthermore, it could be demonstrated that correct, automatic on-site placement of building model data in indoor environments with sufficient spatial accuracy is feasible even with a rather simple marker-based approach. Of course, marker-based approaches are in practice not desirable in many usage scenarios, because they require physical preparation of the building environment via placement of markers. Fonnet et al. [177] for example discuss a marker-based localization approach in the context of on-site AR visualization of cultural heritage BIM data (hBIM), but discard this method because it would not be feasible to equip a large number of different historical buildings with artificial markers.

Anyhow, even if marker-based approaches for localization are not adequate for various scenarios, there certainly are use-cases, where markers can provide simple and efficient means for localization. In the context of historical buildings, a suchlike marker-based localization is suitable, if the number of buildings where the respective AR application should be used is not too large. For example, a single building can easily be equipped with an artificial marker in the entrance area for localizing the pose of the building model data with respect to the real world.

Other usage scenarios where a marker-based approach for augmented reality applications is feasible, concern environments, that are already equipped with marker-like planar signs that can be used for determining the location of a mobile AR client in a large-scale building model. In this context, e.g. the usage of door-plates as markers would be an option for applications in the field of facility management, with pre-known door-plate geometry providing means for pose estimation and room numbers on the door-plates providing for unambiguity of location. Orlosky et al. [432] for example use optical detection of door-plates in combination with magnetic tracking for localization in indoor emergency scenarios.

Irrespective of the way in which the localization is achieved – be it marker-based, via manual adjustment or via some sophisticated method of automatic localization like the approach proposed by Urban et al. [583] – once the room model is localized as holograms in the environment of the user, it stays affixed over time independent of the position of the user inside the respective room [590]. The evaluation results presented in Sec. 4.3.2 show little variation in the mean position error of the virtual room model content over the seven different points of view in the room on which the evaluation procedure was performed. The process of acquiring evaluation data from all seven points took about 30 minutes. During this time, no grave displacements of the apparent position of the holograms presenting the room model were experienced. It was also tested to leave the room through the door and enter it again. In doing so, the virtual room model still proved to match the real room geometry after re-entering.

Ultimately, it can be concluded that the HoloLens as an AR device is principally usable for the on-site visualization of BIM data, if the correct pose of the model data corresponding to the physical building environment it represents can be determined. In contrast to tracking,

this localization procedure does not have to be performed in real-time. It is rather sufficient to localize building model content just initially at the beginning or from time to time, e.g. when entering new rooms or if tracking is lost. Furthermore, it is possible to set so-called 'Spatial Anchors'²⁴ for the holograms representing the BIM data. This allows storing the poses of the respective holograms with respect to the spatial mapping data of the respective room they are placed in. Thus, when the HoloLens application is started a second time in a room, that contains 'Spatial Anchors', the respective holograms can be loaded automatically and are visualized in the correct positions without having to perform a localization procedure with respect to the building model again.

4.5 Conclusion and Outlook

This chapter demonstrated that the Microsoft HoloLens as a mobile augmented reality platform is suited for the spatially correct on-site visualization of building model data. Once the virtual building model content is placed to correctly overlay the physical structures of the indoor building environment it represents, the tracking capacity of the HoloLens is sufficient to keep its apparent position stable over time independent of the movements of the user wearing the device.

So, using the HoloLens as an AR client for the on-site inspection of BIM data comes down to a localization procedure for aligning the virtual building data with its physical counterparts. This chapter presented a simple marker-based approach to this aim. Furthermore, an evaluation procedure was proposed, that measures the spatial precision of hologram placement directly as it is experienced by the user and does not rely on the HoloLens 'Photo/Video' camera that proofed to produce apparent positions of virtual content that differ from the impression the user is perceiving.

Use-cases suited for a marker-based localization approach in the context of AR-based in-situ presentation of BIM data certainly exists. However, automatic localization methods that do not require the use of artificial markers or manual adjustment of the pose of a building model are a promising field of future research. In this context, the range sensor and indoor mapping capability of the HoloLens (see Chap. 2) could be made use of for detecting the part of a building model which the indoor mapping meshes best correspond to. Here, special emphasis has to be laid on how to deal with ambiguities inherent in certain building structures and how to remove those parts of the indoor mapping meshes that are not represented in corresponding building model data (e.g. furniture) and can disturb the correlation with building model structures.

²⁴<https://docs.microsoft.com/en-us/windows/mixed-reality/spatial-anchors> (Last visited on 14/05/2021)

Chapter 5

Pose Normalization of Indoor Mapping Datasets

The indoor reconstruction approach presented in Chap. 3 along with many other recently proposed approaches in this field does not rely on the Manhattan world assumption (see Sec. 1.2.1). However, many other approaches proposed for the task of indoor reconstruction do rely on this assumption about the geometric structure of the building environments to be reconstructed.

The fact that a given indoor reconstruction approach relies on the Manhattan world assumption does not only imply that the building structure to be reconstructed itself must be compliant to the Manhattan world assumption. Rather, this also implies that the geometric representation of the respective building in the indoor mapping data must be correctly aligned with the coordinate axes in accordance with the definition of the Manhattan world assumption, i.e. that the surfaces pertaining to the three main directions (or six when considering oriented directions) are aligned with the three axes of the coordinate system.

In the context of indoor mapping however (see Chap. 2), the pose of the captured building structure with respect to the coordinate system does not necessarily fulfill this requirement. Frequently, the coordinate system is determined by the initial pose of the indoor mapping system at the beginning of the mapping process. Thus, the orientation of the indoor mapping data can deviate from the Manhattan world assumption by a rotation around the vertical coordinate axis even if the mapped building structure itself is totally compliant with the Manhattan world assumption. Moreover, the orientation of the vertical axis itself can also deviate from its optimal orientation according to the Manhattan world assumption, i.e. being perpendicular to horizontal ceiling and floor surfaces. This is generally not the case when a leveled mounting of the respective indoor mapping sensor is used, e.g. in the case of tripod-mounted systems like TLS or trolley-based systems. In the case of hand-held or head-worn indoor mapping systems where a perfectly leveled orientation at the start of the indoor mapping process cannot be guaranteed, an eventual misalignment of the indoor mapping data with respect to the vertical coordinate axis needs to be taken into account.

Aligning an indoor mapping dataset with the coordinate axes - horizontally and depending on the used indoor mapping system also vertically - is thus a necessary preprocessing step for automated indoor reconstruction approaches that rely on the Manhattan world assumption. Moreover, a suchlike alignment process - also known as pose normalization - can still be a reasonable choice, even if the respective indoor reconstruction method does not presuppose a Manhattan world compliant building structure. This is for instance the case, when a respective

indoor reconstruction approach makes use of a voxel grid or octree representation of the input data, like for instance the one proposed in Chap. 3 [174, 201, 124, 242]. Even if a voxel-based indoor reconstruction approach is able to handle building structures deviating from the Manhattan world assumption, having room surfaces aligned with the coordinate axes and thus with the voxel grid will result in a cleaner and visually more appealing reconstruction in voxel space. Furthermore, spatially discretizing data which is not aligned with the coordinate axes can lead to aliasing effects that can impede a successful reconstruction process [370, 371, 646]. Besides, pose normalization often - but not necessarily always, depending on the respective building structure - results in a minimal axis-aligned bounding box circumscribing the indoor mapping data and thus to reduced memory size of the voxel grid structure.

Lastly, pose normalization of indoor mapping data can also be of benefit in the context of the co-registration of multiple datasets representing the same indoor environment that are to be aligned with each other [626, 107, 237]. The respective datasets to be aligned can be acquired by different sensor systems or at different times, e.g. in the context of change detection [52, 294, 360]. While pose normalization with respect to a Manhattan world structure does not entirely solve this problem as an ambiguity of rotations of multiples of 90° around the vertical axis remains, it nonetheless can be reasonable to apply pose normalization when co-registering indoor mapping datasets as it reduces the problem to finding the correct of only four possible states per dataset.

The same arguments speaking in favor of pose normalization - even if an indoor reconstruction approach does not necessarily depend on it - also hold for the case of building structures that are only partly compliant to the Manhattan world assumption. Thus, a pose normalization approach should be robust against a substantial amount of the given indoor mapping geometries deviating from the Manhattan world structure of the building. Particularly in the case of building environments that contain multiple Manhattan world structures (i.e. Atlanta world), the dominant Manhattan world structure (e.g. in terms of the largest fraction of supporting geometries) should be used for alignment with the coordinate axes. In situations, where multiple Manhattan World structures have about the same support, it might be reasonable to detect them all and create multiple solutions for a valid pose normalization.

This chapter presents a novel pose normalization method for indoor mapping point clouds and triangle meshes that is robust to the represented building structures being only partly compliant to the Manhattan world assumption. In case there are multiple major Manhattan world structures present in the data, the dominant one is detected and used for alignment. Besides the horizontal alignment of the Manhattan world structure with the coordinate system axes, vertical alignment is also supported for cases where the deployed indoor mapping system is not leveled and the resulting dataset is thus misaligned with respect to the vertical coordinate axis. In this context, the indoor mapping dataset is assumed to be coarsely leveled to within $\pm 30^\circ$ of the optimal vertical direction which can usually be expected to be the case for hand-carried or head-worn mobile indoor mapping systems. Furthermore, the individual indoor mapping geometries are assumed to have normal vectors which however do not need to be consistently oriented and can thus be easily determined as a preprocessing step for point clouds while triangle meshes do already have normal vectors inherent in the geometries of the individual triangles. The implementation of the proposed pose normalization approach along with the code for the presented quantitative evaluation on publicly available indoor mapping datasets is made available to the community²⁵.

²⁵<https://github.com/huepat/im-posenorm> (Last visited on 15/07/2021)

After discussing related work in Sec. 5.1, the presented approach for pose normalization is described in Sec. 5.2 along with a method to resolve the ambiguity of a rotation of multiples of 90° around the vertical axis and the procedure applied for quantitative evaluation. The results of this evaluation procedure applied to several publicly available indoor mapping point clouds and triangle meshes are subsequently presented in Sec. 5.3 and discussed in further detail in Sec. 5.4. Finally, Sec. 5.5 closes with concluding remarks and an outlook on future research.

This chapter contains material published in [244].

5.1 Related Work

Besides applications in the context of building structures, a range of pose normalization approaches have been presented that aim at aligning arbitrary three-dimensional objects with the coordinate axes. In this more general context, these objects do not necessarily represent building structures [273, 441, 95, 179, 336, 96, 337, 517]. These approaches are mainly motivated by the need to design rotation invariant shape descriptors in the context of shape retrievable, i.e. finding all similar three-dimensional objects to a given query shape from a large database of 3D objects [674, 560].

In this context, variations of the principal component analysis (PCA) algorithm [258] are often made use of [273, 441, 95, 96]. Also, symmetries in the geometry of the respective object are often exploited as well [95, 179, 96]. Other approaches rely on the geometric property of rectilinearity [336, 337] or aim to minimize the size of a surface-oriented bounding box circumscribing the target object [517].

More specifically concerning building structures, a recent pose normalization approach makes use of point density histograms, discretizing and aggregating the points of an indoor mapping point cloud along the direction of one of the horizontal coordinate axes [370, 371]. The optimal horizontal alignment of the point cloud is determined by maximizing the size and distinctness of peaks in this histogram varying with the rotation around the vertical axis.

Other approaches, including the one proposed in this work, do not discretize the data with respect to their position but with respect to their orientation [430, 280, 143, 127]. This is conducted on the extended Gaussian image [231] which consists of the normal vectors of the individual indoor mapping geometries projected on the unit sphere. Besides its application in the context of pose normalization, the extended Gaussian image is also frequently applied to the segmentation of point clouds [615, 547, 523, 546, 684] or plane detection [341], particularly with regard to building structures.

In a straight-forward approach for instance, the points in the extended Gaussian image are subjected to a k-Means clustering [365, 350] to determine three clusters corresponding to the main directions of the Manhattan world structure while disregarding the absolute orientation of the normal vectors (i.e. projecting them all in the same hemisphere) [280, 143]. This however is not robust to deviations of the indoor mapping point cloud from an ideal Manhattan world structure. In contrast, using DBSCAN [163] for clustering on the extended Gaussian image has been proposed [127] which is more robust as it does not fix the number of clusters to exactly three. This allows for the presence of surfaces deviating from an ideal Manhattan world system. The proposed approach however only aims at detecting dominant planes to remove them from the point cloud and does not assemble the detected orientation clusters to Manhattan world structures. In another approach, dominant horizontal directions are detected by projecting the normal vectors to the horizontal plane and binning the resulting angles to a horizontal reference coordinate axis in a similar manner to the approach presented here [430].

All of the approaches mentioned above only concern themselves with determining an orientation around the vertical axis to achieve an alignment of the Manhattan world structure of an indoor mapping dataset with the coordinate axes. To best knowledge, no approach on pose normalization of indoor mapping point clouds or triangle meshes has yet been proposed that aims at determining an optimal alignment with respect to the orientation of the vertical axis as well. Furthermore, the presented approaches do not address the topic of robustness to deviations of the respective building structure from an ideal Manhattan world scenario or the presence of multiple Manhattan world structures in the same building.

5.2 Methodology

This section presents a novel method for automatic pose normalization of indoor mapping point clouds or triangle meshes which represent building structures that are at least partially compliant to the Manhattan world assumption. The presented method aims at rotating the given indoor mapping geometries to a pose with respect to the surrounding coordinate system for which the largest possible fraction of normal vectors is aligned with the three Cartesian coordinate axes. This comprises an optional leveling step to orient horizontal surfaces like floors and ceilings to be orthogonal to a chosen vertical axis if this is not already achieved by the data acquisition process (e.g. by using leveled tripod or trolley mounted acquisition systems). Subsequently, a second step determines the optimal rotation angle around this vertical axis in order to align the largest possible fraction of the building surfaces with the horizontal pair of orthogonal coordinate axes.

The presented method is applicable to all kind of indoor mapping point clouds and triangle meshes. However, it assumes the individual geometric primitives comprising the input data to have normal vectors. While these are intrinsically given for the individual triangles comprising a triangle mesh, the individual points of indoor mapping point clouds do not generally have normal vectors. These can however be easily determined by means of established methods like [391, 75, 666, 500] which are assumed in this work as a necessary preprocessing step. Note that these normal vectors need not be oriented, i.e. only their direction is of importance. Furthermore, the input data is assumed to be at least coarsely levelled, i.e. the represented building structures is assumed to be coarsely aligned with the vertical axis within the range of $\pm 30^\circ$.

In the following, \vec{n}_i denotes the i -th normal vector of N given input geometries (i.e. points or triangles) while $\langle \cdot, \cdot \rangle$ denotes the dot product of two 3D vectors. Furthermore, the vector determining the vertical axis is denoted by \vec{z} . However, it needs to be stated that this vector need not necessarily equal $(0\ 0\ 1)^T$. It can be chosen freely in accordance with the intended coordinate system. However, it must coincide within $\pm 30^\circ$ with the current vertical orientation of the input data. Similarly, a horizontal axis \vec{x} orthogonal to the configured \vec{z} -axis is to be chosen. Lastly, the second horizontal axis completing the Cartesian coordinate system must not be explicitly stated but can be determined as

$$\vec{y} = \vec{z} \times \vec{x} \tag{5.1}$$

Again, note that the horizontal axes need not necessarily equal $(1\ 0\ 0)^T$ and $(0\ 1\ 0)^T$.

In the following, Sec. 5.2.1 first presents the proposed method for determining an optimal rotation around the vertical axis in order to horizontally align the indoor mapping data with the coordinate system in case the dataset is already vertically aligned in relation to the vertical axis. A suitable method for ensuring this vertical alignment that can be applied as a preprocessing step to datasets that are only coarsely aligned with the vertical direction ($\pm 30^\circ$) is subsequently

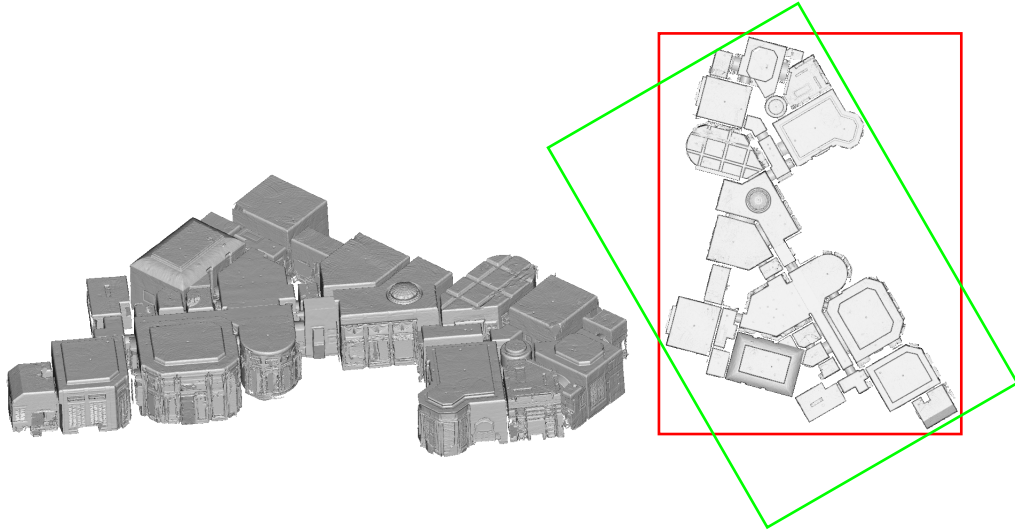


Figure 5.1: Exemplary triangle mesh of a building with multiple Manhattan world systems (dataset ‘mJXqzFtmKg4’ from Matterport3D [94]). The green bounding box on the top-down-view on the right-hand side illustrates the alignment along the dominant Manhattan world structure considered as ground truth pose while the red bounding box illustrates the pose rotated by 30° around the vertical axis as exemplarily used in Sec. 5.2.1. Source: [244].

presented in Sec. 5.2.2. As the proposed method for determining the rotation around the vertical axis is ambiguous with regard to multiples of 90° , Sec. 5.2.3 presents an approach to solve this ambiguity. Lastly, Sec. 5.2.4 presents the evaluation methodology applied in this study.

5.2.1 Rotation around the Vertical Axis

This section preliminarily assumes, that the given indoor mapping data (comprised of triangles or points) is already leveled with regard to a chosen vertical axis \vec{z} (that does not necessarily need to equal $(0\ 0\ 1)^T$). Thus, only one rotation angle around this vertical axis is to be determined in order to align the two horizontal axes of the coordinate system with the horizontal directions of the dominant Manhattan world structure underlying the respective building represented by the input data.

In case the given input data is not entirely compliant to the Manhattan world assumption, a best-possible solution in terms of the alignment of all normal vectors with the horizontal coordinate axes is to be found. Even indoor mapping data that represents building structures entirely compliant to the Manhattan world assumption can have a significant amount of normal vector directions deviating from the directions of the respective Manhattan world system. These deviating normal vector directions can be caused by actual unevenness of walls, by noise inherent in data acquisition and normal determination as well as by clutter like furniture objects being present in the indoor mapping data additionally to the building structure itself.

Besides being robust against these restrictions, the presented method is also applicable to building structures that are only partially Manhattan world conform. Building structures with multiple Manhattan world systems like the one depicted in Fig. 5.1 are aligned according to the respective Manhattan world system supported by the largest fraction of normal vector directions.

Thus, the task at hand is to determine an angle of rotation around the vertical axis \vec{z} that leads to the largest positive fraction of normal vectors being aligned with the horizontal axes \vec{x}

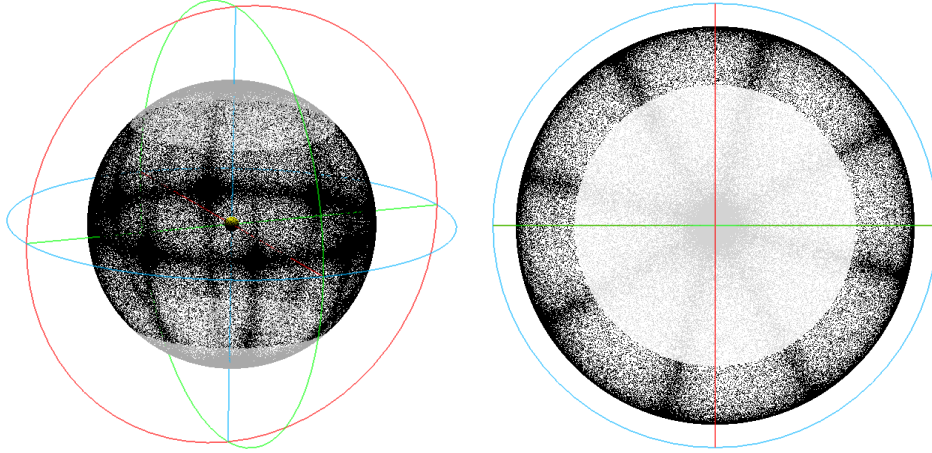


Figure 5.2: The normal vectors \vec{n}_i of the triangle mesh shown in Fig. 5.1 visualized as extended Gaussian image (thinned out by factor 25 for the sake of visibility). The normal vectors \vec{n}_i^h that are horizontal within the range of $\pm 45^\circ$ are visualized in black while the others are visualized in gray. The coordinate axes are visualized in red for \vec{x} , green for \vec{y} and blue for the vertical axis \vec{z} . Source: [244].

and \vec{y} . To this aim, first, the normal vectors that can be considered coarsely horizontal with respect to the vertical axis \vec{z} are filtered. For this, all N^h normal vectors \vec{n}_i^h that are within the range of $\pm 45^\circ$ of a horizontal orientation are considered, thus

$$45^\circ \leq |\angle(\vec{n}_i, \vec{z})| \leq 135^\circ \quad (5.2)$$

where $\angle(\cdot, \cdot)$ denotes the smallest angle between two 3D vectors with respect to any rotation axis. For the indoor mapping mesh depicted in Fig. 5.1, the corresponding horizontal normal vectors \vec{n}_i^h are depicted in the form of an extended Gaussian image in Fig. 5.2. In this example, the triangle mesh of Fig. 5.1 is rotated by 30° around the vertical axis relative to the ground truth pose aligned to the dominant Manhattan world structure.

These horizontal normal vectors \vec{n}_i^h can subsequently be projected in the horizontal plane formed by the horizontal axes \vec{x} and \vec{y} by

$$\vec{n}_i^h = \vec{n}_i^h - \langle \vec{n}_i^h, \vec{z} \rangle \vec{z} \quad (5.3)$$

where their respective angles to the reference direction of \vec{x} around \vec{z} as axis of rotation

$$\gamma_i = \angle_{\vec{z}}(\vec{n}_i^h, \vec{x}) = \arctan \frac{\langle \vec{z}, \vec{n}_i^h \times \vec{x} \rangle}{\langle \vec{n}_i^h, \vec{x} \rangle} \in [-180^\circ, 180^\circ] \quad (5.4)$$

can be determined.

The problem at hand can be formulated as determining the rotation angle $\gamma \in [0^\circ, 90^\circ)$ around the vertical axis that minimizes the sum of angular distances of each horizontal normal vector to

the respectively nearest horizontal coordinate axis, i.e.:

$$\gamma = \arg \min_{\hat{\gamma} \in [0^\circ, 90^\circ)} \sum_{i=0}^{N^h} w_i \min \left\{ \begin{array}{l} |\hat{\gamma} - \gamma_i| \\ |\hat{\gamma} - \gamma_i + 90^\circ| \\ |\hat{\gamma} - \gamma_i + 180^\circ| \\ |\hat{\gamma} - \gamma_i - 90^\circ| \end{array} \right\} \quad (5.5)$$

Here, the angular distances of each angle γ_i to the nearest horizontal axis are weighted by factor w_i . This factor can be constantly set to 1 for the points of an indoor mapping point cloud. In the case of triangle meshes however, it allows to weigh the individual triangles by their respective area as larger triangles imply a larger quantity of points in a corresponding point cloud representation.

Eq. 5.5 is not analytically solvable. It can however be solved numerically by derivative-free minimization methods like e.g. Brent minimization [78]. This, however, does not scale well with the size of the input data, as all the angles derived from the horizontal normal vectors need to be iterated in each step of the respective numeric method. And - particularly in the case of indoor mapping point clouds - the amount of geometric primitives and thus of angles to be processed can reach a tremendous size.

Thus, this work proposes an approach that discretizes the input data into a one-dimensional grid of fixed resolution by means of which the angle of rotation for aligning the input data with the horizontal coordinate system can be determined non-iteratively in one step. In this context, a resolution of 1° proved to be suited for a coarse initial determination of the rotation angle for horizontal alignment that can subsequently be refined. For each angle γ_i , the respective grid cell is determined which is incremented by the respective weight w_i , which again is constantly 1 for points of point clouds but in the case of triangle meshes weights the respective angle by the area of the corresponding triangle.

Fig. 5.3 visualizes a suchlike one-dimensional grid representation of the horizontal angles γ_i over the full circle of 360° for the mesh presented in Fig. 5.1. The peaks in the summarized weights per grid cell correspond to the eight horizontal main directions of the two Manhattan world systems present in the dataset depicted in Fig. 5.1.

To decide about the dominant of the two Manhattan world systems involved and to determine the corresponding rotation angle for an alignment of the input data with it, the weights of the involved grid cells need to be summarized over all peaks pertaining to the same Manhattan world system. To this end, the peaks belonging to the same Manhattan world system and thus having an angular difference of a multiple of 90° between each other need to be identified and associated. Thus, the angles $\gamma_i \in [-180^\circ, 180^\circ)$ are mapped to $[0^\circ, 90^\circ)$ by

$$\gamma_i^* = \begin{cases} \gamma_i + 180^\circ & \gamma_i < 0^\circ \\ \gamma_i & \text{else} \end{cases} \in [0^\circ, 180^\circ) \quad (5.6)$$

and

$$\tilde{\gamma}_i = \begin{cases} \gamma_i^* - 90^\circ & \gamma_i^* > 90^\circ \\ \gamma_i^* & \text{else} \end{cases} \in [0^\circ, 90^\circ) \quad (5.7)$$

The discretized grid representation of the angles $\tilde{\gamma}_i \in [0^\circ, 90^\circ)$ thus needs only a quarter of the size in comparison to discretizing the angles $\gamma_i \in [-180^\circ, 180^\circ)$ with the same resolution. Furthermore, the resulting grid as visualized in Fig. 5.4 enables the coarse initial determination of

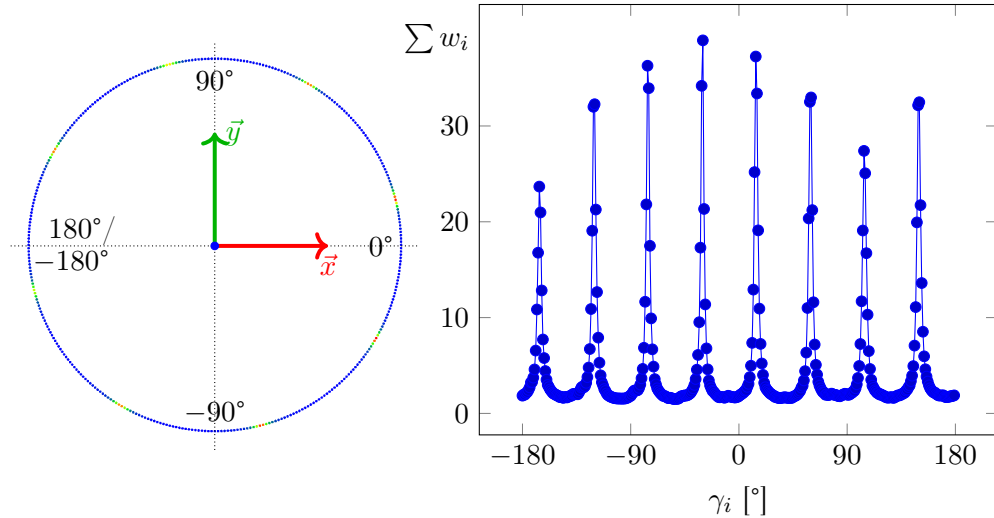


Figure 5.3: Visualization of a one-dimensional 360° grid corresponding to Fig. 5.1. The grid cells contain the summarized weights w_i of the contained angles γ_i with value colorization ranging from blue for low values over green and yellow to red for large values. Source: [244].

the rotation angle γ . To this end, the weight sums per grid cell are thresholded with a threshold value of 0.75 times the maximal weight sum of the whole grid and subsequently clustered. While clustering, the fact that clusters can extend over the discontinuity between 0° and 90° needs to be taken account of.

Finally, the grid cell cluster with the largest weight summarized over the contained cells is selected and γ is determined as the weighted average of the angle values corresponding to the cluster cells (with 1° resolution) weighted by their respective weight sum values. Fig. 5.5(a) shows the horizontal triangle mesh faces of Fig. 5.1 corresponding to the largest peak at 60° in Fig. 5.4 that determines the dominant Manhattan world system of that dataset. The faces corresponding to the second peak at 15° in Fig. 5.4 are visualized in Fig. 5.5(b).

The resulting value for γ can subsequently be further refined by determining the weighted median over all $\tilde{\gamma}_i$ within a certain angular distance of the initial value for γ while applying the weights w_i . A threshold of 5° was found to be suitable for this task.

Finally, the indoor mapping data can be rotated by the thus refined angle γ around the vertical axis to achieve the alignment of the building geometry with the horizontal coordinate axes. In the case of a triangle mesh, it is sufficient to rotate the vertices of the triangles as the respective normal vectors of the rotated triangles can be calculated on the basis of the triangle geometry. In the case of point clouds however, the respective normal vectors of the points need to be explicitly updated along with the coordinates of the points.

5.2.2 Orientation of the Vertical Axis

In the preceding Sec. 5.2.1, the rotation around the vertical axis was determined under the assumption that the vertical axis is perfectly leveled with respect to the building structure, i.e. that it is orthogonal to horizontal floor and ceiling surfaces. In the case of tripod mounted indoor mapping systems like terrestrial laser scanners, this assumption is justified as these devices are typically leveled before usage. However, in the case of mobile indoor mapping systems like hand-carried or head-worn devices, this is generally not the case. In these cases, the coordinate

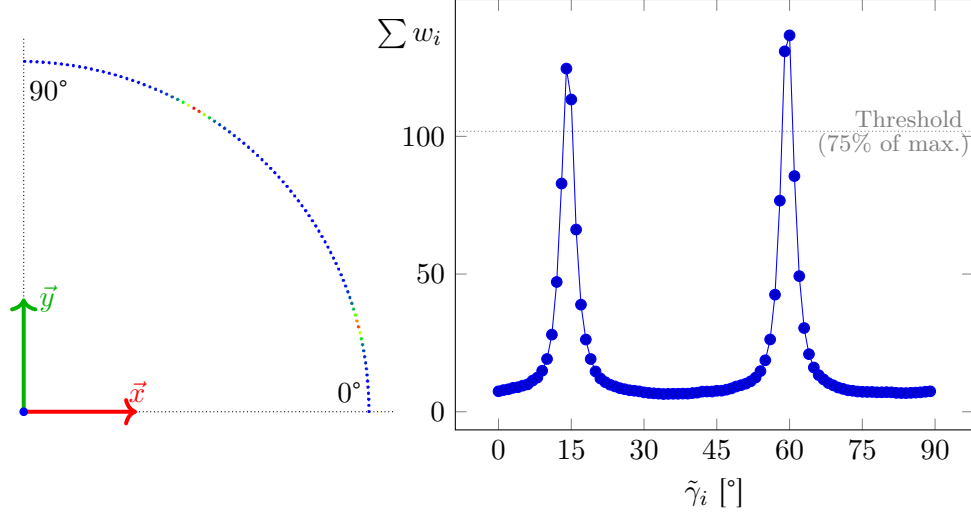


Figure 5.4: Visualization of a one-dimensional 90° grid corresponding to Fig. 5.1. The grid cells contain the summarized weights w_i of the contained angles $\tilde{\gamma}_i$ with value colorization ranging from blue for low values over green and yellow to red for large values. Source: [244].

system of the indoor mapping data is often defined by the initial pose of the mobile mapping device when starting the data acquisition process. In consideration of typical usage postures of such mobile systems, it can be assumed that the respective vertical axis of the coordinate system is still roughly pointing upwards $\pm 30^\circ$. If this is not the case, a coarse leveling within this range can easily be conducted manually.

To justify the assumption made in the previous section, this section presents an approach for automatically leveling indoor mapping point clouds or triangle meshes where a chosen vertical axis \vec{z} corresponds coarsely within $\pm 30^\circ$ with the actual upwards direction of the building structure standing orthogonally on horizontal floor surfaces. As in the preceding section, the input data for conducting this alignment of the input mapping data with the coordinate system are again the N normal vectors \vec{n}_i of the individual geometric primitives comprising the indoor mapping data (i.e. points or triangles).

Analogous to Eq. 5.5, the task of vertically aligning the indoor mapping geometries with the coordinate system axis \vec{z} can be formulated as

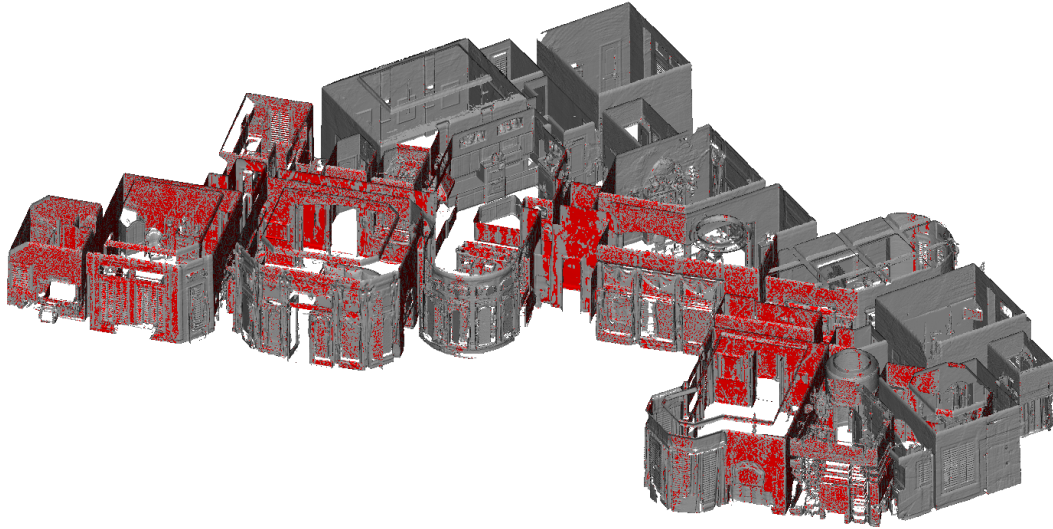
$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \arg \min_{\hat{\alpha}, \hat{\beta} \in [-30^\circ, 30^\circ]} \sum_{i=0}^{N^v} w_i \min \left\{ \begin{array}{l} |\langle (R(\hat{\alpha}, \hat{\beta}) \vec{n}_i^v, \vec{z})| \\ |\langle (R(\hat{\alpha}, \hat{\beta}) \vec{n}_i^v, \vec{z}) - 180^\circ| \end{array} \right\} \quad (5.8)$$

where \vec{n}_i^v are the N^v normal vectors that are vertically oriented within the range

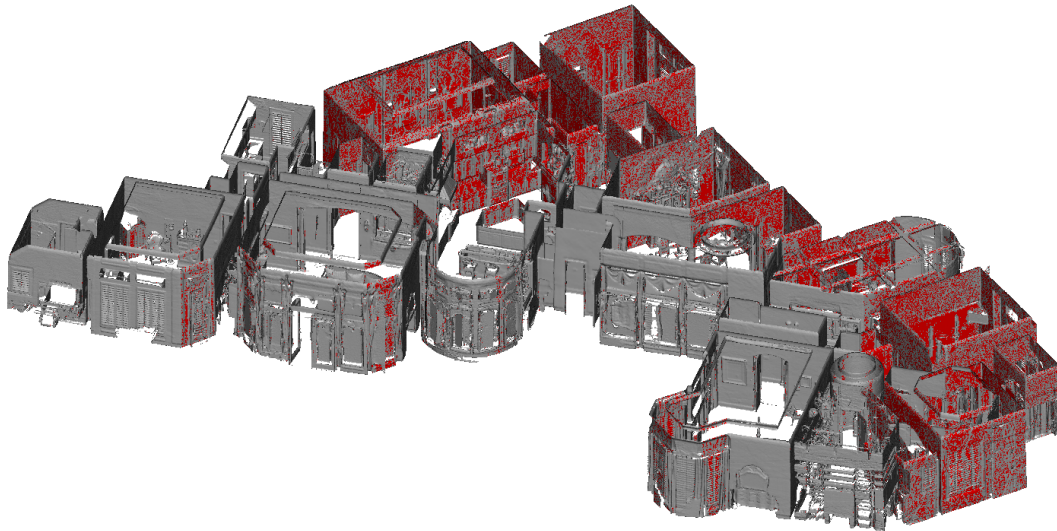
$$|\langle (\vec{n}_i, \vec{z})| \leq 40^\circ \wedge |\langle (\vec{n}_i, \vec{z})| \geq 140^\circ \quad (5.9)$$

and w_i again is a weighting factor being constant for points of a point cloud but corresponding to the respective triangle area for the faces of a triangle mesh. Furthermore, $R(\alpha, \beta)$ denotes a 3×3 rotation matrix determined by two rotation angles α and β around the two horizontal coordinate axes \vec{x} and \vec{y} respectively.

Thus, the aim of Eq. 5.8 is two find the optimal vertical axis \vec{z}^* as a vector



(a) Faces corresponding to the largest peak at 60° in Fig. 5.4 determining the dominant Manhattan World structure.



(b) Faces corresponding to the minor peak at 15° in Fig. 5.4.

Figure 5.5: The vertical faces of the triangle mesh presented in Fig. 5.1 corresponding to the horizontal normal vectors \vec{n}_i^h . The faces corresponding to the two peaks shown in Fig. 5.4 are depicted in red respectively. Source: [244].

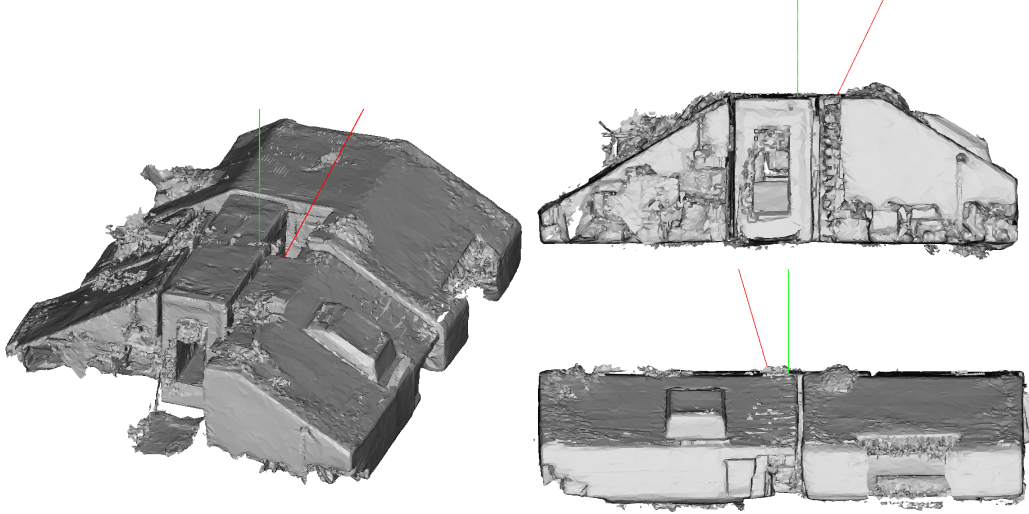


Figure 5.6: Exemplary triangle mesh of a building with partially slanted ceiling (dataset 'Attic' from [243]). The green line visualizes the reference orientation of the vertical axis considered as ground truth while the red line visualizes the vertical axis rotated -25° around the horizontal \vec{x} axis and 15° around the horizontal \vec{y} axis as exemplarily used in Sec. 5.2.2. Source: [244].

$$\vec{z}^* = R(\alpha, \beta)\vec{z} \quad (5.10)$$

in the initially given coordinate system that has a minimal sum of angles to the vertical normals \vec{n}_i^v . This optimal vertical axis \vec{z}^* as well as the initial vertical axis \vec{z} are exemplarily depicted in Fig. 5.6 for a building with slanted ceilings only coarsely aligned with the actual vertical direction.

As it already was the case with Eq. 5.5 in Sec. 5.2.1, Eq. 5.8 is not analytically solvable and solving it numerically is all the more inefficient as this time, a two-dimensional minimization is concerned. Thus, as in the case of determining the rotation angle around the vertical axis in Sec. 5.2.1, we again seek to formulate the problem at hand as the task of searching a maximum peak within a discrete grid representation of the relevant input elements.

The relevant input elements in this case are the three-dimensional vertical normal vectors \vec{n}_i^v . However, the problem at hand is actually two-dimensional as a rotation around the two horizontal axes \vec{x} and \vec{y} by the rotation angles α and β is sufficient for aligning the vertical axis \vec{z} with the optimal vertical direction \vec{z}^* .

In an alternative formulation, this can also be considered as the task of finding the position of the optimal vertical direction \vec{z}^* on the surface of a unit sphere, i.e. within the extended Gaussian image. The orientation of a normal vector with respect to the coordinate system can be expressed via the polar angles azimuth

$$\varphi_i = \arctan \frac{\langle \vec{n}_i, \vec{y} \rangle}{\langle \vec{n}_i, \vec{x} \rangle} \in [-180^\circ, -180^\circ] \quad (5.11)$$

and inclination

$$\theta_i = \arccos \langle \vec{n}_i, \vec{z} \rangle \in [0^\circ, -180^\circ] \quad (5.12)$$

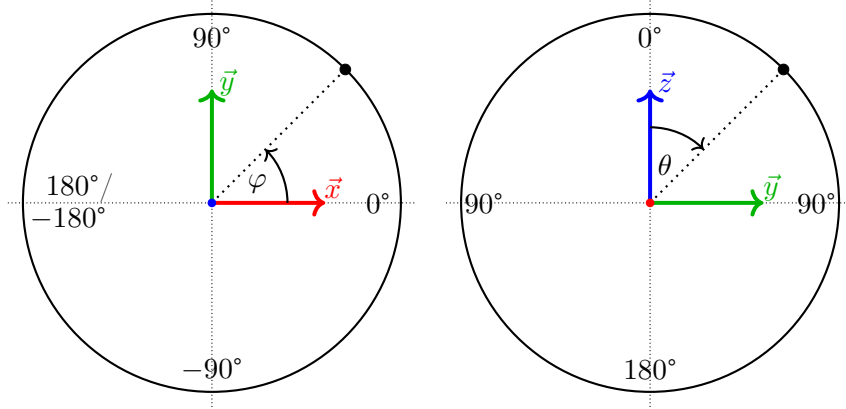


Figure 5.7: Azimuth φ and inclination θ . Source: [244].

indicating the position of a respective normal vector \vec{n}_i on the unit sphere. The definition of azimuth and inclination with respect to the coordinate system is further illustrated in Fig. 5.7. This representation allows to construct a two-dimensional azimuth/inclination grid analogous to the approach presented in Sec. 5.2.1 whose cells are weighted by the summarized weights w_i of the contained normal vectors \vec{n}_i . A suchlike grid of a resolution of 1° extending over the whole unit sphere is depicted in Fig. 5.8 corresponding to the exemplary case presented in Fig. 5.6.

As before in Sec. 5.2.1, this grid over the full range of the sphere surface is transformed to a smaller grid where the weights of cells pertaining to opposing normal vectors get accumulated. This is achieved by

$$\tilde{\varphi}_i = \left| |\varphi_i| - 90^\circ \right| \in [0^\circ, 90^\circ) \quad (5.13)$$

and

$$\tilde{\theta}_i = 90^\circ - |\theta_i - 90^\circ| \in [0^\circ, 90^\circ) \quad (5.14)$$

while restricting the extension of the grid in the dimension of the inclination to the range of $[0^\circ, 40^\circ]$ and thus only considering the vertical normal vectors \vec{n}_i^v . A schematic visualization of this transformation is depicted in Fig. 5.9(a) while Fig. 5.10 shows the resulting two-dimensional azimuth/inclination grid corresponding to the dataset presented in Fig. 5.6.

Subsequently, peaks with cell grid weights above a threshold of 75% of the highest weight value are again clustered like in the case of the one-dimensional grid of Sec. 5.2.1. While doing so however, not only the azimuth discontinuity between 0° and 90° needs to be considered, but also the pole point at 0° inclination where all azimuth values merge to one and the same grid cell.

While in the case of the one-dimensional grid of Sec. 5.2.1, grid cell indices could be directly mapped to angles by multiplication with the grid resolution, here, it is not possible to infer the direction of the optimal vertical axis from grid cell indices as the transformed azimuth values $\tilde{\varphi}$ are ambiguous by multiples of 90° . This ambiguousness did also exist in Sec. 5.2.1. However, it did not affect the correctness of the resulting horizontal alignment as is the case here.

Thus, to be able to deduce correct directions from peaks in the two-dimensional grid, the respective normal vectors \vec{n}_i^v need to be hashed per grid cell. So, the correct direction of the vertical axis can be initialized by a weighted average of all the hashed normal directions weighted

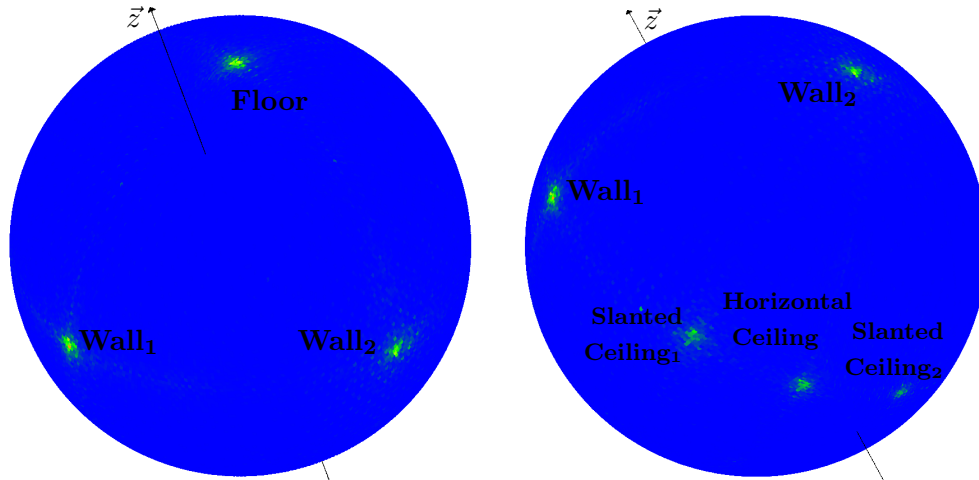


Figure 5.8: Azimuth/inclination grid of 1° resolution over the whole surface of the unit sphere corresponding to Fig. 5.6. The grid cells contain the summarized weights w_i of the contained normal vectors \vec{n}_i at polar angles (φ_i, θ_i) with value colorization ranging from blue for low values over green and yellow to red for large values. Source: [244].

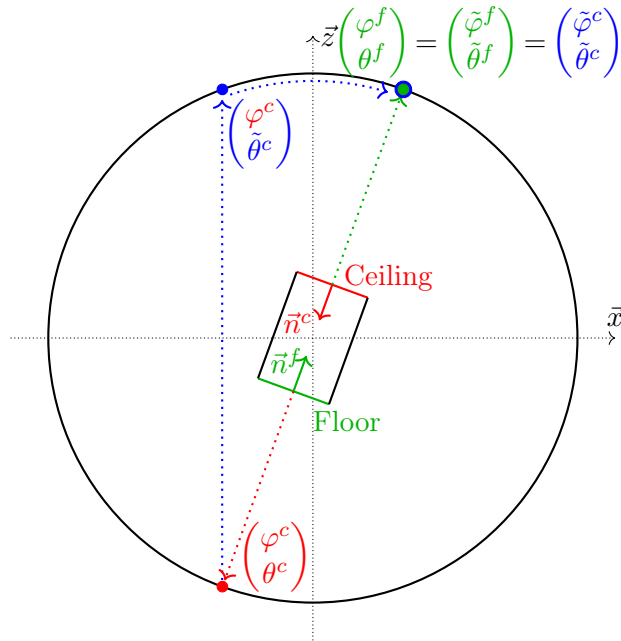
by their respective w_i value of the cluster with the largest summarized weight. In doing so, normal vectors pointing downwards need to be corrected by inverting the direction to point upwards when calculating the weighted average vector. Like in Sec. 5.2.1, the initial result is further refined by a weighted median of all normal vectors within $\pm 5^\circ$ of the coarsely determined resulting vertical axis.

Besides the need to deduce the correct direction from the detected maximum peak grid cells, there is a second reason to hash normal directions per grid cell. As illustrated in Fig. 5.9(b), two normal vectors that are oriented by the same angle around the vertical axis \vec{z} in a way that the axis \vec{z} is the angle bisector between both normals get projected to the same $(\tilde{\varphi}, \tilde{\theta})$ grid cell by Eq. 5.13 and Eq. 5.14. On the one hand, this can distort the weight sums of the individual grid cells that are used for peak detection. On the other hand, the presence of normal vectors with deviating orientations beyond the ambiguity of $\pm 180^\circ$ between opposing surfaces can severely distort the initial determination of the vertical direction from the largest peak in the grid.

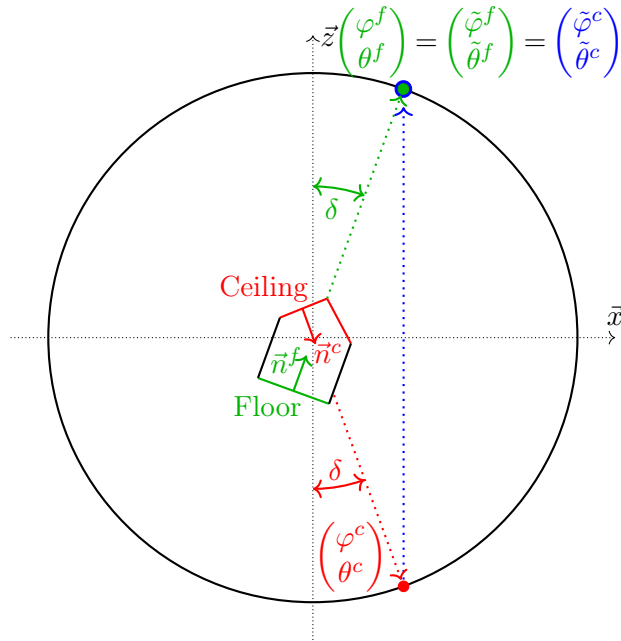
For this reason, a cluster analysis is conducted among the hashed normal vectors per grid cell. In doing so, all the normal vectors in a grid cell are assigned to clusters. A normal vector can be assigned to an existing cluster if its direction coincides within $\pm 2^\circ$ with the average direction of the cluster (with consideration of an ambiguity of $\pm 180^\circ$). Else, the respective normal vector initializes a new cluster. Finally, for each grid cell, only the largest cluster of normals is retained while the others are discarded. The grid cell weights and the hashed normal vectors are adapted accordingly.

5.2.3 Unambiguousness of the Rotation around the Vertical Axis

The alignment of indoor mapping point clouds or triangle meshes along the coordinate axes as described in the preceding sections 5.2.1 and 5.2.2 is ambiguous with respect to a rotation around the vertical axes by multiples of 90° . This is per se not a problem as the aim of the presented approach is to align the indoor mapping data with respect to its Manhattan world



(a) Generally, points corresponding to opposing normal vectors get transformed to the same point.



(b) In case the vertical axis \vec{z} is the angle bisector between the directions of two normal vectors (same angle δ to \vec{z} axis), these get transformed to the same point even if they are not opposed. This needs to be dealt with by means of a cluster analysis per $(\tilde{\varphi}, \tilde{\theta})$ grid cell.

Figure 5.9: Transformation of (φ, θ) positions on the whole unit sphere to $(\tilde{\varphi}, \tilde{\theta})$ positions on one eighth of the unit sphere by Eq. 5.13 and Eq. 5.14. Source: [244].

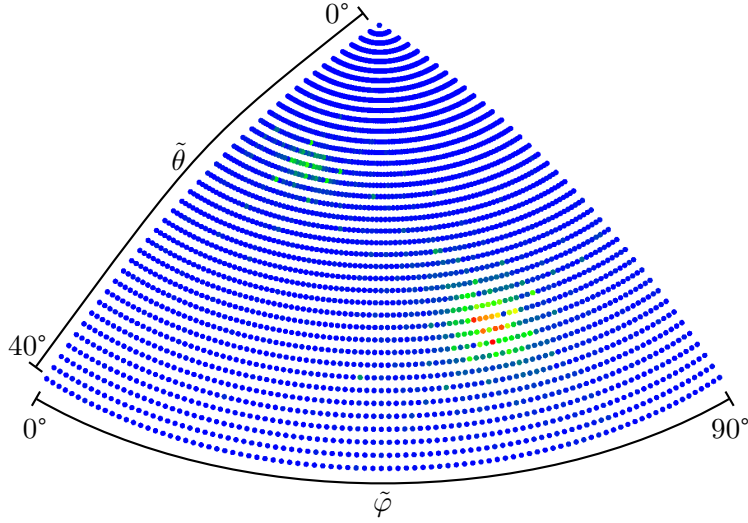


Figure 5.10: Transformed azimuth/inclination grid of 1° resolution corresponding to Fig. 5.6. The grid cells contain the summarized weights w_i of the contained vertical normal vectors \vec{n}_i^v at polar angles $(\tilde{\varphi}_i, \tilde{\theta}_i)$ with value colorization ranging from blue for low values over green and yellow to red for large values. The larger peak corresponds to the floor and the horizontal part of the ceiling while the minor peak corresponds to one of the slanted ceiling surfaces. Source: [244].

structure which inherently implies this ambiguity with respect to four possible rotations around the vertical axis, i.e. all four possible result poses are equally valid with respect to the stated aim.

However, in some situations, it can be desirable to derive an unambiguous pose of the indoor mapping data. For instance, this can be the case when multiple indoor mapping results of the same building environment are to be aligned by the proposed method. These multiple datasets of the same building can e.g. be obtained by different indoor mapping systems or be acquired at different times in the context of change detection.

For this reason, a simple method for resolving the ambiguity in the rotation around the vertical axis by reproducibly choosing one of the four possible horizontal orientations is presented. The proposed method is a straight-forward solution that does not require any semantic interpretation of the indoor mapping data or any elaborate analysis. It can however fail in cases of highly symmetric building layouts with respect to its four inherent Manhattan world directions. Furthermore, it presupposes that two datasets to be aligned unambiguously by this method cover approximately the same section of an indoor environment. If this is not the case, an approach that incorporates semantic knowledge of the represented indoor environment would be more promising.

Here, the unambiguousness between the four possible horizontal orientations is resolved by first aligning the one of the two possible horizontal Manhattan world directions with the chosen reference axis \vec{x} that corresponds to a larger extent of the bounding box of the respective dataset in this horizontal direction, i.e. the longer horizontal edges of the bounding box should be parallel to the \vec{x} axis. This is quite straight forward but can fail in cases where the bounding box is nearly quadratic.

The ambiguity is now reduced to a rotation of 180° . To resolve this, the weighted count of indoor mapping geometries in both proximal 10% sections of the bounding box in \vec{x} direction are considered and the rotation for which the proximal 10% section of the bounding box pointing

towards the positive \vec{x} axis has the higher weight sum is chosen. In this context, the indoor mapping geometries are again weighted by a constant in the case of points of point clouds and by triangle area in the case of triangle mesh faces. This approach fails, when the amount of mapped indoor structures in both proximal sections of the bounding box along the \vec{x} axis is about equal.

5.2.4 Evaluation Method

Quantitatively evaluating the proposed method is fortunately quite straight forward as ground truth data can be easily obtained. If an indoor mapping dataset is not already correctly aligned with the coordinate system axes in the sense of the aim of this study, it can be aligned manually without great effort. A thus aligned dataset can then be rotated to an arbitrary pose within the defined range applicable for the presented method. For this a 3×3 ground truth rotation matrix $R^{GT}(\alpha, \beta, \gamma)$ is created, determined by the rotation angles $\alpha, \beta \in [-30^\circ, 30^\circ]$ around the horizontal axes \vec{x} and \vec{y} respectively and an arbitrary rotation $\gamma \in [-180^\circ, 180^\circ]$ around the vertical axis \vec{z} . For creating R^{GT} , the rotation γ around the vertical axis is applied first and then successively β and α around their respective horizontal axis.

Finally, the method presented in Sec. 5.2.1 and Sec. 5.2.2 is applied to the rotated dataset which should return the rotated dataset back to its aligned state. The resulting 3×3 rotation matrix R^{Test} is constituted by

$$R^{Test} = R_{horizontal}^{Test} R_{vertical}^{Test} \quad (5.15)$$

where first $R_{vertical}^{Test}$ is determined by aligning the rotated dataset vertically with the vertical axis as described in Sec. 5.2.2 and then subsequently, the rotation $R_{horizontal}^{Test}$ around the vertical axis is determined as described in Sec. 5.2.1.

As an evaluation metric, the angular difference δ_v between the vector of the ground truth axis \vec{z} and the resulting vector

$$\vec{z}^{Test} = R^{Test} R^{GT} \vec{z} \quad (5.16)$$

is determined by

$$\delta_v = |\angle(\vec{z}^{Test}, \vec{z})| \quad (5.17)$$

as well as the analogous angular difference δ_h for the horizontal axis \vec{x} . In case of the horizontal deviation δ_h , the ambiguity of valid rotations around the vertical axis by multiples of 90° needs to be considered. To this aim,

$$\delta_h = \begin{cases} \delta_h - 90^\circ & \delta_h \geq 45^\circ \\ \delta_h & else \end{cases} \quad (5.18)$$

is iteratively applied until $\delta_h < 45^\circ$.

The proposed evaluation metrics δ_v and δ_h can be determined for multiple randomly chosen rotations within the mentioned ranges of $[-30^\circ, 30^\circ]$ for the horizontal axes and $[-180^\circ, 180^\circ]$ for the vertical axis in sufficient quantity to allow for a statistical analysis.

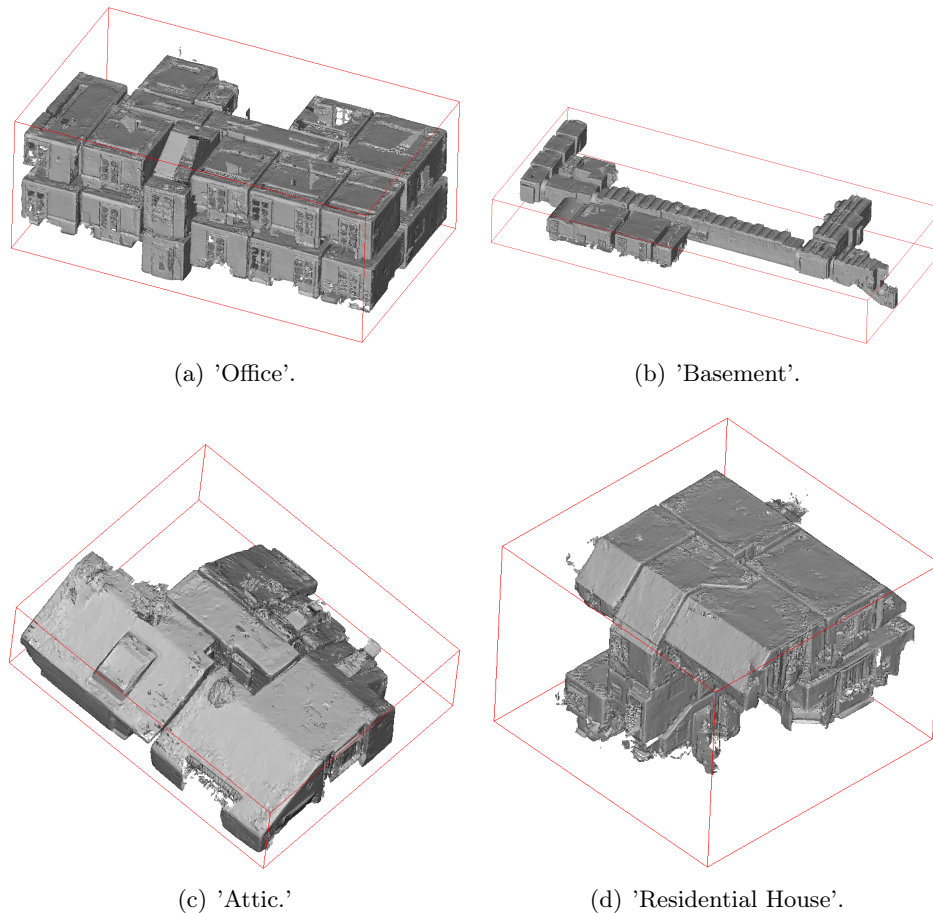


Figure 5.11: The Microsoft HoloLens triangle meshes from Fig. 3.8 used for evaluation. The red box indicates the aligned ground truth pose. Source: [244].

5.3 Results

In order to quantitatively evaluate the approach presented in Sec. 5.2.1 and Sec. 5.2.2, the evaluation procedure proposed in Sec. 5.2.4 was applied to a range of different indoor mapping datasets. Firstly, the four triangle meshes presented in Fig. 3.8 were used for evaluation. These triangle meshes are depicted in Fig. 5.11 along with 3D bounding boxes indicating their respective ground truth pose.

The alignment with the coordinate axes of the HoloLens triangle meshes as presented in Chap. 3 was found to be inaccurate. Actually, in Chap. 3, the presented datasets have been automatically aligned with the coordinate axes by means of an early, inferior version of the approach presented in this chapter. To enable a reasonable evaluation of the proposed approach on these triangle meshes, ground truth poses were determined by manually aligning the datasets with the coordinate axes. The newly aligned datasets along with the implementation of the proposed approach and the evaluation procedure is made publicly available to allow for reproducibility of the presented evaluation results²⁶.

All four represented indoor environments show a clearly defined Manhattan world structure.

²⁶<https://github.com/huepat/im-posenorm> (Last visited on 15/07/2021)

While the dataset 'Office' has mostly horizontal ceiling surfaces with the exception of the stairwell, the datasets 'Attic' and 'Residential House' have slanted ceiling surfaces. The dataset 'Basement' on the other hand shows a range of different barrel-shaped ceilings.

Furthermore, the six indoor mapping point clouds of the ISPRS Indoor Modelling Benchmark dataset presented in [281, 284] were used for evaluation purposes. These point clouds as visualized in Fig. 5.12 were acquired by means of different indoor mapping systems with a broad variety of sensor characteristics regarding accuracy and noise. Furthermore, the represented indoor environments are characterized by varying amounts of clutter.

While the other five datasets mostly adhere to the Manhattan World assumption, the dataset 'Case Study 6' has a high amount of horizontally curved wall surfaces and rooms oriented diagonally with respect to the dominant Manhattan World structure defined by three rooms. Furthermore, the point cloud includes a part of the surrounding outdoor terrain with uneven topography and vegetation. As the dataset 'Case Study 6' is quite challenging with respect to the aim of this work, it is depicted in more detail in Fig. 5.13.

The point clouds of the ISPRS benchmark dataset as they are published are already aligned with the coordinate axes in accordance with the aim of this work. Thus, the poses of the point clouds could directly be used as ground truth poses without any manual adjustment. Contrary to triangle meshes however, point clouds do not intrinsically provide normal vectors per point. This is also the case with the point clouds of the ISPRS Indoor Modelling Benchmark. Normal vectors were thus determined for the points after subsampling the point clouds with a resolution of 2 cm using CloudCompare 2.10-alpha [195].

Lastly, some triangle meshes from the Matterport3D dataset [94] were also considered. Matterport3D includes 90 triangle meshes of various kinds of indoor environments acquired with the trolley-mounted Matterport indoor mapping system consisting of multiple RGBD cameras. Among the represented indoor environments are some for which the proposed alignment approach is not applicable, as they are not subject to any clearly identifiable Manhattan world structure. Many others do have a clearly identifiable Manhattan world structure but are to a large extent comparable to general building layouts already covered by the HoloLens triangle meshes or ISPRS point clouds used in the scope of this evaluation.

Thus, 14 triangle meshes from the Matterport3D dataset were selected that were deemed particularly interesting and challenging in the context of this work. This, for instance, comprises triangle meshes representing indoor environments that contain more than one underlying Manhattan World system like the one already presented in Fig. 5.1. In these cases, the presented alignment method is supposed to align the triangle mesh with the most dominant of the Manhattan World structures at hand being supported by the largest fraction of geometries. The 14 selected triangle meshes from the Matterport3D dataset are depicted in Fig. 5.14.

As with the ISPRS benchmark point clouds, the poses of the triangle meshes as they are published were again treated as ground truth alignments without any manual adjustments. To which extent this decision is justified will be discussed in the subsequent Sec. 5.4.

The different datasets used in the scope of this evaluation are listed in Tab. 5.1 along with the respective number of points or triangles comprising them and the respective evaluation results. For conducting the evaluation, the evaluation procedure described in Sec. 5.2.4 was applied to the individual datasets. In doing so, each dataset was rotated 50 times while each time, the respective rotation consists of a randomly determined rotation angle $\gamma \in [-180^\circ, 180^\circ)$ around the vertical axis and two random rotations $\alpha, \beta \in [-30^\circ, 30^\circ]$ around the respective horizontal coordinate axes.

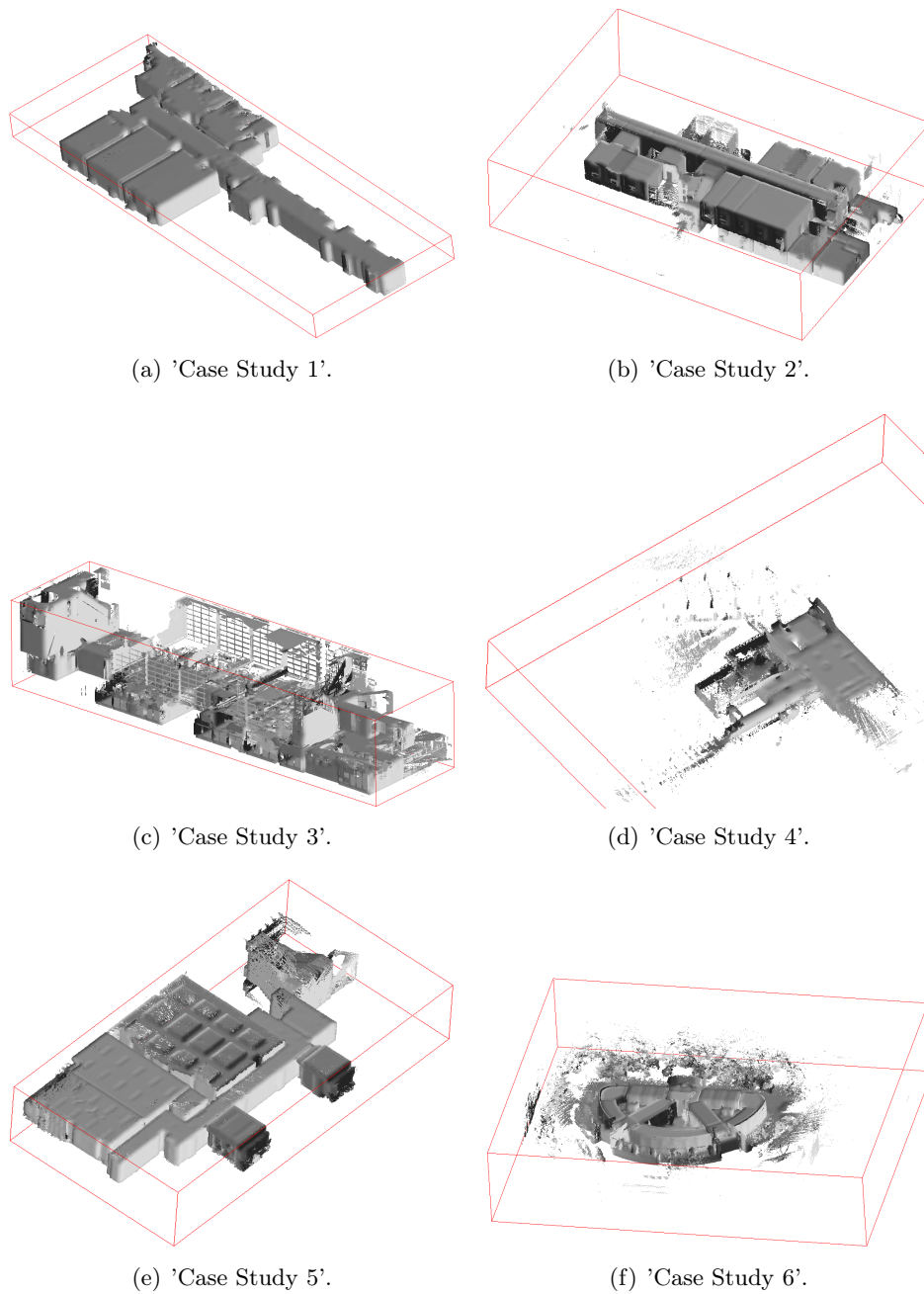
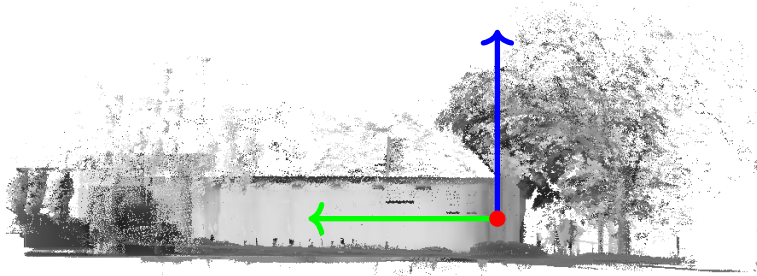
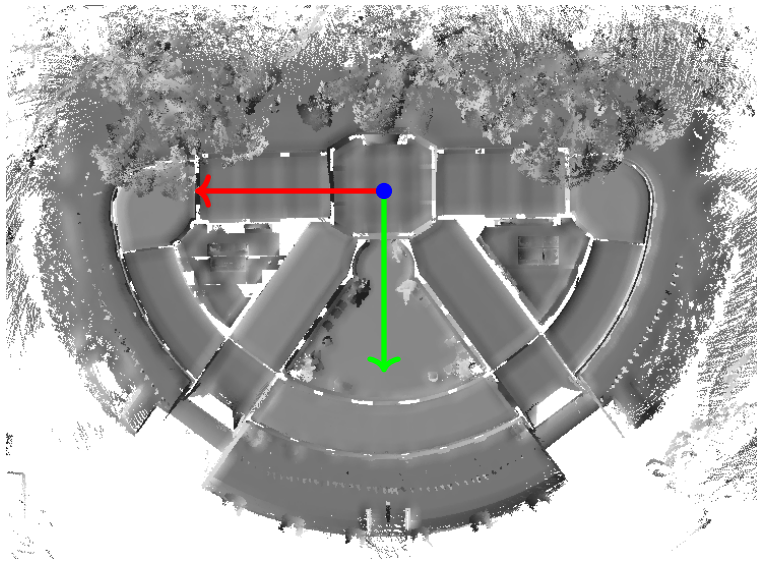


Figure 5.12: The point clouds of the ISPRS Indoor Modelling Benchmark dataset [281, 284] used for evaluation. The red box indicates the aligned ground truth pose. Source: [244].



(a) Side view.



(b) Top down view.

Figure 5.13: Detailed visualization of the dataset 'Case Study 6' from the ISPRS Indoor Modelling Benchmark dataset [284] also depicted in Fig. 5.12(f). The vertical axis is visualized in blue while the two horizontal axes aligned with the dominant Manhattan World structure of the building are depicted in red and green. Source: [244].

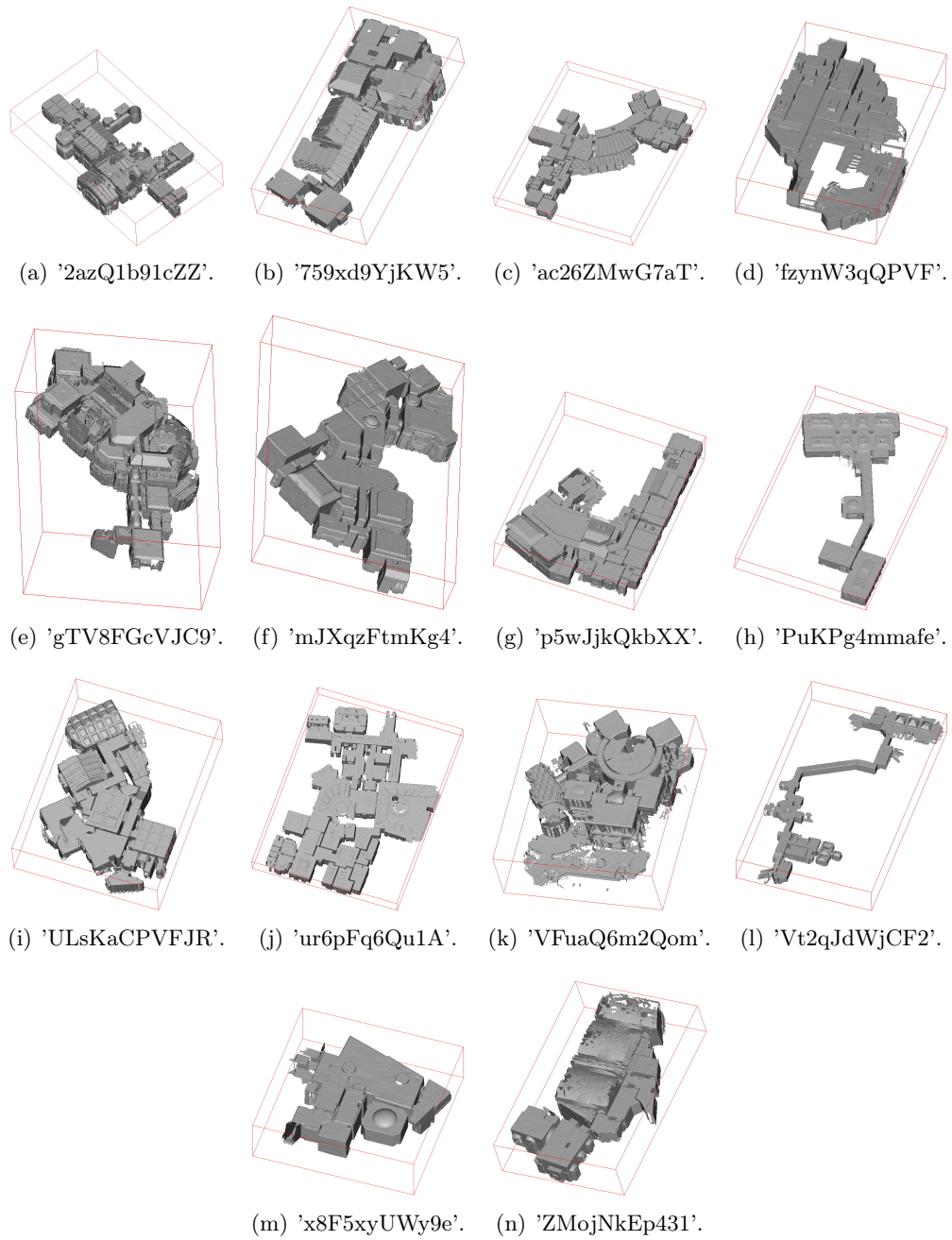


Figure 5.14: The triangle meshes of the Matterport3D dataset [94] used for evaluation. The red box indicates the aligned ground truth pose. Source: [244].

Table 5.1: Evaluation results for the datasets presented in Figures 5.11, 5.12 and 5.14. The presented values represent 50 randomly chosen orientations per dataset within the range of $[-180^\circ, 180^\circ]$ for rotations around the vertical axis and $[-30^\circ, 30^\circ]$ for rotations around the horizontal axes. The reported numbers of points for the point clouds of the ISPRS Indoor Modelling Benchmark refer to point clouds downsampled to a resolution of 2 cm as used in this evaluation. The values marked in red are discussed in more detail in Sec. 5.4. Source: [244].

Source	Type	Dataset	Number of Points/Triangles	Mean δ_v [°]	Std.Dev. δ_v [°]	Mean δ_h [°]	Std.Dev. δ_h [°]	Mean Time [s]	Std.Dev. Time [s]
HoloLens [243]	Triangle Mesh	Office	958,820	0.28	0.25	0.33	0.07	0.68	0.10
		Basement	695,041	0.45	0.06	0.10	0.08	0.50	0.04
		Attic	147,146	3.54	23.86	0.26	0.42	0.13	0.02
		Residential House	252,820	0.16	0.05	0.71	0.42	0.30	0.04
ISPRS Indoor Modelling Benchmark [281, 284]	Point Cloud	Case Study 1	5,014,452	0.01	0.05	0.03	0.16	4.41	0.19
		Case Study 2	8,202,319	0.01	0.02	0.01	0.13	7.40	0.26
		Case Study 3	5,906,718	0.02	0.01	0.04	0.17	5.68	0.29
		Case Study 4	4,846,736	0.01	0.26	0.03	0.44	4.19	0.27
		Case Study 5	4,409,794	0.02	0.07	0.02	0.06	3.96	0.23
		Case Study 6	11,760,325	0.02	0.02	0.06	0.77	8.65	0.53
Matterport3D [94]	Triangle Mesh	2azQ1b91cZZ	9,549,830	0.03	0.02	0.44	0.06	8.24	0.39
		759xd9YjKW5	6,208,440	0.05	0.01	0.18	0.05	5.48	0.35
		ac26ZMwG7aT	10,811,581	0.05	0.09	0.52	0.06	9.84	0.49
		fzynW3qQPVF	9,105,979	0.09	0.02	0.05	0.06	10.75	0.60
		gTV8FGcVJC9	14,436,867	0.05	0.05	0.11	0.07	12.29	0.96
		mJXqzFtmKg4	8,237,802	0.07	0.33	2.73	14.29	6.90	0.54
		p5wJjkQkbXX	10,678,539	0.07	0.02	0.40	0.03	10.35	0.68
		PuKPg4mmafe	1,968,102	0.05	0.01	15.28	20.07	1.83	0.11
		ULsKaCPVFJR	6,612,194	0.05	0.01	44.41	0.04	5.51	0.47
		ur6pFq6Qu1A	9,277,187	0.02	0.01	12.85	0.05	9.42	0.42
		VFuaQ6m2Qom	9,453,891	0.03	0.02	0.13	0.06	8.53	0.37
		Vt2qJdWjCF2	6,429,106	0.10	0.01	0.05	0.09	6.40	0.38
		x8F5xyUWy9e	2,862,858	0.07	0.01	0.21	0.08	2.66	0.16
		ZMojNkEp431	4,690,777	0.06	0.05	0.18	0.08	4.31	0.27

For each of the 50 random input rotations, the alignment procedure described in Sec. 5.2.1 and Sec. 5.2.2 was applied and the resulting vertical and horizontal angular deviations δ_v and δ_h as defined in Sec. 5.2.4 were determined. Tab. 5.1 lists mean values and standard deviations for these evaluation metrics aggregated over all 50 samples per dataset. Furthermore, mean values and standard deviations for the processing time are given as well. The stated values refer to a system with a i7-8550U CPU with 24 GB RAM and do not include data import and export. The implementation is CPU-parallelized.

As can be seen in Tab. 5.1, the resulting averaged vertical and horizontal angular deviations are largely below 1° with the corresponding standard deviations being in a similar range. Some outliers marked in red will be discussed in further detail in the subsequent Sec. 5.4.

5.4 Discussion

Taking a closer look at the evaluation results presented in Tab. 5.1, the overall quite low values for the horizontal and vertical angular deviations δ_h and δ_v with overall equally low standard deviations indicate that the proposed alignment method works overall quite well for a large range of different indoor mapping point clouds and triangle meshes with randomly varying input rotations within the defined bounds. The consistently larger δ_v and δ_h values for the triangle meshes acquired with the Microsoft HoloLens may be attributable to them being less accurate and more affected by noise. Triangles pertaining to an actually smooth planar room surface show a considerable variation in normal vector direction. However, the reported δ_v and δ_h values

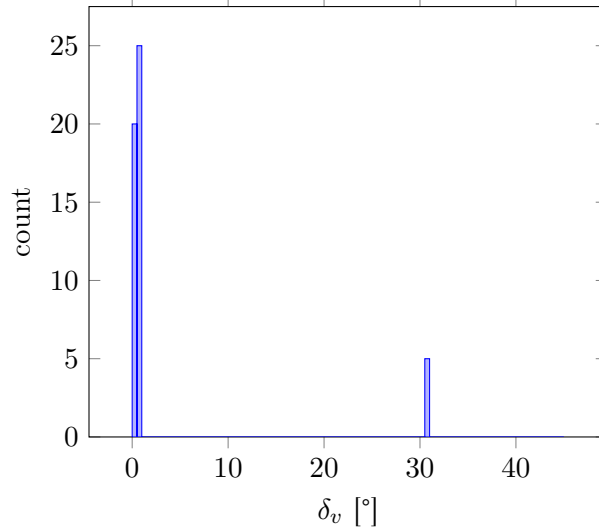


Figure 5.15: Histogram of the 50 δ_v values resulting in the mean value of $3.54^\circ \pm 23.86^\circ$ presented in Tab. 5.1 for the triangle mesh 'Attic' depicted in Fig. 5.11(c). Without the 5 outliers around 31° , mean δ_v results in $0.50^\circ \pm 0.13^\circ$. Source: [244].

for these datasets are still mostly well below 1° .

Some datasets however show significantly higher averaged values for δ_v or δ_h , sometimes with the corresponding standard variation being significantly raised as well. These outliers are marked red in Tab. 5.1 and will be discussed in more detail in the following paragraphs. To analyze these cases, a closer look at the distribution of the individual 50 deviations constituting the respective mean value and standard deviation will be taken.

In the case of the HoloLens triangle mesh 'Attic' for instance, the histogram of δ_v values depicted in Fig. 5.15 indicates, that the heightened mean and standard deviation values for the angular deviation in the vertical alignment are not caused by a large variability in the resulting vertical alignment. The vertical orientations resulting from the evaluated alignment method rather fluctuate between two clearly defined states, one being the correct vertical orientation according to the ground truth pose at around 0° angular deviation δ_v of the vertical axis supported by 45 of the 50 measurements. The other state is a vertical orientation with an angular deviation of about 30° occurring in the remaining five measurements. As visualized by the red box in Fig. 5.16, this corresponds to an alignment where the vertical axis is oriented orthogonally to one of the slanted ceiling surfaces.

This is the only case where the vertical alignment did not work satisfyingly in all 50 samples for all the datasets used in the evaluation. It can be suspected, that the misalignments occurring sporadically on this dataset can be ascribed to the noisy surfaces of the HoloLens triangle meshes. The triangles comprising the large horizontal floor surface for instance differ significantly in the direction of their normal vectors. Thus, only a fraction of the triangles comprising the floor actually corresponds to the proper vertical direction with respect to the applied resolution of 1° . Depending on the input rotation, a peak caused by a slanted ceiling surface with a not insignificant area in comparison to horizontal surfaces like in the case of the dataset at hand representing only the attic story may thus induce a larger peak and consequently a misalignment. In cases like this, applying an angular resolution of more than 1° may be more suited to prevent suchlike misalignments.

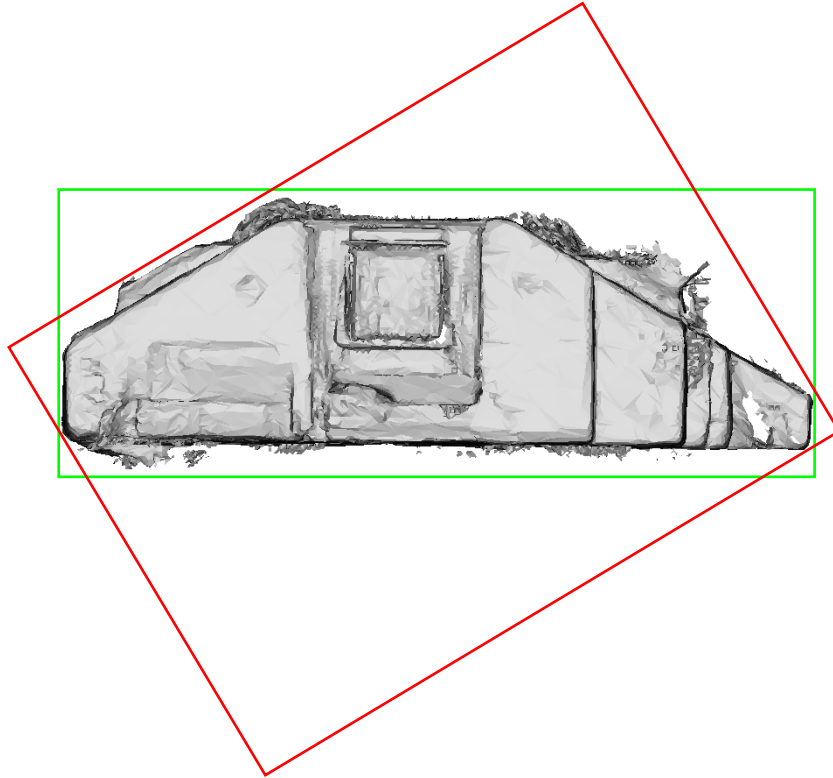


Figure 5.16: Resulting vertical alignments of the triangle mesh 'Attic' from Fig. 5.11(c) for the two peaks in the histogram of δ_v values depicted in Fig. 5.15. The green bounding box corresponds to the peak at $\delta_v \approx 0^\circ$ while the red bounding box corresponds to the minor peak at $\delta_v \approx 30^\circ$. Source: [244].

Besides the discussed outlier in the vertical alignment, some outliers in the horizontal alignment do exist. The Matterport3D datasets 'mJXqzFtmKg4' and 'PuKPg4mmafe' for instance show heightened average δ_h values along with high standard deviations. The histograms showing the distribution of all 50 δ_h values are again depicted in Fig. 5.17 and Fig. 5.18 respectively. Like in the case before, it is apparent that the alignment results fluctuate between two states depending on the input rotation for both cases while each time, one peak at 0° corresponds to the correct horizontal alignment according to the respective ground truth pose. As can be seen in Fig. 5.19 and Fig. 5.20, the respective second peak corresponds in both cases to a valid second Manhattan world structure present in the respective indoor environment.

In the case of the dataset 'mJXqzFtmKg4', this seems immediately plausible, as both Manhattan world structures present in the indoor environment are supported by a comparable amount of geometries, as was already demonstrated in Fig. 5.4 and Fig. 5.5. Thus, different input rotations may result in slightly different discretizations within the grid of 1° resolution, sometimes favoring one and sometimes the other Manhattan world structure as having the largest peak of summarized geometry weights.

In the case of the dataset 'PuKPg4mmafe' however, the two Manhattan world structures present in the indoor environment apparently do not seem to be supported by an approximately equal fraction of geometries. Rather, the upper right section in Fig. 5.20 constituting the one Manhattan world structure seems to be far smaller than the section on the lower left constituting

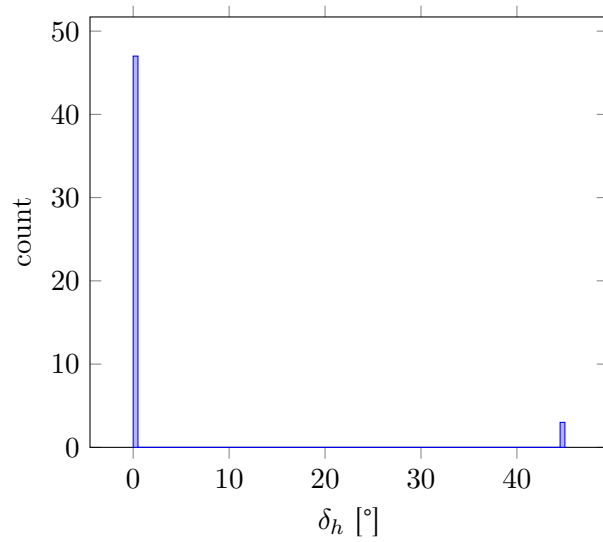


Figure 5.17: Histogram of the 50 δ_h values resulting in the mean value of $2.73^\circ \pm 14.29^\circ$ presented in Tab. 5.1 for the triangle mesh 'mJXqzFtmKg4' depicted in Fig. 5.14(f). Source: [244].

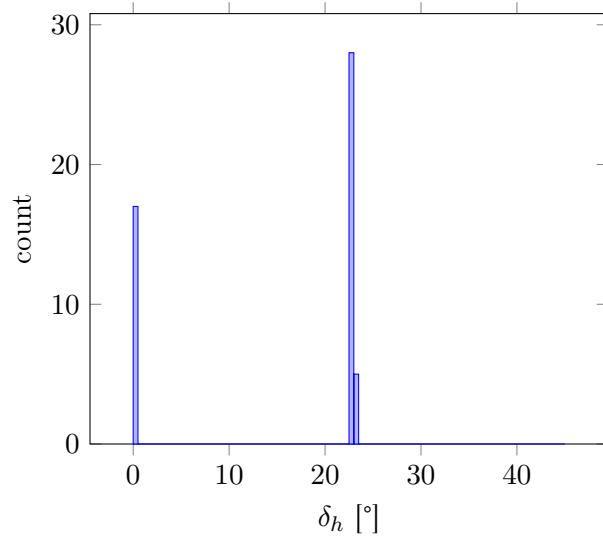


Figure 5.18: Histogram of the 50 δ_h values resulting in the mean value of $15.28^\circ \pm 20.07^\circ$ presented in Tab. 5.1 for the triangle mesh 'PuKPg4mmafe' depicted in Fig. 5.14(h). Source: [244].

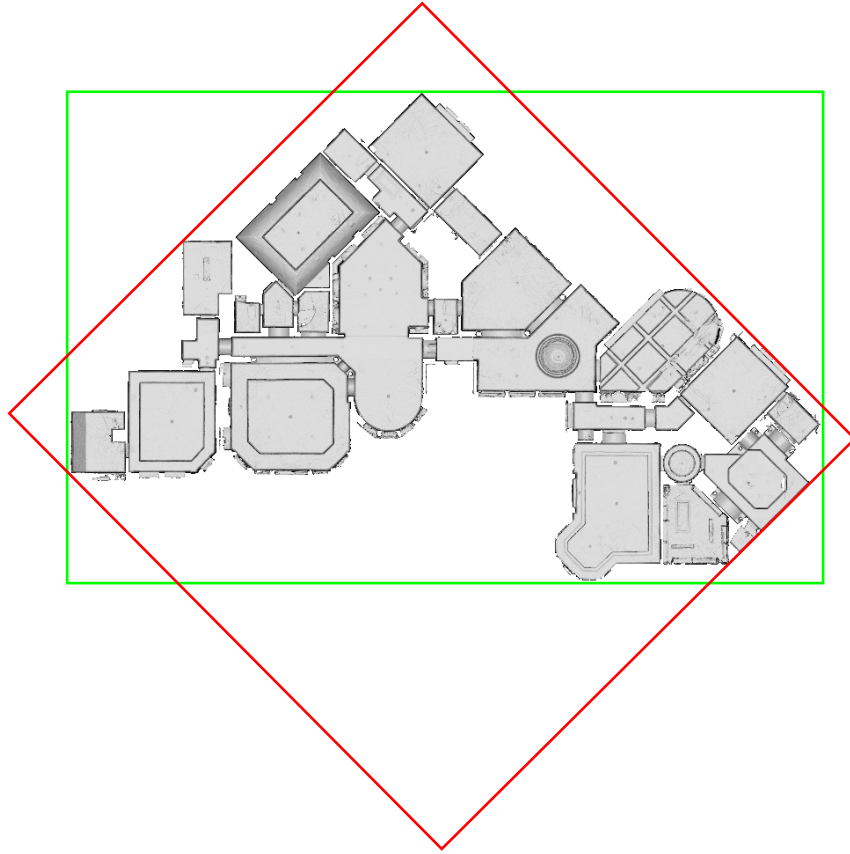


Figure 5.19: Resulting horizontal alignments of the triangle mesh 'mJXqzFtmKg4' from Fig. 5.17 for the two peaks in the histogram of δ_h values depicted in Fig. 5.17. The green bounding box corresponds to the peak at $\delta_h \approx 0^\circ$ while the red bounding box corresponds to the minor peak at $\delta_h \approx 45^\circ$. Source: [244].

the other Manhattan world structure. In this case, the ground truth pose of the triangle mesh as published in [94] is aligned with the apparently smaller Manhattan world structure. It is thus not surprising that in the evaluation, a majority of measurements results in high δ_h deviations as the evaluated alignment method favors the larger Manhattan world structure. However, it is surprising that a not insignificant fraction of 17 of the 50 randomly chosen input rotations results in a horizontal alignment along the apparently significantly smaller Manhattan world structure.

This situation may be explainable by taking a closer look at the walls constituting the respective Manhattan world structures. As can be seen in Fig. 5.21, the smaller Manhattan world section on the right hand side consists of wall surfaces that are generally smooth and completely covered with geometries. The larger section on the left however has a large fraction of open wall surface where there are no geometries due to the walls there actually being openings or glass surfaces that cannot be captured by the Matterport system used for the acquisition of this dataset. Furthermore, large parts of the actually represented wall surfaces are covered with curtains or other structures resulting in inhomogeneous normal vector directions. In consideration of this, it seems plausible that the actual support for both Manhattan World structures present in the building could be approximately equal and the applied alignment method could thus be prone to fluctuate between both Manhattan world systems with varying input rotations.

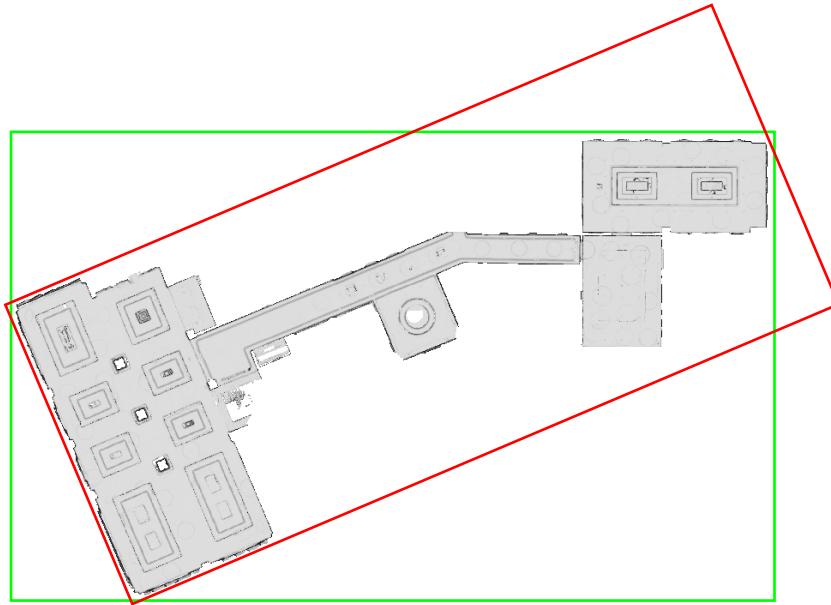


Figure 5.20: Resulting horizontal alignments of the triangle mesh 'PuKPg4mmafe' from Fig. 5.14(h) for the two peaks in the histogram of δ_h values depicted in Fig. 5.18. The green bounding box corresponds to the peak at $\delta_v \approx 0^\circ$ while the red bounding box corresponds to the peak at $\delta_v \approx 23^\circ$. Source: [244].

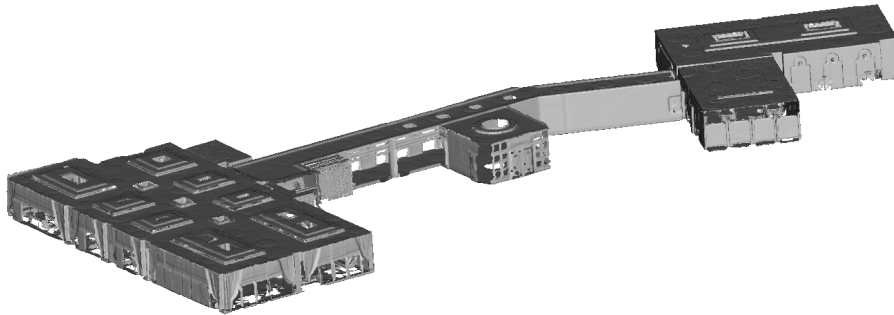


Figure 5.21: Detailed view of the triangle mesh 'PuKPg4mmafe' from the Matterport3D dataset [94] also depicted in Fig. 5.14(h) and Fig. 5.20. Note that in the case of the larger part of the building structure determining the Manhattan world system visualized by the red bounding box in Fig. 5.20, large parts of the wall surfaces are missing as wall openings or constituted by curtains or other structures with inhomogeneous normal direction. The smaller part of the building structure on the right hand side which determines the Manhattan world system visualized by the green bounding box in Fig. 5.20 however has largely closed, smooth wall surfaces. Source: [244].

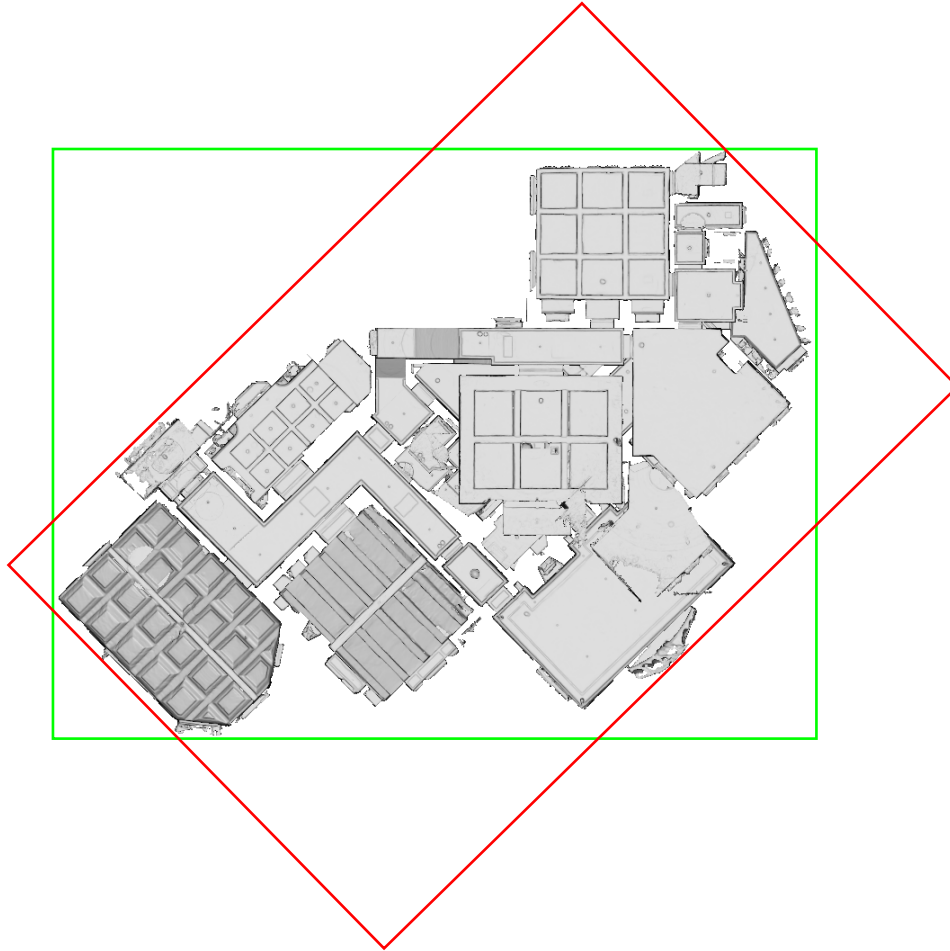


Figure 5.22: The green bounding box represents the horizontal alignment of the triangle mesh 'ULsKaCPVFJR' from Fig. 5.14(i) as it is published in [94] and used as ground truth pose for the evaluation results presented in Tab. 5.1. The red bounding box on the other hand represents the horizontal alignment resulting from our presented approach. Source: [244].

Besides these both cases discussed so far, there are two further datasets with high average horizontal angular alignment deviations in the evaluation results reported in Tab. 5.1. These are the triangle meshes 'ULsKaCPVFJR' and 'ur6pFq6Qu1A' which are also part of the Matterport3D dataset. Unlike the cases discussed before, these however only show heightened mean values for δ_h while the respective standard deviations are low in a range comparable to the other Matterport3D triangle meshes where the evaluated alignment method proofed to be consistently successful.

This suggests that the proposed method consistently results in the same horizontal orientation for all 50 input rotations for both datasets. The respective resulting alignment however deviates from the assumed ground truth pose in the rotation around the vertical axis. This is further illustrated by Fig. 5.22 and Fig. 5.23 where it is easily discernible that the depicted buildings again respectively contain two Manhattan world structures and that the evaluated alignment method consistently chooses the respective other Manhattan world structure that does not coincide with the ground truth pose.

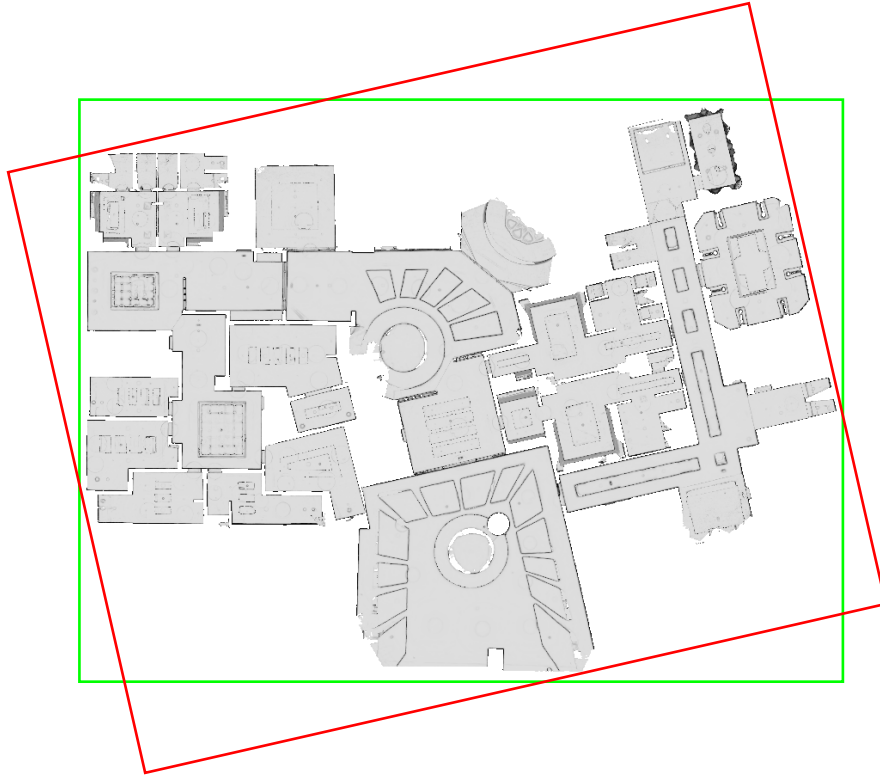


Figure 5.23: The green bounding box represents the horizontal alignment of the triangle mesh 'ur6pFq6Qu1A' from Fig. 5.14(j) as it is published in [94] and used as ground truth pose for the evaluation results presented in Tab. 5.1. The red bounding box on the other hand represents the horizontal alignment resulting from our presented approach. Source: [244].

Arguably, it is disputable which of the two Manhattan world structures respectively present in the datasets is the 'correct' one as again in these two examples, both seem to encompass more or less the same fraction of the represented building environment and it is not readily discernable which is the dominant one. Nevertheless, the proposed method proves to find a reasonable alignment with high accuracy in almost all cases with the only exception being the vertical alignment of the HoloLens triangle mesh 'Attic'. In all other cases where the resulting pose deviates from the ground truth pose, the resulting alignment is still reasonable in the sense that it corresponds to another Manhattan world structure inherent in the respective dataset that is readily identifiable by a human observer even if it may differ from the given ground truth pose corresponding to another alternative Manhattan world structure.

Besides aligning an indoor mapping dataset with the dominant Manhattan world structure supported by the largest fraction of geometries, the proposed method can easily be augmented to identify all major Manhattan World structures along with the respective sets of associated geometries. Among other possible fields of application that will be briefly discussed in the following Sec. 5.5, this allows for providing multiple possible alternatives for alignment to the user to choose from in cases where multiple major Manhattan World structures are present in the dataset at hand and it is not readily apparent which among these to use for alignment.

5.5 Conclusion and Outlook

This chapter presented a novel method for the automated pose normalization of indoor mapping data like point clouds and triangle meshes. The aim of the proposed method is to align an indoor mapping point cloud or triangle mesh along the coordinate axes in a way that a chosen vertical axis points upwards with respect to the represented building structure, i.e. the chosen vertical axis is expected to be orthogonal to horizontal floor and ceiling surfaces. Furthermore, a rotation around this vertical axis is to be determined in a way that aligns the two horizontal coordinate axes with the main direction of the dominant Manhattan world structure of the respective building geometry. In case multiple Manhattan world systems are present in the data, the dominant structure supported by the largest fraction of geometries should determine the horizontal alignment.

For both fundamental steps of the proposed method - determining the correct orientation of the vertical axis and subsequently the correct horizontal rotation around this resulting vertical axis - a theoretical solution is presented. As the proposed formulation of the problem at hand cannot be solved efficiently, an efficient approximate solution for a practical implementation is presented. This encompasses discretizing the input data into a grid with fixed resolution while transforming it in a way that enables the problem to be solved by determining the largest peak within this grid of fixed size and finally refining the resulting coarse result by resorting to the original input data in the vicinity of the detected peak. A CPU-parallelized implementation of the proposed method along with the code for the automated evaluation procedure is made available to the public.

The proposed method is quantitatively evaluated on a range of different indoor mapping point clouds and triangle meshes that are publicly available. The presented results show, that the approach is overall able to consistently produce correct poses for the considered datasets for different input rotations with high accuracy. Furthermore, cases where high deviations with respect to the given ground truth pose occur are presented and discussed.

Concerning potential for future research, it has already been mentioned that the proposed method offers the possibility to not only identify the dominant Manhattan world structure along with the associated geometries in an indoor mapping dataset, but also to detect multiple Manhattan world structures that are sufficiently supported by geometries. Besides enabling to present multiple reasonable alternatives for alignment to choose from, this could potentially also be used in the context of automated indoor reconstruction (see Chap. 3). In particular, knowing the major Manhattan world structures and their associated geometries could be beneficial for abstracting and idealizing indoor surfaces, i.e. reconstructing suitable surfaces as planes that perfectly conform the Manhattan world assumption. In addition, automatically detecting the involved Manhattan world structures in a building may also be of interest in the context of automatically analyzing the architectural structure of buildings [9, 493].

Furthermore, the presented methodology could possibly also be used in the context of indoor environments and indoor mapping in general (see Chap. 2). Here, identifying Manhattan World structures during the mapping process (or in post processing if the individual indoor mapping geometries have associated timestamps to reconstruct the sequence of acquisition) could potentially be used to correct or reduce drift effects by applying the assumption that building structures that apparently seem to deviate only slightly from an ideal Manhattan world system are to be corrected according to the Manhattan world assumption [447, 510, 545, 546, 661, 345].

Chapter 6

Synopsis

In the preceding chapters, the different topics in which contributions were presented within this work were discussed individually. Here, a synoptic discussion encompassing all topics relevant to this work is presented.

First, Sec. 6.1 starts with a concise summary of the content discussed in the previous chapters. Next, Sec. 6.2 presents further discussions on these topics. In this context, special focus is placed on outlining the potential for future research opportunities. Finally, Sec. 6.3 closes this work with concluding remarks.

6.1 Summary

Within the previous chapters, first, the topic of indoor mapping was discussed in Chap. 2. Indoor mapping encompasses the process of the geometric acquisition of indoor building structures by means of adequate sensor systems. While stationary systems such as terrestrial laser scanners (TLS) can be used to this aim, mobile sensor systems such as backpack-mounted or hand-carried laser scanners or range cameras are more favorable in this context due to the geometric complexity of typical indoor environments, requiring a multitude of TLS positions. Besides dedicated indoor mapping systems, specifically tailored towards this task, mobile augmented reality (AR) systems are equipped with suitable sensors as well, in order to acquire the structure of their surrounding environment. Thus, it is conceivable to use indoor augmented reality systems for the task of indoor mapping. In this context, Chap. 2 presented an evaluation of the AR headset Microsoft HoloLens regarding its aptitude to the task of indoor mapping. The evaluation study first focused individually on the range camera and the tracking system of the device, these two being the two main components determining the quality of an indoor mapping system. Both components were evaluated against respective ground truth data and finally, the indoor mapping results in the form of triangle meshes acquired with the HoloLens were evaluated against ground truth data as well. The presented results demonstrate, that the Microsoft HoloLens is overall quite suited to the task of indoor mapping and can readily be deployed to this aim. However, it should be taken into account, that drift effects can occur over large distances. While loop closure errors are detected by the tracking system of the device, they are however not corrected globally in the acquired triangle meshes, as the actual intended usage of the HoloLens as an AR device only requires local consistency of the acquired structures with the physical environment.

Subsequently, Chap. 3 discussed the topic of automated indoor reconstruction, i.e. the extraction of semantically meaningful and geometrically completed and refined indoor building models from unstructured geometries as acquired by the indoor mapping systems discussed in the preceding

chapter. After giving an overview of the diversity of different approaches to address the challenging problem of indoor reconstruction, a novel approach applicable to triangle meshes as acquired by the Microsoft HoloLens was presented. The proposed approach relies on a voxel representation of the input data. Indoor voxel models are reconstructed semantically as well as geometrically by means of rule-based procedures. Besides classifying voxels in the semantic classes 'Wall', 'Wall Opening', 'Floor', 'Ceiling', 'Empty Interior' and 'Interior Object', the voxel space is partitioned into rooms as well. The main advantage of the proposed approach lies in its genericity regarding the structure of indoor environments to be reconstructed. While many indoor reconstruction approaches rely on simplifying assumptions such as the Manhattan world assumption, the proposed approach only requires wall surfaces to be vertically straight. Horizontally, walls can be of arbitrary shape while ceiling and floor surfaces can be arbitrarily shaped as well. Furthermore, the input geometries are required to have consistently oriented normals allowing a distinction between inside and outside. Besides the indoor reconstruction approach itself, adequate evaluation methodology was presented along with the quantitative evaluation results achieved on four different datasets which are made available to the public.

In order to use building models in the context of indoor fused reality scenarios, the augmented reality device needs to be able to determine its initial position within the respective building represented by the model. This indoor localization process has been discussed in Chap. 4. After first addressing the different methods available to this aim, a rather straight-forward, marker-based approach for indoor localization using the Microsoft HoloLens was presented. While a marker-based localization method may be less favored due to the necessity to equip the respective indoor environment with artificial markers, it is nonetheless applicable in many scenarios where unambiguously recognizable, planar objects are present anyway, e.g. in the case of door plates. The presented quantitative evaluation shows, that the Microsoft HoloLens can indeed be deployed for the in-situ visualization of room-scale building model data.

Lastly, Chap. 5 presented a novel method for the pose normalization of indoor mapping datasets. The coordinate system, in which building geometries are acquired is often determined by the initial pose of the respective indoor mapping system when starting the acquisition process. While the orientation of the horizontal coordinate axes is oftentimes arbitrary, the vertical axis of the coordinate system can in most cases be expected to be roughly pointing towards the actual upwards direction due to typical usage poses of hand-held or head-worn indoor mapping devices. The proposed method aims to automatically align indoor mapping datasets with the coordinate axes in a way, that the rough initial leveling is optimally refined and the rotation of the horizontal axes around the thus refined vertical axis is determined so as to optimally align them with the dominant Manhattan world system underlying the structure of the respective indoor environment. A suchlike alignment of indoor mapping data with the coordinate axes is a necessary preprocessing step for indoor reconstruction procedures that rely on the Manhattan world assumption. Indeed, this assumption not only implies that a respective building structure conforms with it, but also that the building structure is aligned with the coordinate axes accordingly. Even in the case of indoor reconstruction procedures that do not rely on the Manhattan world structure, a pose normalization process can nonetheless make sense when grid structures are used as is the case with the indoor reconstruction approach proposed in Chap. 3. The robustness and accuracy of the proposed method was demonstrated by means of quantitative evaluation.

6.2 Discussion and Outlook

As already hinted on in the previous section, all topics covered by this work are interconnected by their usefulness for the realization of the concept of indoor fused reality as presented in Sec. 1.1 and 1.2.5. Indoor mapping systems discussed in Chap. 2 can be used to acquire raw geometric data which is the input to automated indoor reconstruction procedures as discussed in Chap. 3. The resulting building models can in turn be used in the context of indoor fused reality applications to visualize building-related information in-situ after determining the global pose of a mobile augmented reality device within the respective building environment by means of indoor localization as discussed in Chap. 4. Finally, the pose normalization methodology presented in Chap. 5 can be deployed as a preprocessing step for indoor reconstruction approaches besides other potential application fields such as for instance the co-registration of multiple indoor mapping datasets. In the following, the interrelations between these main topics will be further discussed. In this context, special emphasis will be put on the potentials for future research beyond the state which has been realized within the scope of this work.

The feasibility of a continuous workflow from indoor mapping to indoor reconstruction could be clearly demonstrated by using indoor mapping results of the Microsoft HoloLens evaluated in this regard in Chap. 2 as input data for the indoor reconstruction approach presented in Chap. 3. Regarding indoor reconstruction and indoor localization however, only the individual feasibility of both aspects was demonstrated. Using the actual results of the proposed indoor reconstruction method to localize an AR device within the represented building and visualize digital building information spatially registered to this model remains subject of future research.

In this context, an important aspect is the conversion of the indoor building model in voxel representation resulting from the proposed indoor reconstruction method to a more compact and established, vector-based data representation. Due to the generic nature of the presented indoor reconstruction approach where few restricting assumptions about the shape of room surfaces are made, this is not a straight-forward task. While planar room surfaces could be detected among the room surfaces in the form of voxel segments and modeled accordingly as planes, voxel segments with a more complex shape could for instance be represented as parametrized free-form shapes or triangle meshes. Assembling the resulting surfaces to water-tight rooms however, is a complex task requiring further research efforts.

However, it must be stressed, that representing indoor building environments by voxel grids is not only a suitable means for realizing indoor reconstruction tasks, but can in itself be regarded as a promising form of data representation. Combined with pyramid approaches and efficient data structures such as octrees to mitigate the disadvantageous scaling of memory consumption and processing time with increasing size of a given grid, voxel representations do indeed hold great potential for all kinds of analysis tasks. This is especially the case for scenarios where not only given geometries, but also the empty space between, within or around them is the object of analysis. In this regard, it can be conjectured, whether voxel grids could not be a suitable way to represent building environments in the context of indoor localization tasks as well.

Generally, the field of indoor localization still holds potential for future research. While many different approaches were proposed to this aim, their feasibility is most often demonstrated only for small examples lacking the ambiguity typically inherent in large buildings. A promising strategy to tackle ambiguous environments in the context of indoor localization is to use image sequences instead of single query images for pose regression [118, 3]. This allows the mobile unit which is to be localized to explore a sufficiently large part of its environment in order to resolve these ambiguities. A rigorous study of ambiguities and self-similar structures in indoor environments and their effects on indoor localization tasks is however still missing and would certainly be a worthwhile endeavor in the

context of indoor localization research. Another interesting aspect in this regard is to investigate, how ambiguous localization results can be utilized in indoor fused reality approaches. For instance, some useful information can already be displayed, when it can be determined, where in a given storey of a multi-storey building the mobile AR device currently is, even if there still remains an ambiguity about which storey it is in. For displaying more building-related information, the ambiguity concerning the storey needs to be resolved and to this aim, the user could be guided to a part of the building which is unambiguously identifiable, e.g. the stairwell. In this context, simple marker-based localization approaches like the one discussed in Chap. 4 could potentially be integrated into more complex indoor localization schemes relying on images or indoor mapping data. Thus, door plates or signboards with storey numbers in stairwells could be utilized to resolve localization ambiguities due to self-similar building geometry.

Another potentially worthwhile field of future research is to not only apply indoor reconstruction methods to indoor mapping data, but to directly integrate them into the indoor mapping process. That is to say, instead reconstructing building models in a single step from already acquired indoor mapping geometries covering the complete indoor environment to be reconstructed, indoor reconstruction could also be applied dynamically, on-the-fly to the indoor mapping data as it is acquired. Thus, the indoor building model would grow successively with the advancement of the indoor mapping process. Realizing a suchlike dynamic indoor reconstruction is by no means trivial. However, it could for instance prove to be advantageous to the task of indoor localization as well.

Furthermore, in case an augmented reality device like the Microsoft HoloLens is used for indoor mapping, user interaction could be integrated in the dynamic indoor reconstruction process. This could already be realized by just visualizing the current state of the dynamically derived indoor model in-situ overlaying the physical building structures. For instance, when using the HoloLens for indoor mapping, the already captured triangle meshes are visualized, allowing the user to directly assess where the building is already sufficiently captured and which parts need more attention in order to achieve a complete indoor mapping result. Similarly, visualizing the current state of the envisioned on-the-fly indoor reconstruction process would allow the user to specifically focus on acquiring more observations of those parts of the building environment that are not yet sufficiently reconstructed, e.g. due to too sparse acquisition. Furthermore, it is also conceivable for the user to directly intervene into the reconstruction process by correcting errors in the arising building model, e.g. via gestures or voice commands.

In this context, enriching the reconstructed building models with further information which is valuable for indoor fused reality applications but can usually not be reconstructed automatically is also possible. Suchlike information, e.g. concerning the course of pipes and cables within the walls or semantic information not discernible from indoor mapping data, is typically modeled manually using building models derived from manual or automated reconstruction processes as basis. Here, augmented reality devices like the Microsoft HoloLens could again facilitate and improve this process substantially, by allowing the user to conduct the process directly in-situ while capturing the underlying building geometry in place. Investigating different ways of user interaction, e.g. modeling cables by tapping on the walls in specific locations, or automatically integrating measurements from external sensors such as electronic cable finders could be an interesting field of research as well while at the same time offering the means to easily create rich building models for existing buildings that can be deployed in indoor fused reality applications.

Concerning indoor reconstruction (either as a separate postprocessing step applied to indoor mapping data or directly integrated into the indoor mapping process), further potential for future research exists as well as already laid out in Sec. 3.5. In this context, applying deep learning techniques to this task is particularly interesting. While a number of indoor reconstruction approaches using deep learning have been proposed recently [104, 288, 183, 653], these typically apply deep learning

techniques on 2D raster data orthoprojected from the three-dimensional indoor mapping data or use them to classify points in indoor mapping point clouds and then applying indoor reconstruction on the segmented point clouds. Voxel representations like the one used in the indoor reconstruction approach presented in Chap. 3 could be potentially well-suited for applying deep learning methods to indoor reconstruction. While initially being frequently applied in the context of 3D deep learning [375, 637, 516, 563], voxel representations have meanwhile lost in favor for applications in this field [473, 474, 161, 334, 551]. Nonetheless, they may still be a reasonable choice in the context of indoor reconstruction. Here, comparably coarse voxel resolutions are sufficient for understanding the layout of indoor building environments and empty space holds semantic meaning, e.g. when reconstructing occluded room surfaces or partitioning the indoor space in rooms. 3D deep learning on voxel grids in the context of indoor environments has for instance successfully been applied in the field of semantic scene reconstruction [129, 124].

The method for pose normalization of indoor mapping datasets as a preprocessing step for indoor reconstruction procedures presented in Chap. 5 holds potential for further research as well. Besides its usage for pose normalization, the proposed methodology could potentially also be applied in the context of indoor reconstruction itself. Here, it could be deployed to refine extracted wall surfaces to perfect Manhattan world structures if they are found to constitute a local Manhattan world system within the bounds of a certain accuracy threshold determined by the quality of the respective indoor mapping data and the building structure itself. Furthermore, the proposed methodology could potentially also be applied to reduce drift in indoor mapping. Lastly, applications in the context of indoor localization and the co-registration of indoor mapping datasets are conceivable as well.

6.3 Conclusion

This work started with the illustrative example of an imaginary mechanic performing repair tasks in a large building environment while being guided and instructed by an headset-based augmented reality application which draws information from a digital building model. Within the preceding chapters, different technological requirements necessary in order to realize this scenario of indoor fused reality were introduced and discussed. Among these requirements, the need for digital models for existing building structures where suchlike AR applications are to be deployed can be satisfied by acquiring indoor geometry data by means of indoor mapping systems and subsequently converting them to digital building models in a process of automated indoor reconstruction. Furthermore, spatially overlaying physical building structures with virtual information from a corresponding building model can be achieved by a process of indoor localization which determines the global pose of a mobile augmented reality device within the building.

As already discussed in the previous section, all these topics investigated and discussed in this work still hold ample potential for future research. However, the potential of the indoor fused reality scenario enabled by the presented components is worthy of further attention as well. The envisioned combination of digital building models and augmented reality techniques holds great potential for driving future research in both fields and demonstrating their respective worth for building-related application fields such as facility management, energy efficiency, construction, architecture and navigation. In this context, such applications focusing on building environments have the potential of becoming a prime example and demonstration case for the usefulness and economic worth of augmented reality technology applied to geospatial data - a concept defined and discussed in this work as 'fused reality'. Likewise, disciplines related to building environments can benefit tremendously from the prospects, augmented reality technology can offer them.

List of Figures

1.1	Example for the conceptual difference between augmented reality and fused reality	6
1.2	Schematic overview of the concept of indoor fused reality	8
1.3	Schematic overview of the structure of this work	10
2.1	Schematic overview of the evaluation of the Microsoft HoloLens	12
2.2	Overlay of images recorded by the different camera sensors of the Microsoft HoloLens	16
2.3	Pixels of the HoloLens range camera that actually contain values	17
2.4	Range image and depth image	18
2.5	Ground truth data captured for a three-dimensional scene	19
2.6	Rigid body of reflective sphere markers affixed to the HoloLens device	20
2.7	Change in distance measurement of the HoloLens range sensor over time	23
2.8	Noise of the distance measurements of the HoloLens ToF sensor	25
2.9	Depth images for a three-dimensional scene	26
2.10	Noise of the range measurements for a three-dimensional scene	26
2.11	Accuracy of HoloLens range measurements for a three-dimensional scene	26
2.12	Accuracy of the HoloLens triangle mesh of the three-dimensional scene	27
2.13	HoloLens trajectory in a room equipped with a motion capture system	27
2.14	Velocity and RPE values over the course of the trajectory	28
2.15	Closed trajectory of a total length of 287 m	29
2.16	Triangle mesh of an indoor office environment captured by the HoloLens	30
2.17	Accuracy of a HoloLens triangle mesh (registration with fixed scale)	30
2.18	Accuracy of a HoloLens triangle mesh with furniture (registration with estimated scale)	30
2.19	Precision of a HoloLens triangle mesh without furniture	31
2.20	Accuracy of a HoloLens triangle mesh without furniture	31
2.21	Accuracy of a HoloLens triangle mesh without furniture (room-wise registration) . . .	31
2.22	Point cloud of an indoor office environment captured by the HoloLens range sensor . .	32
2.23	Accuracy of a HoloLens point cloud (registration with estimated scale)	32
2.24	Accuracy of a HoloLens point cloud (registration with estimated scale)	32
2.25	Cross section from Fig. 2.11(a)	34
2.26	HoloLens triangle meshes of an indoor environment with and without furniture	35
2.27	Deviations of a triangle mesh of a furniture-less indoor environment	36
3.1	Overview of the processing workflow of the proposed reconstruction method	43
3.2	Overview of the ceiling detection process	44
3.3	Overview of the ceiling and floor reconstruction process	45
3.4	Overview of the voxel classification process	47
3.5	Overview of the wall surface refinement process	49
3.6	Overview of the room segmentation process	50
3.7	Overview of the 2D room segmentation procedure	54

3.8	The Microsoft HoloLens triangle meshes used for quantitative evaluation	57
3.9	Horizontal slice of the dataset 'Residential House' for varying rotation angles around the vertical axis	63
3.10	Horizontal slice of the dataset 'Residential House' for varying voxel resolutions	68
3.11	Dependency of processing time and memory consumption on voxel size	69
3.12	Triangle mesh and results for a Matterport3D dataset	70
3.13	Triangle mesh and results for a Matterport3D dataset	71
4.1	Augmentation of a room with a corresponding model using a Microsoft HoloLens	80
4.2	Room model	81
4.3	The ArUco marker used for localization	82
4.4	The real room augmented by the room model from Fig. 4.2	82
4.5	The ArUco marker used for evaluation	83
4.6	An ArUco marker augmented from different viewing angles	85
4.7	A checkerboard pattern augmented by its virtual counterpart	86
4.8	Placement of seven different points of view that were used during the evaluation	87
5.1	Exemplary triangle mesh of a building with multiple Manhattan world systems	94
5.2	The normal vectors \vec{n}_i of the triangle mesh shown in Fig. 5.1	95
5.3	Visualization of a one-dimensional 360° grid corresponding to Fig. 5.1	97
5.4	Visualization of a one-dimensional 90° grid corresponding to Fig. 5.1	98
5.5	The vertical faces of the triangle mesh presented in Fig. 5.1	99
5.6	Exemplary triangle mesh of a building with partially slanted ceiling	100
5.7	Azimuth φ and inclination θ	101
5.8	Azimuth/inclination grid of 1° resolution over the surface of the unit sphere	102
5.9	Transformation of (φ, θ) positions to $(\tilde{\varphi}, \tilde{\theta})$ positions on one eighth of the unit sphere	103
5.10	Transformed azimuth/inclination grid of 1° resolution corresponding to Fig. 5.6	104
5.11	The Microsoft HoloLens triangle meshes from Fig. 3.8 used for evaluation	106
5.12	The point clouds of the ISPRS Indoor Modelling Benchmark dataset	108
5.13	Detailed visualization of the dataset 'Case Study 6'	109
5.14	The triangle meshes of the Matterport3D dataset [94] used for evaluation	110
5.15	Histogram of 50 δ_v values for the triangle mesh 'Attic'	112
5.16	Vertical alignments of the triangle mesh 'Attic'	113
5.17	Histogram of 50 δ_h values for the triangle mesh 'mJXqzFtmKg4'	114
5.18	Histogram of 50 δ_h values for the triangle mesh 'PuKPg4mmafe'	114
5.19	Horizontal alignments of the triangle mesh 'mJXqzFtmKg4'	115
5.20	Horizontal alignments of the triangle mesh 'PuKPg4mmafe'	116
5.21	Detailed view of the triangle mesh 'PuKPg4mmafe' from the Matterport3D dataset [94]	116
5.22	Horizontal alignment of the triangle mesh 'ULsKaCPVFJR'	117
5.23	Horizontal alignment of the triangle mesh 'ur6pFq6Qu1A'	118

List of Tables

2.1	HoloLens camera sensors and their characteristics	16
2.2	Noise of the distance measurements of the HoloLens ToF sensor	24
2.3	Evaluation of a three-dimensional scene captured by the HoloLens ToF sensor	24
3.1	Color scheme for voxel classes	42
3.2	Size and class distribution of the four provided benchmark datasets	58
3.3	Evaluation of the dataset 'Office' for varying rotations around the up-axis	59
3.4	Evaluation of the dataset 'Attic' for varying rotations around the up-axis	60
3.5	Evaluation of the dataset 'Basement' for varying rotations around the up-axis	61
3.6	Evaluation of the dataset 'Residential House' for varying rotations around the up-axis	62
3.7	Evaluation of the dataset 'Office' for varying voxel resolutions	64
3.8	Evaluation of the dataset 'Attic' for varying voxel resolutions	65
3.9	Evaluation of the dataset 'Basement' for varying voxel resolutions	66
3.10	Evaluation of the dataset 'Residential House' for varying voxel resolutions	67
4.1	Overlay error between the real and the virtual evaluation marker	87
5.1	Evaluation results for the datasets presented in Figures 5.11, 5.12 and 5.14	111

Bibliography

- [1] Acharya, D., Khoshelham, K., and Winter, S. (2019a). BIM-PoseNet: Indoor Camera Localisation Using a 3D Indoor Model and Deep Learning from Synthetic Images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 150:245–258.
- [2] Acharya, D., Ramezani, M., Khoshelham, K., and Winter, S. (2019b). BIM-Tracker: A Model-Based Visual Tracking Approach for Indoor Localisation Using a 3D Building Model. *ISPRS Journal of Photogrammetry and Remote Sensing*, 150:157–171.
- [3] Acharya, D., Singha Roy, S., Khoshelham, K., and Winter, S. (2020). A Recurrent Deep Network for Estimating the Pose of Real Indoor Images from Synthetic Image Sequences. *Sensors*, 20(19):5492.
- [4] Adan, A. and Huber, D. (2011). 3D Reconstruction of Interior Wall Surfaces under Occlusion and Clutter. In *International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, pages 275–281.
- [5] Adán, A., Quintana, B., and Prieto, S. A. (2019). Autonomous Mobile Scanning Systems for the Digitization of Buildings: A Review. *Remote Sensing*, 11(3):306.
- [6] Adán, A., Quintana, B., Prieto, S. A., and Bosché, F. (2018). Scan-to-BIM for ‘Secondary’ Building Components. *Advanced Engineering Informatics*, 37:119–138.
- [7] Adler, S., Schmitt, S., Wolter, K., and Kyas, M. (2015). A Survey of Experimental Evaluation in Indoor Localization Research. In *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–10.
- [8] Adouane, K., Stouffs, R., Janssen, P., and Domer, B. (2019). A Model-Based Approach to Convert a Building BIM-IFC Data Set Model into CityGML. *Journal of Spatial Science*, 65(2):257–280.
- [9] Ahmed, S., Liwicki, M., Weber, M., and Dengel, A. (2011). Improved Automatic Analysis of Architectural Floor Plans. In *International Conference on Document Analysis and Recognition*, pages 864–869.
- [10] Ahmed, S., Liwicki, M., Weber, M., and Dengel, A. (2012). Automatic Room Detection and Room Labeling from Architectural Floor Plans. In *10th IAPR International Workshop on Document Analysis Systems*, pages 339–343.
- [11] Akbarieh, A., Jayasinghe, L. B., Waldmann, D., and Teferle, F. N. (2020). BIM-Based End-of-Lifecycle Decision Making and Digital Deconstruction: Literature Review . *Sustainability*, 12(7):2670.

- [12] Akçayır, M. and Akçayır, G. (2017). Advantages and Challenges Associated with Augmented Reality for Education: A Systematic Review of the Literature. *Educational Research Review*, 20:1–11.
- [13] Akinade, O. O., Oyedele, L. O., Omoteso, K., Ajayi, S. O., Bilal, M., Owolabi, H. A., Alaka, H. A., Ayris, L., and Looney, J. H. (2017). BIM-Based Deconstruction Tool: Towards Essential Functionalities. *International Journal of Sustainable Built Environment*, 6(1):260–271.
- [14] Akponeware, A. O. and Adamu, Z. A. (2017). Clash Detection or Clash Avoidance? An Investigation into Coordination Problems in 3D BIM. *Buildings*, 7(3):75.
- [15] Al-Adhami, M., Rooble, S., Wu, S., Osuna-Yevenes, C., Ruby-Lewis, V., Greatrix, M., Cartagena, Y., and Talebi, S. (2020). An Automated Approach to Digitise Railway Bridges. In *37th International Symposium on Automation and Robotics in Construction (ISARC 2020)*, pages 962–968.
- [16] Aleksandrov, M., Zlatanova, S., Kimmel, L., Barton, J., and Gorte, B. (2019). Voxel-Based Visibility Analysis for Safety Assessment of Urban Environments. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-4/W8:11–17.
- [17] Aleksy, M., Troost, M., Scheinhardt, F., and Zan, G. T. (2018). Utilizing HoloLens to Support Industrial Service Processes. In *IEEE 32nd International Conference on Advanced Information Networking and Applications*, pages 143–148.
- [18] Alizadeh Naeini, A., Ahmad, A., Sheikholeslami, M. M., Claudio, P., and Sohn, G. (2020). An Unsupervised Registration of 3D Point Clouds to 2D CAD Model: A Case Study of Floor Plan. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-2-2020:9–13.
- [19] Alizadehsalehi, S., Hadavi, A., and Huang, J. C. (2020). From BIM to Extended Reality in AEC Industry. *Automation in Construction*, 116:103254.
- [20] Alkhawaja, F., Jaradat, M., and Romdhane, L. (2019). Techniques of Indoor Positioning Systems (IPS): A Survey. In *Advances in Science and Engineering Technology International Conferences (ASET)*, pages 1–8.
- [21] Alzantot, M. and Youssef, M. (2012). CrowdInside: Automatic Construction of Indoor Floorplans. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems (SIGSPATIAL)*, pages 99–108.
- [22] Ambruş, R., Claiçi, S., and Wendt, A. (2017). Automatic Room Segmentation From Unstructured 3-D Data of Indoor Environments. *IEEE Robotics and Automation Letters*, 2(2):749–756.
- [23] Anagnostopoulos, I., Belsky, M., and Brilakis, I. (2016). Object Boundaries and Room Detection in As-Is BIM Models from Point Cloud Data. In *Proceedings of the 16th International Conference on Computing in Civil and Building Engineering*, pages 1–7.
- [24] Anagnostopoulos, I., Brilakis, I., and Vela, P. A. (2015). A Review on Methods for Generating As-built Building Information Models. In *Proceedings of the 32nd CIB W78 Conference on Construction IT*, pages 1–10.

- [25] Anastasiou, D., Avgeri, T., Iliodromitis, A., Pagounis, V., and Tsakiri, M. (2020). 3D Virtual Models for an Early Education Serious Gaming Application. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIV-4/W1-2020:3–9.
- [26] Andriamamonjy, A., Saelens, D., and Klein, R. (2018). An Automated IFC-Based Workflow for Building Energy Performance Simulation with Modelica. *Automation in Construction*, 91:166–181.
- [27] Andriasyan, M., Moyano, J., Nieto-Julián, J. E., and Antón, D. (2020). From Point Cloud Data to Building Information Modelling: An Automatic Parametric Workflow for Heritage. *Remote Sensing*, 12(7):1094.
- [28] Angulo-Fornos, R. and Castellano-Román, M. (2020). HBIM as Support of Preventive Conservation Actions in Heritage Architecture. Experience of the Renaissance Quadrant Façade of the Cathedral of Seville. *Applied Sciences*, 10(7):2428.
- [29] Arici, F., Yildirim, P., Şeyma Caliklar, and Yilmaz, R. M. (2019). Research Trends in the Use of Augmented Reality in Science Education: Content and Bibliometric Mapping Analysis. *Computers Education*, 142:103647.
- [30] Armeni, I., Sax, A., Zamir, A. R., and Savarese, S. (2017). Joint 2D-3D-Semantic Data for Indoor Scene Understanding. *arXiv preprint*, arXiv:1702.01105:1–9.
- [31] Armeni, I., Sener, O., Zamir, A. R., Jiang, H., Brilakis, I., Fischer, M., and Savarese, S. (2016). 3D Semantic Parsing of Large-Scale Indoor Spaces. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1534–1543.
- [32] Arth, C., Grasset, R., Gruber, L., Langlotz, T., Mulloni, A., Schmalstieg, D., and Wagner, D. (2015). The History of Mobile Augmented Reality - Developments in Mobile AR over the last almost 50 years. Technical report, Inst. for Computer Graphics and Vision, Graz University of Technology, Austria.
- [33] Ashraf, I., Hur, S., and Park, Y. (2019). Application of Deep Convolutional Neural Networks and Smartphone Sensors for Indoor Localization. *Applied Sciences*, 9(11):2337.
- [34] Assali, M., Pipelidis, G., Podolskiy, V., Iwaszczuk, D., Heinen, L., and Gerndt, M. (2019). Quantifying the Quality of Indoor Maps. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W13:739–745.
- [35] Aziz, N. D., Nawawi, A. H., and Ariff, N. R. M. (2016). Building Information Modelling (BIM) in Facilities Management: Opportunities to be Considered by Facility Managers. *Procedia - Social and Behavioral Sciences*, 234:353–362.
- [36] Azuma, R. (1993). Tracking Requirements for Augmented Reality. *Communications of the ACM - Special Issue on Computer Augmented Environments: Back to the Real World*, 36(7):50–51.
- [37] Azuma, R. T. (1997). A Survey of Augmented Reality. *Teleoperators and Virtual Environments*, 6:355–385.
- [38] Babacan, K., Jung, J., Wichmann, A., Jahromi, B. A., Shahbazi, M., Sohn, G., and Kada, M. (2016). Towards Object Driven Floor Plan Extraction from Laser Point Cloud. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B3:3–10.

- [39] Babinec, A., Jurišica, L., Hubinský, P., and Duchon, F. (2014). Visual Localization of Mobile Robot Using Artificial Markers. *Procedia Engineering*, 96:1–9.
- [40] Bach, B., Sicat, R., Beyer, J., Cordeil, M., and Pfister, H. (2018). The Hologram in My Hand: How Effective is Interactive Exploration of 3D Visualizations in Immersive Tangible Augmented Reality? *IEEE Transactions on Visualization and Computer Graphics*, 24(1):457–467.
- [41] Bachelder, E. (2006). Virtual Media for Military Applications. In *Virtual Media for Military Applications*, pages 27:1–14.
- [42] Bai, X., Huang, M., Prasad, N. R., and Mihovska, A. D. (2019). A Survey of Image-Based Indoor Localization using Deep Learning. In *22nd International Symposium on Wireless Personal Multimedia Communications (WPMC)*, pages 1–6.
- [43] Barazzetti, L. (2016). Parametric As-Built Model Generation of Complex Shapes from Point Clouds. *Advanced Engineering Informatics*, 30(3):298–311.
- [44] Barazzetti, L. and Banfi, F. (2017). *Historic BIM for Mobile VR/AR Applications*, chapter 10, pages 271–290. Springer International Publishing AG.
- [45] Barbieri, L. and Marino, E. (2019). An Augmented Reality Tool to Detect Design Discrepancies: A Comparison Test with Traditional Methods. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, pages 99–110.
- [46] Barki, H., Fadli, F., Shaat, A., Boguslawski, P., and Mahdjoubi, L. (2015). BIM Models Generation From 2D CAD Drawings And 3D Scans: An Analysis of Challenges and Opportunities for AEC Practitioners. *WIT Transactions on The Built Environment*, 149(12):369–380.
- [47] Bassier, M., Klein, R., Van Genechten, B., and Vergauwen, M. (2018). IFC Wall Reconstruction from Unstructured Point Clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2:33–39.
- [48] Bassier, M. and Vergauwen, M. (2019). Clustering of Wall Geometry from Unstructured Point Clouds Using Conditional Random Fields. *Remote Sensing*, 11(13):1586.
- [49] Bassier, M. and Vergauwen, M. (2020). Topology Reconstruction of BIM Wall Objects From Point Cloud Data. *Remote Sensing*, 12(11):1800.
- [50] Bassier, M., Vergauwen, M., and Poux, F. (2020). Point Cloud vs. Mesh Features for Building Interior Classification. *Remote Sensing*, 12(14):2224.
- [51] Bassier, M., Vergauwen, M., and Van Genechten, B. (2016). Standalone Terrestrial Laser Scanning for Efficiently Capturing AEC Buildings for As-Built BIM. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, III-6:49–55.
- [52] Bassier, M., Vincke, S., Mattheuwsen, L., de Lima Hernandez, R., Derdaele, J., and Vergauwen, M. (2019). Percentage of Completion of In-Situ Cast Concrete Walls using Point Cloud Data and BIM. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-5/W2:21–28.
- [53] Bassier, M., Yousefzadeh, M., and Van Genechten, B. (2015). Evaluation of Data Acquisition Techniques and Workflows for Scan to BIM. In *Geo Business*, pages 1–25.

- [54] Beauregard, S. and Haas, H. (2006). Pedestrian Dead Reckoning: A Basis for Personal Positioning. In *Proceedings of the 3rd Workshop on Positioning, Navigation and Communication*, pages 27–36.
- [55] Becker, R., Falk, V., Hoenen, S., Loges, S., Stumm, S., Blankenbach, J., Brell-Cokcan, S., Hildebrandt, L., and Vallée, D. (2018). BIM - Towards the Entire Lifecycle. *International Journal of Sustainable Development and Planning*, 13(1):84–95.
- [56] Becker, R., Lublasser, E., Martens, J., Wollenberg, R., Zhang, H., Brell-Cokcan, S., and Blankenbach, J. (2019). Enabling BIM for Property Management of Existing Buildings Based on Automated As-Is Capturing. In *36th International Symposium on Automation and Robotics in Construction (ISARC 2019)*, pages 201–208.
- [57] Becker, S., Peter, M., and Fritsch, D. (2015). Grammar-Supported 3D Indoor Reconstruction from Point Clouds for “As-Built” BIM. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-3/W4:17–24.
- [58] Benecke, N., Born, A., Boerner, A., Rapp, S., Stelzer, P., Tsirigotis, N., Weber, M., and Zuev, S. (2016). Mobile Solution for Positioning, 3D-Mapping and Inspection in Underground Mining. In *XVI International Congress for Mine Surveying*, pages 141–146.
- [59] Besl, P. J. and McKay, N. D. (1992). A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256.
- [60] Bi, S., Yuan, C., Liu, C., Cheng, J., Wang, W., and Cai, Y. (2021). A Survey of Low-Cost 3D Laser Scanning Technology. *Applied Sciences*, 11(9):3938.
- [61] Biljecki, F., Stoter, J., Ledoux, H., Zlatanova, S., and Çöltekin, A. (2015). Applications of 3D City Models: State of the Art Review. *ISPRS International Journal of Geo-Information*, 4(4):2842–2889.
- [62] Bimber, O., Encarnaç o, L. M., and Schmalstieg, D. (2003). The Virtual Showcase as a New Platform for Augmented Reality Digital Storytelling. In *EGVE ’03 Proceedings of the workshop on Virtual environments 2003*, pages 87–95.
- [63] Blankenbach, J. (2018). Building Surveying for As-Built Modeling. In Borrmann, A., K onig, M., Koch, C., and Beetz, J., editors, *Building Information Modeling*, pages 393–411. Springer.
- [64] Blaser, S., Cavegn, S., and Nebiker, S. (2018). Development of a Portable High Performance Mobile Mapping System Using the Robot Operating System. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-1:13–20.
- [65] Blaser, S., Meyer, J., Nebiker, S., Fricker, L., and Weber, D. (2020). Centimetre-Accuracy in Forests and Urban Canyons – Combining a High-Performance Image-Based Mobile Mapping Backpack with New Georeferencing Methods. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-1-2020:333–341.
- [66] Blodow, N., Goron, L. C., Marton, Z.-C., Pangercic, D., R uhr, T., Tenorth, M., and Beetz, M. (2011). Autonomous Semantic Mapping for Robots Performing Everyday Manipulation Tasks in Kitchen Environments. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4263–4270.

- [67] Bobkov, D., Kiechle, M., Hilsenbeck, S., and Steinbach, E. (2017). Room Segmentation in 3D Point Clouds Using Anisotropic Potential Fields. In *IEEE International Conference on Multimedia and Expo (ICME)*, pages 727–732.
- [68] Boin, J.-B., Bobkov, D., Steinbach, E., and Girod, B. (2019). Efficient Panorama Database Indexing for Indoor Localization. In *International Conference on Content-Based Multimedia Indexing (CBMI)*, pages 1–7.
- [69] Bonduel, M., Bassier, M., Vergauwen, M., Pauwels, P., and Klein, R. (2017). Scan-to-BIM Output Validation: Towards a Standardized Geometric Quality Assessment of Building Information Models Based on Point Clouds. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W8:45–52.
- [70] Boonbrahm, S., Boonbrahm, P., and Kaewrata, C. (2020). The Use of Marker-Based Augmented Reality in Space Measurement. *Procedia Manufacturing*, 42:337–343.
- [71] Bormann, R., Jordan, F., Li, W., Hampp, J., and Hägele, M. (2016). Room Segmentation: Survey, Implementation, and Analysis. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 1019–1026.
- [72] Borrmann, A., König, M., Koch, C., and Beetz, J. (2018). Building Information Modeling: Why? What? How? In Borrmann, A., König, M., Koch, C., and Beetz, J., editors, *Building Information Modeling*, pages 1–24. Springer International Publishing.
- [73] Borrmann, D., Elseberg, J., Lingemann, K., and Nüchter, A. (2011). The 3D Hough Transform for Plane Detection in Point Clouds: A Review and a New Accumulator Design. *3DR Express*, 2:3.
- [74] Botteghi, N., Schulte, R., Sirmacek, B., Poel, M., and Brune, C. (2021). Curiosity-Driven Reinforcement Learning Agent for Mapping Unknown Indoor Environments. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-1-2021:129–136.
- [75] Boulch, A. and Marlet, R. (2012). Fast and Robust Normal Estimation for Point Clouds with Sharp Features. In Grinspun, E. and Mitra, N., editors, *Eurographics Symposium on Geometry Processing*, volume 31 of 1765-1774.
- [76] Brahmabhatt, S., Gu, J., Kim, K., Hays, J., and Kautz, J. (2018). Geometry-Aware Learning of Maps for Camera Localization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2616–2625.
- [77] Brasche, S. and Bischof, W. (2005). Daily Time Spent Indoors in German Homes – Baseline Data for the Assessment of Indoor Exposure of German Occupants. *International Journal of Hygiene and Environmental Health*, 208(4):247–253.
- [78] Brent, R. P. (1973). *Algorithms for Minimization without Derivatives*, chapter 4, An Algorithm with Guaranteed Convergence for Finding a Zero of a Function. Englewood Cliffs, NJ.
- [79] Brito, C., Alves, N., Magalhães, L., and Guevara, M. (2019). BIM Mixed Reality Tool for the Inspection of Heritage Buildings. *The International Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W6:25–29.
- [80] Budroni, A. and Boehm, J. (2010). Automated 3D Reconstruction of Interiors from Point Clouds. *International Journal of Architectural Computing*, 8(1):55–73.

- [81] Calderon-Hernandez, C. and Brioso, X. (2018). Lean, BIM and Augmented Reality Applied in the Design and Construction Phase: A Literature Review. *International Journal of Innovation, Management and Technology*, 9(1):60–63.
- [82] Calders, K., Disney, M. I., Armston, J., Burt, A., Brede, B., Origo, N., Muir, J., and Nightingale, J. (2017). Evaluation of the Range Accuracy and the Radiometric Calibration of Multiple Terrestrial Laser Scanning Instruments for Data Interoperability. *IEEE Transactions on Geoscience and Remote Sensing*, 55(5):2716–2724.
- [83] Campi, M., di Luggo, A., Monaco, S., Siconolfi, M., and Palomba, D. (2018). Indoor and Outdoor Mobile Mapping Systems for Architectural Surveys. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2:201–208.
- [84] Cardoso, A., do Santos Peres, I. C., Lamounier, E., Lima, G., Miranda, M., and Moraes, I. (2017). Associating Holography Techniques with BIM Practices for Electrical Substation Design. In Fechtelkötter, P. and Legatt, M., editors, *Advances in Human Factors in Energy: Oil, Gas, Nuclear and Electric Power Industries*, pages 37–47.
- [85] Carfagni, M., Furferi, R., Governi, L., Servi, M., Uccheddu, F., and Volpe, Y. (2017). On the Performance of the Intel SR300 Depth Camera: Metrological and Critical Characterization. *IEEE Sensors Journal*, 17(14):4508–4519.
- [86] Caron, G., Dame, A., and Marchand, E. (2014). Direct Model Based Visual Tracking and Pose Estimation Using Mutual Information. *Image and Vision Computing*, 32(1):54–63.
- [87] Castellazzi, G., D’Altri, A. M., Bitelli, G., Selvaggi, I., and Lambertini, A. (2015). From Laser Scanning to Finite Element Analysis of Complex Buildings by Using a Semi-Automatic Procedure. *Sensors*, 15(8):18360–18380.
- [88] Caudell, T. P. and Mizell, D. W. (1992). Augmented Reality: An Application of Headsup Display Technology to Manual Manufacturing Processes. In *Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences*, pages 659–669.
- [89] Čejka, J., Zsíros, A., and Liarokapis, F. (2020). A Hybrid Augmented Reality Guide for Underwater Cultural Heritage Sites. *Personal and Ubiquitous Computing*, 24:815–828.
- [90] Chahrour, R., Hafeez, M. A., Ahmad, A. M., Sulieman, H. I., Dawood, H., Rodriguez-Trejo, S., Kassem, M., Najj, K. K., and Dawood, N. (2021). Cost-Benefit Analysis of BIM-Enabled Design Clash Detection and Resolution. *Construction Management and Economics*, 39(1):55–72.
- [91] Chai, J., Chi, H.-L., Wang, X., Wu, C., Jung, K. H., and Lee, J. M. (2016). Automatic As-Built Modeling for Concurrent Progress Tracking of Plant Construction Based on Laser Scanning. *Concurrent Engineering: Research and Applications*, 24(4):369–380.
- [92] Chalhoub, J. and Ayer, S. K. (2018). Using Mixed Reality for Electrical Construction Design Communication. *Automation in Construction*, 86:1–10.
- [93] Chalhoub, J., Ayer, S. K., and McCord, K. H. (2021). Augmented Reality to Enable Users to Identify Deviations for Model Reconciliation. *Buildings*, 11(2):77.
- [94] Chang, A., Dai, A., Funkhouser, T., Halber, M., Nießner, M., Savva, M., Song, S., Zeng, A., and Zhang, Y. (2017). Matterport3D: Learning from RGB-D Data in Indoor Environments. In *International Conference on 3D Vision (3DV)*, pages 667–676.

- [95] Chaouch, M. and Verroust-Blondet, A. (2008). A Novel Method for Alignment of 3D Models. In *IEEE International Conference on Shape Modeling and Applications*, pages 187–195.
- [96] Chaouch, M. and Verroust-Blondet, A. (2009). Alignment of 3D Models. *Graphical Models*, 71(2):63–76.
- [97] Chatzopoulos, D., Bermejo, C., Huang, Z., and Hui, P. (2017). Mobile Augmented Reality Survey: From Where We Are to Where We Go. *IEEE Access*, 5:6917–6950.
- [98] Chen, C., Tang, L., Hancock, C. M., and Zhang, P. (2019a). Development of Low-Cost Mobile Laser Scanning for 3D Construction Indoor Mapping by Using Inertial Measurement Unit, Ultra-Wide Band and 2D Laser Scanner. *Engineering, Construction and Architectural Management*, 26(7):1367–1386.
- [99] Chen, C., Tian, Z., Li, D., Pang, L., Wang, T., and Hong, J. (2021a). Projection-Based Augmented Reality System for Assembly Guidance and Monitoring. *Assembly Automation*, 41(1):10–23.
- [100] Chen, C., Yang, B., Song, S., Tian, M., Li, J., Dai, W., and Fang, L. (2018a). Calibrate Multiple Consumer RGB-D Cameras for Low-Cost and Efficient 3D Indoor Mapping. *Remote Sensing*, 10(2):328.
- [101] Chen, H., Hou, L., Zhang, G. K., and Moon, S. (2021b). Development of BIM, IoT and AR/VR Technologies for Fire Safety and Upskilling. *Automation in Construction*, 125:103631.
- [102] Chen, J. and Clarke, K. C. (2017). Modeling Standards and File Formats for Indoor Mapping. In *Proceedings of the 3rd International Conference on Geographical Information Systems Theory, Applications and Management (GISTAM 2017)*, pages 268–275.
- [103] Chen, J. and Clarke, K. C. (2019). Indoor Cartography. *Cartography and Geographic Information Science (CaGIS)*, pages 1–22.
- [104] Chen, J., Kira, Z., and Cho, Y. K. (2019b). Deep Learning Approach to Point Cloud Scene Understanding for Automated Scan to 3D Reconstruction. *Journal of Computing in Civil Engineering*, 33(4):04019027.
- [105] Chen, J., Liu, C., Wu, J., and Furukawa, Y. (2019c). Floor-SP: Inverse CAD for Floorplans by Sequential Room-Wise Shortest Path. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2661–2670.
- [106] Chen, J., Mora, O. E., and Clarke, K. C. (2018b). Assessing the Accuracy and Precision of Imperfect Point Clouds for 3D Indoor Mapping and Modelling. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-4/W6:3–10.
- [107] Chen, S., Nan, L., Xia, R., Zhao, J., and Wonka, P. (2020a). PLADE: A Plane-Based Descriptor for Point Cloud Registration With Small Overlap. *IEEE Transactions on Geoscience and Remote Sensing*, 58(4):2530–2540.
- [108] Chen, X.-T., Li, Y., Fan, J.-H., and Wang, R. (2021c). RGAM: A Novel Network Architecture for 3D Point Cloud Semantic Segmentation in Indoor Scenes. *Information Sciences*, 571:87–103.
- [109] Chen, Y., Chen, R., Liu, M., Xiao, A., Wu, D., and Zhao, S. (2018c). Indoor Visual Positioning Aided by CNN-Based Image Retrieval: Training-Free, 3D Modeling-Free. *Sensors*, 18(8):2692.

- [110] Chen, Y., Tang, J., Jiang, C., Zhu, L., Lehtomäki, M., Kaartinen, H., Kaijaluoto, R., Wang, Y., Hyyppä, J., Hyyppä, H., Zhou, H., Pei, L., and Ruizhi (2018d). The Accuracy Comparison of Three Simultaneous Localization and Mapping (SLAM)-Based Indoor Mapping Technologies. *Sensors*, 18(10):3228.
- [111] Chen, Y., Wang, Q., Chen, H., Song, X., Tang, H., and Tian, M. (2019d). An Overview of Augmented Reality Technology. *Journal of Physics: Conference Series*, 1237(2):022082.
- [112] Chen, Y.-J., Lai, Y.-S., and Lin, Y.-H. (2020b). BIM-Based Augmented Reality Inspection and Maintenance of Fire Safety Equipment. *Automation in Construction*, 110:103041.
- [113] Chen, Z. H., Cheb, E., Lia, F. X., Olsen, M. J., and Turkan, Y. (2019e). Web-based Deep Segmentation of Indoor Point Clouds. In *International Symposium on Automation and Robotics in Construction (ISARC)*, pages 552–552.
- [114] Cheng, J. C. P., Chen, K., and Chen, W. (2020). State-of-the-Art Review on Mixed Reality Applications in the AECO Industry. *Journal of Construction Engineering and Management*, 146(2):03119009.
- [115] Chionna, F., Argese, F., Palmieri, V., Spada, I., and Colizzi, L. (2015). Integrating Building Information Modeling and Augmented Reality to Improve Investigation of Historical Buildings. *Conservation Science in Cultural Heritage*, 15:133–148.
- [116] Cho, Y. K., Ham, Y., and Golpavar-Fard, M. (2015). 3D As-Is Building Energy Modeling and Diagnostics: A Review of the State-of-the-Art. *Advanced Engineering Informatics*, 29(2):184–195.
- [117] Choi, S., Zhou, Q.-Y., and Koltun, V. (2015). Robust Reconstruction of Indoor Scenes. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5556–5565.
- [118] Chu, H., Kim, D. K., and Chen, T. (2015). You are Here: Mimicking the Human Thinking Process in Reading Floor-Plans. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2210–2218.
- [119] Chu, M., Matthews, J., and Love, P. E. D. (2018). Integrating Mobile Building Information Modelling and Augmented Reality Systems: An Experimental Study. *Automation in Construction*, 85:305–316.
- [120] Cohen, A., Schönberger, J. L., Speciale, P., Sattler, T., Frahm, J.-M., and Pollefeys, M. (2016). Indoor-Outdoor 3D Reconstruction Alignment. In *European Conference on Computer Vision (ECCV)*, pages 285–300.
- [121] Côté, S., Beauvais, M., Girard-Vallée, A., and Snyder, R. (2014). A Live Augmented Reality Tool for Facilitating Interpretation of 2D Construction Drawings. In *International Conference on Augmented and Virtual Reality (AVR)*, pages 421–427.
- [122] Côté, S., Trudel, P., Desbiens, M.-A., Giguère, M., and Snyder, R. (2013). Live Mobile Panoramic High Accuracy Augmented Reality for Engineering and Construction. In *Proceedings of the 13th International Conference on Construction Applications of Virtual Reality*, pages 262–271.
- [123] Coudron, I., Puttemans, S., and Goedemé, T. (2018). Polygonal Reconstruction of Building Interiors from Cluttered Pointclouds. In *European Conference on Computer Vision (ECCV)*, pages 459–472.

- [124] Coudron, I., Puttemans, S., Goedemé, T., and Vandewalle, P. (2020). Semantic Extraction of Permanent Structures for the Reconstruction of Building Interiors from Point Clouds. *Sensors*, 20(23):6916.
- [125] Cranmer, E. E., tom Dieck, M. C., and Fountoulaki, P. (2020). Exploring the Value of Augmented Reality for Tourism. *Tourism Management Perspectives*, 35:100672.
- [126] Cui, Y., Li, Q., and Dong, Z. (2019). Structural 3D Reconstruction of Indoor Space for 5G Signal Simulation with Mobile Laser Scanning Point Clouds. *Remote Sensing*, 11(19):2262.
- [127] Czerniawski, T., Nahangi, M., Walbridge, S., and Haas, C. (2016). Automated Removal of Planar Clutter from 3D Point Clouds for Improving Industrial Object Recognition. In *Proceedings of the 33rd International Symposium in Automation and Robotics in Construction ISARC*, pages 357–365.
- [128] Czmocho, I. and Pękala, A. (2014). Traditional Design versus BIM Based Design. *Procedia Engineering*, 91:210–215.
- [129] Dai, A., Ritchie, D., Bokeloh, M., Reed, S., Sturm, J., and Nießner, M. (2018). ScanComplete: Large-Scale Scene Completion and Semantic Segmentation for 3D Scans. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4578–4587.
- [130] Dai, F., Rashidi, A., Brilakis, I., and Vela, P. (2013). Comparison of Image-Based and Time-of-Flight-Based Technologies for Three-Dimensional Reconstruction of Infrastructure. *Journal of Construction Engineering and Management*, 139(1):929–939.
- [131] Dalla Mura, M., Zanin, M., Andreatta, C., and Chippendale, P. (2012). Augmented Reality: Fusing the Real and Synthetic Worlds. In *IEEE International Geoscience and Remote Sensing Symposium*, pages 170–173.
- [132] de las Heras, L.-P., Ahmed, S., Liwicki, M., Valveny, E., and Sánchez, G. (2013). Statistical Segmentation and Structural Recognition for Floor Plan Interpretation - Notation Invariant Structural Element Recognition. *International Journal on Document Analysis and Recognition (IJDAR)*, 17:221–237.
- [133] de S. Moreira, L. C., Pontes Mota, P., and Almeida Machado, F. (2020). BIM, IoT and MR Integration Applied on Risk Maps for Construction. In *Proceedings of the 18th International Conference on Computing in Civil and Building Engineering (ICCCB)*, pages 895–906.
- [134] de Vries, B. and Harink, J. M. J. (2007). Generation of a Construction Planning from a 3D CAD Model. *Automation in Construction*, 16(1):13–18.
- [135] Debandi, F., Iacoviello, R., Messina, A., and Montagnuolo, M. (2018). Enhancing Cultural Tourism by a Mixed Reality Application for Outdoor Navigation and Information Browsing Using Immersive Devices. *IOP Conf. Series: Materials Science and Engineering*, 364:1–8.
- [136] Dellaert, F., Fox, D., Burgard, W., and Thrun, S. (1999). Monte Carlo Localization for Mobile Robots. In *IEEE International Conference on Robotics and Automation*, pages 1322–1328.
- [137] Depari, A., Flammini, A., Fogli, D., and Magrino, P. (2018). Indoor Localization for Evacuation Management in Emergency Scenarios. In *Workshop on Metrology for Industry 4.0 and IoT*, pages 146–150.

- [138] Di Filippo, A., Sánchez-Aparicio, L. J., Barba, S., Martín-Jiménez, J. A., Mora, R., and González Aguilera, D. (2018). Use of a Wearable Mobile Laser System in Seamless Indoor 3D Mapping of a Complex Historical Site. *Remote Sensing*, 10(12):1897.
- [139] Diakitè, A. A., Zlatanova, S., Alattas, A. F. M., and Li, K. J. (2020). Towards IndoorGML 2.0: Updates and Case Study Illustrations. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B4-2020:337–344.
- [140] Diao, P.-H. and Shih, N.-J. (2019). BIM-Based AR Maintenance System (BARMS) as an Intelligent Instruction Platform for Complex Plumbing Facilities. *Applied Sciences*, 9(8):1592.
- [141] Díaz-Vilariño, L., Boguslawski, P., Khoshelham, K., and Lorenzo, H. (2019). Obstacle-Aware Indoor Pathfinding Using Point Clouds. *ISPRS International Journal of Geo-Information*, 8(5):233.
- [142] Díaz-Vilariño, L., González-de Santos, L., Verbree, E., Michailidou, G., and Zlatanova, S. (2018). From Point Clouds to 3D Isovists in Indoor Environments. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4:149–154.
- [143] Díaz-Vilariño, L., Khoshelham, K., Martínez-Sánchez, J., and Arias, P. (2015). 3D Modeling of Building Indoor Spaces and Closed Doors from Imagery and Point Clouds. *Sensors*, 15(2):3491–3512.
- [144] Ding, C., Liu, H., and Li, H. (2019a). Stitching of Depth and Color Images from Multiple RGB-D Sensors for Extended Field of View. *International Journal of Advanced Robotic Systems*, pages 1–8.
- [145] Ding, Y., Zheng, X., Zhou, Y., Xiong, H., and Gong, J. (2019b). Low-Cost and Efficient Indoor 3D Reconstruction Through Annotated Hierarchical Structure-from-Motion. *Remote Sensing*, 11(1):58.
- [146] Ding, Z., Liu, S., Liao, L., and Zhang, L. (2019c). A Digital Construction Framework Integrating Building Information Modeling and Reverse Engineering Technologies for Renovation Projects. *Automation in Construction*, 102:45–58.
- [147] Dlesk, A., Vach, K., and Holubec, P. (2019). Analysis of Possibilities of Low-Cost Photogrammetry for Interior Mapping. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-5/W3:27–31.
- [148] Dodge, S., Xu, J., and Stenger, B. (2017). Parsing Floor Plan Images. In *Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*, pages 358–361.
- [149] Döllner, J. (2020). Geospatial Artificial Intelligence: Potentials of Machine Learning for 3D Point Clouds and Geospatial Digital Twins. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 88:15–24.
- [150] Drummond, T. and Cipolla, R. (2002). Real-Time Visual Tracking of Complex Structures with On-Line Camera Calibration. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 24(7):932–946.
- [151] Dumitru, R.-C., Borrmann, D., and Nüchter, A. (2013). Interior Reconstruction using the 3D Hough Transform. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-5/W1:65–72.

- [152] Dunston, P. S. and Wang, X. (2011). A Hierarchical Taxonomy of AEC Operations for Mixed Reality Applications. *Journal of Information Technology in Construction*, 16:433–444.
- [153] Eckert, M., Volmerg, J. S., and Friedrich, C. M. (2019). Augmented Reality in Medicine: Systematic and Bibliographic Review. *JMIR mHealth and uHealth*, 7(4):e10967.
- [154] El Ammari, K. and Hammad, A. (2019). Remote Interactive Collaboration in Facilities Management Using BIM-Based Mixed Reality. *Automation in Construction*, 107:102940.
- [155] Elmoogy, A., Dong, X., Lu, T., Westendorp, R., and Reddy, K. (2020). SURF-LSTM: A Descriptor Enhanced Recurrent Neural Network For Indoor Localization. In *IEEE 92nd Vehicular Technology Conference (VTC2020-Fall)*, pages 1–5.
- [156] Elmqaddem, N. (2019). Augmented Reality and Virtual Reality in Education. Myth or Reality? *International Journal of Emerging Technologies in Learning*, 14(3):234–242.
- [157] Elseicy, A., Nikoohemat, S., Peter, M., and Oude Elberink, S. (2018). Space Subdivision of Indoor Mobile Laser Scanning Data Based on the Scanner Trajectory. *Remote Sensing*, 10(11):1815.
- [158] Elshafey, A., Saar, C. C., Aminudin, E. B., Gheisari, M., and Usmani, A. (2020). Technology Acceptance Model for Augmented Reality and Building Information Modeling Integration in the Construction Industry. *Journal of Information Technology in Construction*, 25:161–172.
- [159] Endres, F., Hess, J., Engelhard, N., Sturm, J., Cremers, D., and Burgard, W. (2012). An Evaluation of the RGB-D SLAM System. In *IEEE International Conference on Robotics and Automation*, pages 1691–1696.
- [160] Engel, J., Schöps, T., and Cremers, D. (2014). LSD-SLAM: Large-Scale Direct Monocular SLAM. In *European Conference on Computer Vision (ECCV)*, pages 834–849.
- [161] Engelmann, F., Kontogianni, T., Hermans, A., and Leibe, B. (2018). Exploring Spatial Context for 3D Semantic Segmentation of Point Clouds. In *IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 716–724.
- [162] Erickson, A., Kim, K., Bruder, G., and Welch, G. (2020). A Review of Visual Perception Research in Optical See-Through Augmented Reality. In *International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*, pages 1–9.
- [163] Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *KDD’96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, page 226–231.
- [164] Faion, F., Zea, A., Noack, B., Steinbring, J., and Hanebeck, U. D. (2016). Camera- and IMU-based Pose Tracking for Augmented Reality. In *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*.
- [165] Fallah, N., Apostolopoulos, I., Bekris, K., and Folmer, E. (2013). Indoor Human Navigation Systems: A Survey. *Interacting with Computers*, 25(1):21–33.
- [166] Fallon, M., Johannsson, H., Kaess, M., and Leonard, J. J. (2014). The MIT Stata Center Dataset. *The International Journal of Robotics Research*, 32(14):1695–1699.
- [167] Fan, S.-L., Skibniewski, M. J., and Hung, T. W. (2014). Effects of Building Information Modeling During Construction. *Journal of Applied Science and Engineering*, 17(2):157–166.

- [168] Fang, H., Lafarge, F., Pan, C., and Huang, H. (2021). Floorplan Generation from 3D Point Clouds: A Space Partitioning Approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 175:44–55.
- [169] Farzaneh, A., Carriere, J., Forgues, D., and Monfet, D. (2018). Framework for Using Building Information Modeling to Create a Building Energy Model. *Journal of Architectural Engineering*.
- [170] Fenais, A., Smilovsky, N., Ariaratnam, S. T., and Ayer, S. K. (2018). A Meta-Analysis of Augmented Reality Challenges in the Underground Utility Construction Industry. In *Construction Research Congress*, pages 1–10.
- [171] Feng, C.-W. and Chen, C.-W. (2019). Using BIM and AR to Improve the Process of Job Site Construction and Inspection. *WIT Transactions on The Built Environment*, 192(12):21–32.
- [172] Feng, G., Liang, S., Wang, H., Liny, J., Liu, J., Lv, H., and Zhao, S. (2018a). An Indoor Localization Simulation Platform for Localization Accuracy Evaluation. In *15th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pages 1–3.
- [173] Feng, Z., González, V. A., Ma, L., Al-Adhami, M. M. A., and Mourgues, C. (2018b). Rapid 3D Reconstruction of Indoor Environments to Generate Virtual Reality Serious Games Scenarios. In *Proceedings of the 18th International Conference on Construction Applications of Virtual Reality (CONVR2018)*, pages 185–195.
- [174] Fichtner, F. W., Diakité, A. A., Zlatanova, S., and Voûte, R. (2017). Semantic Enrichment of Octree Structured Point Clouds for Multi-Story 3D Pathfinding. *Transactions in GIS*, 22:233–248.
- [175] Fischler, M. A. and Bolles, R. C. (1981). Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 24(6):381–395.
- [176] Flikweert, P., Peters, R., Díaz-Vilariño, L., Voûte, R., and Staats, B. (2019). Automatic Extraction of a Navigation Graph Intended for IndoorGML from an Indoor Point Cloud. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W5:271–278.
- [177] Fonnet, A., Alves, N., Sousa, N., Guevara, M., and Magalhães, L. (2017). Heritage BIM Integration With Mixed Reality for Building Preventive Maintenance. In *Proceedings of the 24^o Encontro Português de Computação Gráfica e Interação (EPCGI)*, pages 1–7.
- [178] Frías, E., Balado, J., Díaz-Vilariño, L., and Lorenzo, H. (2020). Point Cloud Room Segmentation Based on Indoor Spaces and 3D Mathematical Morphology. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIV-4/W1-2020:49–55.
- [179] Fu, H., Cohen-Or, D., Dror, G., and Sheffer, A. (2008). Upright Orientation of Man-Made Objects. In *ACM SIGGRAPH*, page 42.
- [180] Fuentes-Pacheco, J., Ruiz-Ascencio, J., and Rendón-Mancha, J. M. (2015). Visual Simultaneous Localization and Mapping: A Survey. *Artificial Intelligence Review*, 43(1):55–81.
- [181] Fürsattel, P., Placht, S., Balda, M., Schaller, C., Hofmann, H., Maier, A., and Riess, C. (2016). A Comparative Error Analysis of Current Time-of-Flight Sensors. *IEEE Transactions on Computational Imaging*, 2(1):27–41.

- [182] Gankhuyag, U. and Han, J.-H. (2020). Automatic 2D Floorplan CAD Generation from 3D Point Clouds. *Applied Sciences*, 10(8):2817.
- [183] Gankhuyag, U. and Han, J.-H. (2021). Automatic BIM Indoor Modelling from Unstructured Point Clouds Using a Convolutional Neural Network. *Intelligent Automation Soft Computing*, 28(1):133–152.
- [184] Gao, X. and Pishdad-Bozorgi, P. (2019). BIM-Enabled Facilities Operation and Maintenance: A Review. *Advanced Engineering Informatics*, 39:227–247.
- [185] Gao, X.-S., Hou, X.-R., Tang, J., and Cheng, H.-F. (2003). Complete Solution Classification for the Perspective-Three-Point Problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8):930–943.
- [186] Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F., and Marín-Jiménez, M. (2014). Automatic Generation and Detection of Highly Reliable Fiducial Markers Under Occlusion. *Pattern Recognition*, 47(6):2280–2292.
- [187] Garwood, T. L., Hughes, B. R., O’Connor, D., Calautit, J. K., Oates, M. R., and Hodgson, T. (2018). A Framework for Producing gbXML Building Geometry from Point Clouds for Accurate and Efficient Building Energy Modelling. *Applied Energy*, 224:527–537.
- [188] Gattullo, M., Scurati, G. W., Fiorentino, M., Uva, A. E., Ferrise, F., and Bordegoni, M. (2019). Towards Augmented Reality Manuals for Industry 4.0: A Methodology. *Robotics and Computer-Integrated Manufacturing*, 56:276–286.
- [189] Gerstweiler, G., Platzer, K., and Kaufmann, H. (2018). DARGs: Dynamic AR Guiding System for Indoor Environments. *Computers*, 7(1):5.
- [190] Getuli, V., Capone, P., Bruttini, A., and Isaac, S. (2020). BIM-Based Immersive Virtual Reality for Construction Workspace Planning: A Safety-Oriented Approach. *Automation in Construction*, 114:103160.
- [191] Getuli, V., Mastrolembo Ventura, S., Capone, P., and Ciribini, A. L. C. (2017). BIM-based Code Checking for Construction Health and Safety. *Procedia Engineering*, 196:454–461.
- [192] Ghaffarianhoseini, A., Tookey, J., Ghaffarianhoseini, A., Naismith, N., Azhard, S., Efimova, O., and Raahemifarb, K. (2017). Building Information Modelling (BIM) Uptake: Clear Benefits, Understanding its Implementation, Risks and Challenges. *Renewable and Sustainable Energy Reviews*, 75:1046–1053.
- [193] Gheisari, M., Goodman, S., Schmidh, J., Williams, G., and Irizarry, J. (2014). Exploring BIM and Mobile Augmented Reality Use in Facilities Management. In *Construction Research Congress*, pages 1941–1950.
- [194] Gheisari, M. and Irizarry, J. (2016). Investigating Human and Technological Requirements for Successful Implementation of a BIM-based Mobile Augmented Reality Environment in Facility Management Practices. *Facilities*, 34(1/2):69–84.
- [195] Girardeau-Montaut, D. (2021). CloudCompare: 3D Point Cloud and Mesh Processing Software. <http://www.cloudcompare.org> (last visited on 12/04/2021).

- [196] Glocker, B., Izadi, S., Shotton, J., and Criminisi, A. (2013). Real-Time RGB-D Camera Relocalization. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 173–179.
- [197] Goldstein, R., Breslav, S., and Khan, A. (2014). Towards Voxel-Based Algorithms for Building Performance Simulation. In *IBPSA eSIM*, pages 1–14.
- [198] Golparvar-Fard, M., Peña-Mora, F., and Savarese, S. (2015). Automated Progress Monitoring Using Unordered Daily Construction Photographs and IFC-Based Building Information Models. *Journal of Computing in Civil Engineering*, 29(1).
- [199] Gonzalez-Jorge, H., Rodríguez-Gonzálvez, P., Martínez-Sánchez, J., González-Aguilera, D., Arias, P., Gesto, M., and Díaz-Vilariñoa, L. (2015). Metrological Comparison between Kinect I and Kinect II Sensors. *Measurement*, 70:21–26.
- [200] Gorte, B., Aleksandrov, M., and Zlatanova, S. (2019a). Towards Egress Modelling in Voxel Building Models. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-4/W9:43–47.
- [201] Gorte, B., Zlatanova, S., and Fadli, F. (2019b). Navigation in Indoor Voxel Models. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W5:279–283.
- [202] Gourlis, G. and Kovacic, I. (2017). Building Information Modelling for Analysis of Energy Efficient Industrial Buildings – A Case Study. *Renewable and Sustainable Energy Reviews*, 68(2):953–963.
- [203] Grieves, M. (2014). Digital Twin: Manufacturing Excellence through Virtual Factory Replication. White Paper.
- [204] Gröger, G. and Plümer, L. (2012). CityGML - Interoperable Semantic 3D City Models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 71:12–33.
- [205] Gu, F., Hu, X., Ramezani, M., Acharya, D., Khoshelham, K., Valaee, S., and Shang, J. (2019a). Indoor Localization Improved by Spatial Context - A Survey. *ACM Computing Surveys*, 52(3):1–35.
- [206] Gu, F., Ramezani, M., Khoshelham, K., Zheng, X., Zhou, R., and Shang, J. (2019b). Fast and Reliable WiFi Fingerprint Collection for Indoor Localization. In *International Conference on Urban Intelligence and Applications (ICUIA 2019)*, pages 1–8.
- [207] Gu, F., Valaee, S., Khoshelham, K., Shang, J., and Zhang, R. (2020). Landmark Graph-Based Indoor Localization. *IEEE Internet of Things Journal*, 7(9):8343–8355.
- [208] Guclu, O., Caglayan, A., and Can, A. B. (2019). RGB-D Indoor Mapping Using Deep Features. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1248–1257.
- [209] Guha, D., Alotaibi, N. M., Nguyen, N., Gupta, S., McFaul, C., and Yang, V. X. (2017). Augmented Reality in Neurosurgery: A Review of Current Concepts and Emerging Applications. *Canadian Journal of Neurological Sciences*, 44(3):235–245.
- [210] Guhl, J., Tung, S., and Kruger, J. (2017). Concept and Architecture for Programming Industrial Robots Using Augmented Reality with Mobile Devices Like Microsoft HoloLens. In *22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, pages 1–4.

- [211] Gustafsson, F. (2010). Particle Filter Theory and Practice with Positioning Applications. *IEEE Aerospace and Electronic Systems Magazine*, 25(7):53–81.
- [212] Ha, I., Kim, H., Park, S., and Kim, H. (2018). Image Retrieval Using BIM and Features from Pretrained VGG Network for Indoor Localization. *Building and Environment*, 140:23–31.
- [213] Haala, N. and Kada, M. (2010). An Update on Automatic 3D Building Reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(6):570–580.
- [214] Hackel, T., Savinov, N., Ladicky, L., Wegner, J. D., Schindler, K., and Pollefeys, M. (2017). SEMANTIC3D.NET: A New Large-Scale Point Cloud Classification Benchmark. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-1/W1:91–98.
- [215] Hajian, H. and Becerik-Gerber, B. (2010). Scan to BIM: Factors Affecting Operational and Computational Errors and Productivity Loss. In *The 27th International Symposium on Automation and Robotics in Construction*, pages 265–272.
- [216] Hakkarainen, M., Woodward, C., and Rainio, K. (2009). Software Architecture for Mobile Mixed Reality and 4D BIM Interaction. In *Proceedings of the 25th CIB W78 Conference*, pages 517–524.
- [217] Han, P. and Zhao, G. (2019). A Review of Edge-Based 3D Tracking of Rigid Objects. *Virtual Reality Intelligent Hardware*, 1(6):580–596.
- [218] Han, X., Li, S., Wang, X., and Zhou, W. (2021). Semantic Mapping for Mobile Robots in Indoor Scenes: A Survey. *Information*, 12(2):92.
- [219] Hanussek, B. (2019). Conducting Archaeogaming Protecting Digital Heritage: Does the Future for Archaeology Lie in The Immaterial. *Art and Science*, 3(1):1–15.
- [220] Hartmann, J., Gupta, A., and Vogel, D. (2020). Extend, Push, Pull: Smartphone Mediated Interaction in Spatial Augmented Reality via Intuitive Mode Switching. In *Symposium on Spatial User Interaction*, pages 1–10.
- [221] He, M., Zhu, C., Huang, Q., Ren, B., and Liu, J. (2020). A Review of Monocular Visual Odometry. *The Visual Computer*, 36:1053–1065.
- [222] Heinz, E., Holst, C., Kuhlmann, H., and Klingbeil, L. (2020). Design and Evaluation of a Permanently Installed Plane-Based Calibration Field for Mobile Laser Scanning Systems. *Remote Sensing*, 12(3):555.
- [223] Henry, P., Krainin, M., Herbst, E., Ren, X., and Fox, D. (2012). RGB-D Mapping: Using Kinect-Style Depth Cameras for Dense 3D Modeling of Indoor Environments. *The International Journal of Robotics Research*, 31(5):647–663.
- [224] Herbers, P. and König, M. (2019). Indoor Localization for Augmented Reality Devices Using BIM, Point Clouds, and Template Matching. *Applied Sciences*, 9(20):4260.
- [225] Herle, S., Becker, R., Wollenberg, R., and Blankenbach, J. (2020). GIM and BIM - How to Obtain Interoperability Between Geospatial and Building Information Modelling? *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 88:33–42.
- [226] Hesch, J. A. and Roumeliotis, S. I. (2007). An Indoor Localization Aid for the Visually Impaired. In *Proceedings 2007 IEEE International Conference on Robotics and Automation*, pages 3545–3551.

- [227] Heya, T. A., Arefin, S. E., Chakrabarty, A., and Alam, M. (2018). Image Processing Based Indoor Localization System for Assisting Visually Impaired People. In *Ubiquitous Positioning, Indoor Navigation and Location-Based Services (UPINLBS)*, pages 1–7.
- [228] Hillemann, M. and Jutzi, B. (2017). UCalMiCeL - Unified Intrinsic and Extrinsic Calibration of a Multi-Camera-System and a Laserscanner. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W3:17–24.
- [229] Hockett, P. and Ingleby, T. (2016). Augmented Reality with HoloLens: Experiential Architectures Embedded in the Real World. *arXiv*, arXiv:1610.04281.
- [230] Hönig, W., Milanes, C., Scaria, L., Phan, T., Bolas, M., and Ayanian, N. (2015). Mixed Reality for Robotics. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5382–5387.
- [231] Horn, B. K. P. (1984). Extended Gaussian Images. In *Proceedings of the IEEE*, volume 72, pages 1671–1686.
- [232] Horn, B. K. P. (1986). Closed-Form Solution of Absolute Orientation Using Unit Quaternions. *Journal of the Optical Society of America A*, 4:629–642.
- [233] Hosseininaveh, A. and Remondino, F. (2021). An Imaging Network Design for UGV-Based 3D Reconstruction of Buildings. *Remote Sensing*, 13(10):1923.
- [234] Howard, R. and Björk, B.-C. (2007). Use of Standards for CAD Layers in Building. *Automation in Construction*, 16(3):290–297.
- [235] Howard-Jenkins, H., Li, S., and Prisacariu, V. (2018). Thinking Outside the Box: Generation of Unconstrained 3D Room Layouts. In *Asian Conference on Computer Vision (ACCV)*, pages 432–448.
- [236] Hu, S., Wang, D., and Xu, S. (2016). 3D Indoor Modeling Using a Hand-Held Embedded System with Multiple Laser Range Scanners. In *Proceedings SPIE 10155, Optical Measurement Technology and Instrumentation*, volume 101552D, pages 1–6.
- [237] Huang, R., Xu, Y., Hoegner, L., and Stilla, U. (2020). Efficient Estimation of 3D Shifts Between Point Clouds Using Low-Frequency Components of Phase Correlation. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-2-2020:227–234.
- [238] Hübner, P., Clintworth, K., Liu, Q., Weinmann, M., and Wursthorn, S. (2020a). Evaluation of HoloLens Tracking and Depth Sensing for Indoor Mapping Applications. *Sensors*, 20(4):1021.
- [239] Hübner, P., Landgraf, S., Weinmann, M., and Wursthorn, S. (2019). Evaluation of the Microsoft HoloLens for the Mapping of Indoor Building Environments. In *Dreiländertagung der DGPF, der OVG und der SGPF in Wien, Österreich – Publikationen der DGPF*, volume 28, pages 44–53.
- [240] Hübner, P., Weinmann, M., Hillemann, M., Jutzi, B., and Wursthorn, S. (2018a). Combining Independent Visualization and Tracking Systems for Augmented Reality. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2:455–462.
- [241] Hübner, P., Weinmann, M., and Wursthorn, S. (2018b). Marker-Based Localization of the Microsoft HoloLens in Building Models. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-1:195–202.

- [242] Hübner, P., Weinmann, M., and Wursthorn, S. (2020b). Voxel-Based Indoor Reconstruction from HoloLens Triangle Meshes. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-4-2020:79–86.
- [243] Hübner, P., Weinmann, M., Wursthorn, S., and Hinz, S. (2021a). Automatic Voxel-based 3D Indoor Reconstruction and Room Partitioning from Triangle Meshes. *ISPRS Journal of Photogrammetry and Remote Sensing*, accepted for publication.
- [244] Hübner, P., Weinmann, M., Wursthorn, S., and Hinz, S. (2021b). Pose Normalization of Indoor Mapping Datasets Partially Compliant to the Manhattan World Assumption. *arXiv*, arXiv:2107.07778:1–39.
- [245] Ijaz, F., Yang, H. K., Ahmad, A. W., and Lee, C. (2013). Indoor Positioning: A Review of Indoor Ultrasonic Positioning Systems. In *15th International Conference on Advanced Communications Technology (ICACT)*, pages 1146–1150.
- [246] Ikehata, S., Yang, H., and Furukawa, Y. (2015). Structured Indoor Modeling. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1323–1331.
- [247] Ingman, M., Virtanen, J.-P., Vaaja, M. T., and Hyyppä, H. (2020). A Comparison of Low-Cost Sensor Systems in Automatic Cloud-Based Indoor 3D Modeling. *Remote Sensing*, 12(16):2624.
- [248] Irizarry, J., Gheisari, M., Williams, G., and Roper, K. (2014). Ambient Intelligence Environments for Accessing Building Information: A Healthcare Facility Management Scenario. *Facilities*, 32(3/4):120–138.
- [249] Ishikawa, T., Kanai, S., and Date, H. (2021). Direct Generation of Cartesian Grid for As-built CFD Analysis from Laser-Scanned Point Clouds. *Computer-Aided Design Applications*, 18(6):1341–1358.
- [250] Iwaszczuk, D., Koppanyi, Z., Pfrang, J., and Toth, C. (2019). Evaluation of a Mobile Multi-Sensor System for Seamless Outdoor and Indoor Mapping. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-1/W2:31–35.
- [251] Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman, D., Davison, A., and Fitzgibbon, A. (2011). KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, pages 559–568.
- [252] Jafari, K. G., Sharyatpanahi, N. S. G., and Noorzai, E. (2021). BIM-Based Integrated Solution for Analysis and Management of Mismatches During Construction. *Journal of Engineering, Design and Technology*, 19(1):81–102.
- [253] Jang, H., Yu, K., and Yang, J. (2020). Indoor Reconstruction from Floorplan Images with a Deep Learning Approach. *ISPRS International Journal of Geo-Information*, 9(2):65.
- [254] Jiang, S., Jiang, L., Han, Y., Wu, Z., and Wang, N. (2019). OpenBIM: An Enabling Solution for Information Interoperability. *Applied Sciences*, 9(24):5358.
- [255] Jin, R., Zhong, B., Ma, L., Hashemi, A., and Ding, L. (2019). Integrating BIM with Building Performance Analysis in Project Life-Cycle. *Automation in Construction*, 106:102861.

- [256] Jin, Y., Seo, J., Lee, J. G., Ahn, S., and Han, S. (2020). BIM-Based Spatial Augmented Reality (SAR) for Architectural Design Collaboration: A Proof of Concept. *Applied Sciences*, 10(17):5915.
- [257] Johansson, M., Roupé, M., and Tallgren, M. V. (2014). From BIM to VR - Integrating Immersive Visualizations in the Current Design Process. In *Fusion, Proceedings of the 32nd International Conference on Education and Research in Computer Aided Architectural Design in Europe*, pages 261–269.
- [258] Jolliffe, I. T. and Cadima, J. (2016). Principal Component Analysis: A Review and Recent Developments. *Philosophical Transactions. Series A, Mathematical, Physical and Engineering Sciences*, 374(2065):20150202.
- [259] Joshi, R., Hiwale, A., Birajdar, S., and Gound, R. (2019). Indoor Navigation with Augmented Reality. In *Proceedings of the 2nd International Conference on Communications and Cyber Physical Engineering (ICCCE)*, pages 159–165.
- [260] Jung, J., Stachniss, C., Ju, S., and Heo, J. (2018). Automated 3D Volumetric Reconstruction of Multiple-Room Building Interiors for As-Built BIM. *Advanced Engineering Informatics*, 38:811–825.
- [261] Jung, J., Stachniss, C., and Kim, C. (2017). Automatic Room Segmentation of 3D Laser Data Using Morphological Processing. *ISPRS International Journal of Geo-Information*, 6(7):206.
- [262] Jung, W. and Lee, G. (2015). The Status of BIM Adoption on Six Continents. *International Journal of Civil, Structural, Construction and Architectural Engineering*, 9(5):406–410.
- [263] Jurado, D., Jurado, J. M., Ortega, L., and Feito, F. R. (2021). GEUINF: Real-Time Visualization of Indoor Facilities Using Mixed Reality. *Sensors*, 21(4):1123.
- [264] Kang, H.-K. and Li, K.-J. (2017). A Standard Indoor Spatial Data Model - OGC IndoorGML and Implementation Approaches. *ISPRS International Journal of Geo-Information*, 6(4):116.
- [265] Kang, T., Patil, S., Kang, K., Koo, D., and Kim, J. (2020a). Rule-Based Scan-to-BIM Mapping Pipeline in the Plumbing System. *Applied Sciences*, 10(21):7422.
- [266] Kang, W. and Han, Y. (2015). SmartPDR: Smartphone-Based Pedestrian Dead Reckoning for Indoor Localization. *IEEE Sensors Journal*, 15:2906–2916.
- [267] Kang, Z., Yang, J., Yang, Z., and Cheng, S. (2020b). A Review of Techniques for 3D Reconstruction of Indoor Environments. *ISPRS International Journal of Geo-Information*, 9(5):330.
- [268] Karam, S. and Alsadik, B. (2021). The Simultaneous Localization and Mapping (SLAM) - An Overview. *Surveying and Geospatial Engineering Journal*, 1(2):34–45.
- [269] Karam, S., Peter, M., Hosseinyalamdary, S., and Vosselman, G. (2018). An Evaluation Pipeline for Indoor Laser Scanning Point Clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-1:85–92.
- [270] Karam, S., Vosselman, G., Peter, M., Hosseinyalamdary, S., and Lehtola, V. (2019). Design, Calibration, and Evaluation of a Backpack Indoor Mobile Mapping System. *Remote Sensing*, 11(8):905.

- [271] Kästner, L. and Lambrecht, J. (2019). Augmented-Reality-Based Visualization of Navigation Data of Mobile Robots on the Microsoft Hololens - Possibilities and Limitations. In *IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM)*, pages 344–349.
- [272] Kato, H. and Billinghurst, M. (1999). Marker Tracking and HMD Calibration for a Video-Based Augmented Reality Conferencing System. In *Proceedings of the 2nd International Workshop on Augmented Reality (IWAR 99)*, pages 85–94.
- [273] Kazhdan, M. (2007). An Approximate and Efficient Method for Optimal Rotation Alignment of 3D Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(29):1221–1229.
- [274] Kendall, A., Grimes, M., and Cipolla, R. (2015). PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2938–2946.
- [275] Kenney, N. (2019). Laser Scan May One Day Aid Notre Dame’s Restorers. the art newspaper. <https://www.theartnewspaper.com/news/interpreting-a-laser-scan-that-may-one-day-aid-notre-dame-s-restorers> (last visited on 05/06/2021).
- [276] Kerle, N., Nex, F., Gerke, M., Duarte, D., and Vetrivel, A. (2020). UAV-Based Structural Damage Mapping: A Review. *ISPRS International Journal of Geo-Information*, 9(1):14.
- [277] Kern, J., Weinmann, M., and Wursthorn, S. (2017). Projector-Based Augmented Reality for Quality Inspection of Scanned Objects. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W4:83–90.
- [278] Khajehzadeh, I. and Vale, B. (2017). How New Zealanders Distribute their Daily Time Between Home Indoors, Home Outdoors and Out of Home. *Kōtuitui: New Zealand Journal of Social Sciences Online*, 12(1):17–31.
- [279] Khan, S. and Panuwatwanich, K. (2021). Applicability of Building Information Modeling Integrated Augmented Reality in Building Facility Management. In *Proceedings of The 16th East Asian-Pacific Conference on Structural Engineering and Construction (EASEC)*, pages 2139–2148.
- [280] Khoshelham, K. and Díaz-Vilariño, L. (2014). 3D Modelling of Interior Spaces: Learning the Language of Indoor Architecture. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-5:321–326.
- [281] Khoshelham, K., Díaz Vilariño, L., Peter, M., Kang, Z., and Acharya, D. (2017). The ISPRS Benchmark on Indoor Modelling. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W7:367–372.
- [282] Khoshelham, K. and Oude Elberink, S. (2012). Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications. *Sensors*, 12(2):1437–1454.
- [283] Khoshelham, K., Tran, H., and Acharya, D. (2019). Indoor Mapping Eyewear: Geometric Evaluation of Spatial Mapping Capability of HoloLens. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W13:805–810.
- [284] Khoshelham, K., Tran, H., Acharya, D., Díaz Vilariño, L., Kang, Z., and Dalyot, S. (2020). The ISPRS Benchmark on Indoor Modelling - Preliminary Results. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B5-2020:207–211.

- [285] Khoshelham, K., Tran, H., Díaz-Vilariño, L., Peter, M., Kang, Z., and Acharya, D. (2018). An Evaluation Framework for Benchmarking Indoor Modelling Methods. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4:297–302.
- [286] Kim, H., Nagao, S., Maekawa, S., and Naemura, T. (2014). MRsionCase: A Glasses-free Mixed Reality Showcase for Surrounding Multiple Viewers. *ITE Transactions on Media Technology and Applications*, 2(3):200–208.
- [287] Kim, K., Billingham, M., Bruder, G., Duh, H. B.-L., and Welch, G. F. (2018). Revisiting Trends in Augmented Reality Research: A Review of the 2nd Decade of ISMAR (2008–2017). *IEEE Transactions on Visualization and Computer Graphics*, 24(11):2947–2962.
- [288] Kim, T., Cho, W., Matono, A., and Kim, K.-S. (2020). PinSout: Automatic 3D Indoor Space Construction from Point Clouds with Deep Learning. In *SIGSPATIAL '20: Proceedings of the 28th International Conference on Advances in Geographic Information Systems*, pages 211–214.
- [289] Kirks, T., Jost, J., Uhlott, T., Püth, J., and Jakobs, M. (2019). Evaluation of the Application of Smart Glasses for Decentralized Control Systems in Logistics. In *IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 4470–4476.
- [290] Klein, G. and Murray, D. (2009). Parallel Tracking and Mapping on a Camera Phone. In *8th IEEE International Symposium on Mixed and Augmented Reality*, pages 83–86.
- [291] Klein, L., Li, N., and Becerik-Gerber, B. (2012). Imaged-Based Verification of As-Built Documentation of Operational Buildings. *Automation in Construction*, 21:161–171.
- [292] Koch, C., Neges, H.-M., König, M., and Abramovici, M. (2014). Natural Markers for Augmented Reality Based Indoor Navigation and Facility Maintenance. *Automation in Construction*, 48:18–30.
- [293] Koch, T., Körner, M., and Fraundorfer, F. (2016). Automatic Alignment of Indoor and Outdoor Building Models Using 3D Line Segments. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 689–697.
- [294] Koeva, M., Nikoohemat, S., Oude Elberink, S., Morales, J., Lemmen, C., and Zevenbergen, J. (2019). Towards 3D Indoor Cadastre Based on Change Detection from Point Clouds. *Remote Sensing*, 6(7):206.
- [295] Kolbe, T. H., Gröger, G., and Plümer, L. (2005). CityGML: Interoperable Access to 3D City Models. In van Oosterom, P., Zlatanova, S., and Fendel, E. M., editors, *Geo-information for Disaster Management*, pages 883–899. Springer.
- [296] Kopsida, M. and Brilakis, I. (2016). BIM Registration Methods for Mobile Augmented Reality-Based Inspection. In *Proceedings of the European Conference on Product and Process Modelling*, pages 1–7.
- [297] Kray, C., Fritze, H., Fechner, T., Schwering, A., Li, R., and Anacta, V. J. (2013). Transitional Spaces: Between Indoor and Outdoor Spaces. In *International Conference on Spatial Information Theory (COSIT)*, pages 14–32.
- [298] Kress, B. C. and Cummings, W. J. (2017). Towards the Ultimate Mixed Reality Experience: HoloLens Display Architecture Choices. In *SID Symposium Digest of Technical Papers*, volume 48, pages 127–131.

- [299] Krinkin, K., Filatov, A., Filatov, A., Huletski, A., and Kartashov, D. (2015). Evaluation of Modern Laser Based Indoor SLAM Algorithms. In *22nd Conference of Open Innovations Association (FRUCT)*, pages 101–106.
- [300] Kritzinger, W., Karner, M., Traar, G., Henjes, J., and Sihm, W. (2018). Digital Twin in Manufacturing: A Categorical Literature Review and Classification. *IFAC-PapersOnLine*, 51(11):1016–1022.
- [301] Krūminaitė, M. and Zlatanova, S. (2014). Indoor Space Subdivision for Indoor Navigation. In *ISA '14: Proceedings of the Sixth ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness*, pages 25–31.
- [302] Kumar, G. A., Patil, A. K., Kang, T. W., and Chai, Y. H. (2019). Sensor Fusion Based Pipeline Inspection for the Augmented Reality System. *Symmetry*, 11(10):1325.
- [303] Kunz, C., Schneider, M., Hlavac, M., Zenth, D., Pylatiuk, C., and Hein, B. (2019). Augmented Reality-Assisted Ventricular Puncture with Marker-Based Scene Registration. *International Journal of Computer Assisted Radiology and Surgery*, 14(1):151–152.
- [304] Kutzner, T., Chaturvedi, K., and Kolbe, T. H. (2020). CityGML 3.0: New Functions Open Up New Applications. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 88:43–61.
- [305] Kuula, T., Piira, K., Seisto, A., Hakkarainen, M., and Woodward, C. (2012). User Requirements for Mobile AR and BIM Utilization in Building Life Cycle Management. In *Proceedings of the 12th International Conference on Construction Applications of Virtual Reality (CONVR2012)*, pages 203–211.
- [306] Kwon, O.-S., Park, C.-S., and Lim, C.-R. (2014). A Defect Management System for Reinforced Concrete Work Utilizing BIM, Image-Matching and Augmented Reality. *Automation in Construction*, 46:74–81.
- [307] Laakso, M. and Kiviniemi, A. (2012). The IFC Standard: A Review of History, Development, and Standardization. *ITcon Journal of Information Technology in Construction*, 17(9):134–161.
- [308] Lachat, E., Macher, H., Mittet, M.-A., Landes, T., and Grussenmeyer, P. (2015). First Experiences with Kinect V2 Sensor for Close Range 3D Modelling. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-5/W4:93–100.
- [309] Lagüela, S., Dorado, I., Gesto, M., Arias, P., González-Aguilera, D., and Lorenzo, H. (2018). Behavior Analysis of Novel Wearable Indoor Mapping System Based on 3D-SLAM. *Sensors*, 18(3):766.
- [310] Laine, T. H. (2018). Mobile Educational Augmented Reality Games: A Systematic Literature Review and Two Case Studies. *Computers*, 7(1):19.
- [311] Landrieu, L., Raguét, H., Vallet, B., Mallet, C., and Weinmann, M. (2017). A Structured Regularization Framework for Spatially Smoothing Semantic Labelings of 3D Point Clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 132:102–118.
- [312] Landrieu, L. and Simonovsky, M. (2018). Large-scale Point Cloud Semantic Segmentation with Superpoint Graphs. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4558–4567.

- [313] Langlois, C., Tiku, S., and Pasricha, S. (2017). Indoor Localization with Smartphones: Harnessing the Sensor Suite in Your Pocket. *IEEE Consumer Electronics Magazine*, 6(4):70–80.
- [314] Laska, M. and Blankenbach, J. (2021). DeepLocBox: Reliable Fingerprinting-Based Indoor Area Localization. *Sensors*, 21(6):2000.
- [315] Lauterbach, H. A., Borrmann, D., Heß, R., Eck, D., Schilling, K., and Nüchter, A. (2015). Evaluation of a Backpack-Mounted 3D Mobile Scanning System. *Remote Sensing*, 7(10):13753–13781.
- [316] Ledoux, H., Arroyo Ogori, K., Kumar, K., Dukai, B., Labetski, A., and Vitalis, S. (2019). CityJSON: A Compact and Easy-to-Use Encoding of the CityGML Data Model. *Open Geospatial Data, Software and Standards*, 4:4.
- [317] Lee, J. H., Park, J. J., and Yoon, H. (2020). Automatic Bridge Design Parameter Extraction for Scan-to-BIM. *Applied Sciences*, 10(20):7346.
- [318] Lee, X. S., Khamidi, M. F., See, Z. S., Lees, T. J., and Chai, C. (2016). Augmented Reality for nDimensional Building Information Modelling: Contextualization, Customization and Curation. In *Proceedings of the 22nd International Conference on Virtual System & Multimedia (VSMM)*, pages 57–64.
- [319] Lehtola, V. V., Kaartinen, H., Nüchter, A., Kaijaluoto, R., Kukko, A., Litkey, P., Honkavaara, E., Rosnell, T., Vaaaja, M. T., Virtanen, J.-P., Kurkela, M., El Issaoui, A., Zhu, L., Jaakkola, A., and Hyyppä, J. (2017). Comparison of the Selected State-Of-The-Art 3D Indoor Scanning and Point Cloud Generation Methods. *Remote Sensing*, 9(8):796.
- [320] Lehtola, V. V., Nikoohemat, S., and Nüchter, A. (2020). Indoor 3D: Overview on Scanning and Reconstruction Methods. In *Handbook of Big Geospatial Data*, chapter 3, pages 55–97. Springer Nature Switzerland AG.
- [321] Leinonen, T., Brinck, J., Vartiainen, H., and Sawhney, N. (2021). Augmented Reality Sandboxes: Children’s Play and Storytelling with Mirror Worlds. *Digital Creativity*, 32(1):38–55.
- [322] Lercari, N. (2016). Terrestrial Laser Scanning in the Age of Sensing. In Forte, M. and Campana, S., editors, *Digital Methods and Remote Sensing in Archaeology*, pages 3–33. Springer.
- [323] Li, C., Fahmy, A., and Sienz, J. (2019a). An Augmented Reality Based Human-Robot Interaction Interface Using Kalman Filter Sensor Fusion. *Sensors*, 19(20):4586.
- [324] Li, J., Wang, C., Kang, X., and Zhao, Q. (2019b). Camera Localization for Augmented Reality and Indoor Positioning: A Vision-Based 3D Feature Database Approach. *International Journal of Digital Earth*, 13(6):727–741.
- [325] Li, J., Yang, B., Chen, D., Wang, N., Zhang, G., and Bao, H. (2019c). Survey and Evaluation of Monocular Visual-Inertial SLAM Algorithms for Augmented Reality. *Virtual Reality Intelligent Hardware*, 1(4):386–410.
- [326] Li, K.-J., Conti, G., Konstantinidis, E., Zlatanova, S., and Bamidis, P. (2019d). OGC IndoorGML: A Standard Approach for Indoor Maps. In Conesa, J., Pérez-Navarro, A., Torres-Sospedra, J., and Montoliu, R., editors, *Geographical and Fingerprinting Data to Create Systems for Indoor Positioning and Indoor/Outdoor Navigation*, chapter 10, pages 187–207. Academic Press.

- [327] Li, K.-J., Zlatanova, S., Torres-Sospedra, J., Perez-Navarro, A., Laoudias, C., and Moreira, A. (2019e). Survey on Indoor Map Standards and Formats. In *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–8.
- [328] Li, L., Su, F., Yang, F., Zhu, H., Li, D., Zuo, X., Li, F., Liu, Y., and Ying, S. (2018a). Reconstruction of Three-Dimensional (3D) Indoor Interiors with Multiple Stories via Comprehensive Segmentation. *Remote Sensing*, 10(8):1281.
- [329] Li, M., Chen, R., Liao, X., Guo, B., Zhang, W., and Guo, G. (2020a). A Precise Indoor Visual Positioning Approach Using a Built Image Feature Database and Single User Image from Smartphone Cameras. *Remote Sensing*, 12(5):869.
- [330] Li, M. and Rottensteiner, F. (2019). Vision-Based Indoor Localization via a Visual SLAM Approach. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W13:827–833.
- [331] Li, N., Becerik-Gerber, B., Krishnamachari, B., and Soibelman, L. (2014). A BIM Centered Indoor Localization Algorithm to Support Building Fire Emergency Response Operations. *Automation in Construction*, 42:78–89.
- [332] Li, R., Liu, Q., Gui, J., Gu, D., and Hu, H. (2017). Indoor Relocalization in Challenging Environments With Dual-Stream Convolutional Neural Networks. *IEEE Transactions on Automation Science and Engineering*, 15(2):651–662.
- [333] Li, X., Yi, W., Chi, H.-L., Wang, X., and Chan, A. P. (2018b). A Critical Review of Virtual and Augmented Reality (VR/AR) Application in Construction Safety. *Automation in Construction*, 86:150–162.
- [334] Li, Y., Bu, R., Sun, M., Wu, W., Di, X., and Chen, B. (2018c). PointCNN: Convolution On χ -Transformed Points. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pages 828–838.
- [335] Li, Y., Li, W., Tang, S., Darwish, W., Hu, Y., and Chen, W. (2020b). Automatic Indoor as-Built Building Information Models Generation by Using Low-Cost RGB-D Sensors. *Sensors*, 20(1):293.
- [336] Lian, Z., Rosin, P. L., and Sun, X. (2008). A Rectilinearity Measurement for 3D Meshes. In *MIR '08: Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval*, page 395–402.
- [337] Lian, Z., Rosin, P. L., and Sun, X. (2010). Rectilinearity of 3D Meshes. *International Journal of Computer Vision*, 89:130–151.
- [338] Liberatore, M. J. and Wagner, W. P. (2021). Virtual, Mixed, and Augmented Reality: A Systematic Review for Immersive Systems Research. *Virtual Reality*, pages 1–27.
- [339] Lim, G. and Doh, N. (2021). Automatic Reconstruction of Multi-Level Indoor Spaces from Point Cloud and Trajectory. *Sensors*, 21(10):3493.
- [340] Lim, G., Oh, Y., Kim, D., Jun, C., Kang, J., and Doh, N. (2020). Modeling of Architectural Components for Large-Scale Indoor Spaces From Point Cloud Measurements. *IEEE Robotics and Automation Letters*, 5(3):3830–3837.

- [341] Limberger, F. A. and Oliveira, M. M. (2015). Real-Time Detection of Planar Regions in Unorganized Point Clouds. *Pattern Recognition*, 48(6):2043–2053.
- [342] Lin, Z., Petzold, F., and Ma, Z. (2019). A Real-Time 4D Augmented Reality System for Modular Construction Progress Monitoring. In *International Association for Automation and Robotics in Construction (IAARC)*, pages 743–748.
- [343] Liu, C., Wu, J., and Furukawa, Y. (2018a). FloorNet: A Unified Framework for Floorplan Reconstruction from 3D Scans. In *European Conference on Computer (ECCV)*, pages 203–219.
- [344] Liu, C., Wu, J., Kohli, P., and Furukawa, Y. (2017). Raster-to-Vector: Revisiting Floorplan Transformation. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2214–2222.
- [345] Liu, J. and Meng, Z. (2020). Visual SLAM With Drift-Free Rotation Estimation in Manhattan World. *IEEE Robotics and Automation Letters*, 5(4):6512–6519.
- [346] Liu, R., Peng, T., Asari, V. K., and Loomis, J. S. (2018b). Real-Time 3D Scene Reconstruction and Localization with Surface Optimization. In *IEEE National Aerospace and Electronics Conference*, pages 280–285.
- [347] Liu, S., Hu, Y., Zeng, Y., Tang, Q., Jin, B., Han, Y., and Li, X. (2018c). See and Think: Disentangling Semantic Scene Completion. In *Conference on Neural Information Processing Systems (NeurIPS)*, pages 1–12.
- [348] Liu, W., Li, Z., Sun, S., Malekian, R., Ma, Z., and Li, W. (2019). Improving Positioning Accuracy of the Mobile Laser Scanning in GPS-Denied Environments: An Experimental Case Study. *IEEE Sensors Journal*, 19(22):10753–10763.
- [349] Liu, Y., Dong, H., Zhang, L., and El Saddik, A. (2018d). Technical Evaluation of HoloLens for Multimedia: A First Look. *IEEE Multimedia*, 25(3):1–7.
- [350] Lloyd, S. (1982). Least Squares Quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137.
- [351] Looser, J., Billingham, M., and Cockburn, A. (2004). Through the Looking Glass: The Use of Lenses as an Interface Tool for Augmented Reality Interfaces. In *GRAPHITE '04 Proceedings of the 2nd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia*, pages 204 – 211.
- [352] López, F. J., Leronés, P. M., Llamas, J., Gómez-García-Bermejo, J., and Zalama, E. (2018). A Review of Heritage Building Information Modeling (H-BIM). *Multimodal Technologies and Interaction*, 2(2):21.
- [353] López Iglesias, J., Díaz Severiano, J. A., Lizcano Amorochó, P. E., Manchado del Val, C., Gómez-Jáuregui, V., Fernández García, O., Preciados Royano, A., and Otero González, C. (2019). Revision of Automation Methods for Scan to BIM. In *International Conference on The Digital Transformation in the Graphic Engineering. INGEGRAF 2019: Advances in Design Engineering*, pages 482–490.
- [354] Lu, Q., Chen, L., Li, S., and Pitt, M. (2020). Semi-Automatic Geometric Digital Twinning for Existing Buildings Based on Images and CAD Drawings. *Automation in Construction*, 115:103183.

- [355] Lu, Q. and Lee, S. (2017). Image-Based Technologies for Constructing As-Is Building Information Models for Existing Buildings. *Journal of Computing in Civil Engineering*, 31(4):04017005.
- [356] Lymberopoulos, D. and Liu, J. (2017). The Microsoft Indoor Localization Competition: Experiences and Lessons Learned. *IEEE Signal Processing Magazine*, 34(5):125–140.
- [357] Lymberopoulos, D., Liu, J., Yang, X., Choudhury, R. R., Handziski, V., and Sen, S. (2015). A Realistic Evaluation and Comparison of Indoor Location Technologies: Experiences and Lessons Learned. In *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*, pages 178–189.
- [358] Ma, L., Jiang, H., Qin, D., and Tan, X. (2020). Vision-Based Indoor Positioning Method by Joint Using 2D Images and 3D Point Cloud Map. In *International Wireless Communications and Mobile Computing (IWCMC)*, pages 1679–1684.
- [359] Ma, Z. and Liu, S. (2018). A Review of 3D Reconstruction Techniques in Civil Engineering and Their Applications. *Advanced Engineering Informatics*, 37:163–174.
- [360] Maalek, R., Lichti, D. D., and Ruwanpura, J. Y. (2019). Automatic Recognition of Common Structural Elements from Point Clouds for Automated Progress Monitoring and Dimensional Quality Control in Reinforced Concrete Construction. *Remote Sensing*, 11(9):1102.
- [361] Maboudi, M., Bánhidi, D., and Gerke, M. (2017). Evaluation of Indoor Mobile Mapping Systems. In *GFaI Workshop 3D North East (20th Application-Oriented Workshop on Measuring, Modeling, Processing and Analysis of 3D-Data)*, pages 125–134.
- [362] Machado, R. L. and Vilela, C. (2020). Conceptual Framework for Integrating BIM and Augmented Reality in Construction Management. *Journal of Civil Engineering and Management*, 26(1):83–94.
- [363] Macher, H., Landes, T., and Grussenmeyer, P. (2015). Point Clouds Segmentation as Base for As-Built BIM Creation. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-5/W3:191–197.
- [364] Macher, H., Landes, T., and Grussenmeyer, P. (2017). From Point Clouds to Building Information Models: 3D Semi-Automatic Reconstruction of Indoors of Existing Buildings. *Applied Sciences*, 7(10):1030.
- [365] MacQueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. In *Berkeley Symposium on Mathematical Statistics and Probability*, volume 5.1, pages 281–297.
- [366] Macé, S., Locteau, H., Valveny, E., and Tabbone, S. (2010). A System to Detect Rooms in Architectural Floor Plan Images. In *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*, pages 167–174.
- [367] Mahmood, B., Han, S., and Lee, D.-E. (2020). BIM-based Registration and Localization of 3D Point Clouds of Indoor Scenes Using Geometric Features for Augmented Reality. *Remote Sensing*, 12(14):2302.
- [368] Marchand, E., Uchiyama, H., and Spindler, F. (2016). Pose Estimation for Augmented Reality: A Hands-On Survey. *IEEE Transactions on Visualization and Computer Graphics*, 22:2633–2651.

- [369] Marchisotti, D. and Zappa, E. (2021). Virtual Simulation Benchmark for the Evaluation of SLAM and 3D Reconstruction Algorithms Uncertainty. *Measurement Science and Technology*, 32(9):095404.
- [370] Martens, J. and Blankenbach, J. (2019). An Automated Approach for Point Cloud Alignment Based on Density Histograms. In *26th International Workshop on Intelligent Computing in Engineering*, pages 1–11.
- [371] Martens, J. and Blankenbach, J. (2020). An Evaluation of Pose-Normalization Algorithms for Point Clouds Introducing a Novel Histogram-Based Approach. *Advanced Engineering Informatics*, 46:101132.
- [372] Maset, E., Magri, L., and Fusiello, A. (2019). Improving Automatic Reconstruction of Interior Walls from Point Cloud Data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W13:849–855.
- [373] Masiero, A., Fissore, F., Guarnieri, A., Pirotti, F., Visintini, D., and Vettore, A. (2018). Performance Evaluation of Two Indoor Mapping Systems: Low-Cost UWB-Aided Photogrammetry and Backpack Laser Scanning. *Applied Sciences*, 8(3):416.
- [374] Mattern, H. and König, M. (2018). BIM-Based Modeling and Management of Design Options at Early Planning Phases. *Advanced Engineering Informatics*, 38:316–329.
- [375] Maturana, D. and Scherer, S. (2015). VoxNet: A 3D Convolutional Neural Network for Real-Time Object Recognition. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 922–928.
- [376] May, M., Clauss, M., and Salzmann, P. (2016). A Glimpse into the Future of Facility and Maintenance Management: A Case Study of Augmented Reality. *Corporate Real Estate Journal*, 6(3):227–244.
- [377] McKinney, K. and Fischer, M. (1998). Generating, Evaluating and Visualizing Construction Schedules with CAD Tools. *Automation in Construction*, 7(6):433–447.
- [378] Meister, S., Kohli, P., Izadi, S., Hämmerle, M., Rother, C., and Kondermann, D. (2012). When Can We Use KinectFusion for Ground Truth Acquisition? In *IEEE/RSJ International Conference on Intelligent Robots and Systems, Workshop on Color-Depth Camera Fusion in Robotics*, pages 1–7.
- [379] Mendez, O., Hadfield, S., Pugeault, N., and Bowden, R. (2019). SeDAR: Reading Floorplans Like a Human - Using Deep Learning to Enable Human-Inspired Localisation. *International Journal of Computer Vision*, 128:1286–1310.
- [380] Mendoza-Silva, G. M., Torres-Sospedra, J., and Huerta, J. (2019). A Meta-Review of Indoor Positioning Systems. *Sensors*, 19(20):4507.
- [381] Meža, S., Žiga Turk, and Dolenc, M. (2014). Component Based Engineering of a Mobile BIM-based Augmented Reality System. *Automation in Construction*, 42:1–12.
- [382] Meža, S., Žiga Turk, and Dolenc, M. (2015). Measuring the Potential of Augmented Reality in Civil Engineering. *Advances in Engineering Software*, 90(C):1–10.

- [383] Michailidis, G.-T. and Pajarola, R. (2015). Automatic Reconstruction of Wall Features under Clutter and Occlusion. In *Computer Graphics International Short Papers*, pages 1–4.
- [384] Microsoft (2020). Sample Code and Documentation for Using the Microsoft HoloLens for Computer Vision Research. <https://github.com/microsoft/HoloLensForCV> (last visited on 12/05/2021).
- [385] Miettinen, R., Kerosuo, H., Metsälä, T., and Paavola, S. (2018). Bridging the Life Cycle: A Case Study on Facility Management Infrastructures and Uses of BIM. *Journal of Facilities Management*, 16(1):2–16.
- [386] Milgram, P. and Kishino, F. (1994). A Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Information Systems*, E77-D(12).
- [387] Mill, T., Alt, A., and Liias, R. (2014). Combined 3D Building Surveying Techniques—Terrestrial Laser Scanning (TLS) and Total Station Surveying for BIM Data Management Purposes. *Journal of Civil Engineering and Management*, 19(1):23–32.
- [388] Millane, A., Oleynikova, H., Lanegger, C., Delmerico, J., Nieto, J., Siegart, R., Pollefeys, M., and Cadena, C. (2021). Freetures: Localization in Signed Distance Function Maps. *IEEE Robotics and Automation Letters*, Early Access:1–8.
- [389] Miller, J., Hoover, M., and Winer, E. (2019). Overcoming Limitations of the HoloLens for Use in Product Assembly. *Electronic Imaging, The Engineering Reality of Virtual Reality*, page 179.
- [390] Mirarchi, C., Pavan, A., De Marco, F., Wang, X., and Song, Y. (2018). Supporting Facility Management Processes through End-Users’ Integration and Coordinated BIM-GIS Technologies. *ISPRS International Journal of Geo-Information*, 7(5):191.
- [391] Mitra, N. J. and Nguyen, A. (2003). Estimating Surface Normals in Noisy Point Cloud Data. In *SCG ’03: Proceedings of the Nineteenth Annual Symposium on Computational Geometry*, page 322–328.
- [392] Mohaimenur Rahman, A. B. M., Li, T., and Wang, Y. (2020). Recent Advances in Indoor Localization via Visible Lights: A Survey. *Sensor*, 20(5):1382.
- [393] Montoliu, R., Sansano, E., Torres-Sospedra, J., and Belmonte, O. (2017). IndoorLoc Platform: A Public Repository for Comparing and Evaluating Indoor Positioning Systems. In *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–8.
- [394] Morar, A., Moldoveanu, A., Mocanu, I., Moldoveanu, F., Radoi, I. E., Asavei, V., Gradinaru, A., and Butean, A. (2020). A Comprehensive Survey of Indoor Localization Methods Based on Computer Vision. *Sensors*, 20(9):2641.
- [395] Moya, B., Badías, A., Alfaro, I., Chinesta, F., and Cueto, E. (2020). Digital Twins that Learn and Correct Themselves. *International Journal for Numerical Methods in Engineering*, pages 1–11.
- [396] Mueller, M. S. and Jutzi, B. (2018). UAS Navigation with SqueezePoseNet—Accuracy Boosting for Pose Regression by Data Augmentation. *Drones*, 2(7):1–21.
- [397] Müller, M. S., Urban, S., and Jutzi, B. (2017). SqueezePoseNet: Image Based Pose Regression With Small Convolutional Neural Networks for Real Time UAS Navigation. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W3:49–57.

- [398] Munzer, B. W., Khan, M. M., Shipman, B., and Mahajan, P. (2019). Augmented Reality in Emergency Medicine: A Scoping Review. *JMIR Journal of Medical Internet Research*, 21(4):e12368.
- [399] Mur-Artal, R., Montiel, J. M. M., and Tardós, J. D. (2015). ORB-SLAM: A Versatile and Accurate Monocular SLAM System. *IEEE Transactions on Robotics*, 31(5):1147–1163.
- [400] Mura, C., Jaspe Villanueva, A., Mattausch, O., Gobbetti, E., and Pajarola, R. (2014a). Reconstructing Complex Indoor Environments with Arbitrary Wall Orientations. In *Eurographics Posters*, pages 1–2.
- [401] Mura, C., Mattausch, O., Jaspe Villanueva, A., Gobbetti, E., and Pajarola, R. (2013). Robust Reconstruction of Interior Building Structures with Multiple Rooms under Clutter and Occlusions. In *International Conference on Computer-Aided Design and Computer Graphics*, pages 52–59.
- [402] Mura, C., Mattausch, O., Jaspe Villanueva, A., Gobbetti, E., and Pajarola, R. (2014b). Automatic Room Detection and Reconstruction in Cluttered Indoor Environments with Complex Room Layouts. *Computers Graphics*, 44:20–32.
- [403] Mura, C., Mattausch, O., and Pajarola, R. (2016). Piecewise-planar Reconstruction of Multi-room Interiors with Arbitrary Wall Arrangements. *Computer Graphics Forum*, 35(7):179–188.
- [404] Mura, C., Pajarola, R., Schindler, K., and Mitra, N. (2021). Walk2Map: Extracting Floor Plans from Indoor Walk Trajectories. In *Eurographics*, volume 40, pages 1–14.
- [405] Murali, S., Speciale, P., Oswald, M. R., and Pollefeys, M. (2017). Indoor Scan2BIM: Building Information Models of House Interiors. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6126–6133.
- [406] Nagel, C., Stadler, A., and Kolbe, T. H. (2009). Conceptual Requirements for the Automatic Reconstruction of Building Information Models from Uninterpreted 3D Models. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 34(XXX):46–53.
- [407] Nakagawa, M. and Nozaki, R. (2018). Geometrical Network Model Generation using Point Cloud Data for Indoor Navigation. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4(4):141–146.
- [408] Natephra, W. and Motamedi, A. (2019). Live Data Visualization of IoT Sensors Using Augmented Reality (AR) and BIM. In *Proceedings of the 36th International Symposium on Automation and Robotics in Construction (ISARC)*, pages 632–638.
- [409] Naticchia, B., Corneli, A., Carbonari, A., Bosché, F., and Principi, L. (2019). Augmented Reality Application Supporting On-Site Secondary Building Assets Management. In Skibniewski, M. J. and Hajdu, M., editors, *Proceedings of the Creative Construction Conference*.
- [410] Nebiker, S., Cavegn, S., and Loesch, B. (2015). Cloud-Based Geospatial 3D Image Spaces - A Powerful Urban Model for the Smart City. *ISPRS International Journal of Geo-Information*, 4(4):2267–2291.
- [411] Neges, H.-M., Koch, C., König, M., and Abramovici, M. (2017). Combining Visual Natural Markers and IMU for Improved AR Based Indoor Navigation. *Advanced Engineering Informatics*, 31:18–31.

- [412] Nessa, A., Adhikari, B., Hussain, F., and Fernando, X. N. (2020). A Survey of Machine Learning for Indoor Positioning. *IEEE Access*, 8:214945–214965.
- [413] Newcombe, R. A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A. J., Kohli, P., Shotton, J., Hodges, S., and Fitzgibbon, A. (2011). KinectFusion: Real-Time Dense Surface Mapping and Tracking. In *IEEE International Symposium on Mixed and Augmented Reality*, pages 127–136.
- [414] Nicał, A. and Wodyński, W. (2016). Enhancing Facility Management through BIM 6D. *Procedia Engineering*, 164:299–306.
- [415] Nießner, M., Zollhöfer, M., Izadi, S., and Stamminger, M. (2013). Real-time 3D Reconstruction at Scale using Voxel Hashing. *ACM Transactions on Graphics (TOG)*, 32(6):169.
- [416] Nikoohemat, S., Diakit , A., Zlatanova, S., and Vosselman, G. (2019). Indoor 3D Modeling and Flexible Space Subdivision from Point Clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W5:285–292.
- [417] Nikoohemat, S., Diakit , A. A., Lehtola, V., Zlatanova, S., and Vosselman, G. (2020a). Consistency Grammar for 3D Indoor Model Checking. *Transactions in GIS*, pages 1–24.
- [418] Nikoohemat, S., Diakit , A. A., Zlatanova, S., and Vosselman, G. (2020b). Indoor 3D Reconstruction from Point Clouds for Optimal Routing in Complex Buildings to Support Disaster Management. *Automation in Construction*, 113:103109.
- [419] Nikoohemat, S., Koeva, M., Oude Elberink, S. J., and Lemmen, C. H. J. (2018a). Change Detection from Point Clouds to Support Indoor 3D Cadastre. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4:451–457.
- [420] Nikoohemat, S., Peter, M., Oude Elberink, S., and Vosselman, G. (2017). Exploiting Indoor Mobile Laser Scanner Trajectories for Semantic Interpretation of Point Clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W4:355–362.
- [421] Nikoohemat, S., Peter, M., Oude Elberink, S., and Vosselman, G. (2018b). Semantic Interpretation of Mobile Laser Scanner Point Clouds in Indoor Scenes Using Trajectories. *Remote Sensing*, 10(11):1754.
- [422] Ning, X., Ma, J., Lv, Z., Xu, Q., and Wang, Y. (2019). Structure Reconstruction of Indoor Scene from Terrestrial Laser Scanner. In *International Conference on E-Learning and Games*, pages 91–98.
- [423] Nocerino, E., Menna, F., Remondino, F., Toschi, I., and Rodr guez-Gonz lvez, P. (2017). Investigation of Indoor and Outdoor Performance of Two Portable Mobile Mapping Systems. In Remondino, F. and Shortis, M. R., editors, *Videometrics, Range Imaging, and Applications XIV*, volume 10332, pages 125–139.
- [424] Nocerino, E., Rodr guez-Gonz lvez, P., and Menna, F. (2019). Introduction to Mobile Mapping with Portable Systems. In Riveiro, B. and Lindenbergh, R., editors, *Laser Scanning: An Emerging Technology in Structural Engineering*, volume 1, chapter 4, pages 37–52. CRC Press.
- [425] Ochmann, S., Vock, R., and Klein, R. (2019). Automatic Reconstruction of Fully Volumetric 3D Building Models from Point Clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 151:251–262.

- [426] Ochmann, S., Vock, R., Wessel, R., and Klein, R. (2015). Automatic Reconstruction of Parametric Building Models from Indoor Point Clouds. *Computers Graphics*, 54:94–103.
- [427] Oesau, S., Lafarge, F., and Alliez, P. (2014). Indoor Scene Reconstruction Using Feature Sensitive Primitive Extraction and Graph-Cut. *ISPRS Journal of Photogrammetry and Remote Sensing*, 90:68–82.
- [428] Oesterreich, T. and Teuteberg, F. (2017). Evaluating Augmented Reality Applications in Construction - A Cost-Benefit Assessment Framework Based on VoFI. In *Proceedings of the 25th European Conference on Information Systems (ECIS)*, pages 1–16.
- [429] Oh, S., Lee, D., Kim, M., Kim, T., and Cho, H. (2021). Building Component Detection on Unstructured 3D Indoor Point Clouds Using RANSAC-Based Region Growing. *Remote Sensing*, 13(2):161.
- [430] Okorn, B., Xiong, X., Akinci, B., and Huber, D. (2010). Toward Automated Modeling of Floor Plans. In *International Symposium 3D Data Processing, Visualization and Transmission 3DPVT*, pages 1–8.
- [431] Oldfield, J., Van Oosterom, P., Beetz, J., and Krijnen, T. F. (2019). Working with Open BIM Standards to Source Legal Spaces for a 3D Cadastre. *ISPRS International Journal of Geo-Information*, 6(11):351.
- [432] Orlosky, J., Toyama, T., Sonntag, D., Sarkany, A., and Lorincz, A. (2014). On-Body Multi-Input Indoor Localization for Dynamic Emergency Scenarios: Fusion of Magnetic Tracking and Optical Character Recognition with Mixed-Reality Display. In *IEEE International Conference on Pervasive Computing and Communication Workshops (PERCOM WORKSHOPS)*, pages 320–325.
- [433] Otero, R., Frías, E., Lagüela, S., and Arias, P. (2020a). Automatic gbXML Modeling from LiDAR Data for Energy Studies. *Remote Sensing*, 12(17):2679.
- [434] Otero, R., Lagüela, S., Garrido, I., and Arias, P. (2020b). Mobile Indoor Mapping Technologies: A Review. *Automation in Construction*, 120:103399.
- [435] Oufqir, Z., El Abderrahmani, A., and Satori, K. (2020). From Marker to Markerless in Augmented Reality. In Bhateja, V., Satapathy, S. C., and Satori, H., editors, *Embedded Systems and Artificial Intelligence*, pages 599–612. Springer.
- [436] Ozturkcan, S. (2021). Service Innovation: Using Augmented Reality in the IKEA Place App. *Journal of Information Technology Teaching Cases*, 11(1):8–13.
- [437] O’Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Velasco-Hernández, G. A., Riordan, D., and Walsh, J. (2019). Adaptive Multimodal Localisation Techniques for Mobile Robots in Unstructured Environments - A Review. In *IEEE 5th World Forum on Internet of Things (WF-IoT)*, pages 799–804.
- [438] Pandey, T., Pena, D., Byrne, J., and Moloney, D. (2021). Leveraging Deep Learning for Visual Odometry Using Optical Flow. *Sensors*, 21(4):1313.
- [439] Pang, Y., Zhang, C., Zhou, L., Lin, B., and Lv, G. (2018). Extracting Indoor Space Information in Complex Building Environments. *ISPRS International Journal of Geo-Information*, 7(321):1–23.

- [440] Panou, C., Ragia, L., Dimelli, D., and Mania, K. (2018). An Architecture for Mobile Outdoors Augmented Reality for Cultural Heritage. *ISPRS International Journal of Geo-Information*, 7(12):463.
- [441] Papadakis, P., Pratikakis, I., Perantonis, S., and Theoharis, T. (2007). Efficient 3D Shape Matching and Retrieval Using a Concrete Radialized Spherical Projection Representation. *Pattern Recognition*, 40:2437–2452.
- [442] Park, C.-S., Lee, D.-Y., Kwon, O.-S., and Wang, X. (2013). A Framework for Proactive Construction Defect Management Using BIM, Augmented Reality and Ontology-Based Data Collection Template. *Automation in Construction*, 33:61–71.
- [443] Pascoe, G., Maddern, W., and Newman, P. (2015). Robust Direct Visual Localisation using Normalised Information Distance. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 70:1–13.
- [444] Pătrăucean, V., Armeni, I., Nahangi, M., Yeung, J., Brilakis, I., and Haas, C. (2015). State of Research in Automatic As-Built Modelling. *Advanced Engineering Informatics*, 29(2):162–171.
- [445] Patti, E., Mollame, A., Erba, D., Dalmaso, D., Osello, A., Macii, E., and Acquaviva, A. (2017). Combining Building Information Modelling and Ambient Data in Interactive Virtual and Augmented Reality Environments. *IT Professional*, 19(3):52–60.
- [446] Payá, L., Gil, A., and Reinoso, O. (2017). A State-of-the-Art Review on Mapping and Localization of Mobile Robots Using Omnidirectional Vision Sensors. *Journal of Sensors*, page 3497650.
- [447] Peasley, B., Birchfield, S., Cunningham, A., and Dellaert, F. (2012). Accurate On-Line 3D Occupancy Grids using Manhattan World Constraints. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5283–5290.
- [448] Pepe, M., Costantino, D., and Garofalo, A. R. (2020). An Efficient Pipeline to Obtain 3D Model for HBIM and Structural Analysis Purposes from 3D Point Clouds. *Applied Sciences*, 10(4):1235.
- [449] Perfetti, L., Polari, C., and Fassi, F. (2017). Fisheye Photogrammetry: Tests and Methodologies for the Survey of Narrow Spaces. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W3:572–580.
- [450] Pertile, M., Chiodini, S., Giubilato, R., Mazzucato, M., Valmorbidia, A., Fornaser, A., Debei, S., and Lorenzini, E. C. (2018). Metrological Characterization of a Vision-Based System for Relative Pose Measurements with Fiducial Marker Mapping for Spacecrafts. *Robotics*, pages 1–29.
- [451] Pfeil, K., Masnadi, S., Belga, J., Sera-Josef, J.-V. T., and LaViola, J. (2021). Distance Perception with a Video See-Through Head-Mounted Display. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 528:1–9.
- [452] Philipp, D., Baier, P., Dibak, C., Dürr, F., Rothermel, K., Becker, S., Peter, M., and Fritsch, D. (2014). MapGENIE: Grammar-Enhanced Indoor Map Construction from Crowd-Sourced Data. In *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 139–147.

- [453] Piasco, N., Sidibé, D., Demonceaux, C., and Gouet-Brunet, V. (2018). A Survey on Visual-Based Localization: On the Benefit of Heterogeneous Data. *Pattern Recognition*, 74:90–109.
- [454] Piniotis, G., Soile, S., Bourexis, F., Tsakiri, M., and Ioannidis, C. (2020). Experimental Assessment of 3D Narrow Space Mapping Technologies. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2020:149–156.
- [455] Pintore, G., Ganovelli, F., Villanueva, A. J., and Gobbetti, E. (2019). Automatic Modeling of Cluttered Multi-Room Floor Plans from Panoramic Images. *Computer Graphics Forum*, 38(7):347–358.
- [456] Pintore, G., Mura, C., Ganovelli, F., Fuentes-Perez, L., Pajarola, R., and Gobbetti, E. (2020). State-of-the-Art in Automatic 3D Reconstruction of Structured Indoor Environments. In *Eurographics*, volume 39, pages 667–699.
- [457] Pipelidis, G. and Prehofer, C. (2018). Models and Tools for Indoor Maps. Technical report, TUM Living Lab Connected Mobility State of the Art Report.
- [458] Placed, J. A. and Castellanos, J. A. (2020). A Deep Reinforcement Learning Approach for Active SLAM. *Applied Sciences*, 10(23):8386.
- [459] Poddar, S., Kottath, R., and Karar, V. (2019). Motion Estimation Made Easy: Evolution and Trends in Visual Odometry. In Hassaballah, M. and Hosny, K. M., editors, *Recent Advances in Computer Vision*, pages 305–331. Springer.
- [460] Ponciano, J.-J., Trémeau, A., and Boochs, F. (2019). Automatic Detection of Objects in 3D Point Clouds Based on Exclusively Semantic Guided Processes. *ISPRS International Journal of Geo-Information*, 8(10):442.
- [461] Potortì, F., Barsocchi, P., Girolami, M., Torres-Sospedra, J., and Montoliu, R. (2015). Evaluating Indoor Localization Solutions in Large Environments Through Competitive Benchmarking: The EvAAL-ETRI Competition. In *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–10.
- [462] Potortì, F., Park, S., Jiménez Ruiz, A. R., Barsocchi, P., Girolami, M., Crivello, A., Lee, S. Y., Lim, J. H., Torres-Sospedra, J., Seco, F., Montoliu, R., Mendoza-Silva, G. M., Del Carmen Pérez Rubio, M., Losada-Gutiérrez, C., Espinosa, F., and Macias-Guarasa, J. (2017). Comparing the Performance of Indoor Localization Systems through the EvAAL Framework. *Sensors*, 17(10):2327.
- [463] Poux, F. and Billen, R. (2019a). A Smart Point Cloud Infrastructure for Intelligent Environments. In Riveiro, B. and Lindenbergh, R., editors, *Laser Scanning - An Emerging Technology in Structural Engineering*, volume 1, chapter 9, pages 127–149. ISPRS Book Series.
- [464] Poux, F. and Billen, R. (2019b). Voxel-based 3D Point Cloud Semantic Segmentation: Unsupervised Geometric and Relationship Featuring vs Deep Learning Methods. *ISPRS International Journal of Geo-Information*, 8(5):213.
- [465] Poux, F., Hallot, P., Neuville, R., and Billen, R. (2016). Smart Point Cloud: Definition and Remaining Challenges. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W1:119–127.

- [466] Poux, F., Mattes, C., and Kobbelt, L. (2020). Unsupervised Segmentation of Indoor 3D Point Cloud: Application to Object-Based Classification. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIV-4/W1-2020:111–118.
- [467] Previtali, M., Barazzetti, L., Brumana, R., and Scaioni, M. (2014a). Towards Automatic Indoor Reconstruction of Cluttered Building Rooms from Point Clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-5:281–288.
- [468] Previtali, M., Díaz-Vilariño, L., and Scaioni, M. (2018). Indoor Building Reconstruction from Occluded Point Clouds Using Graph-Cut and Ray-Tracing. *Applied Sciences*, 8(9):1529.
- [469] Previtali, M., Scaioni, M., Barazzetti, L., and Brumana, R. (2014b). A Flexible Methodology for Outdoor/Indoor Building Reconstruction from Occluded Point Clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-3:119–126.
- [470] Protchenko, K., Dąbrowski, P., and Garbacz, A. (2018). Development and Assessment of VR/AR Solution for Verification During the Construction Process. In *MATEC Web of Conferences. XXVII R-S-P Seminar, Theoretical Foundation of Civil Engineering (27RSP) (TFoCE 2018)*, volume 196, pages 04083:1–6.
- [471] Pu, S. and Vosselman, G. (2009). Knowledge Based Reconstruction of Building Models from Terrestrial Laser Scanning Data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(6):575–584.
- [472] Puente, I., González-Jorge, H., Martínez-Sánchez, J., and Aria, P. (2013). Review of Mobile Mapping and Surveying Technologies. *Measurement*, 46(7):2127–2145.
- [473] Qi, C. R., Su, H., Mo, K., and Guibas, L. J. (2017a). PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 77–85.
- [474] Qi, C. R., Yi, L., Su, H., and Guibas, L. J. (2017b). PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. In *NIPS’17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, page 5105–5114.
- [475] Qian, J., Shamma, D. A., Avrahami, D., and Biehl, J. (2020). Modality and Depth in Touchless Smartphone Augmented Reality Interactions. In *ACM International Conference on Interactive Media Experiences*, page 74–81.
- [476] Qin, G., Zhou, Y., Hu, K., Han, D., and Ying, C. (2021). Automated Reconstruction of Parametric BIM for Bridge Based on Terrestrial Laser Scanning Data. *Advances in Civil Engineering*, page 8899323.
- [477] Qiu, C., Zhou, S., Liu, Z., Gao, Q., and Tan, J. (2019). Digital Assembly Technology Based on Augmented Reality and Digital Twins: A Review. *Virtual Reality Intelligent Hardware*, 1(6):597–610.
- [478] Qu, T. and Sun, W. (2015). Usage of 3D Point Cloud Data in BIM (Building Information Modelling): Current Applications and Challenges. *Journal of Civil Engineering and Architecture*, 9(11):1269–1278.

- [479] Radoi, I. E., Cirimpei, D., and Radu, V. (2019). Localization Systems Repository: A Platform for Open-source Localization Systems and Datasets. In *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–8.
- [480] Raghavan, V., Jung, G. S., and Bollmann, P. (1989). A Critical Investigation of Recall and Precision as Measures of Retrieval System Performance. *ACM Transactions on Information Systems*, 7(3):205–229.
- [481] Ratajczak, J., Marcher, C., Schimanski, C. P., Schweikopfler, A., Riedl, M., and Matt, D. T. (2019). BIM-Based Augmented Reality Tool for the Monitoring of Construction Performance and Progress. In *Proceedings of the 2019 European Conference on Computing in Construction*, pages 467–476.
- [482] Ratajczak, J., Schweikopfler, A., Riedl, M., and Matt, D. T. (2018). Augmented Reality Combined with Location-Based Management System to Improve the Construction Process, Quality Control and Information Flow. In Mutis, I. and Hartmann, T., editors, *Advances in Informatics and Computing in Civil and Construction Engineering*, pages 289–296.
- [483] Rauschnabel, P. A., Rossmann, A., and tom Dieck, M. C. (2017). An Adoption Framework for Mobile Augmented Reality Games: The Case of Pokémon Go. *Computers in Human Behavior*, 76:276–286.
- [484] Real Ehrlich, C. and Blankenbach, J. (2019). Indoor Localization for Pedestrians with Real-Time Capability using Multi-Sensor Smartphones. *Geo-Spatial Information Science*, pages 1–17.
- [485] Reitmayr, G. and Drummond, T. W. (2007). Initialisation for Visual Tracking in Urban Environments. In *6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 161–172.
- [486] Ren, J., Liu, Y., and Ruan, Z. (2016). Architecture in an Age of Augmented Reality: Applications and Practices for Mobile Intelligence BIM-based AR in the Entire Lifecycle. In *Proceedings of the International Conference on Electronic Information Technology and Intellectualization (ICEITI 2016)*, pages 664–674.
- [487] Renaudin, V., Ortiz, M., Perul, J., Torres-Sospedra, J., Jiménez, A. R., Pérez-Navarro, A., Mendoza-Silva, G. M., Seco, F., Landau, Y., Marbel, R., Ben-Moshe, B., Zheng, X., Ye, F., Kuang, J., Li, Y., Niu, X., Landa, V., Hacoheh, S., Shvalb, N., Lu, C., Uchiyama, H., Thomas, D., Shimada, A., Taniguchi, R.-I., Ding, Z., Xu, F., Kronenwett, N., Vladimirov, B., Lee, S., Cho, E., Jun, S., Lee, C., Park, S., Lee, Y., Rew, J., Park, C., Jeong, H., Han, J., Lee, K., Zhang, W., Li, X., Wei, D., Zhang, Y., Park, S. Y., Park, C. G., Knauth, S., Pipelidis, G., Tsiamitros, N., Lungenstrass, T., Morales, J. P., Trogh, J., Plets, D., Opiela, M., Fang, S.-H., Tsao, Y., Chien, Y.-R., Yang, S.-S., Ye, S.-J., Ali, M. U., Hur, S., and Park, Y. (2019). Evaluating Indoor Positioning Systems in a Shopping Mall: The Lessons Learned From the IPIN 2018 Competition. *IEEE Access*, 7:148594–148628.
- [488] Requicha, A. A. G. and Voelcker, H. B. (1977). Constructive Solid Geometry. Technical report, University of Rochester. Production Automation Project. Technical Memorandum 25.
- [489] Riedlinger, U., Oppermann, L., and Prinz, W. (2019). Tango vs. HoloLens: A Comparison of Collaborative Indoor AR Visualisations Using Hand-Held and Hands-Free Devices. *Multimodal Technologies and Interaction*, 3(2):23.

- [490] Rocha, G., Mateus, L., Fernández, J., and Ferreira, V. (2020). A Scan-to-BIM Methodology Applied to Heritage Buildings. *Heritage*, 3(1):47–67.
- [491] Romero-Jarén, R. and Arranz, J. J. (2021). Automatic Segmentation and Classification of BIM Elements from Point Clouds. *Automation in Construction*, 124:103576.
- [492] Rosinol, A., Gupta, A., Abate, M., Shi, J., and Carlone, L. (2020). 3D Dynamic Scene Graphs: Actionable Spatial Perception with Places, Objects, and Humans. In *Robotics: Science and Systems*, pages 1–12.
- [493] Roth, K., Hageny, E., and Gillmann, C. (2020). Shape Analysis and Visualization in Building Floor Plans. In *LEVIA'20: Leipzig Symposium on Visualization in Applications*, pages 1–9.
- [494] Roy, P. and Chowdhury, C. (2021). A Survey of Machine Learning Techniques for Indoor Localization and Navigation Systems. *Journal of Intelligent Robotic Systems*, 101:63.
- [495] Russhakim, N. A. S., Ariff, M. F. M., Majid, Z., Idris, K. M., Darwin, N., Abbas, M. A., Zainuddin, K., and Yusoff, A. R. (2019). The Suitability of Terrestrial Laser Scanning for Building Survey and Mapping Applications. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W9:663–670.
- [496] Ryu, M. W., Oh, S. M., Kim, M. J., Cho, H. H., Son, C. B., and Kim, T. H. (2020). Algorithm for Generating 3D Geometric Representation Based on Indoor Point Cloud Data. *Applied Sciences*, 10(22):8073.
- [497] Sacks, R., Girolami, M., and Brilakis, I. (2020). Building Information Modelling, Artificial Intelligence and Construction Tech. *Developments in the Built Environment*, 4:100011.
- [498] Sahay, P. and Rajagopalan, A. N. (2015). Geometric Inpainting of 3D Structures. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1–7.
- [499] Salgues, H., Macher, H., and Landes, T. (2020). Evaluation of Mobile Mapping Systems for Indoor Surveys. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIV-4/W1-2020:119–125.
- [500] Sanchez, J., Denis, F., Coeurjolly, D., Dupont, F., Trassoudaine, L., and Checchin, P. (2020a). Robust Normal Vector Estimation in 3D Point Clouds through Iterative Principal Component Analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 163:18–35.
- [501] Sanchez, J., Denis, F., Dupont, F., Trassoudaine, L., and Checchin, P. (2020b). Data-Driven Modeling of Building Interiors from LiDAR Point Clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-2-2020:395–402.
- [502] Sanchez, V. and Zakhor, A. (2012). Planar 3D Modeling of Building Interiors from Point Cloud Data. In *IEEE International Conference on Image Processing*, pages 1777–1780.
- [503] Sanchez-Belenguer, C., Wolfart, E., Casado-Coscolla, A., and Sequeira, V. (2020). RISEdb: A Novel Indoor Localization Dataset. In *25th International Conference on Pattern Recognition (ICPR)*, pages 9514–9521.
- [504] Sanhudo, L., Ramos, N. M. M., Poças Martins, J., Almeida, R. M. S. F., Barreira, E., Lurdes Simões, M., and Cardoso, V. (2020). A Framework for In-Situ Geometric Data Acquisition Using Laser Scanning for BIM Modelling. *Journal of Building Engineering*, 28:101073.

- [505] Saracco, R. (2019). Digital Twins: Bridging Physical Space and Cyberspace. *Computer*, 52(12):58–64.
- [506] Sarbolandi, H., Lefloch, D., and Kolb, A. (2015). Kinect range sensing: Structured-light versus Time-of-Flight Kinect. *Computer Vision and Image Understanding*, 139:1–20.
- [507] Satish, A. and Trivedi, J. (2020). Use of BIM and Augmented Reality in Facility Management Phase. In Mahajan, A., Patel, P., and Sharma, P., editors, *Technologies for Sustainable Development*, pages 84–90. CRC Press.
- [508] Sattler, T., Leibe, B., and Kobbelt, L. (2017). Efficient Effective Prioritized Matching for Large-Scale Image-Based Localization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(9):1744–1756.
- [509] Sattler, T., Zhou, Q., Pollefeys, M., and Leal-Taixé, L. (2019). Understanding the Limitations of CNN-Based Absolute Camera Pose Regression. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3297–3307.
- [510] Saurer, O., Fraundorfer, F., and Pollefeys, M. (2012). Homography Based Visual Odometry with Known Vertical Direction and Weak Manhattan World Assumption. In *IROS Workshop on Visual Control of Mobile Robots (ViCoMoR)*, pages 1–6.
- [511] Scargill, T., Hurli, S., Chen, J., and Gorlatova, M. (2021). Will it Move?: Indoor Scene Characterization for Hologram Stability in Mobile AR. In *Proceedings of the 22nd International Workshop on Mobile Computing Systems and Applications*, pages 174–176.
- [512] Schnabel, R., Wahl, R., and Klein, R. (2007). Efficient RANSAC for Point-Cloud Shape Detection. *Computer Graphics Forum*, 26(2):214–226.
- [513] Schöning, J., Rohs, M., Kratz, S., Kratz, S., and Krüger, A. (2009). Map Torchlight: A Mobile Augmented Reality Camera Projector Unit. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, pages 3841–3845.
- [514] Schuster, H.-F. and Weidner, U. (2003). A New Approach Towards Quantitative Evaluation of 3D Building Models. In *ISPRS Comm. IV Joint Workshop "Challenges in Geospatial Analysis"*, pages 1–8.
- [515] Sebastian, R., Olivadese, R., Piaia, E., Di Giulio, R., Bonsma, P., Braun, J.-D., and Riexinger, G. (2018). Connecting the Knowhow of Design, Production and Construction Professionals through Mixed Reality to Overcome Building’s Performance Gaps. *Proceedings*, 2(15):1153.
- [516] Sedaghat, N., Zolfaghari, M., Amiri, E., and Brox, T. (2017). Orientation-boosted Voxel Nets for 3D Object Recognition. In *British Machine Vision Conference (BMVC)*, pages 1–18.
- [517] Sfikas, K., Theoharis, T., and Pratikakis, I. (2011). ROSy+: 3D Object Pose Normalization based on PCA and Reflective Object Symmetry with Application in 3D Object Retrieval. *International Journal of Computer Vision*, 91:262–279.
- [518] Shi, P., Ye, Q., and Zeng, L. (2020a). A Novel Indoor Structure Extraction Based on Dense Point Cloud. *ISPRS International Journal of Geo-Information*, 9(11):660.
- [519] Shi, S., Muralikrishnan, B., and Sawyer, D. (2020b). Terrestrial Laser Scanner Calibration and Performance Evaluation Using the Network Method. *Optics and Lasers in Engineering*, 134:106298.

- [520] Shi, W., Ahmed, W., Li, N., Fan, W., Xiang, H., and Wang, M. (2019). Semantic Geometric Modelling of Unstructured Indoor Point Cloud. *ISPRS International Journal of Geo-Information*, 8(1):9.
- [521] Shigeta, K. and Masuda, H. (2021). Extraction and Recognition of Components from Point Clouds of Industrial Plants. *Computer-Aided Design Applications*, 18(5):890–899.
- [522] Shoushtari, H., Willemsen, T., and Sternberg, H. (2021). Many Ways Lead to the Goal - Possibilities of Autonomous and Infrastructure-Based Indoor Positioning. *Electronics*, 10(4):397.
- [523] Shui, W., Liu, J., Ren, P., Maddock, S., and Zhou, M. (2016). Automatic Planar Shape Segmentation from Indoor Point Clouds. In *ACM SIGGRAPH Conference on Virtual-Reality Continuum and Its Applications in Industry*, volume 1, page 363–372.
- [524] Sidani, A., Duarte, J., Dinis, F., Santos Baptista, J., Poças Martins, J., and Soeiro, A. (2018). The Impact of BIM-Based Virtual and Augmented Reality Interfaces on Health and Safety in Construction Projects: Protocol for a Systematic Review. *IJOOES International Journal of Occupational and Environmental Safety*, 2(1):1–8.
- [525] Sidani, A., Matoseiro Dinis, F., Duarte, J., Sanhudo, L., Calvetti, D., Santos Baptista, J., Poças Martins, J., and Soeiro, A. (2021a). Recent Tools and Techniques of BIM-Based Augmented Reality: A Systematic Review. *Journal of Building Engineering*, 42:102500.
- [526] Sidani, A., Matoseiro Dinis, F., Sanhudo, L., Duarte, J., Santos Baptista, J., Poças Martins, J., and Soeiro, A. (2021b). Recent Tools and Techniques of BIM-Based Virtual Reality: A Systematic Review. *Archives of Computational Methods in Engineering*, 28:449–462.
- [527] Silva, C. S. and Wimalaratne, P. (2017). State-of-Art-in-Indoor Navigation and Positioning of Visually Impaired and Blind. In *Seventeenth International Conference on Advances in ICT for Emerging Regions (ICTer)*, pages 1–9.
- [528] Siniak, N., Źróbek, S., Nikolaiev, V., and Shavrov, S. (2019). Building Information Modeling for Housing Renovation - Example for Ukraine. *Real Estate Management and Valuation*, 27(2):97–107.
- [529] Sithole, G. and Zlatanova, S. (2016). Position, Location, Place and Area: An Indoor Perspective. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, III-4:89–96.
- [530] Soilán, M., Sánchez-Rodríguez, A., del Río-Barral, P., Perez-Collazo, C., Arias, P., and Riveiro, B. (2019). Review of Laser Scanning Technologies and Their Applications for Road and Railway Infrastructure Monitoring. *Infrastructures*, 4(4):58.
- [531] Solla, M., Gonçalves, L. M. S., Gonçalves, G., Francisco, C., Puente, I., Providência, P., Gaspar, F., and Rodrigues, H. (2020). A Building Information Modeling Approach to Integrate Geomatic Data for the Documentation and Preservation of Cultural Heritage. *Remote Sensing*, 12(24):4028.
- [532] Son, H. and Kim, C. (2017). Semantic As-Built 3D Modeling of Structural Elements of Buildings Based on Local Concavity and Convexity. *Advanced Engineering Informatics*, 34:114–124.
- [533] Son, H., Kim, C., and Turkan, Y. (2015). Scan-to-BIM - An Overview of the Current State of the Art and a Look Ahead. In *Proceedings of the 32nd International Symposium on Automation and Robotics in Construction and Mining (ISARC)*, pages 1–8.

- [534] Song, Y., Niu, L., and Li, Y. (2019). Combinatorial Spatial Data Model for Building Fire Simulation and Analysis. *ISPRS International Journal of Geo-Information*, 8(9):408.
- [535] Soudarissanane, S., Lindenbergh, R., Menenti, M., and Teunissen, P. (2011). Scanning Geometry: Influencing Factor on the Quality of Terrestrial Laser Scanning Points. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(4):389–399.
- [536] Speicher, M., Hall, B. D., and Nebeling, M. (2019). What is Mixed Reality? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 537:1–15.
- [537] Staats, B. R., Diakit , A. A., Vo te, R. L., and Zlatanova, S. (2017). Automatic Generation of Indoor Navigable Space Using a Point Cloud and its Scanner Trajectory. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W4:393–400.
- [538] Stachniss, C., Leonard, J. J., and Thrun, S. (2016). Simultaneous Localization and Mapping. In *Springer Handbook of Robotics*, chapter 46, pages 1153–1176. Bruno Siciliano and Oussama Khatib.
- [539] Steptoe, W., Julier, S., and Steed, A. (2014). Presence and Discernability in Conventional and Non-Photorealistic Immersive Augmented Reality. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 213–218.
- [540] Stojanovic, V., Shoushtari, H., Askar, C., Scheider, A., Schuldt, C., Hellweg, N., and Sternberg, H. (2021). A Conceptual Digital Twin for 5G Indoor Navigation. In *MOBILITY 2021 : The Eleventh International Conference on Mobile Services, Resources, and Users*, pages 5–14.
- [541] Stojanovic, V., Trapp, M., Richter, R., Hagedorn, B., and D llner, J. (2018). Towards The Generation of Digital Twins for Facility Management Based on 3D Point Clouds. In *ARCOM 2018: 34th Annual Conference*, pages 1–11.
- [542] Stojanovic, V., Trapp, M., Richter, R., Hagedorn, B., and D llner, J. (2019). Semantic Enrichment of Indoor Point Clouds: An Overview of Progress towards Digital Twinning. In *Architecture in the Age of the 4th Industrial Revolution - Proceedings of the 37th eCAADe and 23rd SIGraDi Conference*, pages 1–10.
- [543] Stoter, J., Arroyo Ohori, K., Dukai, B., Labetski, A., Kumar, K., Vitalis, S., and Ledoux, H. (2020). State of the Art in 3D City Modelling. *GIM International*, 34:1–11.
- [544] Stranner, M., Arth, C., Schmalstieg, D., and Fleck, P. (2019). A High-Precision Localization Device for Outdoor Augmented Reality. In *IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 37–41.
- [545] Straub, J., Bhandari, N., Leonard, J. J., and Fisher III, J. W. (2015). Real-Time Manhattan World Rotation Estimation in 3D. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1913–1920.
- [546] Straub, J., Freifeld, O., Rosman, G., Leonard, J. J., and Fisher III, J. W. (2018). The Manhattan Frame Model - Manhattan World Inference in the Space of Surface Normals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(1):235–249.
- [547] Straub, J., Rosman, G., Freifeld, O., Leonard, J. J., and Fisher III, J. W. (2014). A Mixture of Manhattan Frames: Beyond the Manhattan World. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3770–3777.

- [548] Strauss, W. and Fleischmann, M. (2001). Imagine space fused with data - a model for mixed reality architecture. In *Proceedings of CAST01/Living in Mixed Realities, Conference on Artistic, Cultural and Scientific Aspects of Experimental Media Spaces*, pages 41–45.
- [549] Stumm, S., Schwan, P., Becker, R., Lublasser, E., Blankenbach, J., Vallée, D., Hildebrand, L., and Brell-Cokcan, S. (2017). Towards Life Cycle Complete BIM. In *34th International Symposium on Automation and Robotics in Construction (ISARC 2017)*, pages 632–638.
- [550] Sturm, J., Engelhard, N., Endres, F., Burgard, W., and Cremers, D. (2012). A Benchmark for the Evaluation of RGB-D SLAM Systems. In *Proceedings of the International Conference on Intelligent Robot Systems (IROS)*, pages 573–580.
- [551] Su, H., Jampani, V., Sun, D., Maji, S., Kalogerakis, E., Yang, M.-H., and Kautz, J. (2018a). SPLATNet: Sparse Lattice Networks for Point Cloud Processing. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2530–2539.
- [552] Su, Y., Cai, Z., Shi, L., Zhou, F., Guo, P., Lu, Y., and Wu, J. (2018b). A Multi-Plane Optical See-Through Holographic Three-Dimensional Display for Augmented Reality Applications. *Optik*, 157:190–196.
- [553] Sun, C., Hsiao, C.-W., Sun, M., and Chen, H.-T. (2019). HorizonNet: Learning Room Layout With 1D Representation and Pano Stretch Data Augmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1047–1056.
- [554] Sydora, C. and Stroulia, E. (2018). Augmented Reality on Building Information Models. In *9th International Conference on Information, Intelligence, Systems and Applications (IISA)*, pages 1–4.
- [555] Taira, H., Okutomi, M., Sattler, T., Cimpoi, M., Pollefeys, M., Sivic, J., Pajdla, T., and Torii, A. (2018). InLoc: Indoor Visual Localization with Dense Matching and View Synthesis. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7199–7209.
- [556] Takahashi, H., Date, H., Kanai, S., and Yasutake, K. (2020). Detection of Indoor Attached Equipment from TLS Point Clouds Using Planar Region Boundary. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43:495–500.
- [557] Tamke, M., Zwierzycki, M., Evers, H. L., Ochmann, S., Vock, R., and Wessel, R. (2016). Tracking Changes in Buildings over Time - Fully Automated Reconstruction and Difference Detection of 3D Scan and BIM Files. In *Proceedings of the 34th eCAADe Conference*, pages 643–651.
- [558] Tang, L. and Zhou, J. (2020). Usability Assessment of Augmented Reality-Based Pedestrian Navigation Aid. In *International Conference on Human-Computer Interaction (HCII)*, pages 581–591.
- [559] Tang, P., Huber, D., Akinci, B., Lipman, R., and Lytle, A. (2010). Automatic Reconstruction of As-Built Building Information Models from Laser-Scanned Point Clouds: A Review of Related Techniques. *Automation in Construction*, 19(7):829–843.
- [560] Tangelder, J. W. H. and Veltkamp, R. C. (2007). A Survey of Content Based 3D Shape Retrieval Methods. *Multimedia Tools and Applications*, 39(441):441–471.

- [561] Tao, Y. and Ganz, A. (2020). Simulation Framework for Evaluation of Indoor Navigation Systems. *IEEE Access*, 8:20028–20042.
- [562] Tauscher, H., Lim, J., and Stouffs, R. (2021). A Modular Graph Transformation Rule Set for IFC-to-CityGML Conversion. *Transactions in GIS*, 25(1):261–190.
- [563] Tchaptmi, L. P., Choy, C. B., Armeni, I., Gwak, J., and Savarese, S. (2017). SEGCloud: Semantic Segmentation of 3D Point Clouds. In *International Conference on 3D Vision (3DV)*, pages 537–547.
- [564] Tekavec, J. and Lisec, A. (2020). Cadastral Data as a Source for 3D Indoor Modelling. *Land Use Policy*, 98(104322):1–14.
- [565] Temel, B. A. and Başıağa, H. B. (2020). Investigation of IFC File Format for BIM Based Automated Code Compliance Checking. *Journal of Construction Engineering, Management Innovation*, 3(2):113–130.
- [566] Teulière, C., Marchand, E., and Eck, L. (2015). 3-D Model-Based Tracking for UAV Indoor Localization. *IEEE Transactions on Cybernetics*, 45(5):869–879.
- [567] Thomas, H., Qi, C. R., Deschaud, J.-E., Marcotegui, B., Goulette, F., and Guibas, L. (2019). KPConv: Flexible and Deformable Convolution for Point Clouds. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6410–6419.
- [568] Thomson, C., Apostolopoulos, G., Backes, D., and Boehm, J. (2013). Mobile Laser Scanning for Indoor Modelling. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-5/W2:289–293.
- [569] Thomson, C. and Boehm, J. (2015). Automatic Geometry Generation from Point Clouds for BIM. *Remote Sensing*, 7(9):11753–11775.
- [570] tom Dieck, M. C. and Jung, T. H. (2017). Value of Augmented Reality at Cultural Heritage Sites: A Stakeholder Approach. *Journal of Destination Marketing Management*, 6(2):110–117.
- [571] Torii, A., Taira, H., Sivic, J., Pollefeys, M., Okutomi, M., Pajdla, T., and Sattler, T. (2021). Are Large-Scale 3D Models Really Necessary for Accurate Visual Localization? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(3):814–829.
- [572] Tran, H. and Khoshelham, K. (2019a). A Stochastic Approach to Automated Reconstruction of 3D Models of Interior Spaces from Point Clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W5:299–306.
- [573] Tran, H. and Khoshelham, K. (2019b). Building Change Detection Through Comparison of a Lidar Scan With a Building Information Model. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W13:889–893.
- [574] Tran, H. and Khoshelham, K. (2020). Procedural Reconstruction of 3D Indoor Models from Lidar Data Using Reversible Jump Markov Chain Monte Carlo. *Remote Sensing*, 12(5):838.
- [575] Tran, H., Khoshelham, K., Kealy, A., and Díaz-Vilariño, L. (2017). Extracting Topological Relations Between Indoor Spaces From Point Clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W4:401–406.

- [576] Tsai, T.-H., Chang, C.-H., and Chen, S.-W. (2016). Vision Based Indoor Positioning for Intelligent Buildings. In *2nd International Conference on Intelligent Green Building and Smart Grid (IGBSG)*, pages 1–4.
- [577] Tucci, G., Visintini, D., Bonora, V., and Parisi, E. I. (2018). Examination of Indoor Mobile Mapping Systems in a Diversified Internal/External Test Field. *Applied Sciences*, 8(3):401.
- [578] Turner, E. and Zakhor, A. (2012). Watertight As-Built Architectural Floor Plans Generated from Laser Range Data. In *Second International Conference on 3D Imaging, Modeling, Processing, Visualization Transmission*, pages 316–323.
- [579] Turner, E. and Zakhor, A. (2014). Floor Plan Generation and Room Labeling of Indoor Environments from Laser Range Data. In *International Conference on Computer Graphics Theory and Applications (GRAPP)*, pages 1–12.
- [580] Tuttas, S., Braun, A., Borrmann, A., and Stilla, U. (2016). Evaluation Of Acquisition Strategies For Image-Based Construction Site Monitoring. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B5:734–740.
- [581] Ulrich, M., Wiedemann, C., and Steger, C. (2009). CAD-Based Recognition of 3D Objects in Monocular Images. In *IEEE International Conference on Robotics and Automation*, pages 1191–1198.
- [582] Urban, S. and Jutzi, B. (2017). LaFiDa - A Laserscanner Multi-Fisheye Camera Dataset. *Journal of Imaging*, 3(1):5.
- [583] Urban, S., Leitloff, J., Wursthorn, S., and Hinz, S. (2013). Self-Localization of a Multi-Fisheye Camera Based Augmented Reality System in Textureless 3D Building Models. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-3/W2:43–48.
- [584] Urban, S., Wursthorn, S., Leitloff, J., and Hinz, S. (2016). MultiCol Bundle Adjustment: A Generic Method for Pose Estimation, Simultaneous Self-Calibration and Reconstruction for Arbitrary Multi-Camera Systems. *International Journal of Computer Vision*, 121(2):234–252.
- [585] Usman Ali, M., Hur, S., and Park, Y. (2019). Wi-Fi-Based Effortless Indoor Positioning System Using IoT Sensors. *Sensors*, 19(7):1496:1–20.
- [586] Uygur, I., Miyagusuku, R., Pathak, S., Moro, A., Yamashita, A., and Asama, H. (2020). Robust and Efficient Indoor Localization Using Sparse Semantic Information from a Spherical Camera. *Sensors*, 20(15):4128.
- [587] Väänänen, P. and Lehtola, V. (2019). Inpainting Occlusion Holes in 3D Built Environment Point Clouds. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W17:393–398.
- [588] Valero, E., Adán, A., and Bosché, F. (2016). Semantic 3D Reconstruction of Furnished Interiors Using Laser Scanning and RFID Technology. *Journal of Computing in Civil Engineering*, 30(4):04015053.
- [589] van den Berg, M., Voordijk, H., and Adriaanse, A. (2021). BIM Uses for Deconstruction: An Activity-Theoretical Perspective on Reorganising End-of-Life Practices. *Construction Management and Economics*, 39(4):323–339.

- [590] Vassallo, R., Rankin, A., Chen, E. C. S., and Peters, T. M. (2017). Hologram Stability Evaluation for Microsoft HoloLens. In *Proceedings of SPIE 10136, Medical Imaging 2017: Image Perception, Observer Performance, and Technology Assessment*, page 1013614.
- [591] Vedadi, F. and Valaee, S. (2017). Automatic Visual Fingerprinting for Indoor Image-Based Localization Applications. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 99:1–13.
- [592] Vincke, S., Hernandez, R. D. L., Bassier, M., and Vergauwen, M. (2019). Immersive Visualisation of Construction Site Point Cloud Data, Meshes and BIM Models in a VR Environment Using a Gaming Engine. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-5/W2:77–83.
- [593] Virtanen, J.-P., Daniel, S., Turppa, T., Zhu, L., Julin, A., Hyyppä, H., and Hyyppä, J. (2020). Interactive Dense Point Clouds in a Game Engine. *ISPRS Journal of Photogrammetry and Remote Sensing*, 163:375–389.
- [594] Virtanen, J.-P., Kurkela, M., Turppa, T., Vaaja, M. T., Julin, A., Kukko, A., Hyyppä, J., Ahlavuo, M., Edén von Numers, J., Haggrén, H., and Hyyppä, H. (2018). Depth Camera Indoor Mapping for 3D Virtual Radio Play. *The Photogrammetric Record*, 33(162):171–195.
- [595] Volk, R., Stengel, J., and Schultmann, F. (2014). Building Information Modeling (BIM) for Existing Buildings - Literature Review and Future Needs. *Automation in Construction*, 38:109–127.
- [596] Vosselman, G., Gorte, B. G. H., Sithole, G., and Rabbani, T. (2004). Recognising Structure in Laser Scanner Point Clouds. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVI-8/W2:33–38.
- [597] Wahbeh, W., Kunz, D., Hofmann, J., and Bereuter, P. (2020). Digital Twinning of the Built Environment - an Interdisciplinary Topic for Innovation in Didactics. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-4-2020:231–237.
- [598] Wang, B., Yin, C., Luo, H., Cheng, J. C. P., and Wang, Q. (2021). Fully Automated Generation of Parametric BIM for MEP Scenes Based on Terrestrial Laser Scanning Data. *Automation in Construction*, 125:103615:1–22.
- [599] Wang, C., Cho, Y. K., and Kim, C. (2015). Automatic BIM Component Extraction from Point Clouds of Existing Buildings for Sustainability Applications. *Automation in Construction*, 56:1–13.
- [600] Wang, C., Dai, Y., Elsheimy, N., Wen, C., Retscher, G., Kang, Z., and Lingua, A. (2020a). ISPRS Benchmark on Multisensory Indoor Mapping and Positioning. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-5-2020:117–123.
- [601] Wang, C., Wen, C., Dai, Y., Yu, S., and Liu, M. (2020b). Urban 3D Modeling with Mobile Laser Scanning: A Review. *Virtual Reality Intelligent Hardware*, 2(3):175–212.
- [602] Wang, H., Hu, T., Wang, Z., Kang, Z., Akwensi, P. H., and Yang, J. (2020c). Reconstruction of Power Pylons From LiDAR Point Clouds Based on Structural Segmentation and Parameter Estimation. *IEEE Geoscience and Remote Sensing Letters*, Early Access:1–5.
- [603] Wang, J., Wang, X., Shou, W., and Xu, B. (2014a). Integrating BIM and Augmented Reality for Interactive Architectural Visualisation. *Construction Innovation*, 14(4):453–476.

- [604] Wang, Q., Guo, J., and Kim, M.-K. (2019a). An Application Oriented Scan-to-BIM Framework. *Remote Sensing*, 11(3):365.
- [605] Wang, Q., Zuo, W., Guo, Z., and Li, Q. (2020d). BIM Voxelization Method Supporting Cell-Based Creation of a Path-Planning Environment. *Journal of Construction Engineering and Management*, 146(7):04020080.
- [606] Wang, R., Wan, W., Wang, Y., and Di, K. (2019b). A New RGB-D SLAM Method with Moving Object Detection for Dynamic Indoor Scenes. *Remote Sensing*, 11(10):1143.
- [607] Wang, R., Xie, L., and Chen, D. (2017). Modeling Indoor Spaces Using Decomposition and Reconstruction of Structural Elements. *Photogrammetric Engineering Remote Sensing*, 83(12):827–841.
- [608] Wang, S., Wu, Z., and Zhang, W. (2018). An Overview of SLAM. In *Proceedings of 2018 Chinese Intelligent Systems Conference*, pages 673–681.
- [609] Wang, T.-K. and Piao, Y. (2019). Development of BIM-AR-Based Facility Risk Assessment and Maintenance System. *Journal of Performance of Constructed Facilities*, 33(6):04019068.
- [610] Wang, X., Love, P. E., Kim, M. J., Park, C.-S., Sing, C.-P., and Hou, L. (2013a). A conceptual framework for integrating building information modeling with augmented reality. *Automation in Construction*, 34:37–44.
- [611] Wang, X., Ong, S. K., and Nee, A. Y. C. (2016a). A Comprehensive Survey of Augmented Reality Assembly Research. *Advances in Manufacturing*, 4(1):1–22.
- [612] Wang, X., Truijens, M., Hou, L., Wang, Y., and Zhou, Y. (2014b). Integrating Augmented Reality with Building Information Modeling: Onsite Construction Process Controlling for Liquefied Natural Gas Industry. *Automation in Construction*, 40:96–105.
- [613] Wang, Y. (2020). Multi-Sensor Fusion Tracking Algorithm Based on Augmented Reality System. *IEEE Sensors Journal*, Early Access:1–8.
- [614] Wang, Y., Chen, Q., Zhu, Q., Liu, L., Li, C., and Zheng, D. (2019c). A Survey of Mobile Laser Scanning Applications and Key Techniques over Urban Areas. *Remote Sensing*, 11(13):1540.
- [615] Wang, Y., Hao, W., Ning, X., Zhao, M., Zhang, J., and Shi, Z. (2013b). Automatic Segmentation of Urban Point Clouds Based on the Gaussian Map. *The Photogrammetric Record*, 28(144):342–361.
- [616] Wang, Y., Yang, Q., Zhang, G., and Zhang, P. (2016b). Indoor Positioning System Using Euclidean Distance Correction Algorithm with Bluetooth Low Energy Beacon. *International Conference on Internet of Things and Applications (IOTA)*, pages 243–247.
- [617] Wasenmüller, O. and Stricker, D. (2017). Comparison of Kinect V1 and V2 Depth Images in Terms of Accuracy and Precision. In *Computer Vision – ACCV 2016 Workshops: Asian Conference on Computer Vision*, pages 34–45.
- [618] Weidner, U. and Förstner, W. (1995). Towards Automatic Building Extraction from High-Resolution Digital Elevation Models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 50(4):38–49.
- [619] Weinmann, M. (2016). *Reconstruction and Analysis of 3D Scenes – From Irregularly Distributed 3D Points to Object Classes*. Springer.

- [620] Weinmann, M., Jutzi, B., Hinz, S., and Mallet, C. (2015). Semantic Point Cloud Interpretation Based on Optimal Neighborhoods, Relevant Features and Efficient Classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing*, 105:286–304.
- [621] Weinmann, M., Jäger, M. A., Wursthorn, S., Jutzi, B., Weinmann, M., and Hübner, P. (2020). 3D Indoor Mapping with the Microsoft HoloLens: Qualitative and Quantitative Evaluation by Means of Geometric Features. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-1-2020:165–172.
- [622] Weinmann, M., Wursthorn, S., and Jutzi, B. (2011). Semi-Automatic Image-Based Co-Registration of Range Imaging Data with Different Characteristics. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVIII-3/W22:119–124.
- [623] Werbrouck, J., Pauwels, P., Bonduel, M., Beetz, J., and Bekers, W. (2020). Scan-to-Graph: Semantic Enrichment of Existing Building Geometry. *Automation in Construction*, 119:103286.
- [624] Whelan, T., Kaess, M., Johannsson, H., Fallon, M., Leonard, J. J., and McDonald, J. (2014). Real-Time Large-Scale Dense RGB-D SLAM with Volumetric Fusion. *The International Journal of Robotics Research*, 34(4-5):598–626.
- [625] Wiezel, A. and Becker, R. (1996). CAD Model for Performance Evaluation. *Building and Environment*, 31(4):345–361.
- [626] Wijmans, E. and Furukawa, Y. (2017). Exploiting 2D Floorplan for Building-Scale Panorama RGBD Alignment. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 308–316.
- [627] Williams, T., Szafir, D., Chakraborti, T., and Phillips, E. (2019). Virtual, Augmented, and Mixed Reality for Human-Robot Interaction (VAM-HRI). In *14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 671–672.
- [628] Winter, S., Tomko, M., Vasardani, M., Richter, K.-F., and Khoshelham, K. (2017). Indoor Localization and Navigation Independent of Sensor Based Technologies. *SIGSPATIAL Special*, 9(1):19–26.
- [629] Winter, S., Tomko, M., Vasardani, M., Richter, K.-F., Khoshelham, K., and Kalantari, M. (2019). Infrastructure-Independent Indoor Localization and Navigation. *ACM Computing Surveys*, 52(3):1–24.
- [630] Winterhalter, W., Fleckenstein, F., Steder, B., Spinello, L., and Burgard, W. (2015). Accurate Indoor Localization for RGB-D Smartphones and Tablets given 2D Floor Plans. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3138–3143.
- [631] Wolters, D. (2014). Automatic 3D Reconstruction of Indoor Manhattan World Scenes Using Kinect Depth Data. In *German Conference on Pattern Recognition (GCPR)*, pages 715–721.
- [632] Woodward, C., Hakkarainen, M., Korkalo, O., Kantonen, T., Aittala, M., Rainio, K., and Kähkönen, K. (2010). Mixed Reality for Mobile Construction Site Visualization and Communication. In *Proceedings of the 10th International Conference on Construction Applications of Virtual Reality (CONVR2010)*, pages 35–44.

- [633] Woodward, C., Kuula, T., Honkamaa, P., Hakkarainen, M., and Kemppi, P. (2014). Implementation and Evaluation of a Mobile Augmented Reality System for Building Maintenance. In *Proceedings of the 14th International Conference on Construction Applications of Virtual Reality*, pages 306–315.
- [634] Wu, K., Shi, W., and Ahmed, W. (2020a). Structural Elements Detection and Reconstruction (SEDR): A Hybrid Approach for Modeling Complex Indoor Structures. *ISPRS International Journal of Geo-Information*, 9(12):760.
- [635] Wu, S., Hou, L., and Zhang, G. K. (2021). Integrated Application of BIM and eXtended Reality Technology: A Review, Classification and Outlook. In *Proceedings of the 18th International Conference on Computing in Civil and Building Engineering (ICCCBE)*, pages 1227–1236.
- [636] Wu, Y., Shang, J., Chen, P., Zlatanova, S., Hu, X., and Zhou, Z. (2020b). Indoor Mapping and Modeling by Parsing Floorplan Images. *International Journal of Geographical Information Science*, pages 1–27.
- [637] Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., and Xiao, J. (2015). 3D ShapeNets: A Deep Representation for Volumetric Shape Modeling. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1912–1920.
- [638] Wuest, H., Engekle, T., Wientapper, F., Schmitt, F., and Keil, J. (2016). From CAD to 3D Tracking — Enhancing & Scaling Model-Based Tracking for Industrial Appliances. In *Proceedings of the 2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pages 346–347.
- [639] Xia, H., Zuo, J., Liu, S., and Qiao, Y. (2019). Indoor Localization on Smartphones Using Built-In Sensors and Map Constraints. *IEEE Transactions on Instrumentation and Measurement*, 68(4):1189–1198.
- [640] Xiao, L., Wang, J., Qiu, X., Rong, Z., and Zou, X. (2019). Dynamic-SLAM: Semantic Monocular Visual Localization and Mapping Based on Deep Learning in Dynamic Environment. *Robotics and Autonomous Systems*, 117:1–16.
- [641] Xie, L., Wang, R., Ming, Z., and Chen, D. (2019). A Layer-Wise Strategy for Indoor As-Built Modeling Using Point Clouds. *Applied Sciences*, 9(14):2904.
- [642] Xiong, Q., Zhu, Q., Du, Z., Zlatanova, S., Zhang, Y., Zhou, Y., and Li, Y. (2017). Free Multi-Floor Indoor Space Extraction from Complex 3D Building Models. *Earth Science Informatics*, 10(1):69–83.
- [643] Xiong, X., Adan, A., Akinici, B., and Huber, D. (2013). Automatic Creation of Semantically Rich 3D Building Models from Laser Scanner Data. *Automation in Construction*, 31:325–337.
- [644] Xu, S., Chou, W., and Dong, H. (2019). A Robust Indoor Localization System Integrating Visual Localization Aided by CNN-Based Image Retrieval with Monte Carlo Localization. *Sensors*, 19(2):249.
- [645] Xu, Y. and Stilla, U. (2021). Toward Building and Civil Infrastructure Reconstruction From Point Clouds: A Review on Data and Key Techniques. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:2857–2885.

- [646] Xu, Y., Tong, X., and Stilla, U. (2021). Voxel-Based Representation of 3D Point Clouds: Methods, Applications, and its Potential Use in the Construction Industry. *Automation in Construction*, 126:103675.
- [647] Xue, F., Lu, W., Chen, K., and Zetkucic, A. (2019a). From Semantic Segmentation to Semantic Registration: Derivative-Free Optimization-Based Approach for Automatic Generation of Semantically Rich As-Built Building Information Models from 3D Point Clouds. *Journal of Computing in Civil Engineering*, pages 1–34.
- [648] Xue, F., Wang, X., Li, S., Wang, Q., Wang, J., and Zha, H. (2019b). Beyond Tracking: Selecting Memory and Refining Poses for Deep Visual Odometry. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8567–8575.
- [649] Yan, J., Diakit , A. A., and Zlatanova, S. (2019a). A Generic Space Definition Framework to Support Seamless Indoor/Outdoor Navigation Systems. *Transactions in GIS*, 23(6):1273–1295.
- [650] Yan, J., He, G., Basiri, A., and Hancock, C. (2019b). 3D Passive-Vision-Aided Pedestrian Dead Reckoning for Indoor Positioning. *IEEE Transactions on Instrumentation and Measurement*, pages 1–17.
- [651] Yang, B., Liu, B., Zhu, D., Zhang, B., Wang, Z., and Lei, K. (2020a). Semiautomatic Structural BIM-Model Generation Methodology Using CAD Construction Drawings. *Journal of Computing in Civil Engineering*, 34(3):04020006.
- [652] Yang, F., Zhou, G., Su, F., Zuo, X., Tang, L., Liang, Y., Zhu, H., and Li, L. (2019a). Automatic Indoor Reconstruction from Point Clouds in Multi-Room Environments with Curved Walls. *Sensors*, 19(17):3798.
- [653] Yang, J., Kang, Z., Zeng, L., Akwensi, P. H., and Sester, M. (2021). Semantics-Guided Reconstruction of Indoor Navigation Elements from 3D Colorized Points. *ISPRS Journal of Photogrammetry and Remote Sensing*, 173:238–261.
- [654] Yang, L., Cheng, J. C. P., and Wang, Q. (2020b). Semi-Automated Generation of Parametric BIM for Steel Structures Based on Terrestrial Laser Scanning Data. *Automation in Construction*, 112:103037.
- [655] Yang, L., Dryanovski, I., Valenti, R. G., Wolberg, G., and Xiao, J. (2020c). RGB-D Camera Calibration and Trajectory Estimation for Indoor Mapping. *Autonomous Robots*, 44:485–1503.
- [656] Yang, S., Li, B., Cao, Y., Fu, H., Lai, Y.-K., Kobbelt, L. P., and Hu, S.-M. (2020d). Noise-Resilient Reconstruction of Panoramas and 3D Scenes Using Robot-Mounted Unsynchronized Commodity RGB-D Cameras. *ACM Transactions on Graphics*, 39(5):152.
- [657] Yang, X., Guo, J., Xue, T., and Cheng, K.-T. (2018a). Robust and Real-Time Pose Tracking for Augmented Reality on Mobile Devices. *Multimedia Tools and Applications*, 77:6607–6628.
- [658] Yang, Y., Hancock, C., Kapogiannis, G., Jin, R., de Ligt, H., Yan, J., Chen, C., and Li, C. (2019b). Integrating Indoor Positioning Techniques with Mobile Laser Scanner to Create Indoor Laser Scanning Models. In *FIG Working Week 2019. Geospatial Information for a Smarter Life and Environmental Resilience*, pages 1–16.

- [659] Yang, Y., Jin, S., Liu, R., Kang, S. B., and Yu, J. (2018b). Automatic 3D Indoor Scene Modeling from Single Panorama. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3926–2934.
- [660] Yassin, A., Nasser, Y., Awad, M., Al-Dubai, A., Liu, R., Yuen, C., Raulefs, R., and Aboutanios, E. (2017). Recent Advances in Indoor Localization: A Survey on Theoretical Approaches and Applications. *IEEE Communications Surveys Tutorials*, 19(2):1327–1346.
- [661] Yazdanpour, M., Fan, G., and Sheng, W. (2019). Online Reconstruction of Indoor Scenes With Local Manhattan Frame Growing. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 964–970.
- [662] Yeh, K.-C., Tsai, M.-H., and Kang, S.-C. (2012). On-Site Building Information Retrieval by Using Projection-Based Augmented Reality. *Journal of Computing in Civil Engineering*, 26(3):342–355.
- [663] Yoh, M.-S. (2001). The Reality of Virtual Reality. In *Proceedings Seventh International Conference on Virtual Systems and Multimedia*, pages 1–9.
- [664] Yoon, C., Louie, R., Ryan, J., Vu, M., Bang, H., Derksen, W., and Ruvolo, P. (2019). Leveraging Augmented Reality to Create Apps for People with Visual Disabilities: A Case Study in Indoor Navigation. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, page 210–221.
- [665] Yoon, S., Jung, J., and Heo, J. (2015). Practical Implementation of Semi-Automated As-Built BIM Creation for Complex Indoor Environments. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-4/W5:143–146.
- [666] Yu, Z., Wang, T., Guo, T., Li, H., and Dong, J. (2019). Robust Point Cloud Normal Estimation via Neighborhood Reconstruction. *Advances in Mechanical Engineering*, 11(4):1–19.
- [667] Yung, R. and Khoo-Lattimore, C. (2019). New Realities: A Systematic Literature Review on Virtual Reality and Augmented Reality in Tourism Research. *Current Issues in Tourism*, 22(17):2056–2081.
- [668] Zafari, F., Gkeliass, A., and Leung, K. K. (2019). A Survey of Indoor Localization Systems and Technologies. *IEEE Communications Surveys Tutorials*, 21(3):2568–2599.
- [669] Zaher, M. M., Greenwood, D., and Marzouk, M. M. (2018). Mobile Augmented Reality Applications for Construction Projects. *Construction Innovation*, 18(2):152–166.
- [670] Zamir, A. R. and Shah, M. (2010). Accurate Image Localization Based on Google Maps Street View. In *European Conference on Computer Vision (ECCV)*, pages 255–268.
- [671] Zanuttigh, P., Marin, G., Dal Mutto, C., Dominio, F., Minto, L., and Cortelazzo, G. M. (2016a). Operating Principles of Structured Light Depth Cameras. In Zanuttigh, P., Marin, G., Dal Mutto, C., Dominio, F., Minto, L., and Cortelazzo, G. M., editors, *Time-of-Flight and Structured Light Depth Cameras*, pages 43–79. Springer.
- [672] Zanuttigh, P., Marin, G., Dal Mutto, C., Dominio, F., Minto, L., and Cortelazzo, G. M. (2016b). Operating Principles of Time-of-Flight Depth Cameras. In Zanuttigh, P., Marin, G., Dal Mutto, C., Dominio, F., Minto, L., and Cortelazzo, G. M., editors, *Time-of-Flight and Structured Light Depth Cameras*, pages 81–113. Springer.

- [673] Zavar, H., Arefi, H., Malihi, S., and Maboudi, M. (2021). Topology-Aware 3D Modelling of Indoor Spaces from Point Clouds. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B4-2021:267–274.
- [674] Zhang, D. and Lu, G. (2001). Content-Based Shape Retrieval Using Different Shape Descriptors: A Comparative Study. In *IEEE International Conference on Multimedia and Expo*, pages 317–320.
- [675] Zhang, H. and Xu, F. (2018). MixedFusion: Real-Time Reconstruction of an Indoor Scene with Dynamic Objects. *IEEE Transactions on Visualization and Computer Graphics*, 4(12):3137–3146.
- [676] Zhang, L., Chen, S., Dong, H., and El Saddik, A. (2018). Visualizing Toronto City Data with HoloLens: Using Augmented Reality for a City Model. *IEEE Consumer Electronics Magazine*, 7(3):1–7.
- [677] Zhang, Q., Zhu, M., Zou, L., Li, M., and Zhang, Y. (2020a). Learning Reward Function with Matching Network for Mapless Navigation. *Sensors*, 20(13):3664.
- [678] Zhang, R., Yang, B., Xiao, W., Liang, F., Liu, Y., and Wang, Z. (2019a). Automatic Extraction of High-Voltage Power Transmission Objects from UAV Lidar Point Clouds. *Remote Sensing*, 11(22):2600.
- [679] Zhang, S., Zheng, L., and Tao, W. (2021). Survey and Evaluation of RGB-D SLAM. *IEEE Access*, 9:21367–21387.
- [680] Zhang, X., Lin, J., Li, Q., Liu, T., and Fang, Z. (2020b). Continuous Indoor Visual Localization Using a Spatial Model and Constraint. *IEEE Access*, 8:69800–69815.
- [681] Zhang, Y., Li, D., Wang, H., and Yang, Z.-H. (2019b). Application of Mixed Reality Based on HoloLens in Nuclear Power Engineering. In *International Symposium on Software Reliability, Industrial Safety, Cyber Security and Physical Protection for Nuclear Power Plant*, pages 9–20.
- [682] Zhang, Z. (1994). Iterative Point Matching for Registration of Free-Form Curves and Surfaces. *International Journal of Computer Vision*, 13(2):119–152.
- [683] Zhang, Z. and Scaramuzza, D. (2019). Rethinking Trajectory Evaluation for SLAM: A Probabilistic, Continuous-Time Approach. In *CRA19 Workshop on Dataset Generation and Benchmarking of SLAM Algorithms for Robotics and VR/AR*, pages 1–5.
- [684] Zhao, B., Hua, X., Yu, K., Xuan, W., Chen, X., and Tao, W. (2020a). Indoor Point Cloud Segmentation Using Iterative Gaussian Mapping and Improved Model Fitting. *IEEE Transactions on Geoscience and Remote Sensing*, 58(11):7890–7907.
- [685] Zhao, H., Acharya, D., Tomko, M., and Khoshelham, K. (2020b). Indoor Lidar Relocalization Based on Deep Learning Using a 3D Model. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B1-2020:541–547.
- [686] Zheng, J. M., Chan, K. W., and Gibson, I. (1998). Virtual Reality. *IEEE Potentials*, 17(2):20–23.
- [687] Zheng, Y., Peter, M., Zhong, R., Oude Elberink, S., and Zhou, Q. (2018). Space Subdivision in Indoor Mobile Laser Scanning Point Clouds Based on Scanline Analysis. *Sensors*, 18(6):1838:1–20.
- [688] Zhou, Q., Sattler, T., Pollefeys, M., and Leal-Taixé, L. (2020). To Learn or Not to Learn: Visual Localization from Essential Matrices. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 3319–3326.

- [689] Zhou, Y., Luo, H., and Yang, Y. (2017). Implementation of Augmented Reality for Segment Displacement Inspection During Tunneling Construction. *Automation in Construction*, 82:112–121.
- [690] Zhu, Z., Xu, F., Yan, C., Hao, X., Ji, X., Zhang, Y., and Dai, Q. (2019). Real-Time Indoor Scene Reconstruction with RGBD and Inertial Input. In *IEEE International Conference on Multimedia and Expo (ICME)*, pages 7–12.
- [691] Zlatanova, S., Nourian, P., Gonçalves, R., and Vo, A. V. (2016). Towards 3D Raster GIS: On Developing a Raster Engine for Spatial DBMS. In *ISPRS WG IV/2 Workshop “Global Geospatial Information and High Resolution Global Land Cover/Land Use Mapping”*, pages 45–60.
- [692] Zlatanova, S., Sithole, G., Nakagawa, M., and Zhu, Q. (2013). Problems in Indoor Mapping and Modelling. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-4/W4:63–68.
- [693] Zlatanova, S., Yan, J., Wang, Y., Diakit , A., Isikdag, U., Sithole, G., and Barton, J. (2020). Spaces in Spatial Science and Urban Applications—State of the Art Review. *ISPRS International Journal of Geo-Information*, 9(58):1–28.
- [694] Zou, C., Su, J.-W., Peng, C.-H., Colburn, A., Shan, Q., Wonka, P., Chu, H.-K., and Hoiem, D. (2021). Manhattan Room Layout Reconstruction from a Single 360° Image: A Comparative Study of State-of-the-Art Methods. *International Journal of Computer Vision*, 129:1410–1431.