PFG
https://doi.org/10.1007/s41064-021-00163-y

ORIGINAL ARTICLE

Efficient 3D Mapping and Modelling of Indoor Scenes with the Microsoft HoloLens: A Survey

Martin Weinmann[1] · **Sven Wursthorn**[1] · **Michael Weinmann**[2,3] · **Patrick Hübner**[1]

Received: 22 June 2021 / Accepted: 6 September 2021

Abstract

The Microsoft HoloLens is a head-worn mobile augmented reality device. It allows a real-time 3D mapping of its direct environment and a self-localisation within the acquired 3D data. Both aspects are essential for robustly augmenting the local environment around the user with virtual contents and for the robust interaction of the user with virtual objects. Although not primarily designed as an indoor mapping device, the Microsoft HoloLens has a high potential for an efficient and comfortable mapping of both room-scale and building-scale indoor environments. In this paper, we provide a survey on the capabilities of the Microsoft HoloLens (Version 1) for the efficient 3D mapping and modelling of indoor scenes. More specifically, we focus on its capabilities regarding the localisation (in terms of pose estimation) within indoor environments and the spatial mapping of indoor environments. While the Microsoft HoloLens can certainly not compete in providing highly accurate 3D data like laser scanners, we demonstrate that the acquired data provides sufficient accuracy for a subsequent standard rule-based reconstruction of a semantically enriched and topologically correct model of an indoor scene from the acquired data. Furthermore, we provide a discussion with respect to the robustness of standard handcrafted geometric features extracted from data acquired with the Microsoft HoloLens and typically used for a subsequent learning-based semantic segmentation.

Keywords Microsoft HoloLens · Indoor mapping · Localisation · Geometry acquisition · Scene modelling

Zusammenfassung

Effiziente 3D-Kartierung und -Modellierung von Innenraumszenen mit der Microsoft HoloLens: Ein Überblick. Die Microsoft HoloLens ist ein mobiles Augmented-Reality-System, das als Headset getragen wird. Sie ermöglicht eine Echtzeit-3D-Kartierung ihrer direkten Umgebung und eine Selbstlokalisierung innerhalb der erfassten 3D-Daten. Beide Aspekte sind wesentlich für eine robuste Erweiterung der lokalen Umgebung des Benutzers mit virtuellen Inhalten und für die robuste Interaktion des Benutzers mit virtuellen Objekten. Obwohl die Microsoft HoloLens nicht primär als Indoor-Mapping-System konzipiert ist, bietet sie ein großes Potenzial für ein effizientes und komfortables Erfassen von Innenraumszenen sowohl auf der Basis von einzelnen Räumen als auch auf der Basis von ganzen Gebäuden. In diesem Beitrag wird ein Überblick über das Potenzial der Microsoft HoloLens (Version 1) hinsichtlich einer effizienten 3D-Kartierung und Modellierung von Innenraumszenen gegeben. Insbesondere liegt der Fokus auf den Fähigkeiten der HoloLens hinsichtlich der Lokalisierung (im Sinne einer Posenbestimmung) in Innenräumen sowie der räumlichen Abbildung von Innenraumszenen. Obwohl die Microsoft HoloLens sicherlich nicht mit hochgenauen Systemen wie Laserscannern zur Erfassung von 3D-Daten konkurrieren kann,

✉ Martin Weinmann
martin.weinmann@kit.edu

Sven Wursthorn
sven.wursthorn@kit.edu

Michael Weinmann
mw@cs.uni-bonn.de; m.weinmann@tudelft.nl

Patrick Hübner
patrick.huebner@kit.edu

[1] Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology, Karlsruhe, Germany

[2] Institute of Computer Science II, University of Bonn, Bonn, Germany

[3] EEMCS – Department of Intelligent Systems – Computer Graphics and Visualization Group, Delft University of Technology, Delft, Netherlands

Published online: 19 October 2021

lässt sich zeigen, dass die erfassten Daten eine ausreichende Genauigkeit für die anschließende regelbasierte Rekonstruktion eines semantisch angereicherten und topologisch korrekten Modells einer Innenraumszene aus den erfassten Daten bieten. Darüber hinaus erfolgt eine Diskussion der Robustheit von geometrischen Standardmerkmalen, welche aus mit der Microsoft HoloLens erfassten Daten extrahiert und typischerweise für eine anschließende lernbasierte semantische Segmentierung verwendet werden.

1 Introduction

Due to the technological advancements in recent years, more and more sensor systems have become available for 3D indoor mapping. Recent research particularly focused on the efficient 3D mapping and modelling of indoor scenes, as this enables a rich diversity of applications including scene modelling, navigation and perception assistance, and future use cases like telepresence. Besides 3D reconstruction based on RGB imagery (Remondino et al. 2017; Stathopoulou et al. 2019; Dai et al. 2013), RGB-D data (Zollhöfer et al. 2018) or data acquired via mobile indoor mapping systems (Lehtola et al. 2017; Chen et al. 2018; Nocerino et al. 2017; Masiero et al. 2018), there has also been an increasing interest in the use of Augmented Reality (AR) devices like the Microsoft HoloLens. Although not being primarily designed as an indoor mapping device, such devices also need to satisfy certain constraints regarding indoor mapping, as they need to provide a robust self-localisation (in terms of pose estimation) and a sufficiently accurate spatial mapping to allow for a live augmentation of real scenes with robustly placed virtual contents in the field-of-view of the user. Thus, the Microsoft HoloLens may be used for numerous applications in entertainment, education, navigation, medicine, planning and product design. It may for instance be used for visual guidance during surgery (Gu et al. 2020; Pratt et al. 2018), for visualisation of city models and various types of city data (Zhang et al. 2018) or for the visualisation of 3D objects embedded in the real world (Hockett and Ingleby 2016). Furthermore, the Microsoft HoloLens could be useful for the in-situ visualisation of virtual contents (e.g. Building Information Modelling (BIM) data as shown in Fig. 1 or information directly derived from the acquired data) which, in turn, facilitates numerous applications addressing facility management, cultural heritage documentation or educational services.

To deeper analyse its potential regarding different applications, the Microsoft HoloLens has recently been evaluated regarding its fundamental capabilities as an AR device (Liu et al. 2018; Huang et al. 2018) as well as regarding the spatial stability of holograms (Vassallo et al. 2017). Specifically addressing geometry acquisition within larger indoor environments, further investigations focused on assessing the spatial accuracy of triangle meshes acquired by the Microsoft HoloLens in comparison to highly accurate ground truth data acquired with a terrestrial laser scanning (TLS) system (Khoshelham et al. 2019; Hübner et al. 2019). Beyond geometry acquisition, performance evaluation in recent investigations also addressed the impact of the quality of acquired 3D data on the extraction of geometric features and thus on the results of semantic segmentation (Weinmann et al. 2020), for which both HoloLens data and TLS data were involved.

In this paper, we provide an overview on the capabilities of the Microsoft HoloLens regarding its localisation within indoor environments (Hübner et al. 2020a) and the spatial mapping of indoor environments (Hübner et al. 2019). Beyond geometry reconstruction, we provide a glance view on the reconstruction of models of indoor scenes from unstructured triangle meshes acquired with the Microsoft HoloLens (Hübner et al. 2020b), and we discuss

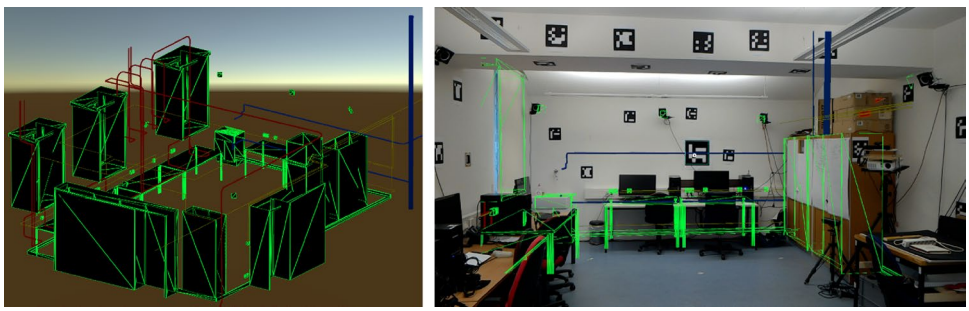

Fig. 1 Augmented reality for indoor scenes (Hübner et al. 2018). The left figure shows a room model where components like tables, cabinets, plug sockets and wall-mounted cameras are depicted with black surface elements and green wireframe, while infrastructure pipelines inside the walls are categorised with respect to heating pipes (red), water pipelines (blue) and power supply lines (yellow). The right figure shows the augmentation of the real room by the room model

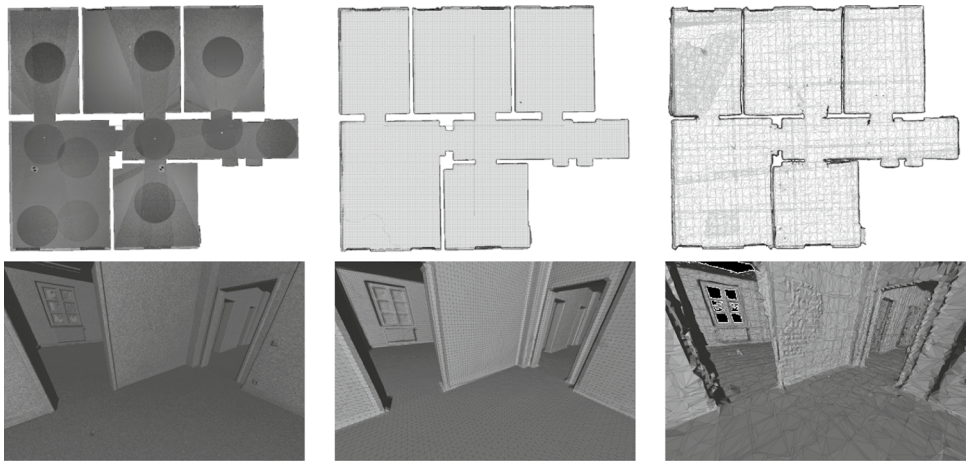

Fig. 2 Visualisation of data acquired for an exemplary indoor scene and shown in nadir view (top row) as well as in detailed oblique view (bottom row) (Weinmann et al. 2020): TLS data acquired with a Leica HDS6000 (left), TLS data downsampled via a voxel-grid filter using a voxel size of 3 cm × 3 cm × 3 cm (center), and HoloLens data (right). While the registration of acquired TLS data was performed manually using artificial markers, the data acquired by the Microsoft HoloLens are directly co-registered during the acquisition stage

the robustness of standard handcrafted geometric features extracted from data acquired with the Microsoft HoloLens (Weinmann et al. 2020).

The paper is organised as follows. We first focus on available sensor systems that can be used for 3D indoor mapping and on the potential of the Microsoft HoloLens in this regard (Sect. 2). In particular, the mapping capabilities of the Microsoft HoloLens rely on a specific sensor design (Sect. 3). Addressing the application of the Microsoft HoloLens for different use cases, we focus on capabilities of the Microsoft HoloLens in terms of localisation within indoor environments (Sect. 4) which are essential for an adequate spatial mapping (Sect. 5) and the subsequent use of the acquired geometric data for rule-based 3D indoor reconstruction (Sect. 6) or learning-based semantic segmentation (Sect. 7). This is followed by a discussion with respect to capabilities regarding localisation, spatial mapping, reconstruction of models of indoor scenes and data-driven learning-based semantic segmentation (Sect. 8). Finally, we provide a summary and concluding remarks (Sect. 9).

2 Sensor Systems for 3D Indoor Mapping

In surveying, Terrestrial Laser Scanning (TLS) systems are often used to achieve a highly accurate geometry acquisition representing the measured counterpart of physical object surfaces. This also holds for indoor environments with weak texture as shown in Fig. 2 for an exemplary indoor scene. While the quality of range measurements generally depends on a variety of influencing factors (Soudarissanane et al. 2011; Weinmann 2016), uncertainties in the range measurements within an indoor scene are mainly caused by (1) characteristics of the observed scene (such as materials, surface reflectivity, surface roughness, etc.), or (2) the scanning geometry (particularly in terms of the relative distance and orientation of object surfaces with respect to the used scanning device). To achieve high scene coverage for an indoor environment with several rooms, a single scan is not sufficient and hence multiple scans have to be acquired from different viewpoints. Since each scan comprises data represented in the local coordinate system of the terrestrial laser scanner, all scans need to be transformed into a common coordinate system (cf. left part of Fig. 2). This process is referred to as point cloud registration and often done manually using artificial markers placed in the scene, which results in a laborious and time-consuming task. Taking furthermore into account a desired data quality, a reasonable number and configuration of viewpoints may be determined based on different dependencies addressing range constraints and/or incidence angle constraints (Soudarissanane and Lindenbergh 2011; Soudarissanane et al. 2011).

A straightforward solution towards more efficient data acquisition is represented by a Mobile Laser Scanning (MLS) system or a Mobile Mapping System[1] (MMS), since all acquired data are directly co-registered on-the-fly. Such sensor systems are meanwhile commonly used for acquiring the geometry of both outdoor scenes (Paparoditis et al. 2012; Gehrung et al. 2017; Roynard et al. 2018; Voelsen et al. 2021) and indoor scenes (Otero et al. 2020). Regarding data acquisition within indoor scenes, different solutions are conceivable such as trolley-based systems (e.g.,

[1] Such systems typically comprise a multi-camera system in combination with one or more multi-profile laser scanners.

the NavVis mobile mapping system (NavVis M6 2021), the Viametris iMS3D mobile mapping system (IMS3D 2021) or the Trimble Indoor Mobile Mapping Solution (TIMMS) (TIMMS Indoor Mapping 2021)), UAV-based systems (Hillemann et al. 2019), backpack-based systems (Nüchter et al. 2015; Filgueira et al. 2016; Blaser et al. 2018) or hand-held systems (e.g., the Leica BLK2GO, Leica BLK2GO 2021)). However, such sensor systems still tend to be rather expensive like TLS systems, and some of the available sensor systems may face major challenges for particular indoor scenes. For instance, trolley-based systems are less practicable for indoor scenes with stairways, while UAV-based systems would typically require an expert to fly the sensor platform through narrow corridors and rooms. In contrast, backpack-based systems need to be carried by the user and have a significant weight, thus reducing applicability when focusing on the mapping of larger indoor environments.

Low-cost solutions for geometry acquisition in indoor scenes are given in the form of RGB-D cameras (e.g., the Microsoft Kinect (Dal Mutto et al. 2012; Smisek et al. 2011) or the Intel RealSense (Intel RealSense Technology 2021)) that can be used as a hand-held device when connected with a unit for data storage and power supply. Similar to mobile laser scanning systems, all data are directly co-registered during the acquisition stage. However, in contrast to laser scanning systems, RGB-D cameras are designed for simultaneously capturing geometric and radiometric information for points on a discrete, regular (and typically rectangular) raster. This can be realised with high frame rates, so that RGB-D cameras also allow an acquisition of dynamic scenes. Addressing both efficient and robust geometry acquisition, KinectFusion (Izadi et al. 2011) and its improved variants (Nießner et al. 2013; Kähler et al. 2016; Dai et al. 2017; Stotko et al. 2019b) are widely used. However, major limitations of RGB-D cameras are typically given regarding the accuracy of geometry acquisition, which might not meet the standard of indoor surveying applications. In particular, errors in geometry acquisition are caused by sensor noise, limited resolution and misalignments due to drift (Zollhöfer et al. 2018), which sometimes necessitates removal of spurious geometry from acquired data. For a detailed survey on geometry acquisition with RGB-D cameras, we refer to the work of Zollhöfer et al. (2018), Dal Mutto et al. (2012), Kolb et al. (2010), and Remondino and Stoppa (2013). For analyses regarding the accuracy of the Microsoft Kinect and Microsoft Kinect v2, we refer to the work of Khoshelham and Oude Elberink (2012) and Lachat et al. (2015), respectively.

Recent technological advancements addressing mobile AR devices have lead to the Microsoft HoloLens (Microsoft HoloLens 2021), a mobile light-weight head-worn AR device. This sensor system allows for live augmentation of real scenes with virtual contents in the field-of-view of the user, which can be helpful for the acquisition and analysis of indoor scenes in terms of an in-situ visualisation of virtual contents like Building Information Modelling (BIM) data or information directly derived from the acquired data, and thus guiding the user in the context of a given application (e.g., about where to look to achieve a complete scene coverage and/or sufficiently dense data representations in the form of point clouds or triangle meshes). Fundamental requirements for an efficient and robust geometry acquisition in indoor scenes address the localisation and the mapping capabilities of the device, both with a reasonable accuracy. In this regard, the Microsoft HoloLens provides the capability to map its direct environment in real-time in the form of triangle meshes and to simultaneously localise itself within the acquired meshes. Knowledge about the geometric structure of the local surrounding and the viewpoint of the device with respect to the geometric structure, in turn, allows for a robust placement of virtual content and enables a realistic interaction with holograms augmenting the real world.

3 The Microsoft HoloLens

The Microsoft HoloLens is a mobile head-worn AR device. To allow for an efficient and robust geometry acquisition in indoor scenes, it is equipped with a variety of sensors. On the one hand, the self-localisation of the device relies on a robust tracking system involving four grey-scale tracking cameras, where two are oriented to the front in a stereo configuration with large overlap, while the other two are oriented to the left and right with nearly no overlap to the center pair. On the other hand, the 3D mapping of the local surrounding is achieved with a time-of-flight camera providing pixel-wise range measurements. In this regard, range images can be queried in two different modes. One mode addresses the range from 0 to 0.8 m ("short throw" mode) and is mostly used for hand gesture recognition, which is for instance important for the interaction of the user with holograms. The other mode addresses the range from 0.8 m to about 3.5 m ("long throw" mode) and is mostly used for geometry acquisition regarding the given indoor scene. Furthermore, a video camera is used for recording videos and imagery, in which the physical environment can be augmented with virtual contents. The respective field-of-view of the involved sensors is illustrated in Fig. 3. For further details, we refer to the work of Hübner et al. (2020a).

4 Localisation Within Indoor Environments

The constraints regarding the self-localisation capabilities of the Microsoft HoloLens are twofold. On the one hand, the device needs to be able to accurately localise itself to

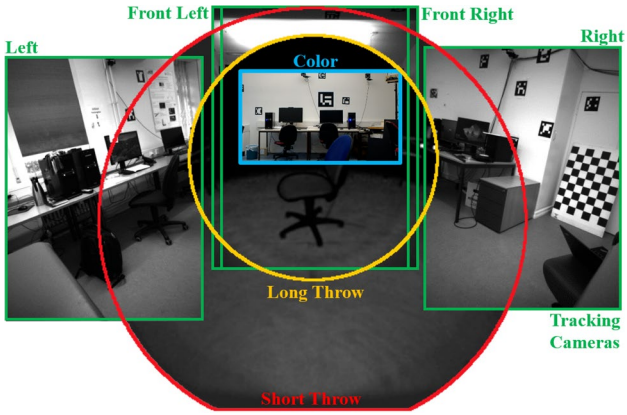

Fig. 3 Overlay of data acquired by the different sensors of the Microsoft HoloLens (Hübner et al. 2020a)

later allow for augmenting indoor environments with virtual room-scale model data with a spatial accuracy of few centimetres. On the other hand, the localisation needs to be robust to allow for aggregating all acquired 3D data in a common coordinate system without significant drift. Consequently, it seems feasible to have a room-scale indoor environment equipped with a motion capture system to allow for quantitative evaluation against a reference trajectory (Sect. 4.1), while a larger (*e.g.*, building-scale) indoor environment seems appropriate for qualitative evaluation in terms of self-induced drift effects (Sect. 4.2).

4.1 Quantitative Evaluation for a Room-Scale Scenario

To obtain a highly accurate ground truth trajectory, we installed the optical motion capture system OptiTrack Prime 17W (OptiTrack 2020) with eight tracking cameras in a laboratory with a size of approximately 8 m × 5 m × 3 m. Furthermore, we equipped the Microsoft HoloLens with a rigid body consisting of five retro-reflecting sphere markers that can easily be tracked by the motion capture system. This allows us to evaluate the performance in terms of localisation for trajectories estimated by the Microsoft HoloLens which is moved by the user within the laboratory.

The spatial offset between the local coordinate system defined by the rigid body and the local coordinate system of the HoloLens is determined via the use of a calibration procedure. The latter relies on the use of a checkerboard pattern observed by the RGB camera of the HoloLens in a static setting. More specifically, the relative pose of the RGB camera of the HoloLens with respect to the local coordinate system of the checkerboard is determined via the Perspective-n-Point (PnP) algorithm (Gao et al. 2003), while the relative pose of the RGB camera with respect to the local coordinate system of the HoloLens is acquired from the Windows 10 SDK. Furthermore, the relative pose between the rigid body and the checkerboard pattern is derived using a tachymeter of type Leica TS06, i.e., the relative poses of the rigid body and the checkerboard with respect to the local coordinate system of the tachymeter are determined via manual measurements of the locations of the sphere targets of the rigid body and the corners of the checkerboard pattern, respectively. From these poses, the spatial offset between the local coordinate system defined by the rigid body and the local coordinate system of the HoloLens may be determined. For more details on the calibration procedure, we refer to Hübner et al. (2020a). To assess the stability of the spatial offset, the distances between the sphere targets were determined with the optical motion capture system before and after a series of measurements conducted within several days, whereby the difference in distance was characterised by a mean of 0.74 mm and a standard deviation of 0.49 mm.

For performance evaluation regarding the self-localisation capabilities of the Microsoft HoloLens, standard evaluation metrics are represented by the Absolute Trajectory Error (ATE) and the Relative Pose Error (RPE) (Sturm et al. 2012) when comparing estimated trajectories against ground truth trajectories. Here, the ATE represents an aggregated measure for tracking quality over a complete trajectory, while the RPE represents a measure accounting for the relative drift between an estimated trajectory and the corresponding ground truth trajectory. For an exemplary movement, the estimated trajectory reveals a mean ATE of about 2 cm and a mean RPE value of about 2 cm in position and about 2° in orientation, as shown in the right part of Fig. 4, whereas the aggregated 3D data are shown in the left part.

4.2 Qualitative Evaluation for a Building-Scale Scenario

We also qualitatively investigated the influence of drift on large-scale trajectories through long corridors typically encountered in large building complexes. To this aim, we focused on the scene depicted in Fig. 5. The user wearing the Microsoft HoloLens started in the basement, walked through the building and used Staircase 1 to get to the ground floor and Staircase 2 to get to the basement again. The total trajectory ended in the same room where it started from, yet the room was re-entered through a different door.

Here, the reference coordinate system of the HoloLens is defined by the conditions when starting the app, (i.e., the origin and orientation of the coordinate system are defined via the pose of the device given when starting the app). For an absolute referencing, e.g., with respect to an existing building model, there is a transformation afterwards which is determined via manually selected tie points and ICP-based refinement. For our scenario, the estimated trajectory and the aggregated 3D data in the form of a triangle

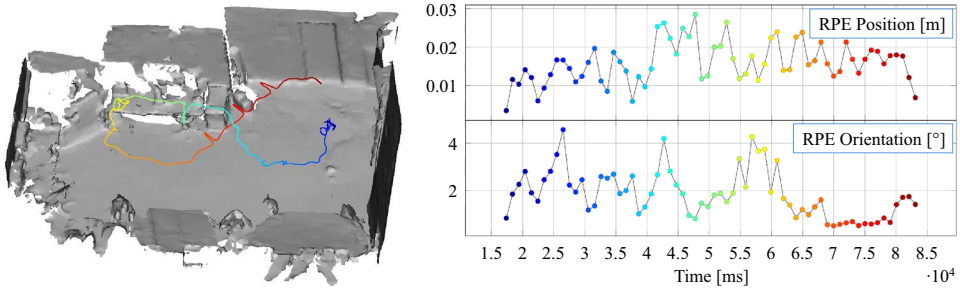

Fig. 4 Trajectory estimated by the Microsoft HoloLens (left) and Relative Pose Error (RPE) with respect to ground truth (right) (Hübner et al. 2020a): the colour encoding of the trajectory represents the acquisition time (blue: beginning of data acquisition; red: end of data acquisition)

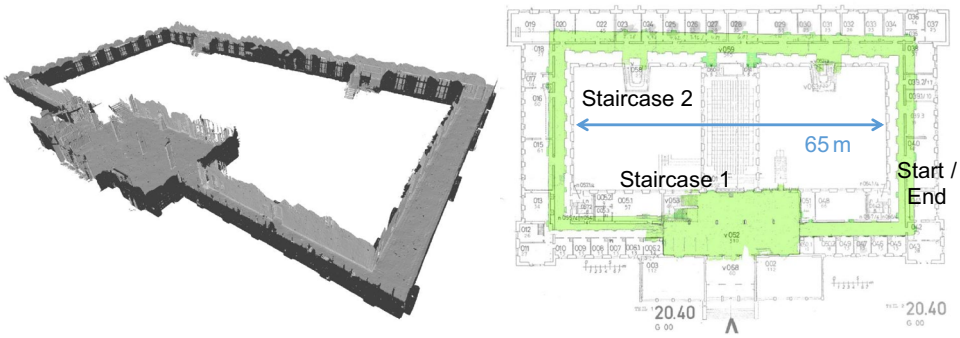

Fig. 5 3D mesh acquired with the Microsoft HoloLens for a building-scale environment (left) and its projection onto the corresponding 2D floor plan of the ground floor (right): yet, for the more complex scenario considered in this paper, the user wearing the Microsoft HoloLens started in the basement, walked through the building and used Staircase 1 to get to the ground floor and Staircase 2 to get to the basement again. The total trajectory ended in the same room where it started from, yet the room was re-entered through a different door

mesh are illustrated in Fig. 6. The estimated trajectory has a total length of 287 m, and the accumulated positional error caused by drift is given by 2.39 m when re-entering the room. After re-entering the room, loop closure is detected by the Microsoft HoloLens and drift-induced errors in position can be corrected for. The corresponding correction of triangle meshes is however not provided as built-in component.

5 Spatial Mapping of Indoor Environments

The accuracy of geometry acquisition with the Microsoft HoloLens (cf. "long throw" mode) depends on different factors. In this regard, the accuracy and stability of range measurements are of particular interest (Sect. 5.1). Furthermore, the accuracy and stability of a 3D mapping of whole indoor environments need to be taken into account (Sect. 5.2).

5.1 Accuracy and Stability of Range Measurements

To analyse the accuracy and stability of range measurements, we used a completely cooled-down Microsoft HoloLens to observe a white and planar wall. The sensor system was placed with a distance of about 1 m to the wall and with an orientation almost perpendicular to the wall. Data were recorded for 100 minutes, whereby the HoloLens app used for data acquisition (HoloLensForCV 2021) was shortly switched off each 25 min in order to avoid the automatic shutdown after 30 min. The automatic shutdown to sleep mode is triggered when the HoloLens is not moved in the meantime, which was exactly the case in this endurance test.

We analysed the temporal variation of the resulting range data relative to the first frame, which is shown in Fig. 7. Here, we may conclude that a warm-up of more than 60 min is required to achieve stable range measurements. For more details as well as further analyses of effects arising from different distances and different incidence angles, we refer to the work of Hübner et al. (2020a).

5.2 Accuracy and Stability of 3D Indoor Mapping

Besides evaluating the accuracy of measurements from a single viewpoint, it is important to also conduct performance evaluation of the Microsoft HoloLens regarding the task of 3D indoor mapping. For this purpose, we consider a specific indoor scene which represents an empty apartment consisting of five rooms of different size and one central hallway as shown in Fig. 2. After a renovation phase, the geometry

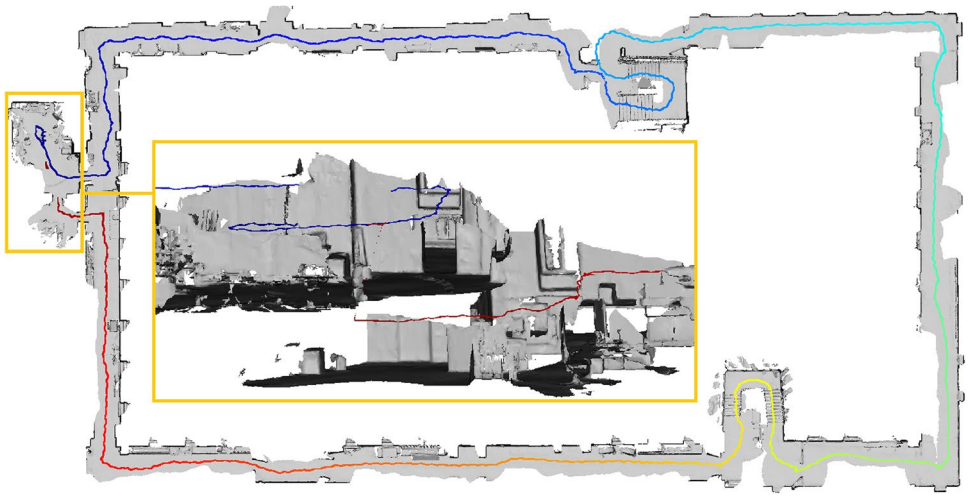

Fig. 6 Trajectory estimated by the Microsoft HoloLens when walking through a building (Hübner et al. 2020a): the trajectory has a length of 287 m and its colour encoding represents the acquisition time (blue: beginning of data acquisition; red: end of data acquisition). The trajectory within the indoor environment is given in nadir view, while a zoom on the room with the start and end position (small orange rectangle) is provided in side view (large orange rectangle) to highlight the drift-induced errors in position and the effect of loop closure (with respect to trajectory only)

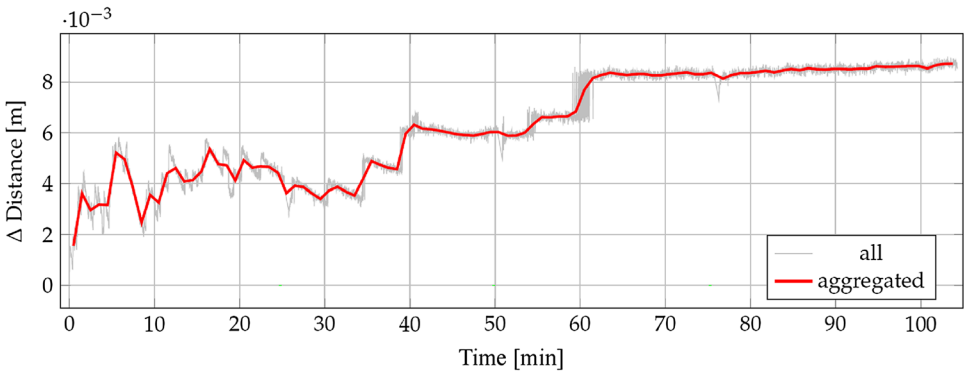

Fig. 7 Warm-up behaviour of the time-of-flight camera integrated in the Microsoft HoloLens (Hübner et al. 2020a): the plot shows the stability of measurements over time and relative to the first frame

of this apartment needed to be acquired in the scope of a project, before the apartment was fully equipped again.

In the scope of our work, we compare the geometric data acquired with the Microsoft HoloLens to the geometric data acquired with a TLS system of type Leica HDS6000. The latter has been mounted on a tripod and provides range measurements with survey-grade accuracy (within a few mm range) in a field-of-view of 360° × 155° in horizontal and vertical direction (i.e., the part below the laser scanner is occluded by the tripod and hence discarded from the scan grid). Thus, the highly accurate geometry acquisition achieved with the TLS system can be considered as ground truth.

For ground truth geometry acquisition with the TLS system, 11 scans were acquired from the positions indicated with a circle in the left part of Fig. 2 to obtain complete scene coverage. Since each scan contains data represented in the local coordinate system of the terrestrial laser scanner, the scans have to be transferred into a common coordinate system. For this purpose, artificial planar and spherical markers were placed in the apartment and used to establish correspondences for the subsequent determination of the transformation parameters (i.e., the relative pose between the single scans). Subsequently, the complete point cloud was manually cleaned in terms of removing minor artefacts in the scans. Furthermore, the data was thinned to an average point distance of 1 cm and meshed with the Poisson surface reconstruction algorithm (Kazhdan et al. 2006) used from the software MeshLab (Cignoni et al. 2008).

For geometry acquisition with the Microsoft HoloLens, a user wearing the device walked through the apartment and thus captured the geometry of the indoor scene within a few minutes. To create the triangle mesh, we used the commercially available SpaceCatcher HoloLens App (SpaceCatcher HoloLens App 2018), which allowed directly visualizing the triangle meshes for the user while

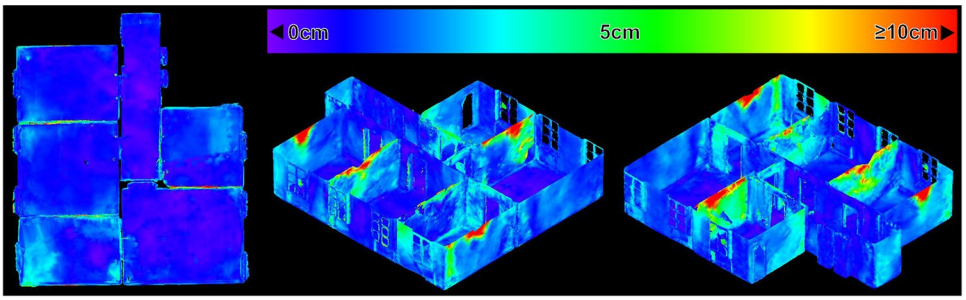

Fig. 8 Mean Hausdorff distance across all possible combinations for comparing two out of five triangle meshes of the complete apartment, where each triangle mesh has been captured with the Microsoft HoloLens (Hübner et al. 2019): nadir view (left) and two side views (center and right)

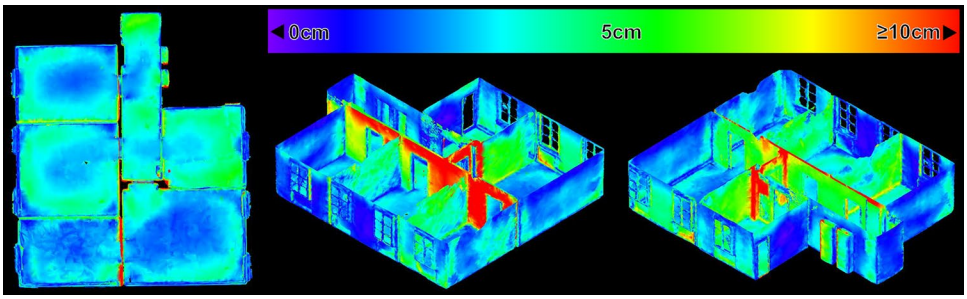

Fig. 9 Mean Hausdorff distance between five triangle meshes captured with the Microsoft HoloLens and the TLS ground truth (Hübner et al. 2019)

they were recorded. A resulting exemplary mesh is visualised in the right part of Fig. 2 and contains 105,200 vertices. To facilitate comparison, the acquired HoloLens meshes were aligned with the TLS ground truth mesh via semi-automatic point cloud registration in CloudCompare (CloudCompare 2018). This was achieved by means of manually selected tie points and subsequent fine registration based on the Iterative Closest Point (ICP) algorithm (Besl and McKay 1992).

For performance evaluation, we focused on two criteria: repeatability and accuracy. For analyses regarding repeatability, the geometry of the indoor scene was acquired five times by the user wearing the Microsoft HoloLens (Hübner et al. 2019). Between consecutive acquisitions, all data on the device were deleted to ensure independent measurements. For comparing two HoloLens meshes with each other, we use the Hausdorff distance (Cignoni et al. 1998) as evaluation metric, which indicates the distance of each point in one point cloud to its nearest point in the other point cloud. As the Hausdorff distance only allows comparing two point clouds or meshes with each other, we calculate it for each possible pair consisting of two out of the five given triangle meshes of the complete apartment, and finally we derive the mean Hausdorff distance across the 10 possible combinations. The mean Hausdorff distance calculated across all pairs of HoloLens meshes is visualised in Fig. 8. For most parts of the scene, the figure reveals deviations of a few centimetres between the compared HoloLens meshes. However, larger deviations are given near the ceiling, where some of the HoloLens meshes exhibit holes (as the ceiling itself was not scanned in the course of all performed acquisitions). Thus, the Microsoft HoloLens performs the spatial mapping of indoor environments with low variation between independent measurements.

For analyses regarding accuracy, the acquired HoloLens meshes were also compared with the TLS ground truth (Hübner et al. 2019), where a scale factor of about 1.012 was observed between HoloLens meshes and TLS ground truth. Again, we use the Hausdorff distance (Cignoni et al. 1998) as evaluation metric for comparing two triangle meshes. The mean Hausdorff distance calculated across all pairs comprising a HoloLens mesh and the TLS ground truth mesh is visualised in Fig. 9. For most parts of the scene, the figure reveals small deviations between the HoloLens meshes and the TLS ground truth. However, for some inner walls parallel to the doors connecting the rooms, larger deviations can be observed. In fact, the transition spaces between neighbouring rooms and the weak texture of the considered scene are challenging for the tracking component of the HoloLens device, thus also affecting the correctness of the spatial mapping. To assess whether the geometry of the rooms is accurately acquired, we extracted the averaged HoloLens mesh for each of the rooms and aligned these meshes with their TLS counterpart based on the same semi-automatic point cloud registration procedure as mentioned before. The resulting mean Hausdorff distance on a per-room basis is visualised in Fig. 10 and indicates low deviations between the HoloLens meshes and the TLS ground truth.

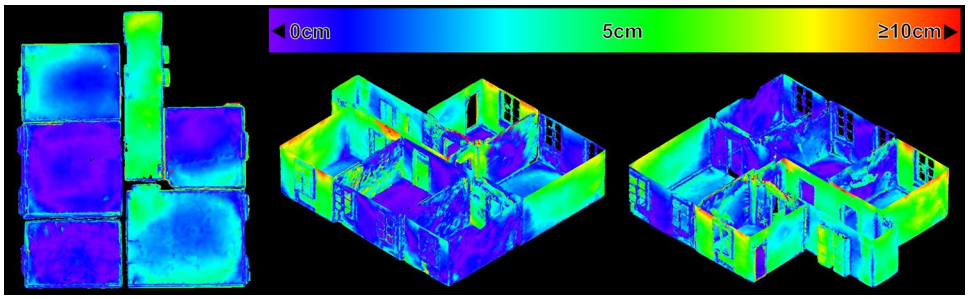

Fig. 10 Mean Hausdorff distance between five triangle meshes captured with the Microsoft HoloLens and the TLS ground truth (Hübner et al. 2019), when considering each room separately and registering the corresponding 3D data against the ground truth data

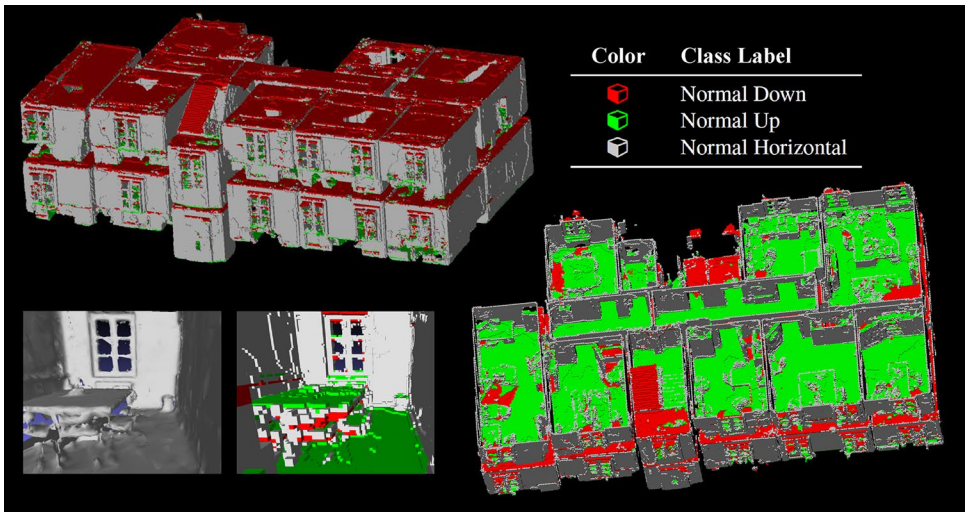

Fig. 11 Voxel representation, where voxels are classified with respect to their normal vector. A close-up view for a room with furniture is depicted on the bottom left part

6 Reconstruction of Scene Models

Beyond the mere acquisition of the geometry of an indoor scene, a topic of great interest is represented by the reconstruction of models of indoor scenes from unstructured 3D data in the form of either point clouds or triangle meshes. In this regard, the resulting structured indoor model includes both semantic information (e.g., with respect to ceiling, floor, walls, wall openings, or furniture) and information about topological relationships (e.g., with respect to room adjacency or accessibility through sufficiently large wall openings).

To advance from an acquired HoloLens triangle mesh to a structured indoor model, we make use of a voxel-based reconstruction approach (Hübner et al. 2020b, 2021) which allows a reconstruction by means of assigning both semantic labels and room labels to the given voxels. The approach thus performs both semantic segmentation, which relies on a cascade of rule-based procedures (for ceiling and floor reconstruction, voxel classification, voxel model refinement regarding wall geometry and wall openings, etc.), and instance segmentation in terms of room partitioning. In this paper, we use a voxel resolution of 5 cm, but the approach is not restricted to this design choice (Hübner et al. 2020b, 2021).

In the following, we consider a dataset acquired with the Microsoft HoloLens by a user walking through an indoor office environment with multiple rooms on two storeys including furniture. The total extent of the scene is given with 13 m × 21 m × 8 m. This indoor scene was selected, because it is much more challenging for state-of-the-art indoor reconstruction approaches than for instance the scene represented in Fig. 2 due to the more complex room layout, the additional furniture, the different storeys and the connecting staircase. Figure 11 shows an initial voxel representation, where voxels are classified with respect to their normal vector, while Fig. 12 shows the reconstructed model of the indoor scene and its room topology. For more technical details and more results, we refer to the work of Hübner et al. (2020b, 2021), while used datasets and implementations can be accessed via links provided in the work of Hübner et al. (2021).

7 Semantic Segmentation

While the reconstruction of indoor models still typically relies on rule-based approaches, a different avenue of research addresses a learning-based semantic segmentation on point level. Here, we may expect a more significant

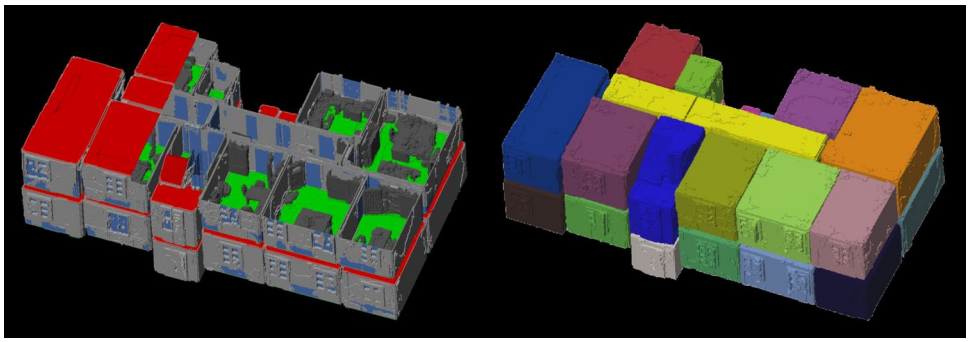

Fig. 12 Reconstructed model of the indoor scene (left) and reconstructed room topology (right): the colour encoding of the reconstructed model addresses the classes *Ceiling* (red), *Floor* (green), *Wall* (grey), *Wall Opening* (blue) and *Interior Object* (dark grey), while the colour encoding of the reconstructed room topology represents different rooms in different colours

impact of the quality of the acquired 3D data on the behaviour and expressiveness of geometric features and thus the results of a subsequent semantic segmentation relying on these features.

For a scenario with few training data available, it is practicable to follow the classic strategy of a point-wise extraction of handcrafted features and the use of these features as input to a standard supervised classification technique which, in turn, delivers a semantic labelling with respect to defined class labels. Since a 3D point has no spatial dimensions, we need to take into account its neighbouring 3D points to describe the local 3D structure. For this purpose, we select a spherical neighbourhood parameterised by the number of nearest neighbours, where the latter is determined locally-adaptive and individually for each 3D point of the point cloud via eigenentropy-based scale selection (Weinmann 2016). Based on these neighbourhoods, we extract a set of 17 standard geometric features used in a diversity of applications (Weinmann et al. 2017b; Weinmann 2016; Weinmann et al. 2017a). Since each of these features represents one single property of the local neighbourhood by a single value, the features are rather intuitive and their behaviour can easily be interpreted. As classifier, we use a Random Forest (Breiman 2001) as representative of standard discriminative classification approaches.

For performance evaluation, we use the data acquired with the Microsoft HoloLens as shown in the right part of Fig. 2 and a downsampled version of the TLS data as shown in the center part of Fig. 2. Here, the downsampling has been achieved by applying a voxel-grid filter using a voxel size of 3 cm × 3 cm × 3 cm and a subsequent Poisson Surface Reconstruction (Kazhdan et al. 2006) resulted in a mesh containing 178,322 vertices. Furthermore, a ground truth labelling has been obtained via manual annotation and it addresses the three classes *Ceiling*, *Floor* and *Wall*, as these information are helpful for subsequent tasks such as guiding the user of an AR device during the acquisition regarding scene completion and densification of sparsely reconstructed areas. We randomly select 1000 points per class for training and all remaining points for performance evaluation. To allow reasoning about the impact of the quality of acquired 3D data on the behaviour and expressiveness of the considered intuitive geometric features, a visualisation of their behaviour across the complete mesh is provided in Fig. 13 for the downsampled TLS data and for the HoloLens data, respectively. Furthermore, the achieved classification results are visualised in Fig. 14. For the downsampled TLS dataset, this corresponds to an OA of 98.10% and F1-scores of 97.58%, 97.57% and 98.44% for classes *Ceiling*, *Floor* and *Wall*, respectively. For the HoloLens dataset, this corresponds to an OA of 93.36% and F1-scores of 90.69%, 92.02% and 94.70% for classes *Ceiling*, *Floor* and *Wall*, respectively. For more detailed analyses, we refer to the work of Weinmann et al. (2020).

8 Discussion

So far, we mainly focused on the capabilities of the Microsoft HoloLens with respect to localisation and spatial mapping. Regarding localisation, results achieved for both room-scale and building-scale indoor environments clearly reveal a robust tracking of the sensor system (Sect. 4) which is a prerequisite for the spatial stability of virtual objects placed in the scene as perceived by the user wearing the Microsoft HoloLens. While small drift effects are accumulated with the travelled distance, corrections in terms of pose estimation can be made via loop closure detection. Regarding spatial mapping, a warm-up behaviour of the Microsoft HoloLens has to be taken into account (Sect. 5.1), where more than 60 min are recommended to achieve stable range measurements, which in turn are required for an accurate geometry acquisition. The results achieved for the spatial mapping of indoor scenes clearly reveal the high potential

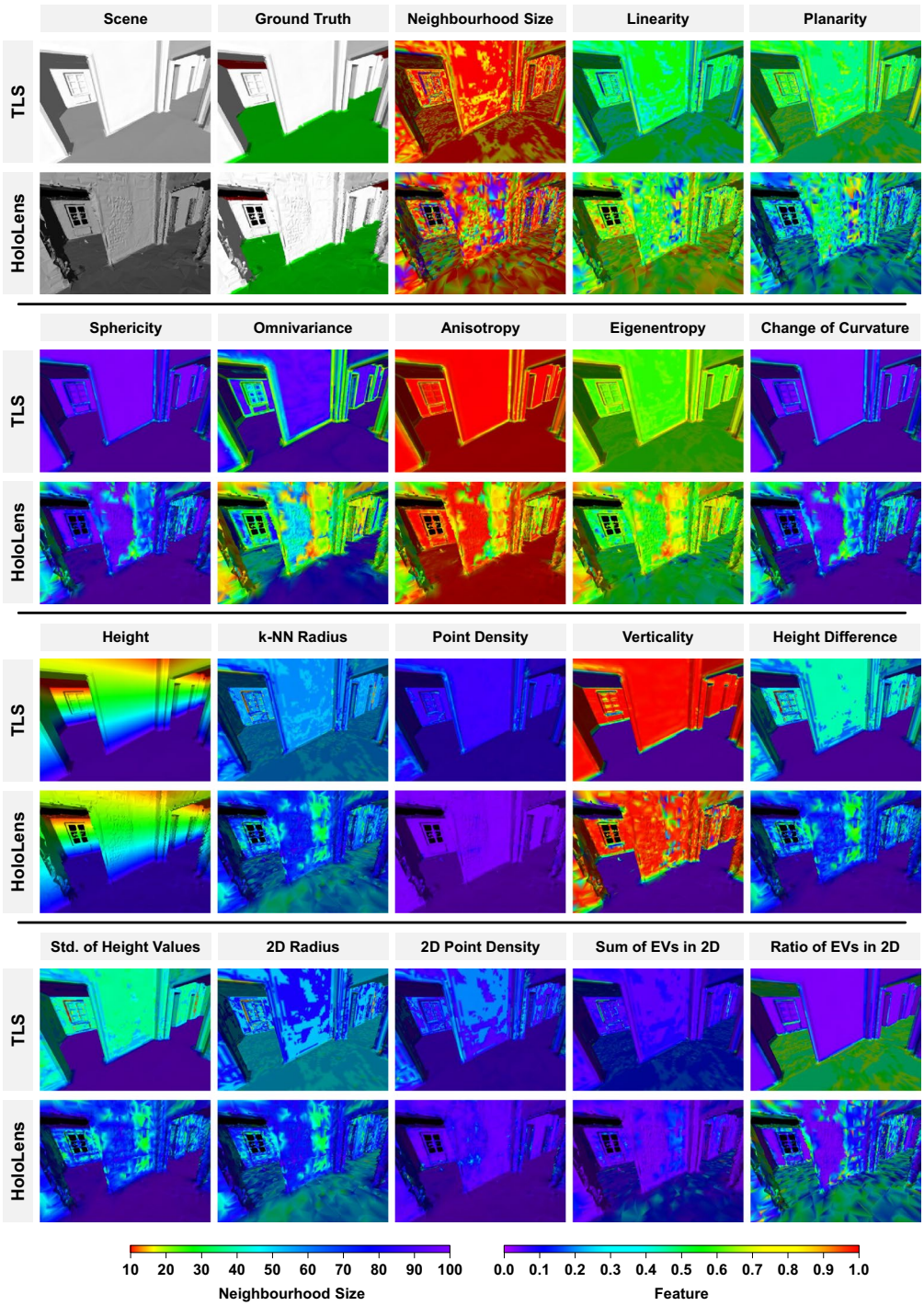

Fig. 13 Visualisation of the scene, a ground truth labelling addressing three classes (*Ceiling*: red; *Floor*: green; *Wall*: white), the locally-adaptive neighbourhood size determined via eigenentropy-based scale selection (Weinmann 2016), and 17 low-level geometric features (Weinmann 2016; Weinmann et al. 2017b) for a downsampled version of data acquired with a TLS system and for data acquired with the Microsoft HoloLens (Weinmann et al. 2020): the neighbourhood size addresses the locally-adaptive number of nearest neighbours determined via eigenentropy-based scale selection, while all features are scaled to the interval [0, 1]

of the Microsoft HoloLens for efficient and easy-to-use mapping of basic indoor building geometry (Sect. 5.2). However, during the spatial mapping, major challenges are faced when the user wearing the device walks through doors and thus through transition spaces between neighbouring rooms. In such situations, the localisation is prone to errors. A further

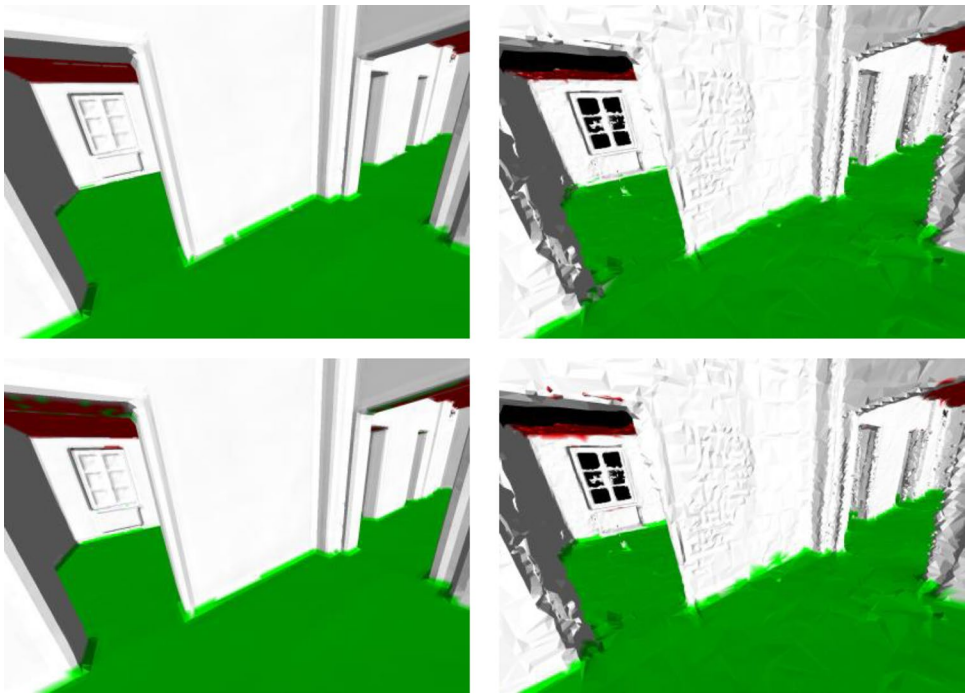

Fig. 14 Visualisation of the ground truth labelling (top row) and the achieved classification results (bottom row) for a downsampled version of data acquired with a TLS system (left) and for data acquired with the Microsoft HoloLens (right) when addressing the classes *Ceiling* (red), *Floor* (green) and *Wall* (white)

reason for such errors might be represented by the weakly-textured indoor scene given for an empty apartment. In additional experiments for the same apartment, but fully equipped, localisation errors were significantly lower and even the thickness of the walls in the acquired data matched well to the ground truth (Hübner et al. 2020a). Furthermore, the measurement accuracy in the range of a few centimetres might not be sufficient to recover fine details of indoor scenes such as light switches, power sockets or door handles for empty apartments and specific lamps, houseplants or thin objects in general for fully equipped apartments.

Beyond geometry acquisition, we focused on the reconstruction of models of indoor scenes from the unstructured 3D data acquired by the Microsoft HoloLens (Sect. 6). Again, achieved results clearly demonstrate the potential of the Microsoft HoloLens, as the efficient geometry acquisition for indoor scenes provides data of sufficient quality for a state-of-the-art voxel-based reconstruction approach relying on a cascade of rule-based procedures.

However, the achieved results also reveal that we may expect challenges for considerations on point level. For instance, applications relying on a data-driven learning-based semantic segmentation (Sect. 7) encounter error propagation from the measurements to subsequent processing steps such as the extraction of geometric features. Such features are known to be sensitive to noise and measurement errors (Dittrich et al. 2017). Regarding the behaviour and expressiveness of such geometric features for 3D data of different quality, the conducted comparison reveals that some features are more and others less affected by the lower quality of the data acquired with the Microsoft HoloLens. Furthermore, the visualisations directly allow a reasoning about expressive features (e.g., height, verticality, or the ratio of the eigenvalues of the 2D structure tensor) and less-expressive features (e.g., radii of the local neighbourhood in 3D and 2D, the local point densities in 3D and 2D, or the sum of the eigenvalues of the 2D structure tensor) with respect to the considered classification task addressing three classes. The suitability of these features might vary for more complex classification tasks (e.g., addressing more classes with a higher similarity) or for more complex indoor scenes (e.g., scenes covering different storeys and/or also containing room inventory). Finally, the comparison of the classification results achieved for both datasets reveals a decrease in OA of about 4.74% when using the HoloLens for data acquisition.

9 Conclusions

In this paper, we provided a survey on the capabilities of the Microsoft HoloLens for efficient 3D indoor mapping. More specifically, we focused on its capabilities regarding the localisation within indoor environments and the spatial mapping of indoor environments. Being not primarily designed as an indoor mapping device, but as a mobile AR headset instead, the capabilities of the Microsoft HoloLens in terms of geometry acquisition are focused on the needs of an AR device. Here, only the geometric structure of the local environment around the user needs to be consistently known

so that it may be augmented with virtual contents. However, the achieved experimental results demonstrate that the Microsoft HoloLens has a high potential for a diversity of applications. As a head-worn AR device, it is easy to use for non-expert users and it allows an efficient and comfortable mapping of basic indoor building geometry. The promising mapping capabilities hold for both room-scale and building-scale indoor environments, and they are mainly due to the interplay of a robust localisation and a spatial mapping with an accuracy in the range of a few centimetres, which is sufficient for many subsequent tasks like the reconstruction of semantically enriched and topologically correct models of indoor scenes or the navigation of a user through an indoor scene.

In future work, we intend to further investigate the potential of the Microsoft HoloLens for the detection of fine-grained object categories with smaller objects that are typical for indoor scenes, e.g., the ones proposed in the work of Chang et al. (2017). Furthermore, to foster research on the processing of data acquired with the Microsoft HoloLens, we recently released several datasets of different complexity that can be used as benchmark datasets for semantic segmentation of 3D data (Hübner et al. 2021). This complements the ISPRS Benchmark on Indoor Modelling (Khoshelham et al. 2017, 2020), where the benchmark data comprise six indoor scenes, each captured with a different sensor (data acquisition was performed using a TLS system, a trolley-based system, a backpack-based system and three hand-held systems). Based on this diversity of datasets, a comprehensive evaluation of the performance of both standard and deep learning approaches seems desirable. Future work might also address automatic scene completion in terms of handling missing or occluded parts of an indoor scene after the acquisition by the user wearing the Microsoft HoloLens. This could be achieved in a similar way as proposed for street-based mobile mapping systems involving 2D range image representations of the acquired 3D point cloud (Biasutti et al. 2018) or via the joint estimation of both the geometry and the semantics of a scene with partially incomplete object surfaces (Roldão et al. 2021).

Besides geometry acquisition, research might also address the visualisation of potential changes in an indoor scene. In this regard, related work focuses on the use of an RGB-D camera for data acquisition within an indoor scene and the subsequent creation of a scene model of the empty room, taking into account light sources, given materials and room geometry (Zhang et al. 2016). In addition to a realistic rendering of the empty room under the same lighting conditions, the proposed framework allows for an editing of the scene in terms of adding furniture, changing material properties of walls, and relighting. This in turn might be interesting for diverse AR applications involving the Microsoft HoloLens for architectural purposes. Well-reconstructed environments with a segmentation into individual object entities may also serve as a key for user interaction with individual objects, e.g., in terms of a gesture-based control of a microcontroller (Schütt et al. 2019) to switch lamps on or off, respectively.

Furthermore, the Microsoft HoloLens may be of great relevance for collaborative AR scenarios (Sereno et al. 2020) and remote collaboration scenarios. Previous approaches on sharing live experiences in a certain user's environment are based on real-time scene capture as well as efficient data streaming and VR-based visualisation for remote users (Stotko et al. 2019a, b) as required for efficient consulting or maintenance purposes that reduce the need for physical on-site presence of experts. This may be extended to also allow for live visualisation of the current state of the acquired scene to the user performing the 3D scene capture. Knowledge about the current state in turn would allow an adaptive acquisition in terms of guiding the user about where to look to acquire data for still missing scene parts or to densify data in areas of low point density (or large triangles, respectively). Besides an in-situ visualisation of virtual contents like the progress of geometry acquisition, information directly derived from the acquired data, or BIM data for the user on-site, it might be interesting to allow remotely immersed users to conduct distance measurements, select objects or perform annotations in the acquired data similar to the work of Zingsheim et al. (2021) and to additionally stream these information to the user on-site performing data acquisition.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Funding Not applicable.

Availability of data and material Not applicable.

Code availability Not applicable.

References

Besl PJ, McKay ND (1992) A method for registration of 3-D shapes. IEEE Trans Pattern Anal Mach Intell 14(2), 239–256

Biasutti P, Aujol J-F, Brédif M, Bugeau A (2018) Disocclusion of 3D LiDAR point clouds using range images. ISPRS Ann Photogramm Remote Sens Spat Inf Sci IV-1/W1, pp 75–82

Blaser S, Cavegn S, Nebiker S (2018) Development of a portable high performance mobile mapping system using the robot operating system. ISPRS Ann Photogramm Remote Sens Spat Inf Sci IV-1, pp 13–20

Breiman L (2001) Random forests. Mach Learn 45(1), 5–32

Chang A, Dai A, Funkhouser T, Halber M, Nießner M, Savva M, Song S, Zeng A, Zhang Y (2017) Matterport3D: learning from RGB-D data in indoor environments. In: Proceedings of the international conference on 3D vision, pp 667–676

Chen Y, Tang J, Jiang C, Zhu L, Lehtomäki M, Kaartinen H, Kaijaluoto R, Wang Y, Hyyppä J, Hyyppä H, Zhou H, Pei L, Chen R (2018) The accuracy comparison of three simultaneous localization and mapping (SLAM)-based indoor mapping technologies. Sensors 18(10):3228

Cignoni P, Callieri M, Corsini M, Dellepiane M, Ganovelli F, Ranzuglia G (2008) MeshLab: an open-source mesh processing tool. In: Proceedings of the Eurographics Italian chapter conference, pp 129–136

Cignoni P, Rocchini C, Scopigno R (1998) Metro: measuring error on simplified surfaces. Comput Graph Forum 17(2), 167–174

CloudCompare (2018) CloudCompare 2.10-alpha. https://www.danielgm.net/cc/. Last accessed Dec 2018

Dai A, Nießner M, Zollhöfer M, Izadi S, Theobalt C (2017) BundleFusion: real-time globally consistent 3D reconstruction using on-the-fly surface reintegration. ACM Trans Graph 36(3):24

Dai F, Rashidi A, Brilakis I, Vela P (2013) Comparison of image-based and time-of-flight-based technologies for three-dimensional reconstruction of infrastructure. J Constr Eng Manag 139(1), 929–939

Dal Mutto C, Zanuttigh P, Cortelazzo GM (2012) Time-of-flight cameras and Microsoft Kinect(TM). Springer, New York

Dittrich A, Weinmann M, Hinz S (2017) Analytical and numerical investigations on the accuracy and robustness of geometric features extracted from 3D point cloud data. ISPRS J Photogramm Remote Sens 126:195–208

Filgueira A, Arias P, Bueno M, Lagüela S (2016) Novel inspection system, backpack-based, for 3D modelling of indoor scenes. In: Proceedings of the international conference on indoor positioning and indoor navigation

Gao X-S, Hou X-R, Tang J, Cheng H-F (2003) Complete solution classification for the perspective-three-point problem. IEEE Trans Pattern Recognit Mach Intell 25(8), 930–943

Gehrung J, Hebel M, Arens M, Stilla U (2017) An approach to extract moving objects from MLS data using a volumetric background representation. ISPRS Ann Photogramm Remote Sens Spat Inf Sci IV-1/W1, pp 107–114

Gu W, Shah K, Knopf J, Navab N, Unberath M (2020) Feasibility of image-based augmented reality guidance of total shoulder arthroplasty using Microsoft HoloLens 1. https://doi.org/10.1080/21681163.2020.1835556

Hillemann M, Weinmann M, Müller MS, Jutzi B (2019) Automatic extrinsic self-calibration of mobile mapping systems based on geometric 3D features. Remote Sens 11(16):1955:

Hockett P, Ingleby T (2016) Augmented reality with HoloLens: experiential architectures embedded in the real world. arXiv:1610.04281v1

HoloLensForCV. https://github.com/microsoft/HoloLensForCV. Last accessed 11 Aug 2021

Huang J, Yang B, Chen J (2018) A non-contact measurement method based on HoloLens. Int J Perform Eng 14(1), 144–150

Hübner P, Landgraf S, Weinmann M, Wursthorn S (2019) Evaluation of the Microsoft HoloLens for the mapping of indoor building environments. In: Kersten TP (ed) Dreiländertagung der OVG, DGPF und SGPF: Photogrammetrie – Fernerkundung – Geoinformation – 2019. DGPF, pp 44–53

Hübner P, Weinmann M, Wursthorn S (2018) Marker-based localization of the Microsoft HoloLens in building models. Int Arch Photogramm Remote Sens Spat Inf Sci XLII-1, pp 195–202

Hübner P, Clintworth K, Liu Q, Weinmann M, Wursthorn S (2020a) Evaluation of HoloLens tracking and depth sensing for indoor mapping applications. Sensors 20(4):1021

Hübner P, Weinmann M, Wursthorn S (2020b) Voxel-based indoor reconstruction from HoloLens triangle meshes. ISPRS Ann Photogramm Remote Sensors Spat Inf Sci V-4-2020, pp 79–86

Hübner P, Weinmann M, Wursthorn S, Hinz S (2021) Automatic voxel-based 3D indoor reconstruction and room partitioning from triangle meshes. ISPRS J Photogramm Remote Sens. (In print)

IMS3D – Trolley mobile scanner. https://viametris.com/trolley-mobile-scanners/. Last accessed: 11 Aug 2021

Intel RealSense Technology. https://www.intel.com/content/www/us/en/architecture-and-technology/realsense-overview.html. Last accessed 11 Aug 2021

Izadi S, Kim D, Hilliges O, Molyneaux D, Newcombe R, Kohli P, Shotton J, Hodges S, Freeman D, Davison A, 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, pp 559–568

Kähler O, Prisacariu VA, Murray DW (2016) Real-time large-scale dense 3D reconstruction with loop closure. In: Proceedings of the European conference on computer vision, pp 500–516

Kazhdan M, Bolitho M, Hoppe H (2006) Poisson surface reconstruction. In: Proceedings of the fourth Eurographics symposium on geometry processing, pp 61–70

Khoshelham K, Díaz Vilariño L, Peter M, Kang Z, Acharya D (2017) The ISPRS benchmark on indoor modelling. Int Arch Photogramm Remote Sens Spat Inf Sci XLII-2/W7, pp 367–372

Khoshelham K, Oude Elberink S (2012) Accuracy and resolution of Kinect depth data for indoor mapping applications. Sens 12(2), 1437–1454

Khoshelham K, Tran H, Acharya D (2019) Indoor mapping eyewear: geometric evaluation of spatial mapping capability of HoloLens. Int Arch Photogramm Remote Sens Spat Inf Sci XLII-2/W13, pp 805–810

Khoshelham K, Tran H, Acharya D, Díaz Vilariño L, Kang Z, Dalyot S (2020) The ISPRS benchmark on indoor modelling–Preliminary results. Int Arch Photogramm Remote Sens Spat Inf Sci XLIII-B5-2020, pp 207–211

Kolb A, Barth E, Koch R, Larsen R (2010) Time-of-flight cameras in computer graphics. Comput Graph Forum 29(1), 141–159

Lachat E, Macher H, Landes T, Grussenmeyer P (2015) Assessment and calibration of a RGB-D camera (Kinect v2 sensor) towards a potential use for close-range 3D modeling. Remote Sens 7(10), 13070–13097

Lehtola VV, Kaartinen H, Nüchter A, Kaijaluoto R, Kukko A, Litkey P, Honkavaara E, Rosnell T, Vaaja MT, Virtanen J-P, Kurkela M, El Issaoui A, Zhu L, Jaakkola A, Hyyppä J (2017) Comparison of the selected state-of-the-art 3D indoor scanning and point cloud generation methods. Remote Sens 9(8):796

Leica BLK2GO. https://blk2go.com. Last accessed 11 Aug 2021

Liu Y, Dong H, Zhang L, El Saddik A (2018) Technical evaluation of HoloLens for multimedia: a first look. IEEE MultiMedia 25(4):8–18

Masiero A, Fissore F, Guarnieri A, Pirotti F, Visintini D, Vettore A (2018) Performance evaluation of two indoor mapping systems: low-cost UWB-aided photogrammetry and backpack laser scanning. Appl Sci 8(3):416

Microsoft HoloLens. https://www.microsoft.com/en-us/hololens. Last accessed 11 Aug 2021

NavVis M6 – Scalable reality capture. https://www.navvis.com/m6. Last accessed 11 Aug 2021

Nießner M, Zollhöfer M, Izadi S, Stamminger M (2013) Real-time 3D reconstruction at scale using voxel hashing. ACM Trans Graph 32(6):169

Nocerino E, Menna F, Remondino F, Toschi I, Rodríguez-Gonzálvez P (2017) Investigation of indoor and outdoor performance of two portable mobile mapping systems. Proc SPIE 10332:125–139

Nüchter A, Borrmann D, Koch P, Kühn M, May S (2015) A man-portable, IMU-free mobile mapping system. ISPRS Ann Photogramm Remote Sens Spat Inf Sci II-3/W5, pp 17–23

OptiTrack. Available online: https://www.optitrack.com/products/prime-17w/. Last accessed 15 Jan 2020

Otero R, Lagüela S, Garrido I, Arias P (2020) Mobile indoor mapping technologies: a review. Aut Constr 120:103399

Paparoditis N, Papelard J-P, Cannelle B, Devaux A, Soheilian B, David N, Houzay E (2012) Stereopolis II: a multi-purpose and multi-sensor 3D mobile mapping system for street visualisation and 3D metrology. Revue Française de Photogrammétrie et de Télédétection 200:69–79

Pratt P, Ives M, Lawton G, Simmons J, Radev N, Spyropoulou L, Amiras D (2018) Through the HoloLens(TM) looking glass: augmented reality for extremity reconstruction surgery using 3D vascular models with perforating vessels. Eur Radiol Exp 2(2), 1–7

Remondino F, Nocerino E, Toschi I, Menna F (2017) A critical review of automated photogrammetric processing of large datasets. Int Arch Photogramm Remote Sens Spat Inf Sci XLII-2/W5, pp 591–599

Remondino F, Stoppa D (2013) TOF range-imaging cameras. Springer, Heidelberg

Roldão L, de Charette R, Verroust-Blondet A (2021) 3D semantic scene completion: a survey. arXiv:2103.07466v1

Roynard X, Deschaud J-E, Goulette F (2018) Paris-Lille-3D: a large and high-quality ground-truth urban point cloud dataset for automatic segmentation and classification. Int J Robot Res 37(6), 545–557

Schütt P, Schwarz M, Behnke S (2019) Semantic interaction in augmented reality environments for Microsoft HoloLens. In: Proceedings of the European conference on mobile robots, pp 1–6

Sereno M, Wang X, Besançon L, McGuffin MJ, Isenberg T (2020) Collaborative work in augmented reality: a survey. IEEE Trans Vis Comput Graph. https://doi.org/10.1109/TVCG.2020.3032761

Smisek J, Jancosek M, Pajdla T (2011) 3D with Kinect. In: Proceedings of the IEEE international conference on computer vision workshops, pp 1154–1160

Soudarissanane S, Lindenbergh R (2011) Optimizing terrestrial laser scanning measurement set-up. Int Arch Photogramm Remote Sens Spat Inf Sci XXXVIII-5/W12, pp 127–132

Soudarissanane S, Lindenbergh R, Menenti M, Teunissen P (2011) Scanning geometry: influencing factor on the quality of terrestrial laser scanning points. ISPRS J Photogramm Remote Sens 66(4), 389–399

SpaceCatcher HoloLens App. http://spacecatcher.madeinholo.com. Last accessed 1 Dec 2018

Stathopoulou E-K, Welponer M, Remondino F (2019) Open-source image-based 3D reconstruction pipelines: review, comparison and evaluation. Int Arch Photogramm Remote Sens Spat Inf Sci XLII-2/W17, pp 331–338

Stotko P, Krumpen S, Hullin MB, Weinmann M, Klein R (2019a) SLAMCast: large-scale, real-time 3D reconstruction and streaming for immersive multi-client live telepresence. IEEE Trans Vis Comput Graph 25(5), 2102–2112

Stotko P, Krumpen S, Weinmann M, Klein R (2019b) Efficient 3D reconstruction and streaming for group-scale multiclient live telepresence. In: Proceedings of the IEEE international symposium on mixed and augmented reality, pp 19–25

Sturm J, Engelhard N, Endres F, Burgard W, Cremers D (2012) A benchmark for the evaluation of RGB-D SLAM systems. In: Proceedings of the international conference on intelligent robot systems, pp 573–580

TIMMS Indoor Mapping. https://www.applanix.com/products/timms-indoor-mapping.htm. Last accessed 11 Aug 2021

Vassallo R, Rankin A, Chen ECS, Peters TM (2017) Hologram stability evaluation for Microsoft HoloLens. Proc SPIE 10136:295–300

Voelsen M, Schachtschneider J, Brenner C (2021) Classification and change detection in mobile mapping LiDAR point clouds. PFG - J Photogramm Remote Sens Geoinf Sci. https://doi.org/10.1007/s41064-021-00148-x

Weinmann M (2016) Reconstruction and analysis of 3D scenes–From irregularly distributed 3D points to object classes. Springer International Publishing, Cham

Weinmann M, Hinz S, Weinmann M (2017a) A hybrid semantic point cloud classification-segmentation framework based on geometric features and semantic rules. PFG - J Photogramm Remote Sens Geoinf Sci 85(3):183–194

Weinmann M, Jutzi B, Mallet C, Weinmann M (2017b) Geometric features and their relevance for 3D point cloud classification. ISPRS Ann Photogramm Remote Sens Spat Inf Sci IV-1/W1, pp 157–164

Weinmann M, Jäger MA, Wursthorn S, Jutzi B, Weinmann M, Hübner P (2020) 3D indoor mapping with the Microsoft HoloLens: qualitative and quantitative evaluation by means of geometric features. ISPRS Ann Photogramm Remote Sens Spat Inf Sci V-1-2020:165–172

Zhang L, Chen S, Dong H, El Saddik A (2018) Visualizing Toronto City data with HoloLens: using augmented reality for a city model. IEEE Consum Electron Mag 7(3), 73–80

Zhang E, Cohen MF, Curless B (2016) Emptying, refurnishing, and relighting indoor spaces. ACM Trans Graph 35(6):174

Zingsheim D, Stotko P, Krumpen S, Weinmann M, Klein R (2021) Collaborative VR-based 3D labeling of live-captured scenes by remote users. IEEE Comput Graph Appl. https://doi.org/10.1109/MCG.2021.3082267

Zollhöfer M, Stotko P, Görlitz A, Theobalt C, Nießner M, Klein R, Kolb A (2018) State of the art on 3D reconstruction with RGB-D cameras. Comput Graph Forum 37(2), 625–652