

INDOOR MAPPING EYEWEAR: GEOMETRIC EVALUATION OF SPATIAL MAPPING CAPABILITY OF HOLOLENS

K. Khoshelham^{1,*}, H. Tran¹, D. Acharya¹

¹ Dept. of Infrastructure Engineering, University of Melbourne, Parkville 3010, Australia
(k.khoshelham, ha.tran, debaditya.acharya)@unimelb.edu.au

Commission IV, WG IV/5

KEY WORDS: Indoor Mapping, Depth Camera, SLAM, Point Cloud, 3D Mesh Model, Accuracy, Mixed Reality

ABSTRACT:

Existing indoor mapping systems have limitations in terms of time efficiency and flexibility in complex environments. While backpack and handheld systems are more flexible and can be used for mapping multi-storey buildings, in some application scenarios, e.g. emergency response, a light-weight indoor mapping eyewear or head-mounted system has practical advantages. In this paper, we investigate the spatial mapping capability of Microsoft HoloLens mixed reality eyewear for 3D mapping of large indoor environments. We provide a geometric evaluation of 3D mesh data captured by the HoloLens in terms of local precision, coverage, and global correctness in comparison with terrestrial laser scanner data and a reference 3D model. The results indicate the high efficiency and flexibility of HoloLens for rapid mapping of relatively large indoor environments with high completeness and centimetre level accuracy.

1. INTRODUCTION

Indoor spaces are dynamic environments. In addition to the movement of people, redecorations, and reconfigurations of the furniture, the layout of the indoor space itself is also subject to frequent changes and modifications. Consequently, existing floor plans and 3D indoor models are often outdated and do not represent the as-is condition of the environment. To maintain an up-to-date indoor spatial information system, there is a need for efficient and flexible systems for mapping indoor environments.

Photogrammetry is perhaps the most convenient approach for mapping indoor environments due to the wide availability and low cost of cameras. The process of 3D reconstruction from imagery is largely automated thanks to recent developments in Structure from Motion (SfM) (Furukawa and Ponce, 2010) and dense matching (Hirschmuller, 2008). However, both SfM and dense matching require rich texture and are challenged in environments with poorly textured surfaces (Khoshelham, 2018).

In contrast, laser scanners and range cameras are independent of the texture and can operate in poorly textured indoor environments. Terrestrial laser scanning is the most accurate technique providing millimetre level accuracy for point clouds captured indoors (Khoshelham, 2018). However, terrestrial laser scanners are typically mounted on a tripod, making the mapping process inefficient, cumbersome, and inflexible. Other tripod-mounted systems that use range cameras, e.g. Matterport (Matterport, 2019), are faster than laser scanners, but are still quite cumbersome because of the tripod.

Mobile indoor mapping systems provide a higher level of efficiency, which is a great advantage especially in mapping large indoor environments. Lehtola et al. (2017) provide a comparison of several mobile indoor mapping systems. These include trolley systems, backpack systems, and handheld sensors. Trolley systems, such as NavVis M6 (NavVis, 2019) and Viametris iMS3D (Viametris, 2019), operate on a flat floor only and have

limitations in staircases and when mapping multi-storey buildings. Backpack and handheld systems, such as the UC Berkeley backpack (Liu et al., 2010), UVigo backpack (Figueira et al., 2016), Leica Pegasus (Leica, 2019), and the different versions of the Zeb scanner (GeoSLAM, 2019), are more flexible and can be used for mapping multi-storey buildings. However, in certain applications, such as emergency response, carrying a backpack or holding a scanner in hand may not be practical. In such applications, a light-weight indoor mapping eyewear or head-mounted system has practical advantages.

The HoloLens released recently by Microsoft is a head-mounted mixed reality (Milgram et al., 1995) system with spatial mapping capability. Equipped with a suite of low-cost sensors including a depth camera, the HoloLens is primarily designed to enable the user to interact with 3D graphical models, the so-called holograms, in a mixed reality experience. Proper interaction with holograms requires mapping the surrounding environment, for which Microsoft offers the spatial mapping capability. While the spatial mapping capability of HoloLens has been demonstrated for scanning small objects and small spaces such as a room, the potential of HoloLens for mapping large indoor environments has not been explored.

In this paper we investigate whether HoloLens can be used for mapping large indoor environments. We describe an experiment where a HoloLens is used as an indoor mapping eyewear to scan an environment of approximately 300 m² in size. We provide a geometric evaluation of the acquired 3D data in terms of local precision, global correctness, and coverage, through comparison with terrestrial laser scanner data and a reference 3D model.

The paper proceeds with a description of the HoloLens and its spatial mapping capability in Section 2. The experimental setup is described in Section 3, and the evaluation results are discussed in Section 4. A summary and concluding remarks are presented in Section 5.

* Corresponding author

2. SPATIAL MAPPING BY HOLOLENS

2.1 The HoloLens

The HoloLens is a head-mounted mixed reality device consisting of multiple sensors, a computer, and a stereo pair of display panels. The built-in sensors include four cameras on the sides, a video camera at the front, a depth camera, and an inertial measurement unit (IMU). The computer includes a custom-built Holographic Processing Unit (HPU) with 2GB RAM designed for fast processing of holograms. The stereo display panels are designed for 3D mixed reality viewing of holograms. Figure 1 shows the main components of the HoloLens.

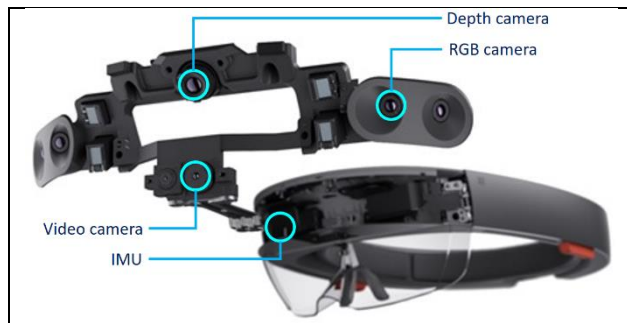


Figure 1. Built-in sensors of the Microsoft HoloLens (adapted from Microsoft (2019)).

2.2 Spatial Mapping

The spatial mapping capability of the HoloLens is based on Microsoft's proprietary software and its algorithms are largely unpublished. Microsoft researchers have previously developed KinectFusion (Izadi et al., 2011), which is a simultaneous localisation and mapping (SLAM) algorithm for real-time 3D scene reconstruction from Kinect depth data. Early experiments with KinectFusion showed that the method is suitable for small-scale reconstruction of room-sized scenes (Meister et al., 2012). More recent research at Microsoft has focused on large-scale scene reconstruction using voxel hashing (Nießner et al., 2013), and RGB-D camera relocalisation (Glocker et al., 2014) for recovering from tracking failures. These developments improve the scalability and robustness of the RGB-D SLAM method. The spatial mapping capability likely combines these algorithms in a customised software for the HoloLens.

The spatial mapping software allows real-time display of the constructed 3D mesh on the stereo display panels of HoloLens creating a mixed reality visualisation of the map. Figure 2 shows an example of mixed reality visualisation of the 3D mesh captured by the HoloLens.

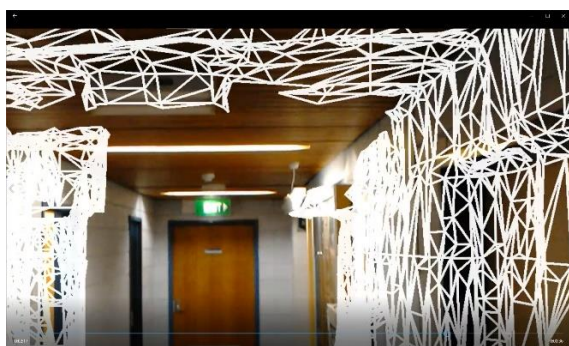


Figure 2. Mixed reality view of the 3D mesh captured by the HoloLens.

2.3 Evaluation Method

To evaluate the quality of the 3D mesh data captured by the HoloLens we focus on three main geometric aspects: local precision, global correctness, and coverage. Local precision describes the precision of individual 3D points captured by the HoloLens. To measure local precision, we fit planes to points captured on planar surfaces. The root mean squared error (RMSE) of plane fitting is taken as the measure of local precision:

$$M_{LP} = \left(\frac{1}{n} \sum_{i=1}^n d_i^2 \right)^{\frac{1}{2}} \quad (1)$$

where n is the number of points on the plane and d_i is the perpendicular distance from point p_i to the fitted plane π , that is:

$$d_i = \pi^T p_i \quad (2)$$

where p_i and π are homogeneous representations of the point and the plane respectively (Khoshelham, 2016).

Global correctness indicates to what extent the global shape of the environment captured in the data is consistent with the actual layout and dimensions of the environment. To measure global correctness, we compare the HoloLens mesh with terrestrial laser scanner data and a reference 3D model of the building. Laser scanner data are highly accurate, but inherently contain many gaps due to occlusion and the static nature of data acquisition. In contrast, 3D models are more complete, but less accurate, since manual generation of these models from point clouds involves visual interpretation of the location of structural elements.

To enable the comparison, the HoloLens mesh is first registered accurately to the terrestrial laser scanner data and the 3D model respectively. The comparison of the HoloLens mesh with the terrestrial laser scanner data is based on computing the distances between the HoloLens mesh vertices and the closest point in the laser scanner point cloud. Following Lehtola et al. (2017), we apply a cut-off distance to reduce the influence of gaps and coverage discrepancies between the two data sets. We analyse the median point-point distance against various cut-off values.

For the comparison with the 3D model, we use the point-model distances recommended by Tran et al. (2019). Specifically, we compute the median of unsigned distances between the HoloLens mesh vertices and the corresponding planar surface in the 3D model for distances that are smaller than a cut-off value. The vertex-surface distance is computed according to Eq. (2), and the correspondence is established by finding the closest surface in the 3D model that contains the projection of the vertex within its boundary (Oude Elberink and Khoshelham, 2015).

In addition to the local precision and global correctness, we compare the efficiency of the HoloLens with the terrestrial laser scanner in terms of coverage. The coverage indicates to what extent the surfaces of the reference 3D model are covered by points in a point cloud or mesh. To measure the coverage, we use the method developed by Tran and Khoshelham (2019), which defines the coverage M_{cov} of a surface π as the ratio of the area of the alpha shape α of the points to the area of the surface:

$$M_{cov}(\pi) = \frac{area(\alpha)}{area(\pi)} \quad (3)$$

3. EXPERIMENTAL SETUP

The Hololens was used to map an indoor environment of approximately 300 m² for which a reference 3D model was available. The same environment was also scanned by a terrestrial laser scanner to enable comparison and geometric evaluation of the Hololens data. The following sections describe the test environment and the datasets.

3.1 Test Environment

The test environment is part of the 3rd floor of the engineering building block B in the Parkville campus of the University of Melbourne. It features a corridor with several turns, an office room, a small kitchen room, and a large lecture room with high level of clutter (desks and chairs). A reference 3D model of the environment was already available as part of the ISPRS Benchmark on Indoor Modelling (Khoshelham et al., 2018; Khoshelham et al., 2017). However, at the time of the experiment some of the rooms were not accessible, and hence, the 3D model was modified to contain only the spaces that were scanned. Figure 3 shows a top view of the modified 3D model of the test environment and its approximate dimensions. This model was created from a Zeb-1 scan with a nominal accuracy of 3 cm (Khoshelham et al., 2017).

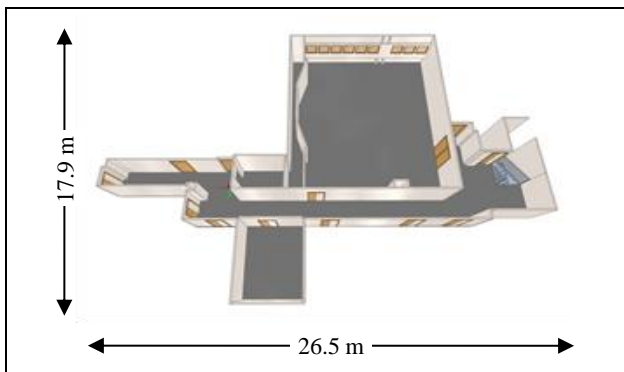


Figure 3. 3D model of the test environment. The ceiling is removed to give a better impression of the interior.

3.2 Data Acquisition

To scan the test environment using the Hololens, a test participant mounted the Hololens on his head and walked along a trajectory while scanning the walls, floors, and ceilings at a distance between 0.8 m and 3.1 m corresponding to the range of the depth camera. The scanning trajectory started in the office room (bottom of the model in Figure 3), and then covered the corridor to the right, the large lecture room, the entire corridor from right to left (see Figure 3), the kitchen, and finally ended back in the office room. As such, the trajectory contained several loops needed to generate a globally consistent map by the SLAM algorithm. The total length of the scanning trajectory was approximately 95 m. The entire scanning process took less than 10 minutes.

The holographic spatial mapping code sample from Microsoft samples portal was used for data acquisition and mixed reality visualisation of the generated mesh in real time. This helped the test participant see the gaps and perform a complete scan of the environment. The recorded data were then downloaded from the Windows Device Portal as a 3D mesh file containing about 180,000 vertices and 320,000 triangular faces. Figure 4(a) shows the top view of the 3D mesh captured by the Hololens.

For the laser scanning, a Faro Focus 3D S120 terrestrial laser scanner was used. In total four scans were recorded. The scans were registered using the Faro Scene software which was able to automatically recognise and measure the spherical targets we had placed in the environment before the scanning. The registration accuracy as measured by mean distance between the target centres after the registration was 1.4 mm. Based on previous experiments with the same laser scanner (Khoshelham, 2018), the accuracy of the point cloud is 5 mm. The registered scans were sampled down to a point spacing of 1 cm resulting in a point cloud of about 4 million points. Figure 4(b) shows the top view of the registered laser scanner point cloud. The laser scanning process and the registration took approximately 2 hours.

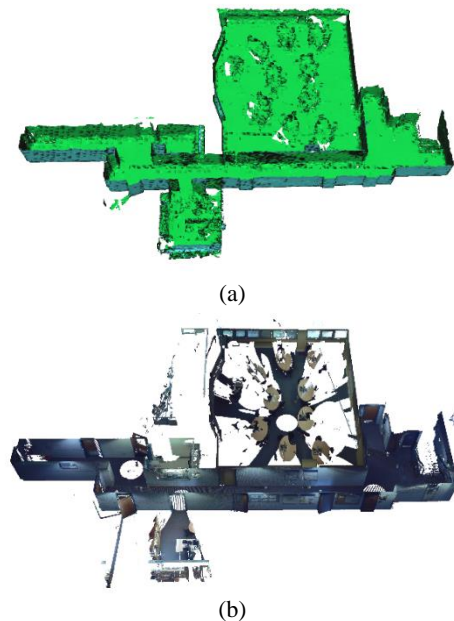


Figure 4. The acquired data: (a) the 3D mesh captured by the Hololens; (b) the registered point clouds captured by the terrestrial laser scanner. In both visualisations the ceiling is removed to give a better impression of the interior.

4. RESULTS

The Hololens 3D mesh data was evaluated in terms of local precision, global correctness, and coverage.

4.1 Local Precision

A total of 10 planar surfaces from across the Hololens mesh data was selected for plane fitting. Figure 5 shows the distribution of point-plane distances for the 10 fitted planes. The local precision of the Hololens mesh as measured by the overall plane fitting RMSE over 4002 points on 10 planes was 2.25 cm.

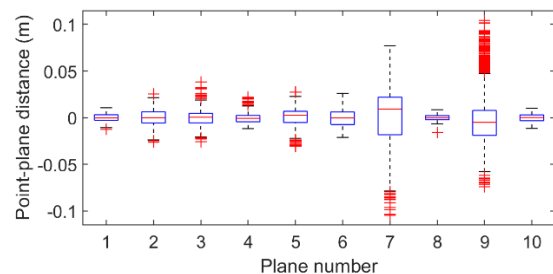


Figure 5. Distribution of signed point-plane distances for 10 planes extracted from the Hololens mesh.

4.2 Global correctness

The global correctness of the Hololens mesh was analysed through comparison with the laser scanner point cloud and the reference 3D model.

4.2.1 Comparison with laser scanner point cloud

To enable comparison of the Hololens mesh with the laser scanner point cloud the two data sets were first registered. The registration was performed by selecting 10 corresponding planar surfaces from across the two datasets. The transformation parameters were estimated by minimising the distance between the Hololens mesh vertices within the selected planes and their corresponding planes in the laser scanner point cloud using the closed form solution of Khoshelham (2015, 2016). The registration accuracy as measured by root mean squared distance between 3409 Hololens mesh vertices and their 10 corresponding laser scanned planes after the registration was 5.68 cm. Figure 6 shows the registered Hololens and laser scanner data.

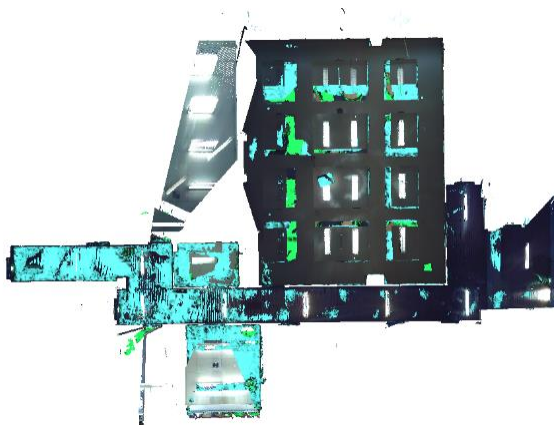


Figure 6. Hololens mesh (in cyan colour) registered with the terrestrial laser scanner point cloud.

Once the two datasets were registered the closest point distances were computed. Figure 7 shows the Hololens point cloud coloured according to closest point distances. While the majority of the Hololens points are less than 5 cm away from the laser scanner points, parts of the office room, the kitchen, and the floor of the lecture room have larger distances. As it can be seen in Figure 4(b), these are the areas that were not covered in the laser scanner point cloud due to occlusion and the small number of scans.

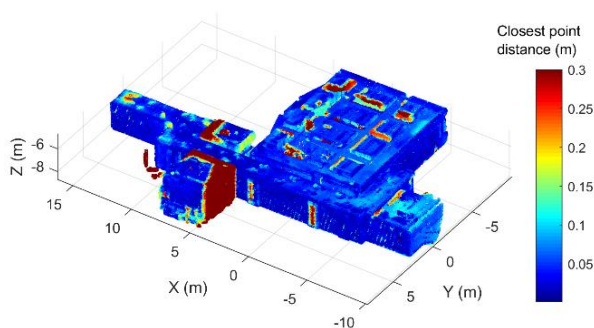


Figure 7. Closest point distances between the Hololens and the laser scanner point cloud.

Figure 8 shows the median point-point distance between the Hololens data and the laser scanner point cloud plotted against increasing cut-off values. As it can be seen, the median point-point distance approaches 5 cm with increasing cut-off distance.

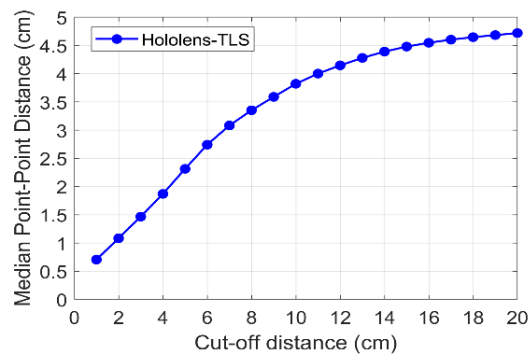


Figure 8. Median point-point distance between the Hololens mesh and the registered laser scanner point cloud.

4.2.2 Comparison with reference 3D model

The Hololens mesh was also registered with the 3D model to enable comparison between the two data sets. The registration was performed by selecting 10 corresponding planar surfaces from across the two datasets. The transformation parameters were estimated by minimising the distance between the Hololens mesh vertices within the selected planes and their corresponding planes in the 3D model using the closed form solution of Khoshelham (2015, 2016). The registration accuracy as measured by root mean squared distance between 3403 Hololens mesh vertices and their 10 corresponding model planes after the registration was 5.42 cm. Figure 9 shows the Hololens data registered with the 3D model.

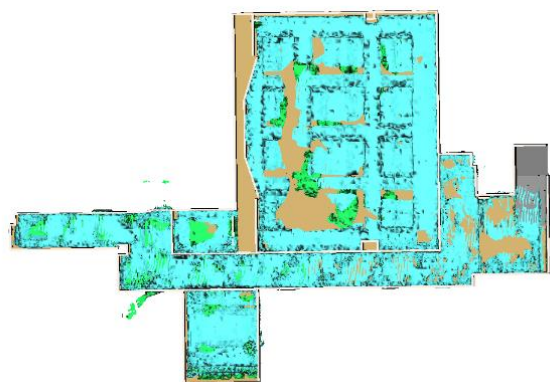


Figure 9. Hololens mesh (in cyan colour) registered with the reference 3D model.

After the registration, point-model distances were computed as described in Section 2.3. Figure 10 shows the Hololens point cloud coloured according to point-model distances. While most points are within 5 cm distance from the model, parts of the office room and the ceiling of the lecture room have large distances. These large distances are due to the low level of detail of the 3D model. The actual test environment features a large protrusion at the location of windows in the office room and a number of false ceilings in the lecture room. While the Hololens data captures these details, they are missing in the 3D model (see Figure 3). It is worth noting that the Hololens data at these locations agrees very well with the terrestrial laser scanner data (see Figure 7).

Figure 11 shows the median point-model distance between the Hololens data and the 3D model plotted against increasing cut-off values. As it can be seen, the median point-model distance is smaller than 3 cm even at large cut-off values.

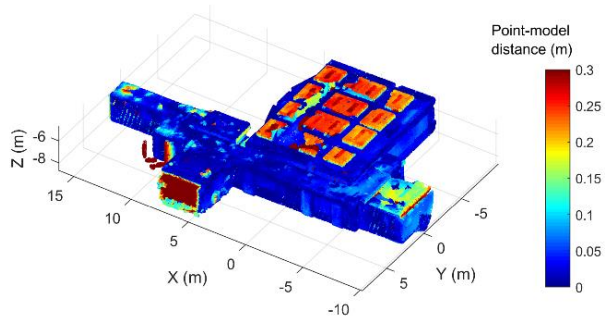


Figure 10. Point-model distances between the Hololens mesh vertices and the 3D model.

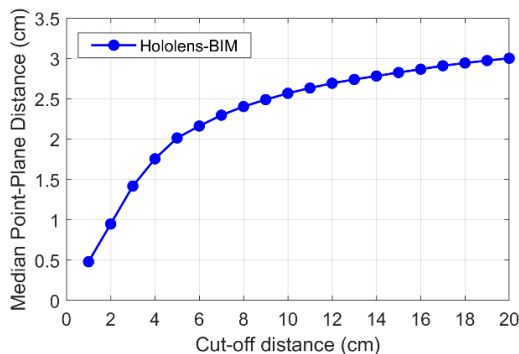


Figure 11. Median point-model distance between the Hololens mesh and the 3D model.

4.3 Coverage

Figure 12 shows the coverage of the Hololens data as measured on the surfaces of the 3D model compared to the coverage of the laser scanner data. Note that only the interior surfaces of walls have a coverage and the exterior surfaces as well as the staircase have a coverage of zero as these were not scanned. The laser scanner point cloud has a lower coverage in the office room and the kitchen, which were scanned only partially from the corridor, and the floor of the lecture room, which was largely occluded by desks and chairs. In comparison, the Hololens data has a relatively high coverage across the test environment. Considering that the data acquisition time was only a few minutes for the Hololens and several hours for the laser scanner, these results indicate the significantly higher efficiency and flexibility of the Hololens for mapping indoor environments.

5. CONCLUSIONS

In this paper we investigated the spatial mapping capability of Microsoft Hololens mixed reality eyewear for 3D mapping of large indoor environments. We performed a geometric evaluation of 3D mesh data captured by the Hololens in terms of local precision, global correctness, and coverage. The results of plane fitting showed a local precision of 2.25 cm for the Hololens mesh data. From the comparison of the Hololens data with a laser scanner point cloud and a reference 3D model of the test

environment, it was found that the Hololens mesh is globally correct with no evident deformation and a mean distance of about 5 cm to the registered laser scanner point cloud and the 3D model. The coverage analysis showed that the Hololens mesh has a higher coverage, and is therefore more complete, than the laser scanner point cloud.

Overall, the results of the experiments indicate the great potential of the Hololens for efficient and flexible mapping of indoor environment with high completeness and an accuracy of a few centimetres. Future research will focus on mapping larger environments, such as multi-storey buildings, and applications in practical scenarios such as emergency response.

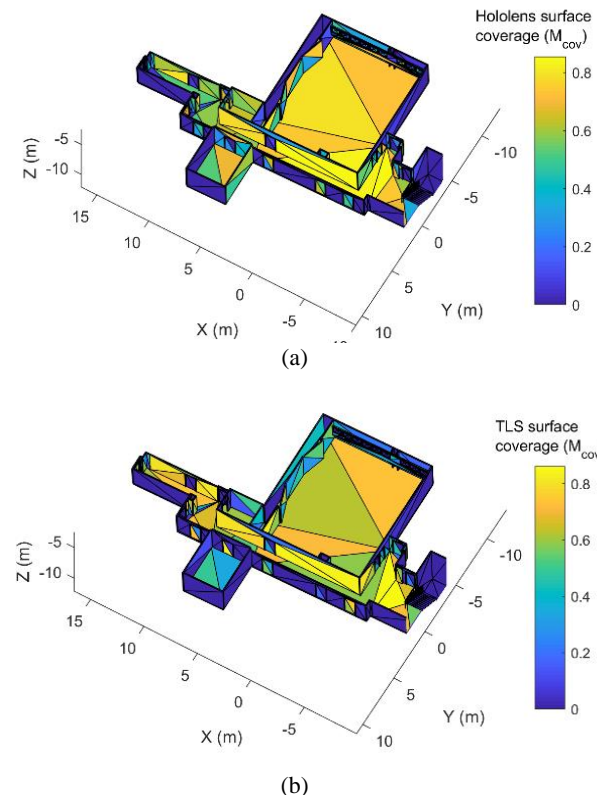


Figure 12. Coverage of the surfaces of the 3D model by the Hololens mesh (a) and the laser scanner point cloud (b).

ACKNOWLEDGEMENTS

This research was supported by the Melbourne School of Engineering Strategic Investment Funds. The support by the Australian Research Council grant DP170100109 is also acknowledged.

REFERENCES

- Filgueira, A., Arias, P., Bueno, M., Lagüela, S., 2016. Novel inspection system, backpack-based, for 3D modelling of indoor scenes, Proceedings of the International Conference on Indoor positioning and Navigation, Alcalá de Henares, Spain, pp. 4-7.
- Furukawa, Y., Ponce, J., 2010. Accurate, Dense, and Robust Multiview Stereopsis. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 1362-1376.
- GeoSLAM, 2019. Zeb Revo RT. <https://geoslam.com/zeb-revo-rt/>, (Accessed 1 Feb. 2019).

- Glocker, B., Shotton, J., Criminisi, A., Izadi, S., 2014. Real-Time RGB-D Camera Relocalization via Randomized Ferns for Keyframe Encoding. *IEEE Transactions on Visualization and Computer Graphics* 21, 571-583.
- Hirschmuller, H., 2008. Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 328-341.
- 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, *ACM Symposium on User Interface Software and Technology*, Santa Barbara, California.
- Khoshelham, K., 2015. Direct 6-DoF Pose Estimation from Point-Plane Correspondences, *Digital Image Computing: Techniques and Applications (DICTA)*, 2015 International Conference on, pp. 1-6.
- Khoshelham, K., 2016. Closed-form solutions for estimating a rigid motion from plane correspondences extracted from point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 114, 78-91.
- Khoshelham, K., 2018. Smart Heritage: Challenges in Digitisation and Spatial Information Modelling of Historical Buildings, in: Belussi, A., Billen, R., Hallot, P., Migliorini, S. (Eds.), *2nd Workshop On Computing Techniques For Spatio-Temporal Data in Archaeology And Cultural Heritage*. *CEUR Workshop Proceedings*, Melbourne Australia, pp. 7-12.
- Khoshelham, K., Tran, H., Díaz-Vilariño, L., Peter, M., Kang, Z., Acharya, D., 2018. An Evaluation Framework for Benchmarking Indoor Modelling Methods, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. XLII-4. Copernicus Publications, Delft, The Netherlands, pp. 297-302.
- Khoshelham, K., Vilariño, L.D., Peter, M., Kang, Z., Acharya, D., 2017. The ISPRS Benchmark on Indoor Modelling, *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, Vol. XLII-2/W7, pp. 367-372.
- Lehtola, V., Kaartinen, H., Nüchter, A., Kajaluoto, R., Kukko, A., Litkey, P., Honkavaara, E., Rosnell, T., Vaaja, M., Virtanen, J.-P., Kurkela, M., El Issaoui, A., Zhu, L., Jaakkola, A., Hyypä, J., 2017. Comparison of the Selected State-Of-The-Art 3D Indoor Scanning and Point Cloud Generation Methods. *Remote Sensing* 9, 796.
- Leica, 2019. Pegasus: Backpack. <https://leica-geosystems.com/products/mobile-sensor-platforms/capture-platforms/>, (Accessed 1 Feb. 2019).
- Liu, T., Carlberg, M., Chen, G., Chen, J., Kua, J., Zakhor, A., 2010. Indoor localization and visualization using a human-operated backpack system, *2010 International Conference on Indoor Positioning and Indoor Navigation*, pp. 1-10.
- Matterport, 2019. Pro2 3D Camera. <https://matterport.com/pro2-3d-camera/>, (Accessed 1 Feb. 2019).
- Meister, S., Izadi, S., Kohli, P., Hammerle, M., Rother, C., Kondermann, D., 2012. When Can We Use KinectFusion for Ground Truth Acquisition?, *IEEE/RSJ International Conference on Intelligent Robots and Systems, Workshop on Color-Depth Camera Fusion in Robotics*, Vilamoura, Portugal.
- Microsoft, 2019. HoloLens (1st gen) hardware details. <https://docs.microsoft.com/en-us/windows/mixed-reality/hololens-hardware-details>, (Accessed 20 March 2019).
- Milgram, P., Takemura, H., Utsumi, A., Kishino, F., 1995. *Augmented reality: a class of displays on the reality-virtuality continuum*, *Photonics for Industrial Applications*. SPIE, Boston, MA, United States, pp. 282-292.
- NavVis, 2019. M6. <https://www.navvis.com/m6>, (Accessed 1 Feb. 2019).
- Nießner, M., Zollhöfer, M., Izadi, S., Stamminger, M., 2013. Real-time 3D Reconstruction at Scale using Voxel Hashing. *ACM Transactions on Graphics* 32, 1-11.
- Oude Elberink, S., Khoshelham, K., 2015. Automatic Extraction of Railroad Centerlines from Mobile Laser Scanning Data. *Remote Sensing* 7, 5565-5583.
- Tran, H., Khoshelham, K., 2019. Building Change Detection Through Comparison of a Lidar Scan with a Building Information Model, *ISPRS Geospatial Week 2019, Enschede, The Netherlands*.
- Tran, H., Khoshelham, K., Kealy, A., 2019. Geometric comparison and quality evaluation of 3D models of indoor environments. *ISPRS Journal of Photogrammetry and Remote Sensing* 149, 29-39.
- Viametris, 2019. iMS3D. <https://www.viametris.com/3d-mobile-scanner/>, (Accessed 1 Feb. 2019).