

HOLOLENS 2 SPATIAL MAPPING CAPABILITIES IN VAST MONUMENTAL HERITAGE ENVIRONMENTS

S. Teruggi¹, F. Fassi¹

¹ 3D Survey Group, ABC Dep, Politecnico di Milano, Via Ponzio 31, 20133 Milano, Italy.
(simone.teruggi, francesco.fassi)@polimi.it

Commission II

KEY WORDS: HoloLens2, Cultural Heritage, Spatial Mapping, Point Cloud, Indoor Mapping.

ABSTRACT:

HoloLens 2 is a state-of-the-art of Mixed Reality device that can augment the natural environment with virtual content in the form of holograms. Spatial mapping is the capability of the device to scan the environment and provide the user with a detailed 3D mesh model representation of its surroundings. The device can place holograms about real-world objects, perform occlusions and simulate physical interactions only having the geometric knowledge of its operating environment. HoloLens 2 mapping capabilities have been extensively tested. Nevertheless, no test was performed on monumental heritage buildings, where dimensions far exceed the typical sensor working range of the device. The authors want to evaluate the capability of the device to map monumental spaces. In these environments, dimensions, repetition of elements and heterogeneity of areas provide a real challenge for out-of-the-box Microsoft localization and mapping algorithms. The test case will be Milan Cathedral, inside the research project to build an MR application to help and enhance the systematic and planned maintenance intervention and inspections. Mapping capabilities have been tested in very different spaces: i) the south nave; ii) the empty volumes between the internal vaults and the external roof ("Sordine"); iii) the narrow spiral staircase connecting the south nave with the roofs lower level.

1. INTRODUCTION

HoloLens 2 (Microsoft, 2020) is a state-of-the-art of Mixed Reality (MR) device that can augment the natural environment with virtual content in the form of holograms. The user can intuitively manipulate these 3D objects in the real world by using their hands and voice. This opens a world of possibilities. Mixing digital and natural allows increasing knowledge, understanding, and collaboration in long-term, complex and cooperative projects. This type of situation is typical of Cultural Heritage (CH) world that can benefit from mixed reality technology both for dissemination and professional practices. Localizing information and registering referenced operations can be very useful to facilitate collaborations in the maintenance processes creating a 1:1 information system directly connected with the natural object and directly usable in the yard.

This is achieved with an acceptable registration between the natural world and the displayed 3D model. This alignment process is possible thanks to different sensors with which the device is provided (Figure 1).

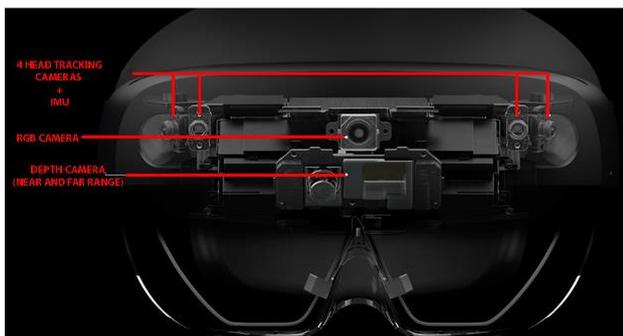


Figure 1. HoloLens 2 head mounted display front view.

Spatial mapping is the device's capability to scan the environment and provide the user with a detailed 3D mesh model representation of its surroundings (Microsoft, 2021a). The device can place holograms on real-world objects, occlude virtual models with their surroundings, and simulate physics interactions between the two worlds only having the geometric knowledge of its operating environment. The accuracy of this knowledge directly affects the whole alignment and augmentation processes. HoloLens mapping capabilities have been extensively tested to acquire human-scale environments (Hübner et al., 2019). Nevertheless, no test was performed on monumental heritage buildings. Here the dimensions far exceed the typical sensor working range of the device affecting, in fact, the localization accuracy of digital information on the real object.

Here, the environmental dimensions, the repetition of the same architectonic elements and the lighting conditions provide a real challenge for out-of-the-box Microsoft localization and mapping algorithms. The authors want to evaluate the device's capability to map monumental spaces.

The case study will be Milan Cathedral, inside the research project to build an MR application to help and enhance the systematic and planned activities performed by the Veneranda Fabbrica del Duomo di Milano (VF) (Teruggi et al., 2021a; 2021b). Having the possibility to see the maintenance and inspection data as holograms, in real time and directly on site, would be a valuable asset to enhance and support the virtual understanding and the decision-making process inside the monumental heritage buildings. It allows for better information management and communication exchange between all actors involved and for an enhanced rationalization of the building site activities (Teruggi et al. 2021c).

1.1 Related works

Many previous works have been devoted to evaluating the accuracy of HoloLens mapping capabilities. Liu et al. designed a

series of experiments to “*quantitatively evaluate HoloLens performance in head localization, real environment reconstruction, spatial mapping, hologram visualization, and speech recognition*” (Liu et al., 2018). Their tests illustrated the astonishing capabilities of the device to reconstruct and align virtual content to natural objects in a controlled small environment.

It results that the HoloLens device could do more than just augmenting the real-world with 3D models. In fact, Huang et al. developed an application that allows users to contactless measure their environment. They exploit the ability of the device to reconstruct the real world in a spatial mesh. Measurement accuracy reached those obtained with a measuring tape (± 0.01 m) (Huang et al., 2017).

The experiments prove HoloLens high accuracy in reconstructing its surroundings, but they are confined to single room environments. Thanks to the work of Hübner et al. and of Khoshelham et al., instruments boundaries start to be stressed. HoloLens device capability to simultaneously track its position and map its environment in real-time is seen as a means to acquire coarse point clouds and meshes of entire structures. These could be used in the context of BIM (Building Information Modelling) substituting the long and manual modelling process, based on acquired survey data, typical for existing buildings (Hübner et al., 2019, 2020; Khoshelham et al., 2019). Captured meshes have been evaluated against TLS (Terrestrial Laser Scanner) point cloud data to assess device mapping accuracy. Results show that deviations stay in the range of a few centimetres.

Data proved to be sufficiently accurate for subsequent reconstruction of a semantically enriched and topologically correct model of an indoor scene (Wienmann et al., 2021). However, all studies presented locate HoloLens field of use in “human” scale environments where surfaces are within the device sensor working range and environmental conditions are fairly constant, making it easier for the device to find and measure its position in space.

The work will enhance state of the art with a better understanding of the device mapping behaviour in monumental and complex ambiances, providing a starting point for position tracking and accuracy improvements. The process constitutes a critical point in understanding how and with which accuracy monumental heritage environments could be augmented through holographic head-mounted displays to localize information and data.

2. TEST AREAS

2.1 The ultimate survey of Milan Cathedral

Milan Cathedral constitutes an emblematic example for its periodic maintenance and inspection activities performed by the VF.

An intensive survey campaign conducted by the 3D Survey Group from Politecnico di Milano produced a complete 3D point cloud to support sites activities (Achille et al., 2020). The incredible complexity of the Cathedral and its heterogeneity of spaces imposed the use of different measurement techniques to capture the geometry of the monument. The main idea was to build a complete, accurate and uniformly resolute 3D point cloud model and use it as a base of an HBIM model to be used directly inside VF activities.

The survey was conducted mainly using close-range photogrammetry for the exterior and narrow passages and terrestrial laser scanner (TLS) for the interior spaces. Totally 1283 scans were necessary for the complete survey. The result is a complete digitalization of the monument in the form of a point cloud of almost 60 billion points with an average resolution of 5 mm.

Due to its vast heterogeneity of spaces, the Cathedral constitute the best monumental subject to assess, which are the mapping limits of the HoloLens 2 device. Three very different types of ambiances have been selected for these tests: i) the south nave of the Cathedral; ii) the empty volumes between the internal vaults and the external roof (“Sordine”), and iii) the narrow spiral staircase connecting the south nave with roofs lower level (Figure 2).

2.1.1 The south nave. The area has considerable dimensions and single architectonic elements repetition. It is characterized by eight times reiteration of a single bay module (9.5 x 9.5 x 24 m) to create a vast space 76 m long, 9.5 m wide, and 24 m high. It is delimited on one side by the wall, with the big windows towards the exterior, and on the other side by a series of pillars dividing it from the adjacent nave. Its survey required 16 scan stations on two levels in height performed with the Leica C10 TLS. This was necessary to avoid shadow area produced by overhangs and to retain a uniform 5 mm resolution (about 173 million points) everywhere on the high part of the vaults.

2.1.2 The “Sordine”. These spaces are empty voids that result between the extrados of the south nave vaults and the intrados of the external tile lower roof. Each “Sordina” corresponds to one bay in the nave below and the space between two buttresses on the upper roof. They result from the construction of transversal brick barrel vaults. They are characterized by the repetition of the same module connected with small door passages. Today they serve as technical spaces and are accessible from the roof level. The survey has been performed using the Leica HDS 7000 TLS for 50 scan stations resulting in a point cloud of about 49 million points with an average resolution of 5 mm.

2.1.3 The “Minguzzi” spiral staircase. It is a marble spiral staircase located in the rightmost corner of the Milan Cathedral façade. It has a total height of 25 m and is characterized by a narrow, walkable space of about 0.7 m wide. It is a complex and very narrow service space that connects the south nave entrance to the central balcony of the façade and the south roof lower level. The space is extremely dark with poor artificial illumination. Natural light comes in from openings (2 m depth and 0.85 m wide towards the interior; 0.3 m width at the exterior) placed at regular intervals. The central part of the stair hole is occupied by a stone pillar (0.4 m in diameter) to which the ramp rolls up. The survey was performed using close-range fisheye photogrammetry for a total of 889 pictures acquired using a Nikon D810 DSLR camera equipped with a Samyang 12mm diagonal fisheye lens. Residual errors on topography optimization reach 0.014 m (Perfetti et al., 2017).

3. METHODOLOGY

All test areas have been carefully mapped with the HoloLens 2 device. The scan process has been performed to ensure completeness and hole filling walking slowly around the areas while looking in different directions. The process has been repeated, crossing the spaces both ways for their entire length. This allows the depth sensor of the device to scan the whole space reconstructing a uniform 3D model. The HoloLens continuously create the 3D mesh of the maps thanks to its depth sensor and allows to download the triangles models from the device cache via the web interface. Anyway, it is advantageous to use a Unity3D (Unity, 2022) custom-designed app exploiting the Scene Understanding (SU) SDKs (Microsoft 2021c). These allow to query and download the entire spatial mesh data generated from the device. All acquired maps couldn't be memorized in the cache memory due to the limitation in size. Furthermore, the SU SDKs allow to set a pre-defined resolution

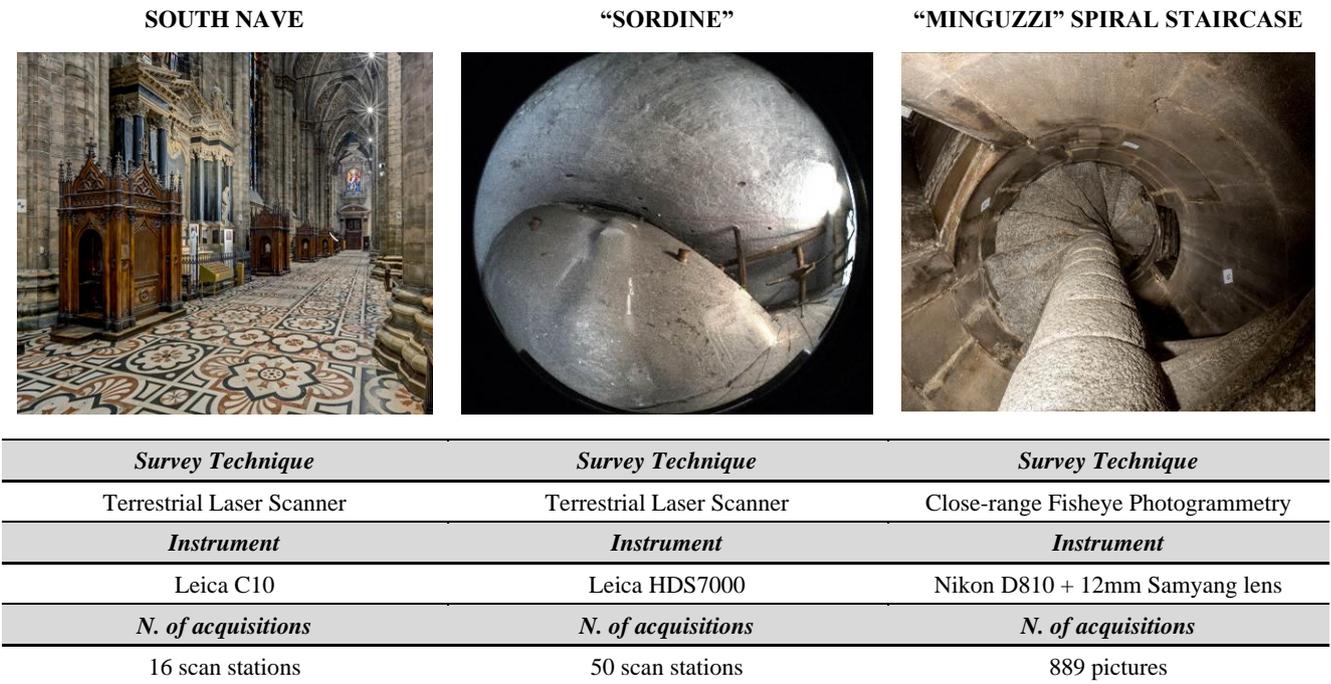


Figure 2. Milan Cathedral test areas with survey techniques used.

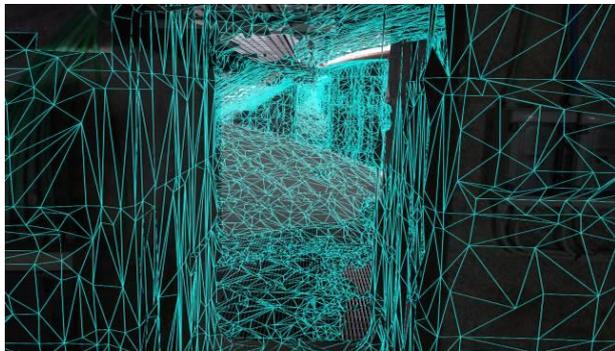


Figure 3. View of the “Sordine” with spatial map overlay.

for the triangle models and permit the operator to see such maps directly superimposed to the real environment and in real-time allowing to tune the scanning process as needed (Figure 3).

For the purpose of this work, the model quality has been set to "fine" during all measurements (20 - 30 cm maximum triangle dimensions). Spatial maps data have been reset between all measurements to make the survey of each area independent. The obtained spatial maps are then compared against ground truth data acquired during the Milan Cathedral survey campaign. The evaluation has been performed elaborating the triangular models and point cloud data inside CloudCompare (CC) software (CloudCompare, 2020). The steps are as follows:

1. the spatial maps have been transformed into point cloud models and sampled to have an average resolution of 1 cm;
2. the obtained point clouds have been aligned to the ground truth data (subsamped at 1 cm resolution) of the corresponding area fixing 4 points in the initial part of the measured space;
3. ground truth data and spatial maps have been cleaned from all non-marble elements (furniture, technical plants and confessionals);

4. Using Cloud to Cloud (C2C) algorithms (CloudCompare, 2020), deviations from ground truth and spatial maps have been computed in terms of absolute distances.

5. Distances in x, y, z direction have been evaluated, taking sample measures with respect to each area local coordinate system.

3.1 The south nave

In this high and long environment, the Holo-scanning process has been performed walking slowly through the space. The goal was to create loops around each bay, as for a classical mobile mapping activity, ensuring that each module was correctly and entirely covered before moving to the next one. The process has been repeated walking in both directions along the nave (from bay 1 to 8 and from 8 to 1). The whole process took up to 70 minutes.

The long-throw depth sensor on HoloLens plays a crucial role in spatial map generation. Due to the nave huge dimensions, it is impossible for the depth sensor to get the complete mapping of the space. Furthermore, in such scenario, the device could fail to track its position. It can frame only very similar portions of walls, floors and pillars leading to errors in the spatial map (holes, hallucinations, wormholes and bias Figure 4b).

As it is possible to see in Figure 4a, the mesh model generated stops approximately at 4 m height. 3D mesh models have been registered on the ground-truth coordinate system fixing four evenly distributed points on bay 1 with a residual error of 0.03 m. After sampling the raw mesh data in point clouds (0.01 m resolution) and cleaning all non-marble elements, deviation from ground truth (0.01 m resolution) data have been computed using C2C algorithms. Figure 4c reports the shifts in terms of cloud to cloud absolute distances. It is possible to note that more significant drifts from the original point clouds are along the x-direction (Figure 4d).

In bay 8, these vary between 0.42 and 0.59 m. The vertical deviation reaches a maximum value of 0.08 m at the end of bay 8 (Figure 4e). The deviations are negligible along with the y-direction, varying around 0.01 m in bay 1 and 0.04 m in bay 8.

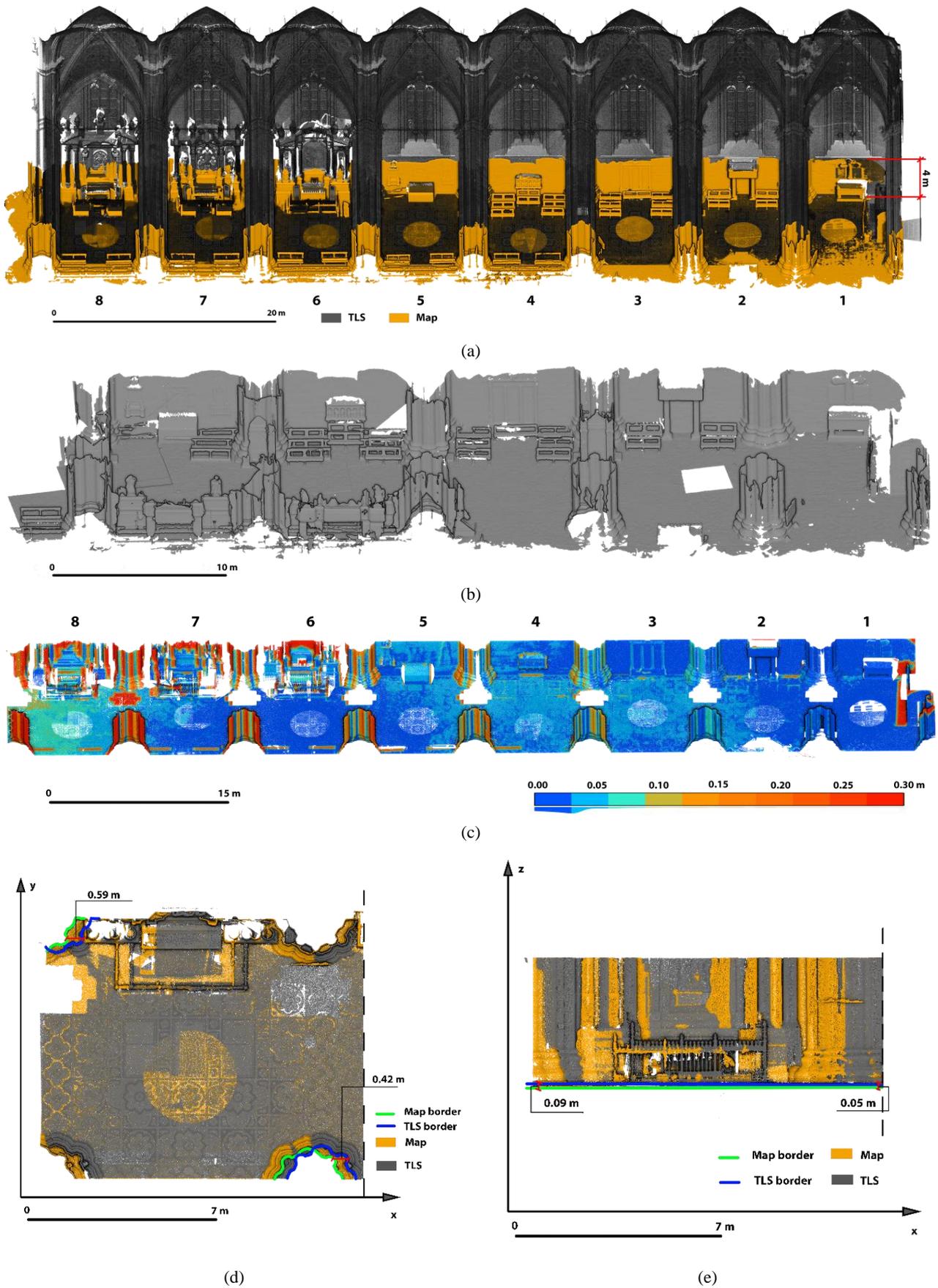


Figure 4. (a) Ground truth and spatial map comparison. (b) Spatial maps suffering from holes, hallucinations, wormholes and bias. (c) Ground truth with colour-coded deviations from spatial map, (d) xy plane detail (bay 8) and (e) xz plane detail (bay 8).

3.2 The "Sordine"

In the "Sordine" area, the mapping process proved challenging for two main reasons: i) during the scanning process, it was impossible to walk at a constant speed around the area, and ii) the lighting condition was not uniform.

Furthermore, due to the reduced height of the ceiling (2 m in its highest point), the device proved to be too near to it to be able to properly scan the surface. The mapping process has been performed scanning each room completely before moving to the next one. A total of two passages in each room were necessary. The second passage was devoted to ceiling mapping while walking crouched to the ground. The process has been performed

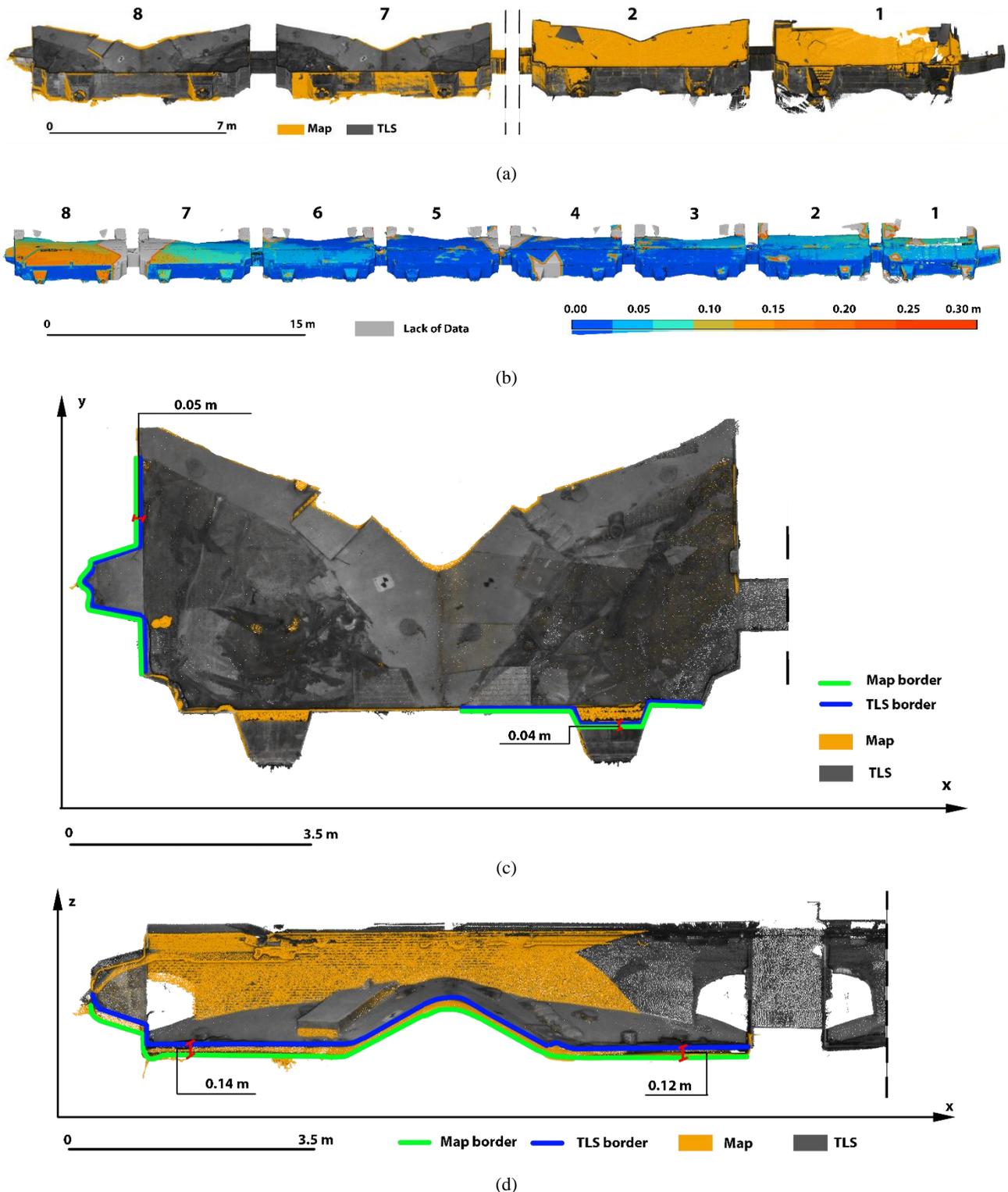


Figure 5. (a) Ground truth and spatial map comparison. (b) Ground truth with color-coded deviations from spatial map. (c) xy plane detail (bay 8) and (e) xz plane detail (bay 8).

crossing the area both ways and took up to 60 minutes. The obtained data, sampled as point clouds at 1 cm resolution, have been oriented on ground truth data fixing four points on bay 1 (residual error 0.03 m) (Figure 5a). Comparisons with the TLS survey (Figure 5b) highlighted that deviations are really encouraging with only few centimetres shifts varying between 0.04 – 0.05 m in the xy plane in bay 8 and between 0.12 – 0.14 m in the vertical direction (bay 8) (Figure 5c,d).

3.3 The “Minguzzi” spiral staircase

The “Minguzzi” spiral staircase proved to be the most difficult of the three test areas to map, being a poorly lit and highly complex service space with a helicoidal shape.

Due to the darkness of the space, the device could not keep track of its position during the walk down the staircase.

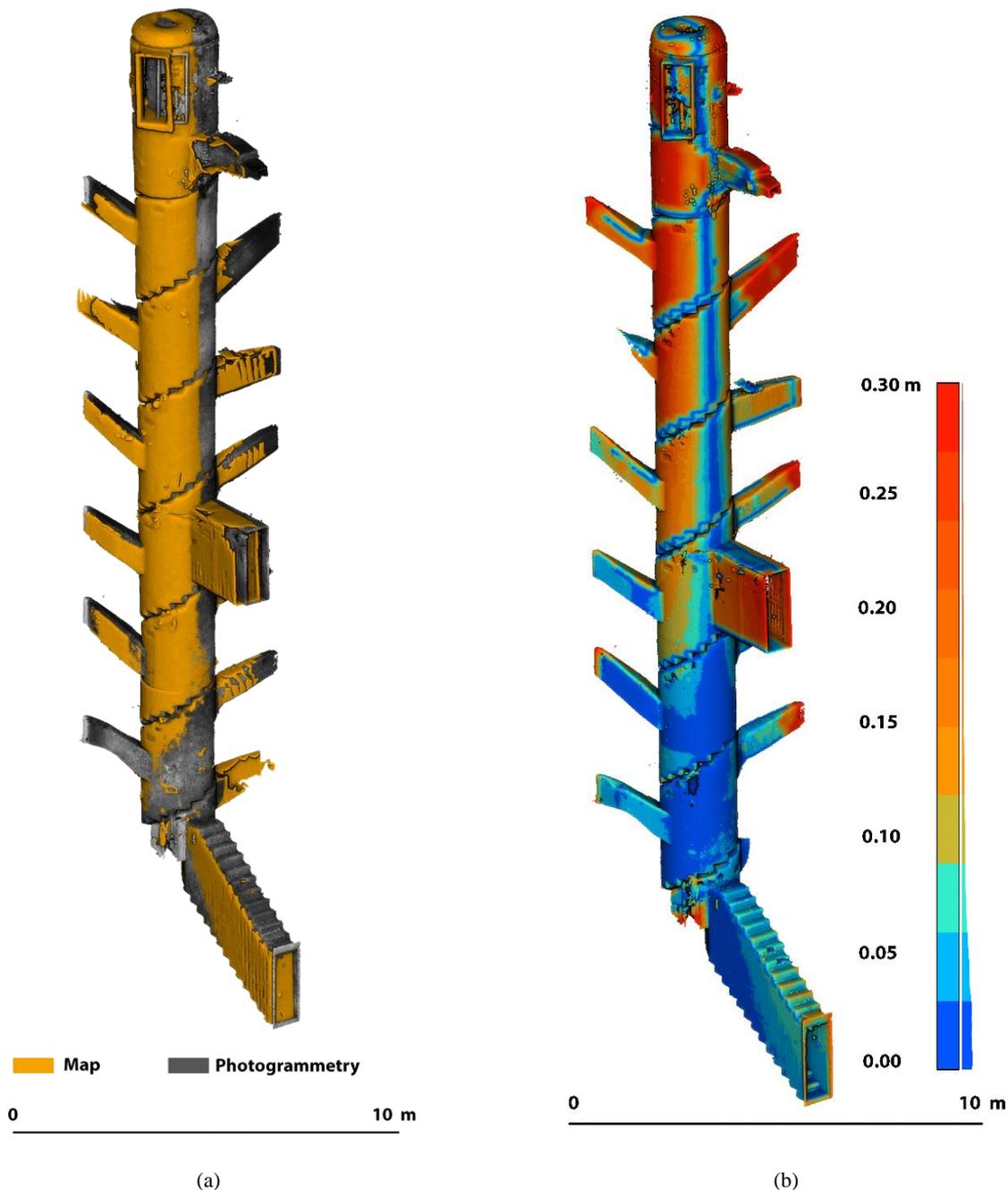


Figure 6. (a) Ground truth and spatial map comparison. (b) Ground truth with color-coded deviations from spatial map.

The problem was solved with an external portable illumination device. Furthermore, the staircase was too narrow to be scanned, looking directly at the walls. It proved impossible to obtain a complete map of the area with a single pass and, crossing the

space in the opposite direction was essential to get a full mesh model of the staircase.

The obtained map has been sampled at 0.01 m resolution and registered on the photogrammetry point cloud fixing four points

at the base of the staircase. The test here shows the highest deviations from the ground truth, reaching 0.4 – 0.48 m at the topmost part of the stair in xz plane and 0.45 m in the xy direction (Figure 7a,b). Figure 6a shows the result of the superimposition process with residual errors of 0.02 m and in Figure 6b it is possible to see a colour coded version of the ground truth point cloud in terms of absolute distances from the map.

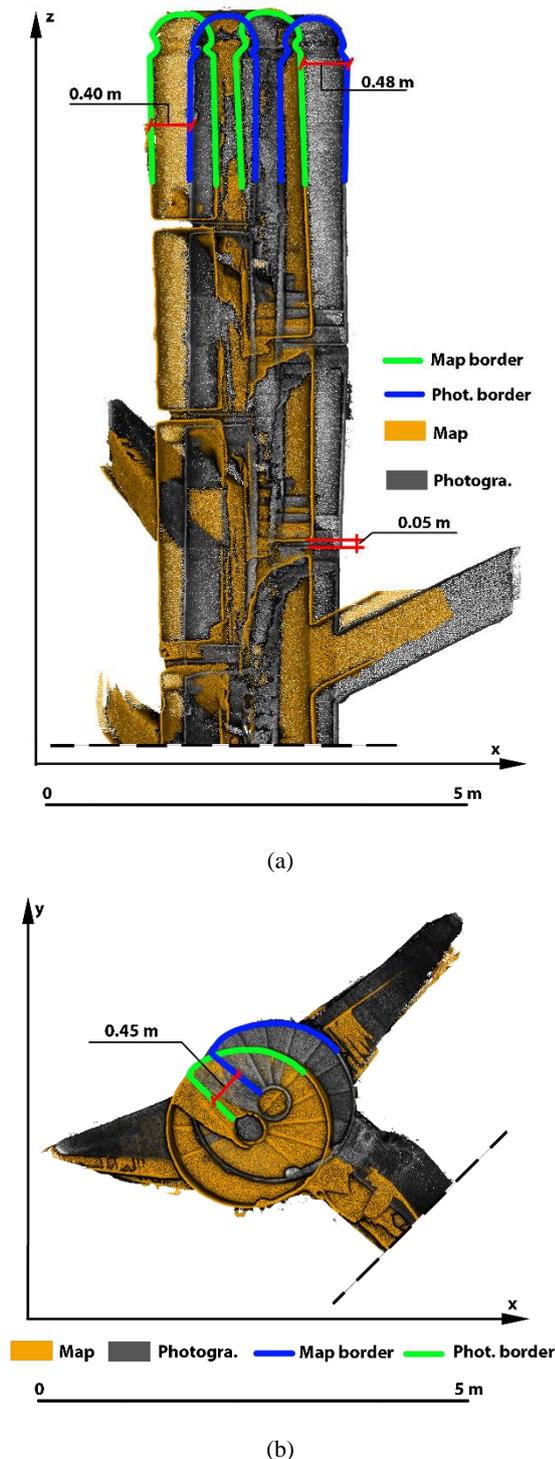


Figure 7. Spatial map ground truth deviations (a) on xy plane. and (b) on the xz plane.

4. DISCUSSION

Testing HoloLens 2 mapping capabilities on all three case study areas proved the effectiveness of the device as a mobile mapping instrument. The produced 3D mesh model, when compared with ground truth data, proved to have a good accuracy while mapping huge and complex spaces. Deviations from reference data are negligible locally on the point of registration and increase linearly with the distance. Experiments showed that environmental conditions highly influence the accuracy of the device with which 3D mesh maps are computed.

In fact, it exploits its four visible-light tracking cameras for real-time visual-inertial SLAM, allowing it to compute the device position in respect of the environment. At the same time, the long throw, low framerate depth sensor, is used to calculate the spatial map on the device (Ungureanu et al. 2020).

While it can capture frames illuminated from infrareds and is unaffected from ambient visible lights, the positional device tracking is highly dependent on rooms illumination. Furthermore, the device could fail to track its position in such scenarios characterized by strong single architectural element repetition and uniform surface patterns. It can frame only very similar portions of walls, floors and pillars, leading to errors in the spatial map (example at Figure 4b).

Experiments in the south nave of the Cathedral proved that the device cannot completely map the space. This is due to the huge dimension of the environment and highly depends on the depth sensor range limit at about 4 m distance. Highest deviations from ground truth in this area are visible in bay 8 reaching a maximum of 0.59 m on the xy plane and 0.1 m in the vertical direction.

The “Sordine”, a very different type of space, prove to be the best environment where to use the HoloLens inside the Cathedral. Being a more human-scale space, the device was able to map it accurately with maximum deviations of 0.05 m along the xy plane and 0.14 m in the vertical direction (bay 8).

Very narrow spaces, like the “Minguzzi” spiral staircase, proved the most challenging environments. Here, the depth sensor lower working range (about 0.5 m) made it impossible to survey parts of the environment closer to the device. In this area, deviations from the ground truth reach 0.48 m on the horizontal plane and 0.05 m along z-direction. The scanning process was mostly affected by bad environmental lighting and by the uniformity of the texture. Having a flashlight illuminating the environment was essential to avoid losing device position and spatial map placement while scanning the dark narrow space. In all three areas, deviation near the points used to register the spatial map on ground truth data are negligible (residual errors in the range of 0.02 – 0.03 m).

3D holograms are positioned in the real-world referred to the spatial maps acquired. However, while the user is using the HoloLens 2 to look at the virtual content, all deviations appreciated during map analysis performed in this work do not affect holograms stability near the device. This is because the device continuously analyses and maps the environment, adjusting map superimposition locally on the ambience. As seen from the comparisons performed, deviations are negligible while near the device, thus resulting in a good hologram superimposition and stability in a 6 to 7 m radius sphere that follows the user while walking across the space. This poses a real problem in areas with huge monumental dimensions like the south nave, where the user can and want to appreciate far holograms (e.g., information placed on the vaults highest point is at 24 m height).

5. CONCLUSIONS

This work presented a quantitative evaluation of the HoloLens 2 device as a mapping instrument inside huge and complex monumental environments.

For comparison, three test areas in Milan Cathedral have been selected. After carefully scanning all ambiances, 3D mesh models are compared against ground truth point cloud data produced during the survey activities performed inside the monument. The highest deviations from ground truth data can be appreciated in the monumental environment of the south nave (0.59 m along the x-axis). Furthermore, the "Minguzzi" spiral staircase proved to be the most challenging ambiance to map due to its narrow dimensions and to its bad lighting conditions. However, the instrument proved very effective in building an accurate local spatial map (3D mesh model) in all test areas. Device sensor limits played a crucial role in acquiring the geometry of the mapped environment. In fact, even if the device proved effective in mapping small but complex ambiances, it was impossible to provide a complete 3D map of the vast monumental areas of the Cathedral. Higher parts of the south nave are simply beyond HoloLens 2 long-throw depth sensor range.

The maps produced from the device are used as a reference to which the virtual content is positioned while augmenting the real environment. Since deviation from the ground truth data increases with the distance, this results in increasing hologram drifts that can be appreciated by the user. The device continuously refine map superimposition locally on the ambiance. As it results from the tests performed, deviations are negligible in 6 to 7 m radius sphere centred on the user. If this is not a problem locally and in small scale environments, it surely leaves an open research question in the quest to develop a Mixed Reality system to enhance the virtual fruition inside monumental spaces (e.g., south nave) where the user could appreciate far holograms as well. Different works are already in progress to solve the presented problems. World locking tools with their space pins features try to adjust the scale error generated with hologram positioning in large environments. Furthermore, the possibility to provide the HoloLens 2 with previously acquired complete spatial maps of monumental ambiances should be investigated.

REFERENCES

Achille, C., Fassi, F., Mandelli, A., Perfetti, L., Rechichi, F., Teruggi, S., 2020. From a traditional to a digital site: 2008- 2019. The history of Milan Cathedral surveys. In: *Daniotti, B., Gianinetta, M., DellaTorre, S. (eds.) Digital Transformation of the Design, Construction and Management Processes of the Built Environment. RD*, pp. 331–341. Springer, Cham. https://doi.org/10.1007/978-3-030-33570-0_30.

CloudCompare, 2020. CloudCompare version 2.11.3 Anoaia [GPL software]. <http://www.cloudcompare.org/> (24 January 2022).

Huang, J., Yang, B., Chen, J., 2017. Non-contact Measurement Method Research Based on HoloLens. *International Conference on Virtual Reality and Visualization (ICVRV)*, 267 – 271. <https://doi.org/10.1109/ICVRV.2017.00061>.

Hübner, P., Landgraf, S., Weinmann, M., & Wursthorn, S., 2019. Evaluation of the Microsoft HoloLens for the Mapping of Indoor Building Environments. *Dreiländertagung Der DGPF, Der OVG Und Der SGPF*, 44–53.

Hübner, P., Clintworth, K., Liu, Q., Weinmann, M., Wursthorn, S., 2020. Evaluation of HoloLens Tracking and Depth Sensing for Indoor Mapping Applications. *Sensors*, 20, 1021. <https://doi.org/10.3390/s20041021>.

Khoshelham, K., Tran, H., and Acharya, D., 2019. Indoor Mapping Eyewear: Geometric Evaluation of Spatial Mapping Capability of HoloLens. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W13, 805–810, <https://doi.org/10.5194/isprs-archives-XLII-2-W13-805-2019>.

Liu, Y., Dong, H., Zhang, L., Saddik, A. E., 2018. Technical Evaluation of HoloLens for Multimedia: A First Look. In *IEEE MultiMedia*, vol 25, no. 4, 8 –18. <https://doi.org/10.1109/MMUL.2018.2873473>.

Microsoft, 2020. HoloLens 2. <https://www.microsoft.com/en-us/hololens/hardware> (24 January 2022).

Microsoft, 2021a. HoloLens 2: Spatial Mapping. <https://docs.microsoft.com/en-us/windows/mixed-reality/design/spatial-mapping> (24 January 2022).

Microsoft, 2021b. Map physical spaces with HoloLens. <https://docs.microsoft.com/en-us/hololens/hololens-spaces>. (24 January 2022).

Microsoft, 2021c. Developing with Scene Understanding. <https://docs.microsoft.com/en-us/windows/mixed-reality/design/scene-understanding> (24 January 2022).

Perfetti, L., Polari, C., and Fassi, F., 2017. Fisheye Photogrammetry: Tests and Methodologies for the Survey of Narrow Spaces. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W3, 573–580. <https://doi.org/10.5194/isprs-archives-XLII-2-W3-573-2017>.

Teruggi, S., Fassi, F., 2021a. Machine Learning for Mixed Reality. The Milan Cathedral from Survey to Holograms. In: Del Bimbo A. et al. (eds) *Pattern Recognition. ICPR International Workshops and Challenges. ICPR 2021. Lecture Notes in Computer Science*, vol 12663. Springer, Cham., https://doi.org/10.1007/978-3-030-68796-0_44.

Teruggi, S., Fassi, F., 2021b. Mixed Reality for the monumental heritage. A first test. In *Proc. of the joint event 9th Arqueologica 2.0 and 3rd geores*, Valencia, Spain, 538 – 541.

Teruggi, S., Grilli, E., Fassi, F., and Remondino, F., 2021c. 3D Surveying, Semantic Enrichment and Virtual Access of Large Cultural Heritage. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, VIII-M-1-2021, 155–162, <https://doi.org/10.5194/isprs-annals-VIII-M-1-2021-155-2021>

Ungureanu, D., Bogo, F., Galliani, S., Sama, P., Meekhof, C., Stühmer, J., Cashman, T. J., Tekin, B., Schönberger, J. L., Olszta, P., Pollefeys, M., 2020. HoloLens 2 research mode as a tool for computer vision research. *arXiv preprint arXiv:2008.11239*.

Unity 3D, 2021. Unity 3D Software. <https://unity.com/> (24 January 2022).

Weinmann, M., Wursthorn, S., Weinmann, M., Hübner, P., 2021. Efficient 3D Mapping and Modelling of Indoor Scenes with the Microsoft HoloLens: A Survey. *PFG* 89, 319–333, <https://doi.org/10.1007/s41064-021-00163-y>.