

3D SURVEYING, SEMANTIC ENRICHMENT AND VIRTUAL ACCESS OF LARGE CULTURAL HERITAGE

Simone Teruggi ^a, Eleonora Grilli ^b, Francesco Fassi ^a, Fabio Remondino ^b

^a 3DSurveyGroup – ABCLab – Politecnico di Milano, via Ponzio 31, 20133 Milano, Italy
E-mail: <simone.teruggi>< francesco.fassi >@polimi.it; Web: <http://www.sitech-3dsurvey.polimi.it>

^b 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy
E-mail: <grilli>< remondino>@fbk.eu; Web: <http://3dom.fbk.eu>

KEY WORDS: 3D Point cloud, 3D classification, Augmented Reality, Mixed Reality, HoloLens

ABSTRACT: In recent years, Artificial Intelligence (AI) methods for 3D point cloud classification has assumed an essential role in the heritage field. The association of semantic information to 3D representations became a valuable instrument for measurement, analysis, education or maintenance, in particular in the Cultural Heritage (CH) sector. Moreover, the recent availability and reliability of head-mounted displays and glasses are allowing extraordinary immersive virtual experiences. This paper presents an end-to-end framework to handle large and complex 3D point clouds, from acquisition to semantic segmentation and final access in a Mixed Reality (MR) environment. Three completely different heritage scenarios are considered: the Temple of Neptune in Paestum, the Milan Cathedral and a large portion of Bologna's porticoes. Mixed Reality experiences are described and shown based on the Microsoft HoloLens 2 device. A video of the results is available at https://youtu.be/Kd_3s0tIX04.

1. INTRODUCTION

In recent years, 3D point cloud data have been extensively used to record, document, valorise, restore and visualise cultural heritage. Despite this, point clouds - the first product of a 3D survey and the input for many different products - continue to be rarely used by a wider audience, including experts. In addition, despite the high metric quality, 3D data generally lack additional information such as semantics and elements hierarchy. These issues could be considered the main reasons why point clouds are not used directly in daily and production processes but have instead to be treated, synthesised and transformed into formats usable in other applications (3D meshes, parametric models, 2D representation, etc.). These post-processing products, which are typically performed with manual or poorly automated processes, try to keep geometric details and quality but usually deteriorate surveying data. Therefore, surveying 3D data are primarily used for basic data visualisations or otherwise handled by experts.

Based on these considerations, current research in the Cultural Heritage (CH) field seeks to make direct use of point clouds in the development, management and knowledge cycle.

This paper presents an innovative end-to-end framework that starts from high-resolution 3D surveying data and delivers mixed reality (MR) experiences using a Holographic Head Mounted Display – HHMD system. Firstly, point clouds are semantically enriched with advanced machine learning methods (Section 3.1), then they are further processed and visualised within a HoloLens 2 device (Section 3.2). The pioneering contribution is to handle large amounts of 3D heritage point clouds in HHMD system, exploiting 3D classification results.

1.1 3D data classification

The term semantic segmentation (or simply classification) for point clouds refers to the grouping of similar data into subsets (called segments) that have characteristics/features (i.e., geometric, radiometric) useful to distinguish and identify in classes the different parts of the point cloud (Grilli et al., 2017). The semantic subdivision of 3D data leads to a hierarchy that facilitates data access, decision-making, and design processes. In addition, it can be preparatory to successive applications such as the reconstruction phase of simplified 3D models (CAD or BIM).

At the same time, working with a classified point cloud could speed-up architecture's analysis and learning, maintenance operations or conservation plans, removing the complicated and time-consuming scan to BIM processes. In fact, the ability to convert the point cloud into an abacus of semantically enriched segments allows them to be associated with morphological and material details and used to define areas, volumes, and masses (if materials are homogeneous). As a result, a metric measurement of the architecture is made, which serves as a valuable tool for restoration and maintenance projects. Furthermore, the process of defining, highlighting, and isolating the main architectural elements allows for data reduction and quicker spatial analysis of the structure, enabling 3D information to be used on mobile devices such as tablet phones or virtual devices. In this way, it is possible to conduct studies on individual elements, check the sizing and positioning in space according to modular rules, and analyse and compare recurring elements and singularities. Recently, standard machine learning approaches and more elaborated and complex deep learning methods showed impressive results when employed for 3D point cloud classification, in particular in the heritage field (Grilli and Remondino, 2019; Malinverni et al., 2019; Teruggi et al. 2020; Matrone et al., 2020b).

1.2 VR, AR, MR, XR

For the sake of clarity, following Margetis et al. (2021) and Skarbez et al. (2021), the following definitions should hold (Figure 1):

- *Virtual Reality* (VR): VR is an immersive experience based on headsets that generate realistic 3D contents, sounds and other sensations to replicate a real environment or create an imaginary world.
- *Augmented Reality* (AR): AR is a live view of a real-world environment with augmented and super-imposed contents. The augmentation is achieved utilising devices like smartphones, tablets or custom headsets and dedicated apps that overlay digital contents onto the seen real environment (without interaction).
- *Mixed Reality* (MR): MR – or hybrid reality, is the merging of real and virtual worlds to produce a new environment where physical and digital objects interact and cooperate in real-time.

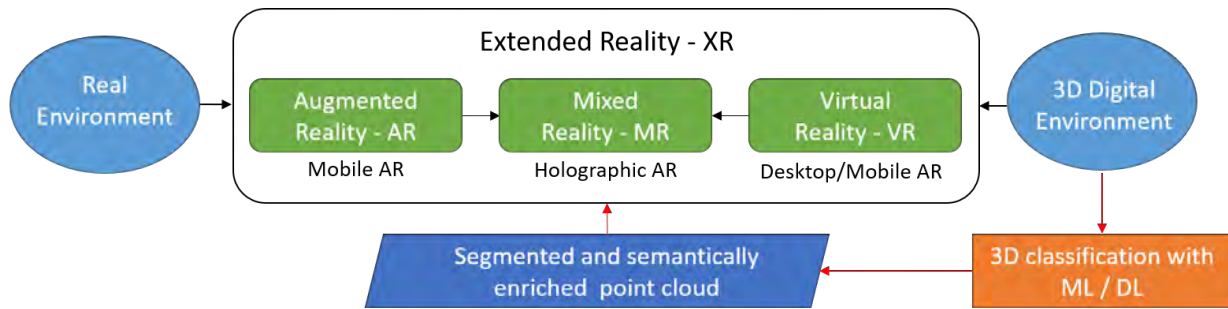


Figure 1. The reality-virtuality continuum of technologies (adapted from connected Milgram and Kishino, 1994) connected to 3D point cloud classification methods.

In Milgram et al. (1995), MR refers to all that is comprised between the physical world and the completely virtual one.

- *Extended Reality (XR)*: XR refers to an umbrella that brings all three Realities (AR, VR, MR) under a unique term. Indeed, all these new technologies create bridges of interaction between us and the surrounding environment, and this mixture between real and virtual worlds is often referred to X-Reality (Mann et al., 2018; Wallgrün et al., 2018).

The distinction between AR and MR has become blurry. However, thanks to the diffusion of devices such as the Microsoft HoloLens 2, MR is now used as a synonym of Holographic AR (Pedersen et al., 2017). The distinction between AR and MR lies in how the user perceives the virtual content. In MR, 3D objects are displayed as holograms. These are objects made of light that seem to be part of the physical world with the same spatial presence as real objects. The user can see them using special glasses (Holographic Head-Mounted Displays - HHMDs) that projects these holograms into the physical world. In addition, these special HHMDs, enable the user to interact directly with the holographic content through hand gesture recognition, gaze, and voice commands. Since projected 3D models have the same spatial presence as real objects, one movement inside the 3D world corresponds exactly to the same movement into the physical world. This blend combines capabilities from the worlds of VR (in which the user is fully immersed in the computer graphic world) and standard AR, resulting in an environment where digital and physical coexists and interact (Teruggi and Fassi, 2021).

MR and AR have been used in a variety of fields: assembly and disassembly (De Pace et al., 2018), human-robot interaction (Avalle et al., 2019), navigation (Fraga-Lamas et al., 2018), medicine (Desselle et al., 2020; Brun et al., 2019), Architecture Engineering Construction and Operations - AECO (Bae et al., 2015; Golparvard-Fard and Ham, 2014; Alam et al., 2017) as well as cultural heritage (De Carolis et al., 2018; Carrozzino et al., 2019).

1.3 3D surveying data in XR

Using surveyed 3D data in the field can facilitate the interpretation, comprehension and storytelling of geometrical shapes in real-time. The opportunity to consult surveying results overlaid to real scenarios, schedule maintenance interventions, conduct in-situ analysis or access and update technical documentation on-site and in real-time is essential (but still quite futuristic) for improving the building management process. Tourism could also benefit more from VR/AR/MR solutions to allow a broad audience to understand history/architecture, proposing virtual guides and educational storytelling.

The main idea is to apply concepts from Industry 4.0, where AR and MR systems are used to reduce production costs, increase efficiency and ease working processes. Unfortunately, the

cultural heritage (CH) world is far from operational scenarios, with applications limited to tourism and education (Blanco-Pons et al., 2019; Voinea et al., 2019). One explanation could be that creating appropriate 3D contents to support an AR system is complicated, especially when applied to very large and complex cases. When dealing with CH scenarios, a complicated modelling phase is necessary to create 3D models that can be displayed through MR devices. While the modelling process is generally easy and straightforward with simple objects or indoor scenarios, it becomes nearly impossible, time-consuming and uneconomic with CH scenarios. In fact, in the CH field, even repetitive objects are different from others when scale and metric accuracy must be retained (Fassi et al., 2011). Furthermore, the modelling phase introduces an unsupervised degree of subjectivity and a reduction in geometric detail that can lower the metric quality intrinsic to the surveyed point cloud. These simplifications make an MR system generally not usable as they would hamper precise referencing information and hologram positioning in the physical world.

This work proposes to skip manual object modelling in favour of directly using semantically enriched point clouds in an MR framework, demonstrating that complex CH architectures can also be used in this innovative process.

2. CASE STUDIES AND NEEDS FOR XR

The proposed methodology (Section 3) is tested and validated using three different case studies (Figure 2):

- the Temple of Neptune in Paestum, Italy: the 450 BC Temple of Neptune is the best-preserved temple in the ancient Greek city of Paestum, one of the most famous Doric monuments of the ancient world. It is approximately 24,5 x 60 m in size, with 6 frontal and 14 lateral columns and two rows of double ordered columns in the interior. The used point cloud consists of 32 million points, and it is a combination of UAV and terrestrial photogrammetric survey (Fiorillo et al., 2013).
- the Cathedral in Milan, Italy: the construction began in 1386 and was finished in 1805, with some final details completed in 1965. The monument spans some 158 x 93m, and it is approximately 108 m in height (till the last golden statue). Like other Gothic cathedrals, it features many sculptures, with 135 gargoyles and 700 figures that decorate the interior spaces, such as capitals, altars and windows. Some 135 spires and 30 decorative reverse arches on the exterior facades seem to support the impressive building's flanks. Several focused survey operations were performed in the last ten years to support special maintenance activities (Achille et al., 2020): Terrestrial Laser Scanning (TLS) for interior spaces and photogrammetry for exteriors were used, collecting some 60 billion 3D points. The point cloud used for the presented

experiment (30 million points) covers the Cathedral's main noble areas and the nearby exteriors.



Neptun temple in Paestum, Italy	Cathedral in Milan, Italy	Porticoes in Bologna, Italy
3D Surveying		
UAV and terrestrial photogrammetry	Terrestrial Laser Scanning and close-range photogrammetry	Close-range photogrammetry
Amount of 3D points		
32 mil.	30 mil.	107 mil.

Figure 2. The large heritage monuments considered in the MR experiments and their specs.

- a portion of porticoes in Bologna, Italy: the old porticoes, which extend approximately 40 kilometres within the old town, were constructed between the 11th and 20th centuries. These structures, including a range of geometric shapes, materials, and architectural features, were digitised using terrestrial photogrammetry as part of a project to designate the porticoes as a UNESCO "world heritage site" (Remondino et al., 2016). The 3D dataset used for this study contains about 107 million points.

The three case studies are significant examples of archaeological, monumental and architectural buildings of cultural interest. They have been subjected to maintenance and conservation activities over time. For years now, there has been a consolidated awareness that an accurate and complete 3D digitalisation is indispensable for various maintenance activities. More recently, there has been a growing need to make better use of this data to share information, enhance the asset's value and disseminate information. Hence the idea of developing ad-hoc MR systems based on advanced HHMD, would allow on-site interactive data access, visualisation and knowledge exchange and sharing.

- For the Temple of Neptune, an HHMD system would provide, besides a virtual touristic guide on-site, valuable real-time information about structural monitoring activities or the previous state of conservation directly in the field during inspections and maintenance activities. At the same time, the system could allow the collection and storage of new and updated information about the monument (e.g. new locations on intervention, etc.).

- For the gigantic Milan Cathedral, an MR framework for on-site workers would be an invaluable tool during the continuous and repeated inspections to verify the structural stability and the state of conservation of the surfaces. Furthermore, MR would promote a better and more immediate understanding of the evolution of deterioration phenomena and a constant and straightforward updating of information.

- For architectures like Bologna's porticoes, MR could give on-site access to heritage information and municipal overlay databases related to the land registry, road cadastre, etc.

3. METHODOLOGY

This section presents the machine learning (ML) multi-level multi-resolution (MLMR) classification framework (Section 3.1) applied to large heritage point clouds in order to facilitate their use within mixed reality (MR) applications (Section 3.2).

3.1 3D classification

Machine and deep learning approaches are based on four main steps: manual data annotation/labelling, feature extraction, model training and prediction. When applied to 3D point clouds, traditional methods generally include all the classes of interest at the same level of training and work at a constant geometric resolution. Even though it has been demonstrated that these methods are effective for heritage classification (Grilli and Remondino, 2020; Matrone et al., 2020a), in large and complex scenarios, a single-phase classification becomes challenging as:

- the amount of data makes the computational process difficult. A sub-sampling of the point cloud could help to manage large quantities of points but would lower the level of detail of the original dataset, and small details (geometric elements) could get lost.
- the high number of semantic classes to be identified at the same time can lead to classification errors.

To overcome these issues, a multi-level multi-resolution (MLMR) approach is here adopted (Teruggi et al., 2020) for subdividing large and complex heritage 3D scenarios into their compositive elements. The MLMR approach proved that hierarchically classifying 3D data at different geometric resolutions can facilitate the learning process and boost semantic segmentation results. Therefore, to prepare 3D data for MR application, the following steps are adopted:

1. Subsampling of the full-resolution point cloud, manual annotation of a data portion, and training of an ML classifier to identify different macro-categories.
2. Back interpolation of the classification results on a higher resolution version of the dataset;
3. Focussing the attention on some limited portions that need to be sub-classified (e.g., columns can be divided into base, shaft, and capitals).

According to classification needs, steps 2 and 3 can be reiterated until the full geometric resolution is reached.

3D classification results must achieve a high level of accuracy to handle and visualise 3D data within MR systems correctly. Therefore, the commonly used accuracy metrics (Overall Accuracy and F1 scores) are used to measure the reliability of the classification results.

3.2 Mixed Reality development

MR environment: the MR applications are created using the HoloLens 2 device (Microsoft, 2021). Microsoft HoloLens 2 is

an HHMD computer that refined its 1st generation to provide a more comfortable and immersive experience paired with more options for collaborating in MR. A user can see and manipulate 3D objects as holograms in the physical world through a pair of transparent screens and exploiting the various onboard sensors (Figure 3). The MR application is developed within Unity3D (Unity3D, 2021) with the MRTK v2 libraries (MRTK, 2021), which enable Microsoft HoloLens 2 devices to import and handle point clouds (PCX, 2021).

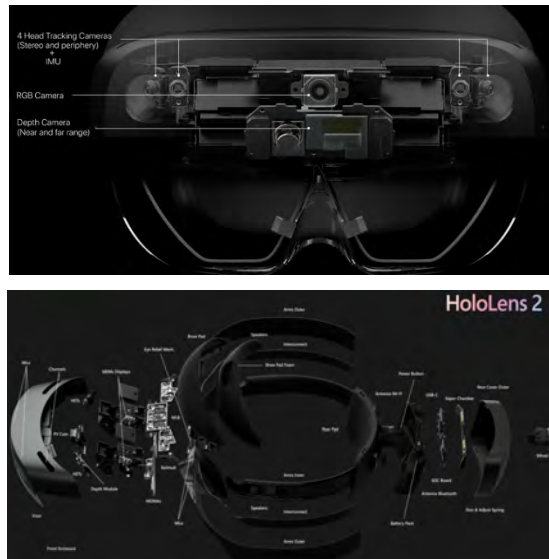


Figure 3. The HHMD HoloLens 2 device with its various components (Microsoft, 2021).

Application design: Although HoloLens 2 is a powerful device, it suffers from computational and battery life limits. It can handle and render 600,000 points inside its field of view before a noticeable frame drop. Its battery life is set around 2-3 hours, depending on the workload. Starting from these considerations, it would be impossible to use, visualise and interact with full-resolution 3D datasets composed of hundreds of millions or even billions of points – typical of large heritage 3D surveying. In this context, the classification process (Section 3.1) directly influences the way the MR application is structured:

1. Different geometric resolutions related to the level of classification can be accessed and visualised. Low-resolution point clouds are used for representing oversized objects in their entirety (groups of columns, friezes, facades, etc.). Higher densities are employed for viewing single instances (single columns, single building facades, etc.). Finally, full resolutions are used only for sub-elements (e.g., single capitals, single shaft) that can be visualised with a great level of detail. Therefore, the point cloud resolution is proportional to the zoom level at which the object is represented and visualised. In this way, the computational power necessary to render a particular scene is reduced, increasing battery life and maintaining a fluent experience while navigating, orbiting and loading different parts of the scene. The application's frame rate must be equal to or greater than 60 frames per second (fps). Otherwise, the experience would result in hologram's instability and flickering, destroying the complete MR user experience.
2. The MLMR classification results directly affect the structuring of the MR application within Unity. As loading the entire 3D dataset in a single MR scene would cause a long loading process and low performance, the dataset is divided into different scenes. Each set contains only elements related

to a specific construction area and classification level, limiting the need for object loading to only some parts.

Point cloud interaction: The HoloLens 2 can recognise hand gestures, gaze and voice commands. Four primary types of hand gestures are used to interact with the 3D heritage point cloud: hand ray / near pointer, hand touch, air tap, air tap & hold. When a user's hand approaches the field of view of the HoloLens 2, one of its depth sensors begins to monitor the user's hand. It is then interpreted as a skeletal model, and a white dashed line appears projecting forward. This hand ray is used as a pointer to pick and highlight various 3D objects that are far away from the user. When this ray is pointed at an object, the air tap gesture can be used to confirm the pick. On the other hand, when objects are close, the dashed ray transforms into a near pointer circle displayed on the index finger fingertip. It is possible to pick holograms simply by rubbing them with the pointer circle. Instead, the air tap and air tap & hold motion can be made on both close and far holograms. It allows for pinching 3D objects and shifting/orbiting them around the scene with one hand or increasing/decreasing their scale with both hands (pinching together the index and the thumb finger).

Our interaction solution maintains the above structure, demonstrating that the technique can be applied to objects with different dimensions and complexity. When the user starts the application, the relevant parts of the 3D dataset are loaded. When the user looks at her / his hand, a small navigation menu appears, which allows performing operations like splitting the point cloud into its various architectural elements, resetting elements positions or going back to the previous classification level. Single macro elements can be selected and isolated at higher resolution using the hand ray or close pointer, confirming the selection with an air tap gesture. If the split button is pressed a second time, macro elements are split into architectural parts that can be highlighted, picked and visualised at a higher resolution. The process is iterative up to single instances of objects displayed at full resolution. In addition, it is possible to attach different types of information (e.g., images, tables, intervention procedures) to each level. This information is displayed along with the 3D elements and can be query while navigating the point cloud.

4. EXPERIMENTS

4.1 Temple of Neptune in Paestum

The 3D point cloud of the temple is divided into his compositive elements using the MLMR classification approach (Section 3.1). Three levels of details and 12 semantic classes are used (Table 1). Part of the low-level resolution (first classification level) at 10 cm sampling distance is the starting point within the MR application. This low-resolution 3D temple (Figure 4a) is used as a keymap where it is possible to select the different macro-elements that need further inspection.

The holographic hand menu gives the possibility to perform actions on the point cloud at the various classification levels. It is possible to i) split the point cloud into its constituting elements reaching the next level of classification (Figure 4b-c); ii) go back to the previous classification level; iii) reset all 3D elements to their original position and scale. The whole point cloud at the first classification level (10 cm resolution) can be interactively split into its macro-elements (grass, crepidoma, floor, etc.) and, using the hand ray/near pointer gestures, it is possible to highlight and select with an air tap each macro element.

The single architectural objects segmented in the successive classification levels can also be overlayed and queried - at higher resolution - within the holographic hand menu. Single column instances (nearly 15,000 points per column) are displayed at 5 cm

resolution (classification level 2). Their architectural elements (shaft, echinus, abacus) can be inspected and queried (Figure 4c).

	LEVEL 1	LEVEL 2	LEVEL 3
Classes	Grass Crepidoma Floor Column Wall Trabeation Pediment	Shaft / Echinus / Abacus Architrave / Frieze Cornice/ Tympanum	Triglyph / Metope
Point number	462,649	2,209,521	32,470,628
Resolution (cm)	10	5	1.5 (Full resolution)
Number of classes	7	11	12
Overall Accuracy (%)	96.40	92.04	97.80
F1 score (%)	96.22	92.13	97.77

Table 1. Temple of Neptune: classification levels, metrics and point cloud resolutions at each level.



Figure 4. MR access to the first level of classification of the Neptune temple in Paestum (a). Selection of the class “column” (b) and virtual interaction with the sub-elements abacus and associated textual information (c).

4.2 Milan Cathedral

Starting from the acquired 3D data of the Cathedral’s main noble areas and the nearby exteriors (ca 30 mil. points), the MLMR classification produces 19 semantic classes, as summarised in Table 2.

Given the complexity of the 3D heritage dataset and the high heterogeneity of objects and spaces, it is impossible to exploit all 3D elements simultaneously. Even if a fairly low geometric resolution of the point cloud is used, it isn’t easy to distinguish all architectural details on a small-scale visualisation. Therefore, the macro-objects identified during the classification process had to be grouped to describe different sectors of the Cathedral (each sector corresponds to a bay within the monument).

The MR experience starts from the first classification level (5 cm resolution and nine semantic classes), where all macro elements are loaded and displayed as a single point cloud. It is possible to move, rotate and zoom the point model using hand gestures (air tap or touch) and, with a split button, the point cloud can be divided into its constituting elements (pillars, vaults, etc.).

Iteratively, classification level 2 is reached with the split button and isolating single architectural components (Figure 5, resolution of 2 cm, average of 200,000 points depending on the visualised component).

Finally, the single objects can be displayed and queried at full resolution (classification level 3, resolution of 5 mm, up to 50,000 points for a single statue).

As shown in the results, different types of information can be attached to the point cloud data and displayed in MR, independently from the classification level considered. Up to now, data are stored locally on the HoloLens, but a structured dataset in the Cloud could become the base for a multi-user on-site informative system that the *Veneranda Fabbrica del Duomo di Milano* could adapt to display, inquiry, update and store all information related to the maintenance and restoration activities. It is planned to use the system primarily in the inspection phases to keep track of the state of preservation, risk status, and hazard of Gothic structures. In addition, it is intended to link the digital archive of the Cathedral so that historical documents are spatially referenced within the Cathedral.

	LEVEL 1	LEVEL 2	LEVEL 3
Classes	Floor Column Chain Vault Roof Wall Buttress Pavement Choir	Base / Shaft / Capital Wall / Basement / Window / Falconatura*	Capital/Statue Wall / Molding / Statue / Gothic Pyramid / Balustrade
Point number	32,341,884	224,367,181	3,562,360,653
Resolution (cm)	5	2	0.5 (Full resolution)
Number of classes	9	14	19
Overall Accuracy (%)	94.74	99.44	91.47
F1 score (%)	94.72	99.43	85.01

Table 2. Milan Cathedral: classification levels, metrics and point cloud resolutions at each level.

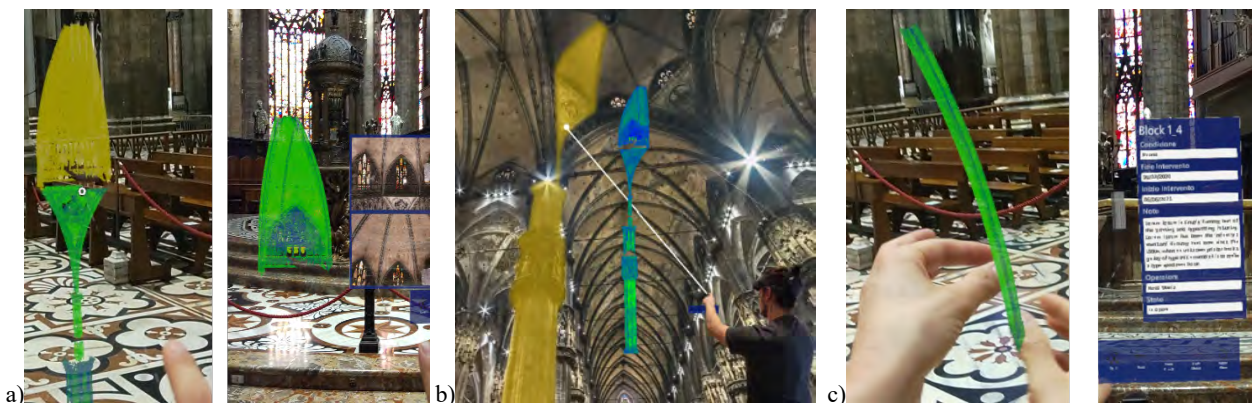


Figure 5. MR access to the segmented 3D point cloud of the Milan Cathedral. Pillar element visualised in front of the main altar (a), pillar, pendentive and vault section (b) and single marble element (with associated attributes about maintenance) isolated and displayed at higher resolution.

4.3 Porticoes in Bologna

The three-level classification achieved with the MLMR approach (Table 3) is essential to exploit the Bologna point cloud inside the HoloLens 2 device. The whole point at first is displayed at a 10 cm resolution (376,017 points). As “façade” is a predominant class in terms of the number of points (the sub-class “wall” at 2 cm would display some 936,817 points), it was important to organise the dataset by grouping all macro-elements classified at the first classification level into different blocks. Using the hand ray / near pointer gesture, it is possible to isolate one part of the dataset in the low-resolution point cloud (Figure 6a). The MR visualisation process continues iteratively up to the different architectural elements present in the classification level. Through the dedicated menu button, the 3D data can be split in its macro-

element and visualised in higher resolution (Figure 6b). In addition, single instances can also be manipulated (Figure 6c).

5. CONCLUSIONS

The paper presented an end-to-end framework to exploit 3D surveying point cloud in an MR system. Accessing 3D survey data on-site using an HHMD system - like the HoloLens 2 headset - would undoubtedly benefit storytelling and knowledge sharing, e.g., for tourism purposes or maintenance intervention, building management processes, as well as access and update of technical documentation in real-time. The presented framework - contrary to other solutions which use 3D polygonal meshes - is based on acquired 3D point clouds and then applies a multi-level

	LEVEL 1	LEVEL 2	LEVEL 3
Classes	Floor Column Vault Facade	Base / Shaft / Capital Wall / Not Wall	Moulding / Windows
Point number	376,017	8,0166,641	107,693,026
Resolution (cm)	10	2	0.5 (Full resolution)
Number of classes	4	7	8
Overall Accuracy (%)	95.41	93.02	88.83
F1 score (%)	97.24	93.37	91.65

Table 3. Porticoes in Bologna: classification levels, metrics and point cloud resolutions at each level.



Figure 6. MR access to an entire building block (a), selection and visualisation of a segmented object in the first classification level (b) and virtual interaction with the segment "vault" class (c).

multi-resolution (MLMR) classification approach to semantically split and enrich the 3D data. This step is essential to structure the 3D datasets and use them in the holographic device. Indeed, the classification procedure improves the application's computational usability, maintains a fluent hologram experience and creates a logical subdivision of the architecture components. The MLMR process produced different point clouds with increasing resolutions and semantic information. The final classification level (full resolution) is used only to inspect single elements from a close-up point of view, allowing accurate positioning of punctual, informative hotspots. Many future works are already planned. First of all, it has to be improved the referencing process of the holograms in the physical world. This operation is mandatory to correctly orient and super-impose 3D holograms to physical world objects, thus giving the correct spatial meaning to information and data; correctly referenced to the point model. Furthermore, the ability to add, change, and consult the data in real-time and on-site through an internet connection via an online DB would certainly improve system usability for all actors working on the CH object. Concluding, far from being usable in everyday practice, the developed prototype showed the feasibility of the research project. Point cloud data could become a standard instrument to be used directly on-site, and CH could benefit from the innovation brought by new technologies such as HoloLens 2.

ACKNOWLEDGEMENTS

The authors would like to thank all the colleagues who have participated in the past and are now collaborating in the Milan Cathedral project. A special thanks to Ing. Francesco Canali, yard director of "Veneranda Fabbrica del Duomo di Milano". This research was partly funded by the project "Artificial Intelligence for Cultural Heritage" (AI4CH - <https://ai4ch.fbk.eu/>), a joint Italy-Israel lab which was funded by the Italian Ministry of Foreign Affairs and International Cooperation (MAECI).

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 *Digital Transformation of the Design, Construction and Management Processes of the Built Environment*, pp. 331-341, Springer.
- Alam, M., F., Katsikas, S., Beltramello, O., Hadjiefthymiades, S., 2017. Augmented and virtual reality-based monitoring and safety system: a prototype IoT platform. *J. Netw Comput Appl*; 89:109-19.

- Avalle, G., De Pace, F., Fornaro, C., Manuri, F., Sanna, A., 2019. An Augmented Reality System to Support Fault Visualisation in Industrial Robotic Tasks. *IEEE Access*, Vol. 7, pp. 132343-132359.
- Bae, H., Golparvar-Fard, M., White, J., 2015. Image-based localisation and content authoring in structure-from-motion point cloud models for real-time field reporting applications. *J Comput Civ Eng*; 29(4):1-13.
- Blanco-Pons, S., Carrion-Ruiz, B., Duong, M., Chartrand, J., Fai, S., Luis Lerma, J., 2019. Augmented Reality Markerless Multi-Image Outdoor Tracking System for the Historical Buildings on Parliament Hill. *Sustainability*, 11, 4268.
- Brun, H., Bugge, R., A., B., Suther, L., K., R., Birkeland, S., Kumar, R., Pelanis, E., Elle, O., J., 2019. Mixed reality holograms for heart surgery planning: First user experience in congenital heart disease. *Eur Heart J Cardiovasc Imaging*; 20(8):883-888.
- Carrozzino, M., Voinea, G.-D., Duguleană, M., Boboc, R. G., Bergamasco, M., 2019. Comparing innovative XR systems in cultural heritage. A case study. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W11, 373-378.
- De Carolis, B. N., Gena, C., Kuflik, T., Origlia, A., Raptis, G. E., 2018. AVI-CH 2018: Advanced Visual Interfaces for Cultural Heritage. In Proc. *ACM Int. Conf. on Advanced Visual Interfaces*.
- De Pace, F., De, Manuri, F., Sanna, A., 2018. Augmented Reality in Industry 4.0. *Am J Comput Sci Inf Technol*; Vol.6 No.1:17.
- Desselle, M., R., Brown, R., A., James, A., R., Midwinter, M., J., Powell, S., K., Woodruff, M., A., 2020. In *Computing in Science & Engineering*, vol. 22, no. 3, pp. 18-26.
- Fassi, F., Achille, C., Fregonese, L., 2011. Surveying and Modelling the Main Spire of Milan Cathedral Using Multiple Data Sources. *The Photogrammetric Record*, Vol. 26, 462-487.
- Fiorillo, F., Fernández-Palacios, B. J., Remondino, F., Barba, S., 2013. 3D Surveying and modelling of the Archaeological Area of Paestum, Italy. *Virtual Archaeology Review*, 4(8), pp. 55-60.
- Fraga-Lamas, P., Fernandez-Carames, T., M., Blanco-Novoa, O., Vilar-Montesinos, M., A., 2018. A Review on Industrial Augmented Reality Systems for the Industry 4.0 Shipyard. In *IEEE Access*, vol. 6, pp. 13358-13375.
- Golparvar-Fard, M., Ham, Y., 2014. Automated diagnostics and visualisation of potential energy performance problems in existing buildings using energy performance augmented reality models. *J Comput Civ Eng*; 28(1):17-29.
- Grilli, E., Menna, F., Remondino, F., 2017. A review of point clouds segmentation and classification algorithms. *ISPRS Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W3, 339-344.
- Grilli, E., Remondino, F., 2019. Classification of 3D Digital Heritage. *Remote Sens.* 11, 847.
- Grilli, E., Remondino, F., 2020. Machine Learning Generalisation across Different 3D Architectural Heritage. *ISPRS Int. J. Geo-Inf.* 9, 379.
- Malinverni, E. S., Pierdicca, R., Paolanti, M., Martini, M., Morbidoni, C., Matrone, F., Lingua, A., 2019. Deep Learning for Semantic Segmentation of 3D Point Cloud. *ISPRS Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W15, 735-742.
- Mann, S., Furness, T., Yuan, Y., Iorio, J., Wang, Z., 2018. All Reality: Virtual, Augmented, Mixed (X), Mediated (X,Y), and Multimediated Reality. *arXiv preprint*.
- Margetis, G., Apostolakis, K.C., Ntoa, S., Papagiannakis, G., Stephanidis, C., 2021. X-Reality Museums: Unifying the Virtual and Real World Towards Realistic Virtual Museums. *Appl. Sci.*, 11, 338.
- Matrone, F., Grilli, E., Martini, M., Paolanti, M., Pierdicca, R., Remondino, F., 2020a. Comparing Machine and Deep Learning Methods for Large 3D Heritage Semantic Segmentation. *ISPRS Int. J. Geo-Inf.* 9, 535.
- Matrone, F., Lingua, A., Pierdicca, R., Malinverni, E. S., Paolanti, M., Grilli, E., Remondino, F., Murtyoso, A., Landes, T., 2020b. A benchmark for large-scale heritage point cloud semantic segmentation. *ISPRS Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLIII-B2-2020, 1419-1426.
- Microsoft, 2020. HoloLens 2: <https://www.microsoft.com/it-it/hololens/hardware>. Last access: Jun 29th, 2021
- Milgram, P., Kishino, F., 1994. A taxonomy of mixed reality visual displays. *IEICE Trans. Inf. Syst.*, 77, 1321-1329.
- Milgram, P., Takemura, H., Utsumi, A., Kishino, F., 1995. Augmented Reality: A class of displays on the reality-virtuality continuum. *Telemanipulator Telepresence Technol.* 2351, 282-292.
- MRTK, 2021. GitHub - microsoft MixedRealityToolkit-Unity. <https://github.com/microsoft/MixedRealityToolkit-Unity>. Last access: Jun 29th, 2021.
- PCX, 2021. GitHub – Keijiro, PCX. <https://github.com/keijiro/Pcx>. Last access: Jun 30th, 2021.
- Pedersen, I., Gale, N., Mirza-Babaei, P., & Reid, S., 2017. More than Meets the Eye: The Benefits of Augmented Reality and Holographic Displays for Digital Cultural Heritage. *ACM Journal on Computing and Cultural Heritage*, Vol. 10, No. 2, Article 11.
- Remondino, F., Gaiani, M., Apollonio, F., Ballabeni, A., Ballabeni, M., Morabito, D., 2016. 3D documentation of 40 kilometers of historical porticoes - the challenge. *ISPRS Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLI-B5, 711-718.
- Skarbez, R., Smith, M., Whitton, M.C., 2021. Revisiting Milgram and Kishino's Reality-Virtuality Continuum. *Front. Virtual Real.*, 2:647997.
- Teruggi, S., Grilli, E., Russo, M., Fassi, F., Remondino, F., 2020. A hierarchical machine learning approach for multi-level and multi-resolution 3D point cloud classification. *MDPI Remote Sensing*, 12(16), 2598.
- Teruggi, S., Fassi, F., 2021. 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.
- Unity3D, 2021. <https://unity.com/>. Last access: Jun 29th, 2021.
- Voinea G.D., Gîrbacia F., Postelnicu C.C., Marto A., 2019. Exploring Cultural Heritage Using Augmented Reality Through Google's Project Tango and ARCore. In *VR Technologies in Cultural Heritage*, 93(106), Springer International Publishing.
- Wallgrün, J.O., Huang, J., Zhao, J., Ebert, C., Roddy, P., Awe, J., Murtha, T., Klippel, A., 2018. Immersive Technologies and Experiences for Archaeological Site Exploration and Analysis. *Lecture Notes in Geoinformation and Cartography*, 198809, pp. 307-314.