Evaluation of Point Cloud Generation from 360° Videos of Indoor Spaces and Heritage Sites

Richard Honti[1], Ján Erdélyi[1], Tomáš Funtík[2]

[1] Department of Surveying, Faculty of Civil Engineering, Slovak University of Technology in Bratislava, Radlinského 11, 810 05 Bratislava, Slovakia – richard.honti@stuba.sk, jan.erdelyi@stuba.sk

[2] Department of Building Technology, Faculty of Civil Engineering, Slovak University of Technology in Bratislava, Radlinského 11, 810 05 Bratislava, Slovakia – tomas.funtik@stuba.sk

Keywords: Indoor Mapping; Point Cloud Generation; 360° Videogrammetry; Low-Cost 3D Capture; TLS; CupixVista.

Abstract

Affordable 360° cameras combined with cloud- or desktop-based photogrammetry have made image-based documentation of complex interiors widely accessible. However, the metric reliability of point clouds reconstructed from 360° video remains inconsistently reported, particularly when compared to survey-grade terrestrial laser scanning (TLS). This study evaluates point-cloud generation from 360° videos captured along guided walking paths using consumer cameras, with high-overlap frame extraction and Structure-from-Motion processing using the CupixVista solution.

Comparisons are performed using cloud-to-cloud (C2C) distance analysis, control-point analysis (CPA), and visual inspection of noise, surface roughness, and edge definition, with efficiency also considered. Point clouds are registered to the TLS reference using a two-step rigid alignment (manual coarse registration followed by ICP (Iterative Closest Point)). C2C distances, point density, and CPA errors are computed in CloudCompare after spatial subsampling. Control points are defined as intersections of locally fitted planes to improve precision.

Several indoor test scenes with varying geometry, scale, and visual characteristics are analyzed. Results show average deviations around 50 mm for C2C and 35 mm for CPA, with visual inspection confirming higher noise levels and deformation of sharp features in 360°-derived point clouds compared to TLS.

1. Introduction

High-quality 3D documentation of indoor environments and heritage assets is essential for conservation planning, condition assessment, deformation monitoring, scan-to-BIM workflows, and long-term digital archiving. For the last decade, survey-grade terrestrial laser scanning has been the dominant "as-built" capture technology because it provides dense geometry with well-characterized accuracy and repeatability. However, practical constraints such as equipment cost, field logistics, line-of-sight requirements and acquisition time often limit the deployment of TLS in dense interiors, confined spaces or time-critical heritage contexts.

In addition to TLS, mobile laser scanning (MLS) has become increasingly popular as a modern alternative (Conti et al., 2024). MLS systems, operating on the SLAM (Simultaneous Localization and Mapping) principle, enable scanning while in motion rather than from static positions. This approach significantly reduces the time required for data acquisition and allows more flexible surveying in complex or cluttered environments. However, similar to TLS, the main limitation of MLS remains its high cost, which can restrict accessibility for smaller-scale projects or institutions with limited budgets.

In parallel, consumer 360° cameras have become widely available and are operationally attractive: a single station captures the entire surrounding scene, reducing the need for careful camera aiming and substantially accelerating fieldwork in complex interiors. The photogrammetric processing of spherical imagery (typically via equirectangular projection) uses the same core pipeline as conventional close-range photogrammetry, involving feature extraction and matching, bundle adjustment / Structure-from-Motion (SfM), and Multi-View Stereo (MVS) densification, while requiring a camera model that respects the spherical observation geometry. Previous studies have described the orientation of spherical images as a geometry that can be expressed through angular measurements (i.e. mapping pixels to rays) and solved using bundle adjustment analogies similar to those used in geodetic network adjustment.

Despite the rapid uptake of 360° capture for indoor and heritage reality capture, the reliability of metrics remains inconsistent across reported studies and processing solutions, particularly when 360° imagery is acquired as video along a walking trajectory (with frames extracted at high overlap) and processed through increasingly "black box" cloud pipelines. The key technical reasons are well known to practitioners: (i) limited effective resolution per direction because a fixed sensor resolution spans a full 360° x 180° field of view; (ii) stitching artefacts and non-ideal multi-lens geometry; (iii) motion blur, rolling-shutter effects, compression, and exposure variations in video; and (iv) weak photogrammetric geometry in corridors or loop-free trajectories, which can exacerbate drift and local deformation. These factors make independent benchmarking against TLS reference data essential before 360° video-based point clouds can be trusted for tasks other than visualization.

1.1 Literature review

Cao et al. (2022) investigated the feasibility of generating point clouds from 360° video sequences in urban and heritage environments. Their workflow employed standard Structure-from-Motion and Multi-View Stereo pipelines, combined with a deep-learning classifier for semantic segmentation of the resulting dense point cloud. The study demonstrated that, with proper calibration and video frame sampling, the generated point clouds achieved an RMSE of 15–28 mm when compared to a TLS reference model captured with a Leica ScanStation.

Although TLS provided sub-centimetre precision, the 360° video-based reconstructions maintained sufficient geometric fidelity for heritage visualization and surface documentation. The classification accuracy reached 96% across structural categories (e.g., walls, ground, roofs), indicating strong potential for efficient and low-cost heritage mapping. However, the authors noted limitations in capturing fine architectural details due to the resolution constraints of consumer-grade 360° cameras (Cao et al., 2022).

Sharma et al. (2024) developed an automated 3D reconstruction method for panoramic imagery, focusing on algorithmic optimization for panoramic data alignment and depth consistency. Their study benchmarked the image-based approach against TLS reference data using control points collected in both indoor and outdoor environments. The RMSE values between the panoramic photogrammetric point clouds and TLS data ranged from 8 mm to 16 mm, depending on lighting conditions and surface texture. While TLS consistently achieved a mean absolute error below 5 mm, the photogrammetric results were considered adequate for architectural modelling and cultural documentation. The authors concluded that panoramic photogrammetry offers a practical balance between precision and efficiency, especially in contexts where access restrictions or budget constraints limit TLS use (Sharma et al., 2024).

A comprehensive investigation by Pepe et al. (2022) extended this comparison to heritage site applications, evaluating spherical videogrammetry and SLAM systems against TLS reference data. Using a Leica ScanStation 2 TLS scanner as the baseline, the researchers assessed point clouds generated from both spherical video and SLAM workflows. The RMSE for spherical videogrammetry ranged from 12 mm to 22 mm, depending on feature density and camera motion speed, while SLAM-based models achieved higher precision, with RMSE values between 7 mm and 10 mm. Although TLS remained the most accurate technique, the study found that spherical videogrammetry provides a low-cost, time-efficient alternative, capable of producing geometrically consistent models suitable for visualization and conservation planning. The total acquisition and processing time for spherical videogrammetry was nearly 70% shorter than that of TLS (Pepe et al., 2022).

Cui et al. (2021) introduced a deep neural framework for 3D point cloud generation from 360° images. The study proposed a method that integrates spherical projection and convolutional depth estimation to improve depth inference accuracy in panoramic imagery. The model outperformed conventional SfM-MVS baselines, reducing RMSE in depth estimation by 27.4% and improving reconstruction completeness by 22.6%. When compared with TLS ground-truth scans of indoor scenes, the approach achieved an average RMSE of 10 mm, effectively narrowing the accuracy gap between photogrammetry and TLS. Moreover, inference time per frame was significantly reduced, making the method well-suited for real-time 3D reconstruction using consumer-grade 360° cameras (Cui et al., 2025).

The study by Skarlatos et al. (2024) compared low-cost photogrammetric methods-including handheld and structured frame camera setups and a 360° multi-lens camera-against a high-precision TLS reference (Faro Focus 3D S70) for documenting confined underground spaces at the Tombs of the Kings, Cyprus. The 360° camera produced high-density point clouds but with greater noise, higher reprojection error (~2.3 pixels), and more variable cloud-to-cloud (C2C) distances from the TLS model, reflecting reduced geometric precision due to lens stitching and sensor limitations. In contrast, traditional frame-based photogrammetry achieved lower reprojection errors (0.7–1.0 pixels), smoother surfaces, and smaller C2C deviations, yielding better alignment with the TLS reference. While the 360° method offered faster and simpler acquisition-ideal for narrow or hard-to-access environments-it sacrificed some accuracy, suggesting that a hybrid workflow combining 360° and traditional imagery could balance efficiency with geometric reliability for heritage documentation (Skarlatos et al., 2024).

The study by Perfetti et al. (2024) introduced and tested Ant3D, a fisheye multi-camera photogrammetric system designed for 3D surveying of narrow or confined spaces and compared its outputs against TLS reference data. The best Ant3D reconstruction achieved a cloud-to-cloud deviation of approximately 40 mm from the reference after best-fit alignment, demonstrating sufficient accuracy for 1:100 architectural documentation. The system also exhibited highly stable drift error control, averaging 5 mm per meter in heritage corridors, 9 mm per meter in a 2 km mining tunnel, and 2 mm per meter in a 3.5 km mountain trail. In comparison, a similar range-based mobile mapping system showed drift errors near 20 mm per meter (Perfetti et al., 2024).

For complex heritage documentation with 360° cameras at the Santa Marta Belltower, the study by Teppati Losè et al. reports accuracy primarily through 18 independently measured distances used as metrical checks. The reconstructed model produced an average absolute distance error of 0.007 m with a standard deviation of 0.008 m, and a maximum deviation of 0.040 m; when expressed as relative error, the mean was 0.001 with a maximum of 0.004 (Losè et al., 2021).

Barazzetti et al. showed that 5K 360° video can enable a largely automated SfM reconstruction workflow when approximate exterior orientation (EO) is provided. For large frame sets extracted from continuous video, unguided ("blind") processing often leads to unreliable or incomplete orientations, whereas EO estimates stabilize alignment by constraining frame-to-frame visibility and reducing the feature-matching search space. The authors therefore argue that EO-aided processing is a practical prerequisite for scaling 360° video photogrammetry to extended trajectories and complex scenes, improving robustness and efficiency while retaining low-cost, rapid acquisition (Barazzetti et al., 2022). Further articles that evaluate the use of low-cost solutions and sensors for object digitization can be found (Alessandri et al., 2019; Gottardi, et al., 2018; Herban et al., 2022; Mandelli et al., 2017; Pérez-García et al., 2024; Honti et al., 2026).

Accordingly, this article focuses on evaluating point clouds generated from 360° videos in indoor spaces and exterior sites, using TLS as the reference baseline. In addition to general distance statistics, the study considers practical quality dimensions that are important for architectural and heritage documentation, such as noise behaviour on planar surfaces, edge and corner preservation, and local roughness around high-curvature elements, while also taking operational efficiency into account.

2. Test environments and the evaluated systems for data capturing

For the evaluation, three test scenarios were selected, each differing in spatial dimensions, geometric complexity, structural layout, object distribution, and visual characteristics such as colour and lighting. These variations were chosen to represent a broad spectrum of real-world conditions typically encountered in building digitization projects. The selected sites include an interior of a large room, an interior space within a heritage building, and an exterior facility.

The first test scene (Dataset 1) corresponds to an elongated room characterized by a repetitive geometric structure, serving as the Laboratory of Surveying (Figure 1.) at a university. The laboratory measures approximately 29.5 m in length, 9.8 m in width, and 3.6 m in height.

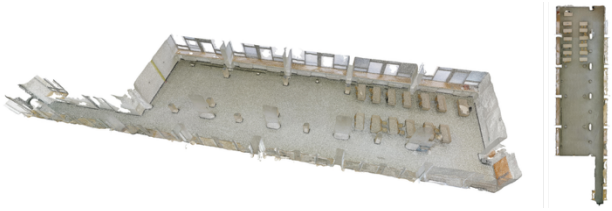

Figure 1. Point cloud generated from 360° videos for the first test scene – Laboratory of Surveying.

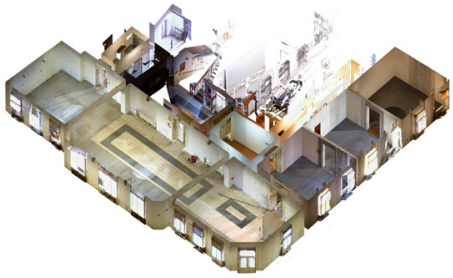

Figure 2. TLS point cloud for the second test scene – interior part of a historical building.

The second test scene (Dataset 2) represents an interior space of a historical building (Figure 2.). This environment is a part of a complex heritage site comprising several interconnected rooms, narrow corridors, and vaulted ceilings with historical decorative elements. The third test scene (Dataset 3) depicts the exterior of a rectangular storage facility (Figure 3.), with dimensions of approximately 34.5 m in length, 13.0 m in width, and 3.5–5.5 m in height.

Figure 3. TLS point cloud of the third test scene – storage facility.

2.1 Reference data capturing

To establish a reliable reference for evaluation, each of the three test scenes was documented using a Trimble X9 terrestrial laser scanner (Figure 4. - left). The Trimble X9 supports a maximum range of 150 meters and a scanning speed of up to 2 million points per second. According to the manufacturer's specifications, the system achieves a range noise of less than 1 mm at 10 meters, a range measurement accuracy of 2 mm, and an angular accuracy better than 16 arcseconds, making it highly suitable as a ground-truth reference for evaluating the geometric accuracy of 3D point cloud data. The scanner integrates three calibrated 10 MP cameras, capable of capturing up to 316 MP of panoramic imagery per setup. It also features survey-grade self-levelling within a ±10° compensation range and includes automated point cloud registration using the ICP algorithm, ensuring precise alignment of multiple scans.
Scanning was performed with a resolution setting of 15 mm at a distance of 10 meters from several stations, providing a high level of geometric detail suitable for precision analysis. The final scans from multiple stations were merged into a single homogeneous dataset using the ICP algorithm, with the transformation error remaining below 2 mm for each dataset.

Figure 4. Sensors used for data capturing: TLS Trimble X9 (left), 360° camera Ricoh Theta X (right).

2.2 CupixVista – the evaluated system

For the evaluation, the CupixVista system was selected, which is a mobile 3D capture and mapping solution based on 360° panoramic imaging and photogrammetric reconstruction principles. The system utilizes a high-resolution 360° camera, mounted on a mobile pole or tripod, to capture continuous panoramic imagery while the operator moves through the environment. These panoramic images are processed through the Cupix cloud-based platform, where Structure-from-Motion and Multi-View Stereo algorithms are applied to reconstruct a dense 3D point cloud. The workflow is designed to generate geometrically coherent and colorized spatial data suitable for building documentation, indoor mapping, and facility management tasks, while requiring minimal expertise or specialized hardware.
CupixVista relies on spherical image data and precise camera path estimation, which are automatically derived during processing to ensure accurate geometric reconstruction. The system supports the integration of GPS/IMU data, if available, to enhance absolute positioning accuracy. However, even without external georeferencing, CupixVista can produce internally consistent 3D models by optimizing the relative alignment of consecutive 360° frames. The resulting point clouds are typically colorized using the imagery captured by the 360° camera.
The principal advantage of the CupixVista workflow lies in its speed, portability, and cost-effectiveness. Unlike TLS, which requires multiple static setups and time-intensive registration, CupixVista enables continuous data acquisition during movement, substantially reducing field time (Table 1). However, the accuracy and precision of the generated point clouds are influenced by factors such as image resolution, lighting conditions, camera trajectory stability, and surface texture variation and also the capabilities of the sensor used for capturing.

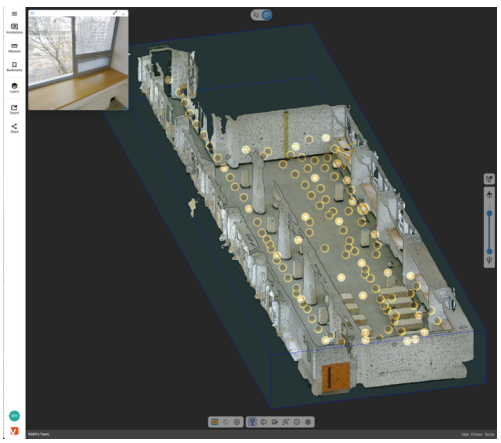

Figure 5. The CupixVista online 3D Map viewer.

From a data processing perspective, a notable limitation of the CupixVista system is the so-called "black-box effect." After capturing the environment in the form of a 360° video, the data are uploaded to the Cupix cloud platform, where processing

occurs automatically without user intervention or control. The video is divided into individual 360° frames, which are then used to create a photogrammetric model. The result is an automatically generated 3D model and point cloud, which can be viewed online through the Cupix web viewer (Figure 5.). Additionally, users can export the resulting point cloud or mesh model for further analysis.

For image acquisition, the Ricoh Theta X 360° camera (Figure 4. - right) was utilized. This camera is equipped with dual 48 MP sensors and f/2.4 fisheye lenses, capable of capturing high-resolution spherical imagery (up to 11K) with a dynamic range optimized for indoor and outdoor lighting conditions. The device features built-in stabilization, fast image stitching, and automatic exposure balancing, ensuring consistency between consecutive panoramic frames.

The data capturing process followed the software manufacturer's official recommendations to achieve optimal reconstruction accuracy and texture quality. The operator moved slowly through the test scenes while maintaining a constant camera height and ensuring adequate overlap between successive 360° frames. The complete data capture for each test site required few minutes, allowing the measurement to be repeated several times to verify consistency. Among the repeated acquisitions, the dataset exhibiting the best image quality, overlap, and geometric stability was selected for subsequent processing and comparison.

Table 1. Comparison of data acquisition times.

	Dataset 1	Dataset 2	Dataset 3
TLS	45 min	70 min	30 min
360° camera	7,5 min	5 min	2,5 min

For this study, the CupixVista generated point clouds were compared with reference TLS datasets to assess geometric accuracy, structural completeness, and spatial consistency. The comparison focused on 3 methodologies described in the next section,

3. Methodology

To enable a meaningful comparison between the two datasets, the point clouds generated from 360° imagery were transformed into the coordinate system of the reference TLS-derived point cloud. This transformation followed a two-step registration process. First, a coarse alignment was performed using a set of manually identified corresponding points visible in both datasets. This was followed by a fine registration stage employing the ICP algorithm to refine the alignment. During registration, only rigid-body transformations (rotations and translations) were applied, ensuring that the geometry and scale of the point clouds remained unchanged.

Three evaluation methods were used in this study:

- Cloud-to-Cloud (C2C) distance analysis,
- Control Point Approach (CPA),
- Visual inspection of noise characteristics.

The C2C analysis, performed in CloudCompare (v2.13.1) (CloudCompare, 2026), quantified the spatial deviations between the 360° photogrammetric and TLS-derived point clouds by calculating the nearest-point distances from the 360° data to the TLS reference surface. To maintain computational efficiency and consistency, both datasets were spatially subsampled, with the TLS cloud reduced to 5 mm point spacing and the 360° imagery cloud to 10 mm spacing.

For each evaluation, the mean distance (MEAN), standard deviation (STD), and Root Mean Square Deviation (RMSD) were computed to quantify geometric differences. The RMSD was calculated as:

$$RMSD = \sqrt{\frac{1}{n}\sum_{i=1}^{n} d_i^2} \quad (1)$$

where n represents the number of evaluated points in the point cloud captured with 360° camera, and d_i is the smallest Euclidean distance between the i-th point and the locally interpolated TLS surface. This surface was approximated by a Triangulated Irregular Network (TIN) constructed from the 10 nearest neighboring TLS points within a spherical radius.

Additionally, the average surface point density (PD) was calculated using the Cloud Density function in CloudCompare, which counts neighboring points within a 0.1 m radius sphere around each point. To further interpret the geometric resolution, the average point spacing (PS), representing the typical distance between adjacent points was computed as the square root of the reciprocal of PD. These metrics reflect each system's capacity to capture fine surface details.

The Control Point Approach (CPA) was also applied, where distinctive geometric features such as corners, column edges, and openings were selected as control points and manually identified in both datasets. The Euclidean distances between corresponding control points were measured to assess structural consistency and positional accuracy. To enhance CPA precision, control points were defined as the intersections of three locally fitted planes, obtained through regression analysis of nearby point neighbourhoods, providing a robust geometric basis for point localization.

Finally, a qualitative visual inspection was conducted to assess noise and surface irregularities within the 360° camera captured dataset. Special attention was given to planar surfaces and sharp edges, where reconstruction artifacts and noise are more pronounced.

4. Results and discussion

This section presents and interprets the results obtained from the comparison of 360° imagery-based point clouds with the reference TLS data. The analysis focuses on geometric accuracy, surface deviations, and point cloud characteristics across all test scenes, followed by a discussion of the observed differences and their implications for practical 3D documentation workflows.

4.1 Cloud-to-Cloud analysis

In this section, the results of the evaluation are presented. The first metric used for comparison is the Cloud-to-Cloud distance, with the results illustrated in Figure 6. - Figure 8. for each test scene. In Figure 6. and Figure 7., the ceilings of the rooms were removed during visualization to provide a clearer view of the interior spaces. The points are color-coded based on the C2C distances, with a color scale ranging from 0 to 300 mm, allowing for a visual interpretation of spatial deviations between the 360°-based and TLS-derived point clouds.

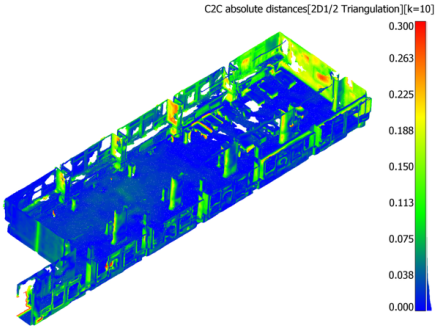

Figure 6. C2C analysis for the Dataset 1.

As can be seen from the figures, and supported by the parameters presented in Table 2, the best results were achieved in the first test scene. This can be attributed to the fact that the scene represents a geometrically complex but enclosed room with a repetitive structural pattern, which enhances image matching stability and overall reconstruction accuracy.

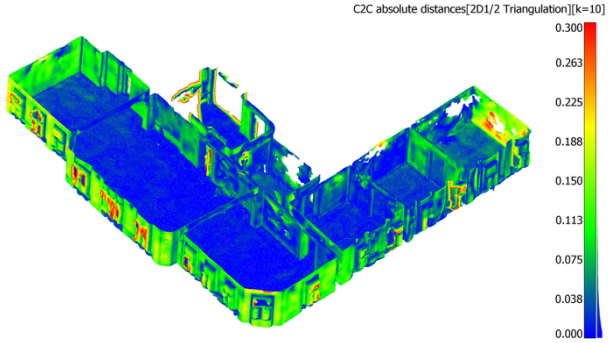

Figure 7. C2C analysis for the Dataset 2.

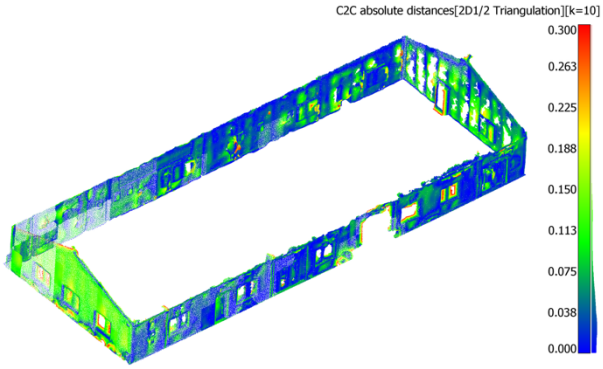

Figure 8. C2C analysis for the Dataset 3.

Table 2. Results from the C2C analysis

	MEAN [m]	STD [m]	RMSD [m]	PS [m]
Dataset 1	0.042	0.050	0.065	0.027
Dataset 2	0.056	0.075	0.098	0.027
Dataset 3	0.057	0.080	0.092	0.029

Table 2 presents the average C2C distance metrics for all three test scenes. The second column lists the mean distances, the third shows the standard deviations, the fourth provides the RMSD values, and the fifth reports the calculated point spacing.

The best point cloud metrics were obtained for the first dataset, while the largest variations were observed in the standard deviation values for datasets 2 and 3. This indicates greater variability in those datasets and suggests a higher level of noise in the reconstructed point clouds. The surface density and the corresponding average point spacing exhibit very similar values across all datasets, ranging from 27 to 29 mm. The mean C2C distance for all test scenes is around 50 mm, which, considering the diversity and complexity of the environments tested, as well as the sensor's price and overall system complexity, can be regarded as a relatively good result.

The graph on Figure 9. shows the percentage of points within specific C2C distance intervals, indicating the geometric accuracy of each dataset. The laboratory scene achieved the best results, with 72.4% of points showing deviations below 50 mm, followed by the Heritage interior with 60.0% and the Storage facility with 56.1%. In the 50–100 mm range, the shares increase to 14.9%, 21.5%, and 23.9%, respectively. Larger deviations occur less frequently, with 6.6–11.6% of points between 100–150 mm, and less than 5% beyond 150 mm for all scenes. Only a very small fraction of points exceeded 200 mm. These results once again confirm that the best outcomes were achieved for Dataset 1 – the Laboratory scene, which produced the most accurate and consistent reconstruction.

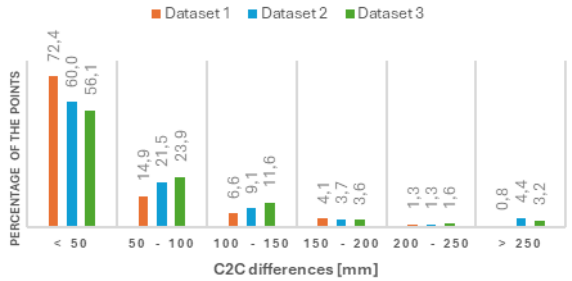

Figure 9. Distribution of C2C differences for all test scenes.

In addition to the C2C analysis, surface roughness was calculated for selected parts of each dataset to further assess the quality of the reconstructed point clouds. Roughness quantifies the local variability of surface geometry, reflecting both the texture of the scanned surfaces and the presence of noise or reconstruction irregularities. Higher roughness values typically indicate increased noise levels or less accurate surface fitting, while lower values correspond to smoother and more geometrically stable surfaces. A comparison of surface roughness on a selected wall area is presented in Figure 10. illustrating differences in surface quality between the datasets.

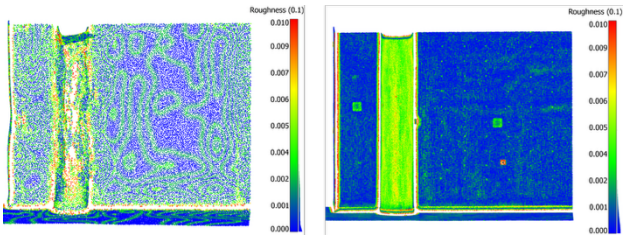

Figure 10. Surface roughness comparison of the point cloud generated from 360° video (left) and from TLS (right).

4.2 Control Point Approach Analysis

The Control Point Approach (CPA) analysis began with the selection of 20 pairs of control points distributed across the test objects in all three test scenes. Each control point was modelled using locally fitted planes to ensure precise identification and positional consistency. The modelling process was based on regression analysis, where three mutually perpendicular planes were computed from neighbouring points, and the control point was defined as the intersection of these planes. Subsequently, Euclidean distances between corresponding control points were measured in both the evaluated and reference (TLS) point clouds, yielding values ranging from 0.152 m to 38.880 m.

Table 3. Results from the CPA analysis.

	MEAN [m]	STD [m]	RMSD [m]
Dataset 1	0.028	0.044	0.050
Dataset 2	0.045	0.057	0.071
Dataset 3	0.044	0.047	0.063

Table 3 summarizes the results of the CPA analysis, showing the mean, standard deviation, and Root Mean Square Deviation of Euclidean distances between corresponding control points for all three test scenes. The Laboratory dataset achieved the lowest deviations (mean 0.028 m, STD 0.044 m, RMSD 0.050 m), confirming its high geometric accuracy. The Heritage interior showed the largest deviations (RMSD 0.071 m), likely due to its complex geometry and lighting conditions, while the Storage facility yielded intermediate results (RMSD 0.063 m). Overall, the findings are consistent with the C2C analysis, confirming that the Laboratory scene (Dataset 1) provided the most accurate reconstruction.

4.3 Visual evaluation of the noise level

This final verification step focuses on the visual evaluation of the 360° camera generated data, assessing geometric deformations, noise levels, and the clarity and recognizability of individual features and structures within the scanned scenes.

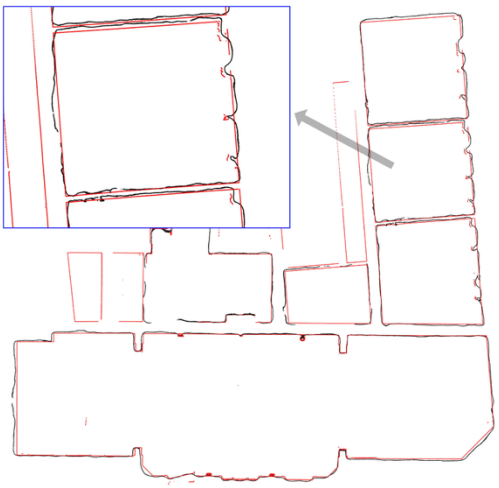

Figure 11. Horizontal section of the point clouds for the Dataset 2 captured with 360° camera (black), and TLS (red).

The Figure 11. - Figure 13. shows a plan view (horizontal section) illustrating the geometric correspondence between the two datasets. The black contour represents the section derived from the 360° imagery-based point cloud, while the red contour corresponds to the TLS reference model. As can be observed, the overall alignment between both datasets is consistent, with deviations predominantly in the centimeter range, as confirmed by the results of the C2C and CPA analyses.

However, a more detailed inspection reveals noticeable differences in edge definition and sharpness. The contours derived from the evaluated point cloud display blurred and deformed boundaries, where edges appear stretched or rounded rather than sharply defined. This effect is particularly visible along wall intersections and openings. Such irregularities are primarily attributed to the photogrammetric reconstruction process applied to frames extracted from 360° video data, where image stitching errors, limited pixel resolution, and optical distortions can introduce minor geometric inconsistencies.

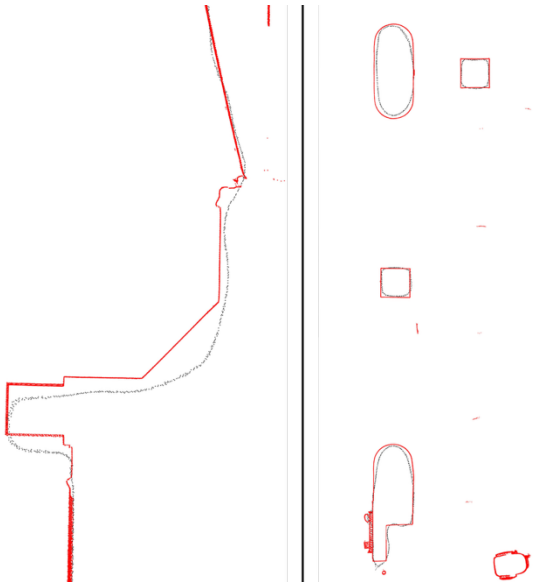

Figure 12. Comparison of horizontal section line details extracted from the point clouds generated using the 360° camera (black) and the reference TLS dataset (red).

Figure 14. illustrates the capture of a decorative wall detail in Dataset 2 (heritage site). The comparison confirms that fine architectural details are difficult to clearly identify in point clouds generated from 360° video.

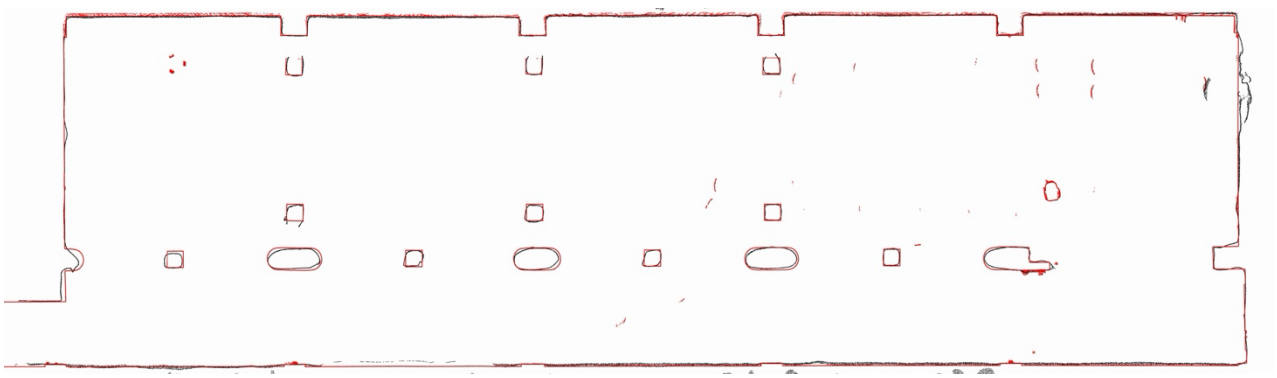

Figure 13. Horizontal section of the point cloud for the Dataset 1: from the 360° camera (black), TLS (red)

Figure 14. Example of capturing a historical decorative wall element in Dataset 2 using a 360° camera (left) and TLS (right)

5. Conclusions

This study evaluated the metric reliability and practical usability of point clouds generated from guided-walk 360° video acquisitions processed with the CupixVista system, using survey-grade TLS as a geometric reference. Three representative environments: an enclosed laboratory interior, a complex heritage interior, and an exterior storage facility were analysed through cloud-to-cloud (C2C) distance metrics, a feature-based Control Point Approach (CPA), and qualitative inspection of noise, surface roughness, and edge definition. The results consistently show that 360° video–based photogrammetry enables extremely fast and operationally simple data acquisition, reducing field time by approximately one order of magnitude compared to TLS, while still providing geometrically coherent and complete 3D representations of indoor and near-building-scale environments.

From a quantitative perspective, the evaluated datasets exhibited mean C2C deviations on the order of 42–57 mm and RMSD values between 65 and 98 mm, with CPA results showing mean discrepancies of 28–45 mm and RMSD values of 50–71 mm. The best performance was achieved in the laboratory scene, characterized by repetitive geometry and favourable imaging conditions, while the heritage interior produced the largest deviations due to its complex layout, variable lighting, and more challenging photogrammetric geometry. Across all scenes, point spacing values of approximately 27–29 mm indicate a relatively uniform geometric sampling density, confirming that the observed accuracy differences are primarily attributable to reconstruction quality rather than point density alone. Visual analysis further revealed that, although global alignment and overall geometry are well preserved, point clouds derived from 360° video display higher surface noise, increased roughness on planar elements, and noticeable deformation or blurring of sharp edges and corners when compared to TLS data.

Taken together, these findings indicate that 360° video–based point cloud generation represents a viable low-cost and time-efficient alternative for applications where centimetre-level accuracy is sufficient, such as rapid indoor mapping, facility documentation, visualization, preliminary scan-to-BIM workflows, and general heritage recording. However, the method remains unsuitable for tasks that demand millimetre-level precision, sharp edge delineation, or high-fidelity representation of fine architectural details, where TLS continues to provide unmatched accuracy, repeatability, and geometric clarity. The largely automated and “black-box” nature of current cloud-based processing solutions further limits user control over reconstruction parameters, making independent benchmarking against reference data essential for informed deployment.

Acknowledgements

Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V05-00005 and under the project No. 09I02-03-V01-00029.

References

Alessandri, L., Baiocchi, V., Del Pizzo, S., Rolfo, M.F., Troisi, S., 2019. P Photogrammetric Survey with Fisheye Lens for the Characterization of the La Sassa Cave. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-2-W9, 25–32. https://doi.org/10.5194/isprs-archives-XLII-2-W9-25-2019

Barazzetti, L., Previtali, M., Roncoroni, F., 2022. 3D Modelling With 5K 360 Videos. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVI-2-W1-2022, 65–71. https://doi.org/10.5194/isprs-archives-XLVI-2-W1-2022-65-2022

Cao, Y., Previtali, M., Barazzetti, L., Scaioni, M., 2022. Integration of Point Clouds from 360° Videos and Deep Learning Techniques for Rapid Documentation and Classification in Historical City Centers, in: Mazzeo, P.L., Frontoni, E., Sclaroff,

S., Distante, C. (Eds.), Image Analysis and Processing. ICIAP 2022 Workshops. Springer International Publishing, Cham, pp. 254–265. https://doi.org/10.1007/978-3-031-13324-4_22

CloudCompare, 2026. CloudCompare (version 2.13.1) [WWW Document]. CloudCompare (version 2.13.1). URL https://www.cloudcompare.org/ (accessed 1.15.26).

Conti, A., Pagliaricci, G., Bonora, V., Tucci, G., 2024. A Comparison Between Terrestrial Laser Scanning and Hand-Held Mobile Mapping for the Documentation of Built Heritage. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVIII-2-W4-2024, 141–147. https://doi.org/10.5194/isprs-archives-XLVIII-2-W4-2024-141-2024

Cui, Y., Chang, Q., Liu, Q., Yang, X., Huang, Y., Chen, S., Ren, F., Stricker, D., 2025. 3D Reconstruction With Spherical Cameras. IEEE Access 13, 120852–120868. https://doi.org/10.1109/ACCESS.2021.3119367

Gottardi, C., & Guerra, F. 2018. Spherical images for cultural heritage: Survey and documentation with the Nikon KM360. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-2, 385–390. https://doi.org/10.5194/isprs-archives-XLII-2-385-2018

Herban, S., Costantino, D., Alfio, V.S., Pepe, M., 2022. Use of Low-Cost Spherical Cameras for the Digitisation of Cultural Heritage Structures into 3D Point Clouds. Journal of Imaging 8. https://doi.org/10.3390/jimaging8010013

Honti, R., Erdélyi, J., Funtík, T. 2026. Evaluation of Apple's LiDAR Sensor for Rapid Indoor 3D Capture. In: Contributions to International Conferences on Engineering Surveying. INGEO 2025. Springer Proceedings in Earth and Environmental Sciences. Springer, Cham. https://doi.org/10.1007/978-3-032-12070-0_5

Losè, L.T., Chiabrando, F., Tonolo, F.G., 2021. Documentation of Complex Environments Using 360° Cameras. The Santa Marta Belltower in Montanaro. Remote Sensing 13. https://doi.org/10.3390/rs13183633

Mandelli, A., Fassi, F., Perfetti, L., & Polari, C. 2017. Testing different survey techniques to model architectonic narrow spaces. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-2/W5, 505–511. https://doi.org/10.5194/isprs-archives-XLII-2-W5-505-2017

Pepe, M., Alfio, V.S., Costantino, D., Herban, S., Pepe, M., Alfio, V.S., Costantino, D., Herban, S., 2022. Rapid and Accurate Production of 3D Point Cloud via Latest-Generation Sensors in the Field of Cultural Heritage: A Comparison between SLAM and Spherical Videogrammetry. Heritage 5, 1910–1928. https://doi.org/10.3390/heritage5030099

Pérez-García, J.L., Gómez-López, J.M., Mozas-Calvache, A.T., Delgado-García, J., 2024. Analysis of the Photogrammetric Use of 360-Degree Cameras in Complex Heritage-Related Scenes: Case of the Necropolis of Qubbet el-Hawa (Aswan Egypt). Sensors 24. https://doi.org/10.3390/s24072268

Perfetti, L., Fassi, F., Vassena, G., 2024. Ant3D—A Fisheye Multi-Camera System to Survey Narrow Spaces. Sensors 24. https://doi.org/10.3390/s24134177

Sharma, S.K., Jain, K., Shukla, A.K., 2024. 3D point cloud reconstruction using panoramic images. Appl Geomat 16, 575–592. https://doi.org/10.1007/s12518-024-00563-w

Skarlatos, D., Cuca, B., Kafataris, G., Previtali, M., Agapiou, A., 2024. Low-cost photogrammetry solutions for surveying confined underground spaces: testing the traditional set-up against 360° camera on Tombs of the Kings archaeological site. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVIII-2-W8-2024, 425–430. https://doi.org/10.5194/isprs-archives-XLVIII-2-W8-2024-425-2024