

PHOTOGRAMMETRIC SURVEY OF NARROW SPACES IN CULTURAL HERITAGE: COMPARISON OF TWO MULTI-CAMERA APPROACHES

N. Bruno¹, L. Perfetti^{2*}, F. Fassi³, R. Roncella¹

¹ Università degli Studi di Parma, dept. of Engineering and Architecture (DIA), Italy – (nazarena.bruno, riccardo.roncella)@unipr.it

² Università degli Studi di Brescia, dept. of Civil Engineering, Architecture, Territory, Environment and Mathematics (DICATAM), Italy – luca.perfetti@unibs.it

³ Politecnico di Milano, dept. of Architecture, Built environment and Construction engineering (ABC), Italy – francesco.fassi@polimi.it

Commission II

KEY WORDS: Multi-camera, Spherical camera, 360° camera, Ant3D, Insta 360 Pro2, Photogrammetry, Cultural heritage

ABSTRACT:

Multi-camera devices are increasingly popular in various metrological applications, including cultural heritage digitalisation, where these devices are adopted as low-cost alternatives to more traditional methods or mobile mapping systems. They can be of two types: panoramic and non-panoramic configurations, with the former usually more compact and ready-made off-the-shelves and the latter usually custom-developed for metrological applications. In the paper, we compare the accuracy and reliability performance of two types of multi-camera: the spherical camera INSTA 360 Pro2 and the custom multi-camera rig Ant3D. The case study is a challenging spiral staircase environment, typical in many cultural heritage survey projects. The processed image datasets were evaluated in the most common constrain scenario (GCPs at both ends of the staircase) and the worst-case scenario (open-ended path, GCPs at the start). The datasets were processed with precalibrated IO and various degrees of multi-camera constraints up to precalibrated relative orientations. The results highlight that the nominal scale 1:50 can be achieved, e.g. an accuracy of <2 cm plus complete and precise point clouds and mesh results.

1. INTRODUCTION

Close-range photogrammetry is a well-established technique in digitization of complex architectural spaces. Technological improvements in this field are constantly increasing, making new instruments and processing methodologies available.

Recently, various multi-camera devices, including off-the-shelves compact spherical cameras, have become increasingly popular in metrological applications due to their ease of use, portability, cost-effectiveness, and productivity. Overall, these devices reduce the complexity of the field operation in imaging extensive environments, allowing faster acquisitions in both outdoor and indoor environments. Many academic papers testify to the growing interest and effectiveness in several applications, especially when narrow spaces are to be surveyed (Barazzetti et al., 2018; Meyer et al. 2020; Panella et al., 2020).

Such systems are generally based on rigid multi-camera rigs with fixed baselines and relative orientations equipped with fisheye lenses. They can be grouped into two types: (i) panoramic configurations (Teo, 2015; Barazzetti et al., 2017; Fangi et al., 2018; Teppati Losè et al., 2021) and (ii) non-panoramic configurations (Nocerino et al., 2018; Torresani et al., 2021; Perfetti et al., 2022). In the former group follows practically all off-the-shelves devices. These are conceived to output, directly or after a post-processing stage, a 360° spherical image (e.g., in equirectangular projection), to be used in visualization frameworks such as Google Streetview. To this aim, their design minimises the distances between the cameras' optical centers. In the latter group follows custom devices and prototypes specifically meant for metrological applications designed with a larger baseline between the cameras. The former cameras are generally ready to use and more compact, while the latter have the advantage of allowing inter-rig image triangulation and scale reconstruction based on the rig's known baselines.

1.1 Paper objective

The paper compares two devices: the spherical camera INSTA 360 Pro2, and the non-panoramic multi-camera rig Ant3D (patent No. 10202100000812). The former is a commercial product marketed mainly at creative video production, but its technical specifications also make it suitable for metric purposes. The latter is a prototype designed explicitly for photogrammetric acquisitions, especially in meandering and narrow spaces (Perfetti et al., 2022).

The comparison aims to assess the accuracy and reliability of the two systems, particularly their robustness to drift errors in long acquisitions with few ground constraints and their effectiveness for Cultural Heritage (CH) digitization.

In CH digitization, it is usual to face narrow and meandering spaces, such as paths in the wall thickness or spiral staircases, which are difficult to constrain with a reference topographic network and where the traditional close-range photogrammetry and terrestrial laser scanning techniques are complex to be applied. On the other hand, accurate representations at typical architectural scales (e.g., 1:50 or 1:100) are usually required for these spaces.

The test has been performed in the Milan Cathedral, surveying the Minguzzi spiral staircase, which represents a challenging case study due to its complexity and limited accessibility. It is a marble spiral staircase approximately 28 m high and 70 cm wide, almost entirely dark since the natural illumination is provided only by small, tapered windows that connect the staircase with the outside by penetrating a wall thickness of more than two meters. It can be considered an open-ended path suitable for evaluating error propagation along the acquisition. The same test field has been used in past evaluations to assess the performance of other approaches (Perfetti et al., 2017; Teruggi et al., 2022).

* Corresponding author

1.2 Related works

Spherical cameras are increasingly investigated for their metric accuracy and achievable level of detail, and the availability of even low-cost consumer-grade cameras has further raised the interest in these devices.

Generally, considering entry- and medium-level cameras, such devices are composed of two opposite-facing fisheye lenses that frame the entire 360° Field Of View (FOV). Higher-level cameras account for more sensors, such as the iSTAR Fusion 360 (4 cameras), the Insta360 Titan (8 cameras) or the Pano 360° (36 cameras). All the instruments typically provide ready-to-use panoramas obtained by stitching the individual images (Fangi et al., 2018).

The literature reports tests conducted in different environments and applications, including mainly CH heritage field (Barazzetti et al., 2017; Gottardi and Guerra, 2018; Teppati Losè et al., 2021; Herban et al., 2022; Gómez-López et al., 2023), urban context (Bruno and Roncella, 2019; Chiappini et al., 2020; Barazzetti et al., 2022; Cera and Campi, 2022; Martino et al., 2023), narrow spaces, such as the experiments presented in (Barazzetti et al., 2019), who used 360° imagery to connect the inside and outside central perspective images blocks by passing through doors, or the case study in (Teppati Losè et al., 2021), who surveyed a bell tower's spiral staircase, similar to the one presented in this work. Image blocks are generally constrained using well-distributed GCPs (Ground Control Points) or GNSS (Global Navigation Satellite System) camera position (Barazzetti et al., 2022), while the drift with few ground constraints seems not to be investigated. In most cases, the processing is performed directly on the automatically stitched panoramas provided by the camera dedicated software. On the contrary, (Barazzetti et al., 2017) investigated fisheye image calibration to improve the stitching and tested different stitching software; (Perfetti et al., 2018; Teppati Losè et al., 2021) directly processed the front-rear raw fisheye images acquired by the individual sensors of the camera. Images are acquired as single shots or extracted frames from video, resulting in the fastest and most continuous acquisition procedure. In this context (Gottardi and Guerra, 2018) and (Fangi et al., 2018) evaluated the influence of the number and location of images w.r.t. completeness and smoothing of the dense cloud, while (Teppati Losè et al., 2021) tested different frame extraction intervals to get images from video.

Processing generally involves only the 360° imagery, but (Gómez-López et al., 2023) and (Martino et al., 2023) investigated the integrated processing with other data, respectively, to produce reliable texture over terrestrial laser scanning data and to improve the orientation.

The metrics considered for the accuracy evaluation are commonly based on the evaluation of Check Points (CPs) residuals, tie points reprojection error and completeness and deviation of dense point clouds w.r.t. a reference dataset. Currently, considering the achievable accuracies, most of the analyzed literature shows that 360° photogrammetry is suitable for metric applications up until 1:100 or 1:200 nominal scale (Barazzetti et al., 2017; Gottardi and Guerra, 2018; Fangi et al., 2018; Barazzetti et al., 2019; Teppati Losè et al., 2021). Meshes and dense clouds are usually noisy and coarse, prevalently for the high reprojection errors related to inaccurate image stitching (Barazzetti et al., 2017) and/or low resolution of the farthest parts from the sensor. GCPs are necessary to orient the block correctly, and, especially with a large sequence of frames, approximated initial Exterior Orientation (EO) parameters are necessary to complete the image orientation step (Barazzetti et al., 2022).

Multi-camera systems designed with significant baselines between the cameras have the advantage of producing an already

scaled 3D reconstruction. This solution is increasingly popular and has been tested in many acquisition scenarios as an alternative to terrestrial laser scanning and mobile mapping systems. Initial iterations of low-cost multi-camera devices were assembled using multiple off-the-shelves action cameras: Koehl et al., 2016 assembled 4 GoPro cameras on a rigid bar, while Holdener et al., 2017 designed a multi-camera composed of five GitUp cameras. More recently, authors proposed custom devices assembled from more specialised hardware that provides precise synchronisation and global shutter sensors: Ortiz-Coder and Sánchez-Ríos, 2019 designed a handheld device housing two cameras meant for agile survey in the archaeological and CH field; Perfetti, 2020 designed a handheld device with 5 cameras meant for narrow spaces survey evaluating between different possible arrangement. The same system named Ant3D was later tested for accuracy in the CH (Perfetti and Fassi, 2022, Perfetti et al., 2023) and other fields. Torresani et al., 2021 designed a modular handheld stereo system named GuPho, implementing software features to guide the field capture and improve the reconstruction. A variation of said system named FROG was presented in Menna et al., 2023, for underwater acquisitions. Custom devices such as the aforementioned handheld multi-cameras generally mount fisheye lenses to take advantage of the wide FOV and acquire synchronized images at various frame rates or video sequences. Generally, no positioning sensors are used, and the image orientation is achieved through offline Structure from Motion (SfM) (Perfetti and Fassi, 2022) or visual SLAM (Simultaneous Localisation And Mapping) real-time processing that can be later refined with SfM (Ortiz-Coder and Sánchez-Ríos, 2019; Torresani et al., 2021). Camera calibration and relative orientation constraints are essential to achieve accurate results. Interior Orientation (IO) and Relative Orientation (RO) parameters are mandatory input data for most systems that use visual SLAM and must be calibrated beforehand. Nocerino et al., 2018, proposed a methodology to calibrate IO and RO of a multi-camera system, while Perfetti and Fassi, 2022, used self-calibration to compute IO parameters and baselines between the cameras.

The accuracy achievable with multi-camera devices is usually evaluated by comparison against other established survey methodologies. Ortiz-Coder and Sánchez-Ríos, 2019, compared the multi-camera performance against a classical close-range photogrammetric survey performed with a DSLR (Digital Single Lens Reflex) comparing the two approaches on CPs. Torresani et al., 2021, compared the dense point cloud derived from the multi-camera acquisition to reference terrestrial laser scanning by computing the signed Euclidian distance to the reference for each point. On the other hand, Perfetti and Fassi, 2022, evaluated the drift error that accumulates in long, unconstrained, open-ended acquisition path by performing a 7 parameters transformation on targets at the beginning of the path and checking the drift error on CPs along the surveyed trajectory.

2. MATERIALS AND METHODS

2.1 INSTA 360 Pro2

The INSTA 360 Pro2 is a professional 360° camera. It has a spherical shape with a diameter of 143 mm and comprises 6 equidistant cameras around its equator, rotated relative to each other by 60° (Figure 1 - left). Each camera has a sensor of 4000x3000 pixels resolution and is equipped with F2.4 fisheye lenses with a fixed focal length of 1.88 mm. The lenses have 200° FOV along the maximum sensor dimension. The different shooting modes allows capturing 360° still-images, videos and timelapses, which make the device usable also for dynamic acquisitions. Equirectangular images can be stabilized (so that



Figure 1. Schematics of the two multi-camera systems and acquisition setups. On the left the INSTA 360 Pro2, on the right Ant3D.

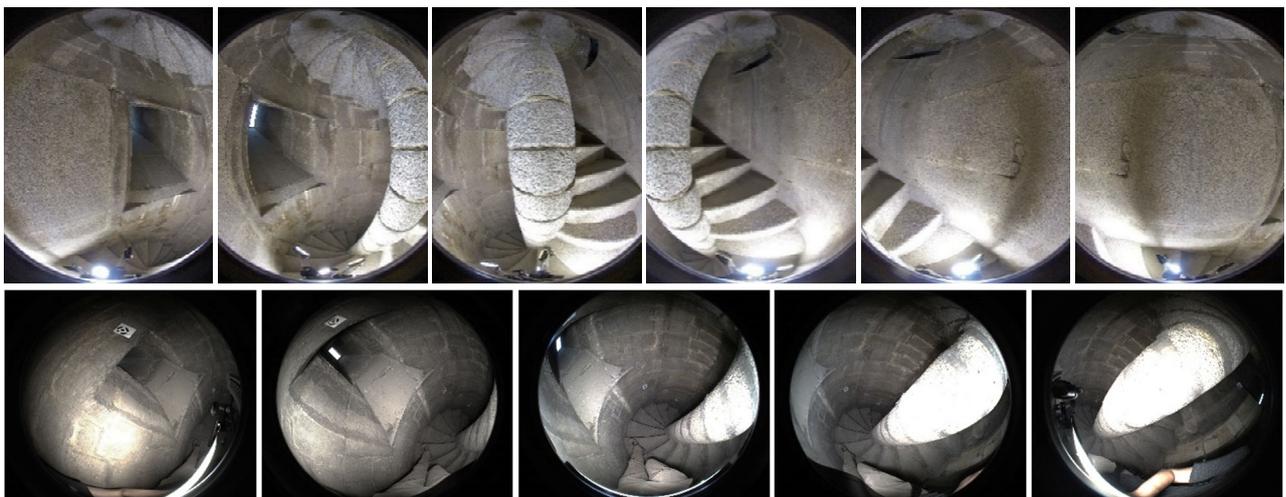


Figure 2. Fisheye images acquired by the two systems: top INSTA 360 Pro2, bottom Ant3D.

they can always point in the same direction) with a 9-axis gyroscope, and the camera is equipped with a built-in GNSS module. It is possible to set acquisition parameters such as ISO, shutter speed and white balance.

The camera records both the raw fisheye images acquired by each sensor (Figure 2 – top) and (if enabled by the user) the equirectangular images (7680 x 3840 pixels resolution) obtained by real-time stitching. Stitched images can also be obtained in post-processing using the dedicated Insta360stitcher software, other open-source or commercial software, such as PTGui and Autopano, or in-house codes to get higher control over distortion corrections.

2.2 Ant3D

The Ant3D multi-camera is a device meant for dynamic acquisitions that is operated handheld and captures sequences of synchronized images (Figure 2 – bottom). It comprises 5 cameras with a resolution of 2448x2048 pixels equipped with fisheye lenses with a focal length of 2.7mm. The arrangement of the cameras has been designed to be ideal for narrow spaces acquisition. Specifically, it optimises agility, transportation, manoeuvrability, FOV, and base length between the cameras. The cameras lie roughly on a plain and aim towards the outside with a single symmetry axis (Figure 1 - right). The FOV of the multi-camera covers the whole front hemisphere (the back side is not imaged to avoid including the operator). The device includes 3 illuminators, allowing the acquisition of dark areas and is

connected to a backpack that houses a controlling computer unit and a battery (Perfetti et al., 2022).

2.3 Camera calibration

For both cameras, an initial pre-calibration has been performed to compute the IO, distortion and RO (e.g., shifts and rotations between the cameras) parameters of each sensor.

A small room of about 1.8 x 0.8 x 2 m (i.e., width comparable to that of the Minguzzi staircase) and a wall with high texture contrast were set up with black and white photogrammetric targets. Some targets have been measured with a total station, and all coordinates were estimated, with an accuracy of 0.5 mm, from a redundant photogrammetric acquisition performed using a DSLR (Digital Single Lens Reflex) Nikon D810 with 24 mm lenses. More detail on the calibration test-field setup can be found in (Perfetti et al., 2022), as the same procedure has been replicated for the present investigation.

The calibration room was then surveyed with the two multi-camera systems from multiple shooting positions at different heights from the ground and rotating the sensors in all directions to decouple IO and EO parameters.

222 images were acquired with the INSTA 360 Pro2 (37 x 6), fixing the camera on a tripod and 90 x 5 = 450 images were acquired with Ant3D handholding the instrument.

Image processing was performed on the raw fisheye images acquired using Agisoft Metashape v2.0. Each dataset was processed following two pipelines:

- In the first pipeline, images were processed without considering the multi-camera constraint. First, targets were automatically detected and verified before the SfM process; then, SfM was run. The targets' reference coordinates were imported, treating them half as GCPs and half as CPs; bundle adjustments were re-run considering the GCPs. After that, the tie points were filtered twice, removing high collinearity residuals tie-points: at each of the two filtering steps, around 10% of all tie points was discarded and a new bundle adjustment was computed. Finally, the quality of the process was verified with the CPs' RMSE (Root Mean Squared Error), and the computed IO parameters were saved. For Ant3D only, from this result, the baselines between the multi-camera sensors were estimated and saved, according to Perfetti et al., 2022.
- In the second pipeline, images were processed considering the multi-camera constraint. First, an initial estimate of slave sensors' RO was specified with a weight of corresponding pseudo-observation of 5 mm for translations and 1 deg for rotations. After that, the same processing steps as pipeline 1 were followed. In addition to the IO parameters, the adjusted RO parameters were recorded at the end.

Table 1 and Table 2 report the shifts between the projection centres of the sensors w.r.t. the master camera for the INSTA 360 and the Ant3D systems (pipeline 2). It is worth noting that, for the INSTA 360 Pro2, the hypothesis of center of perspective coincidence between the cameras is not satisfied and significant geometric errors in equirectangular image generation should be expected, especially for short camera-object distances.

Camera	X [mm]	Y [mm]	Z [mm]
Sensor 1 (master)	0	0	0
Sensor 2 (slave)	-0.277	-59.573	-34.877
Sensor 3 (slave)	-0.306	-59.531	-103.847
Sensor 4 (slave)	-0.075	0.410	-138.291
Sensor 5 (slave)	0.229	60.025	-103.496
Sensor 6 (slave)	0.327	60.008	-34.474

Table 1. Baselines of the INSTA Pro2 camera.

Camera	X [mm]	Y [mm]	Z [mm]
Sensor 1 (master)	0	0	0
Sensor 2 (slave)	103.428	12.248	-236.562
Sensor 3 (slave)	-90.221	40.129	-22.216
Sensor 4 (slave)	83.699	40.278	-17.413
Sensor 5 (slave)	-99.954	12.731	234.131

Table 2. Baselines of the Ant3D multi-camera.

2.4 Acquisition

For image acquisition, slightly different setups were followed for the two instruments.

The INSTA 360 Pro2 was used in single-shot mode and mounted on a tripod to avoid framing the operator and to prevent motion blur due to the significantly reduced lighting conditions. Three illuminators anchored to the tripod legs and one placed on the ground upwards were used to light the scene (Figure 1 - left). Images were acquired at each staircase step, with an average base length between the subsequent acquisition of 30-40 cm, for a total of $145 \times 6 = 870$ images (10.4 Gpixels). Since the camera captures 360° images, the acquisition was performed in one direction (down the staircase) and took ca. one hour.

Ant3D was handheld, and the acquisition was carried out on the move. Since the multi-camera does not acquire backwards, the acquisition was performed both in the downward and upward direction. The illumination was provided by three illuminators

attached to the handheld device (Figure 1 - right). Images were acquired with a frame rate of 1 fps, resulting in $825 \times 5 = 4125$ images (20.7 Gpixels). The base length between consecutive captures ranges from 10 to 20 cm. The acquisition was completed in around 13 minutes.

2.5 Image processing and evaluation methodology

Well-recognizable architectural features, whose coordinates were obtained from a previous photogrammetric survey (Perfetti et al., 2017) with an accuracy of 1 cm, were used as GCPs and CPs to check the accuracy of the reconstruction, precisely the drift error that accumulates in the staircase path. Two different ground control point distributions were considered (Figure 3):

- 4 GCPs were placed at the bottom of the staircase, while the whole path was left unconstrained. It was the worst constraint scenario, where higher drift and scale errors should be expected.
- In addition to the 4 GCPs at the bottom, 2 GCPs were placed at the top of the staircase.

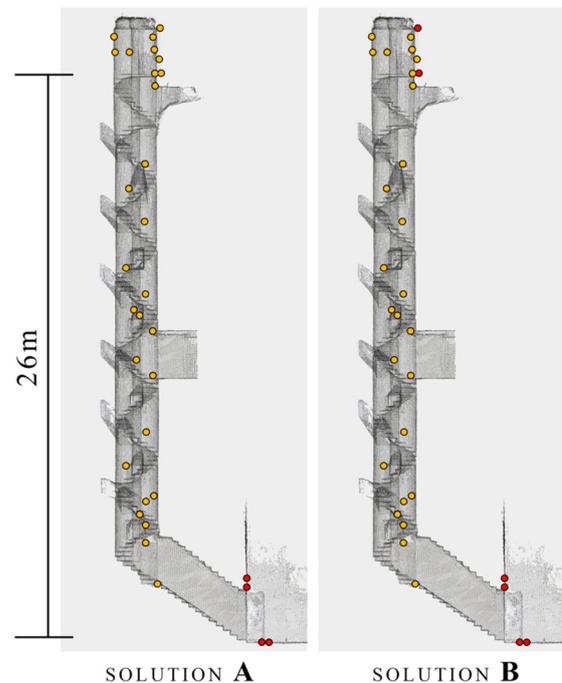


Figure 3. Schematic location of the GCPs (in red) and CPs (in yellow) along the staircase. On the left, ground constraint scenario A; on the right, scenario B

As for the calibration datasets, images were processed using Agisoft Metashape v2.0, using initial precalibrated IO parameters and testing different relative constraint solutions between the sensors (Figure 4) for each multi-camera system:

- Free fisheye images: each image was processed without considering the presence of a multi-camera system. No relative constraints were imposed with the other images acquired from the same position.
- (Only for Ant3D) Baselines constraints: distance constraints were imposed between fisheye images taken from the same position.
- Multi-camera relative constraints with on-the-job calibration: the sensors of the camera rigs were constrained using a multi-camera bind. One sensor was set as master and the other as slaves, with defined

relative angles and shifts resulting from pre-calibration. The precalibrated parameters were used as input for both the RO and IO parameters. While IO parameters were free, RO ones were constrained using a weight of 0.02 mm for shifts and 0.002 deg for rotations.

- iv. Multi-camera relative constraints with fixed IO parameters: the sensors were bound with the multi-camera constraint, and the precalibrated set of parameters was used to fix IO.
- v. (Only for INSTA 360) equirectangular images: the equirectangular images obtained using the camera stitching tools were processed.

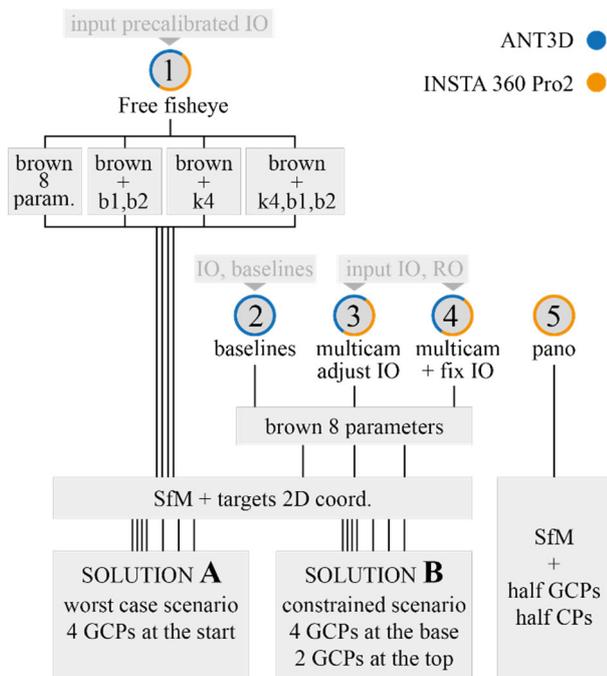


Figure 4. Scheme of the tests.

Image orientation was executed for all cases by initializing the processing with initial IO parameters. No GCPs were used in the initial SfM. The bundle block adjustment was performed to optimize the orientation, with control points set as tie points (only 2D image coordinates, without ground coordinates). Finally, the reference system was set up by entering the ground coordinates of the points and setting GCPs and CPs according to the constraint scenarios A and B described before.

All the tests have been performed considering an equidistant fisheye camera model with the Brown distortion model, estimating f , c_x , c_y , k_1 , k_2 , k_3 , p_1 , and p_2 parameters. Exclusively for the free fisheye image processing (case i), b_1 , b_2 , and k_4 parameters have also been considered. More specifically, the following set of test cases has been considered:

- Brown 8-parameter (cases i-iv)
- Brown 8-parameter + b_1 and b_2 (only case i)
- Brown 8-parameter + k_4 (only case i)
- Brown 8-parameter + b_1 , b_2 and k_4 (only case i).

A qualitative evaluation of the completeness, smoothness and reliability of the obtained point clouds and meshes has been performed to highlight the effectiveness of the proposed methodologies in CH digitization.

3. RESULTS AND DISCUSSION

Table 3 shows the RMSE on CPs for all the configurations tested, evaluating the effect of different ground control solutions and relative constraints on the accuracy of the reconstruction.

3.1 INSTA 360 Pro2 – accuracy results

For the INSTA 360 Pro2, as expected, it can be seen that using only 4 GCPs at the bottom of the staircase (Ground Control solution A) causes high drift errors at the top. This drift can be mainly attributed to the incorrect estimation of the base lengths along the path without ground constraints and increases progressively from the bottom to the top of the staircase, with a predominant effect on the Z direction. Applying relative constraints between sensors inside the multi-camera, particularly the base length estimation, improves the orientation and reduces the global drift. Free fisheye processing (case i) provides the worst results, with RMSE on CPs equal to 29 cm and a maximum of 50.5 cm. The multi-camera approach reduces the RMSE up to 19.4 cm (on-the-job calibration – case iii) and 4.4 cm (fixed precalibrated IO and distortion parameters – case iv), with maxima respectively at 28 cm and 6 cm.

Estimating b_1 and b_2 IO parameters in (case i) reduces the RMSE by 45 % up to 16 cm, with a maximum of 27 cm. Estimating k_4 , on the other hand, does not provide significant variations.

Adding 2 GCPs at the top of the staircase (Ground Control solution B) constrains the overall path length, absorbing most of the drift errors, regardless of the type of RO constraint imposed between the sensors. The average and maximum drift errors are much lower in all the datasets, although the positive effect given using multi-camera relative constraint still emerges. Free fisheye processing provides RMSE equal to 7.1 cm with a maximum of 9.7 cm in the middle of the staircase. Applying the multi-camera constraint, whether using fixed (case iv) or adjusted IO parameters (case iii), RMSE is very low, reaching 1.6 cm and 1.3 cm, respectively. Such values are comparable to the accuracy of the used GCPs and CPs. Also, in these tests, the estimation of b_1 and b_2 IO parameters improves the accuracy of free fisheye processing (case i), reducing the RMSE from 7.1 cm to 4.6 cm.

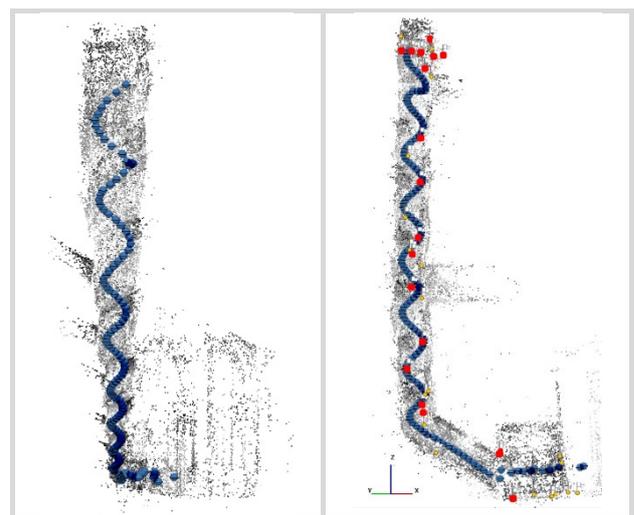


Figure 5. Results of the equirectangular processing (case v). On the left, the solution without GCPs in the initial SfM. On the right, the solution with half GCPs and half CPs in the SfM.

As far as the equirectangular images are concerned (case v), it was not possible to follow the same processing procedure used for the other test cases because the initial orientation without any

GCPs does not converge, with considerable drift error not only in the Z direction but also in the XY plane (Figure 5 – left). In this case, GCPs are needed in the initial SfM phase. The use of 6 GCPs (scenario B) still proved to be insufficient (RMSE equal to 1.90 m) while increasing the GCPs number to half of the points made it possible to reconstruct the staircase more reliably (RMSE on CPs equal to 20 cm and reprojection error of 11 pixels). Such high residuals make the model very noisy.

3.2 Ant3D – accuracy results

For Ant3D, the drift errors obtained by free fisheye processing are one order of magnitude lower than with the INSTA 360 with an RMSE on the CPs equal to 5.3 cm and a maximum error of 8.3 cm. As expected, the error in the Z direction is also higher than the error in the XY plane. Similar to what was observed for the INSTA 360, the test on estimating additional distortion parameters highlighted an improvement in the accuracy of the reconstruction by estimating b1 and b2. The datasets processed with brown + k4 resulted in RMSE that is almost identical to the regular Brown model. On the other hand, the two datasets with

b1-b2 performed similarly and better than regular Brown. The best result was obtained by Brown + b1, b2 and k4 with an RMSE of 3.4 cm and a maximum error of 5.2 cm.

As for the other tests with added relative constraints, we can observe that the drift error is lower than expected. All constraint scenarios produced a very similar result, with the best performer being the multi-camera constraint with fixed IO that shows an RMSE of 4.6 cm and a maximum error of 6.8 cm. However, the magnitude of the improvement is much lower than the one observed for the INSTA 360.

In solution B, adding GCPs at the opposite end of the path drastically improves results, as expected. The free fisheye test results in an RMSE of 1.3 cm and a maximum error of 2.2 cm. The relative constraint tests also improve, with the two multi-camera constraints that almost tie with the free fisheye test at an RMSE and maximum error of 1.6 cm and 2.3 cm for the multi-camera constraint and 1.3 cm and 2.5 cm for the multi-camera with fixed IO. Looking at solution B, adding GCPs at the opposite end of the path drastically improves results, as expected. The free fisheye test results in an RMSE of 1.3 cm and a maximum error of 2.2 cm.

INSTA 360 PRO 2							
Ground control solution A							
RO constraints	Camera calibration model	Reproj. error [pix]	CPs RMSE [m]				
			X	Y	Z	XYZ	max XYZ
(i)	Brown	0.5	0.038	0.016	0.287	0.290	0.505
	Brown + b1,b2	0.5	0.014	0.022	0.158	0.160	0.274
	Brown + k4	0.5	0.034	0.016	0.289	0.291	0.507
	Brown + b1,b2,k4	0.6	0.011	0.029	0.174	0.177	0.305
(iii)	Brown	0.5	0.065	0.055	0.174	0.194	0.281
(iv)	Brown	0.7	0.033	0.017	0.024	0.044	0.060
Ground control solution B							
RO constraints	Camera calibration model	Reproj. error [pix]	CPs RMSE [m]				
			X	Y	Z	XYZ	max XYZ
(i)	Brown	0.5	0.026	0.030	0.059	0.071	0.097
	Brown + b1,b2	0.5	0.023	0.025	0.031	0.046	0.066
	Brown + k4	0.5	0.026	0.031	0.062	0.074	0.101
	Brown + b1,b2,k4	0.6	0.021	0.029	0.038	0.052	0.074
(iii)	Brown	0.5	0.005	0.012	0.004	0.013	0.022
(iv)	Brown	0.8	0.004	0.015	0.004	0.016	0.026
ANT 3D							
Ground control solution A							
RO constraints	Camera calibration model	Reproj. error [pix]	CPs RMSE [m]				
			X	Y	Z	XYZ	max XYZ
(i)	Brown	0.6	0.028	0.023	0.038	0.053	0.083
	Brown + b1,b2	0.6	0.020	0.018	0.027	0.039	0.059
	Brown + k4	0.6	0.029	0.022	0.042	0.056	0.086
	Brown + b1,b2,k4	0.6	0.019	0.018	0.022	0.034	0.052
(ii)	Brown	0.7	0.024	0.023	0.035	0.048	0.076
(iii)	Brown	0.7	0.015	0.039	0.026	0.049	0.076
(iv)	Brown	0.7	0.039	0.024	0.007	0.046	0.068
Ground control solution B							
RO constraints	Camera calibration model	Reproj. error [pix]	CPs RMSE [m]				
			X	Y	Z	XYZ	max XYZ
(i)	Brown	0.6	0.005	0.010	0.007	0.013	0.022
	Brown + b1,b2	0.6	0.006	0.009	0.004	0.011	0.020
	Brown + k4	0.6	0.005	0.010	0.005	0.012	0.021
	Brown + b1,b2,k4	0.6	0.006	0.009	0.003	0.011	0.020
(ii)	Brown	0.7	0.014	0.012	0.016	0.024	0.036
(iii)	Brown	0.8	0.005	0.005	0.015	0.016	0.023
(iv)	Brown	0.8	0.008	0.008	0.006	0.013	0.025

Table 3. Stats of RMSE on CPs in all the processing configurations, for both the instruments.

The relative constraint tests also improve, with the two multi-camera constraints that almost tie with the free fisheye test at an RMSE and maximum error of 1.6 cm and 2.3 cm for the multi-camera constraint and 1.3 cm and 2.5 cm for the multi-camera with fixed IO.

Although the results are good, it is unexpected that the relative constraint tests, especially the multi-camera ones, do not outperform the free fisheye. This can be explained by the relative constraints conflicting with the scale established by the GCPs and suggesting, therefore, that calibration can be improved, or with the relative constraints being perhaps too strong with respect to the actual fluctuation of the shifts and rotation values that could be due to thermal expansions, vibration, and varying entrance pupil.

An additional test was run with lower RO weight constraints: 0.2 mm for shifts (instead of 0.02 mm) and 0.02 deg for relative rotations (instead of 0.002 deg). This test improved the performance with an RMSE of 4.7 cm and a maximum error of 7.4 cm for solution A and an RMSE of 1.1 cm and a maximum error of 2.1 cm for solution B.

3.3 Point cloud and mesh comparison

Both multi-cameras could produce complete and precise point clouds and meshes (Figure 6). The only geometries not completely reconstructed are the niches of the tapered windows that would have required a dedicated acquisition. Although similar, the INSTA 360 results are more detailed (due to the higher resolution images) and more consistent (due to more even illumination). Ant3D mesh shows some artefacts in the less illuminated areas (Figure 6 – bottom right), suggesting that more images could be acquired and that the illumination setup has room for improvements.

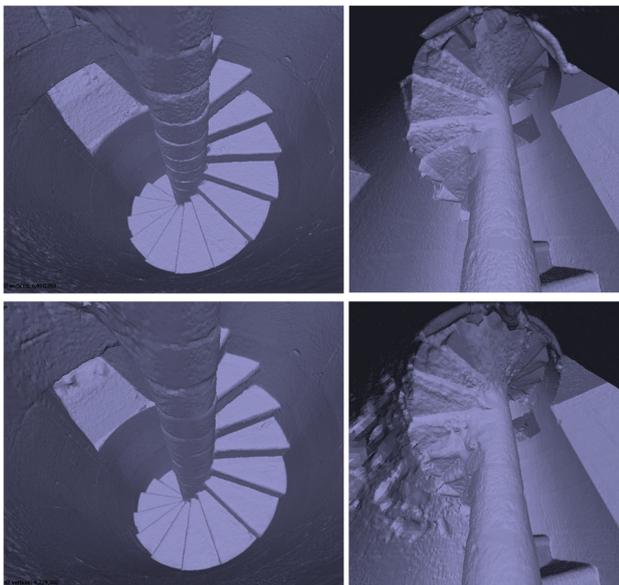


Figure 6. Mesh comparison, INSTA (top), Ant3D (bottom).

3.4 Discussion

Both the multi-camera systems tested proved to be effective in surveying long and complex paths even with few ground constraints. In these contexts, it is usually unfeasible (or even impossible) to materialize and survey a strong network of GCPs or to obtain GNSS camera position data.

Constraining the path at the start and at the end (constraint solution B), both the systems reached the best accuracy of ca 1.3–1.6 cm on CPs and also provided, aside from little artefacts,

complete and well-defined 3D models, suitable to represent architectural CH at 1:50 nominal scale.

The test showed that the Ant3D camera arrangement better constrains the block geometry, strengthening image orientation. The larger base lengths between sensors (varying between 10 to 25 cm) ensure stable matchings even between images taken from the same shooting position. On the contrary, the base lengths inside the INSTA 360 Pro2 system (ranging between 7 and 14 cm) do not provide the same constraint. This is particularly evident when comparing the outcomes of free fisheye image processing (case i). Using GCPs only at the bottom of the staircase (Ground control solution A) INSTA 360 pro2 performs 82% worse than Ant3D (0.290 cm and 0.053 cm, respectively). Constraining the path both at the start and at the end (Ground control solution B) considerably improves the global accuracy, but while Ant3D reaches the target accuracy on CPs (1.3 cm, value comparable to the accuracy of the GCPs used), INSTA 360 Pro2 performs 6 time worse (7.3 cm). Leaving the images free from RO constraints makes the orientation solution of the INSTA less stable.

On the contrary, when a strong RO constraint between the multi-cameras sensors is imposed (case iv), the two systems reach the same accuracy on CPs, i.e. 4.4 cm and 4.6 cm for INSTA and Ant3D, respectively, in Ground control solution A and 1.6 cm and 1.3 cm in Ground control solution B.

The weaker base-length geometry of the INSTA thus requires a stronger constraint, which is, in this case, provided by setting shifts and relative angles between the sensors. On the contrary, the more robust geometry of Ant3D also allows the use of free images and suggests not over-binding the system by imposing almost absolute constraints for the RO of the multi-camera.

A further discussion deserves the processing of equirectangular images obtained using the INSTA 360 stitching tool. With 360° cameras, it is generally customary to work directly on equirectangular images that can be processed using a spherical camera model, reducing processing time. The tests performed in this work highlighted the inadequacy of this method for metric purposes where high accuracy is required.

A multi-camera rig system, such as Ant3D, allows for better and more stable geometric reconstruction, even against drift errors in open-ended paths. On the other hand, using spherical cameras, such as the INSTA 360 Pro2, allows for good geometric reconstruction as long as RO constraints between sensors are imposed. It acquires spherical images that allow the entire scene to be framed in a single shot (i.e., no need for a backward path) and can provide spherical image reconstruction, but just for visualization or virtual navigation purposes.

4. CONCLUSION AND FUTURE WORKS

The paper described a test for accuracy and reliability evaluation of two multi-camera systems (INSTA 360 Pro2 and Ant3D) for the survey of narrow spaces in CH with few ground constraints. The tests have been performed processing the raw fisheye images acquired by the two instruments, and (INSTA 360 pro2 only) also the stitched equirectangular images. Different RO constraints have been tested as well.

The results highlighted that the drift errors constraining the surveyed path only at the start can be significant, reaching in our case study, a maximum of 50 cm for the INSTA and 8 cm for Ant3D. The drift error has been shown to be mainly linked to the robustness of the geometric structure of the instrument. The multi-camera system of Ant3D, with base-length among the sensors up to 25 cm is better able to constrain image orientation and provides higher accuracy. Nevertheless, the use of strong RO constraints between the sensors (multi-camera relative angles and shift constraints) allows both instruments to perform the same

and reach RMSE on CPs of 1.6 and 1.3 cm for the INSTA and Ant3D, respectively.

This accuracy is also noticeable in the 3D reconstruction (dense cloud and mesh model), which, unlike what has been found by other authors working on spherical cameras, is not noisy and can be suitable for a 1:50 scale level of detail. These systems have, therefore, been demonstrated to be suitable and reliable for CH surveys.

Future works can consider repeatability tests on the same environment, accuracy and completeness evaluation of the dense cloud compared to a reference ground truth, further investigation on the IO and distortion parameters and the influence of the camera-to-object distance.

REFERENCES

- Barazzetti, L., Previtali, M., Roncoroni, F., 2017: 3D modelling with the Samsung Gear 360. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W3, 85–90.
- Barazzetti, L., Previtali, M., Roncoroni, F., 2018: Can we use low-cost 360 degree cameras to create accurate 3D models? *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2, 69–75.
- Barazzetti, L., Previtali, M., Roncoroni, F., Valente, R., 2019: Connecting inside and outside through 360° imagery for close-range photogrammetry. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W9.
- Barazzetti, L., Previtali, M., Roncoroni, F., 2022: 3D modeling with 5k 360° videos. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVI-2/W1, 65–71.
- Bruno, N., Roncella, R., 2019: Accuracy assessment of 3D models generated from Google Street View imagery. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W9, 181–188.
- Cera, V., Campi, M., 2022: Fast survey procedures in urban scenarios: some tests with 360° cameras. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-2/W1, 45–50.
- Chiappini, S., Fini, A., Malinverni, E.S., Frontoni, E., Racioppi, G., Pierdicca, R., 2020: Cost effective spherical photogrammetry: A novel framework for the smart management of complex urban environments. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLIII-B4, 441–448.
- Fangi, G., Pierdicca, R., Sturari, M., Malinverni, E.S., 2018: Improving spherical photogrammetry using 360°OMNI-Cameras: Use cases and new applications. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* XLII-2, 331–337.
- Gómez-López, J.M., Pérez-García, J.L., Mozas-Calvache, A.T., Vico-García, D., 2023: Documentation of Cultural Heritage through the fusion of geomatic techniques. Case study of the cloister of "Santo Domingo" (Jaén, Spain). *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, XLVIII-M-2, 677–683.
- Gottardi, C., Guerra, F., 2018: Spherical images for cultural heritage: Survey and documentation with the NIKON KM360. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, XLII-2, 385–390. 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. *J. Imaging* 8.
- Holdener, D., Nebiker, S., Blaser, S., 2017. Design and implementation of a novel portable 360° stereo camera system with low-cost action cameras. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W8, 105–110.
- Koehl, M., Delacourt, T., and Boutry, C., 2016: Image capture with synchronized multiple-cameras for extraction of accurate geometries. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLIB1, 653–660.
- Martino, A., Breggion, E., Balletti, C., Guerra, F., Renghini, G., Centanni, P., 2023: Digitization approaches for urban cultural heritage: last generation MMS within Venice outdoor scenarios. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, XLVIII-1/W1, 265–272.
- Menna, F., Battisti, R., Nocerino, E., Remondino, F., 2023. FROG: a portable underwater mobile mapping system. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-1/W1-2023, 295–302.
- Meyer, D. E., Lo, E., Klingspon, J., Netchaev, A., Ellison, C., and Kuester, F., 2020. TunnelCAM- a HDR spherical camera array for structural integrity assessments of dam interiors. *Electronic Imaging*, 2020(7), 227–1.
- Nocerino, E., Nawaf, M. M., Saccone, M., Ellefi, M. B., Pasquet, J., Royer, J.-P., Drap, P., 2018. Multi-camera system calibration of a low-cost remotely operated vehicle for underwater cave exploration. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-1, 329-337.
- Ortiz-Coder, P., Sánchez-Ríos, A., 2019. A Self-Assembly Portable Mobile Mapping System for Archeological Reconstruction Based on VSLAM-Photogrammetric Algorithm. *Sensors* 2019, 19, 3952.
- Panella, F., Roecklinger, N., Vojnovic, L., Loo, Y., and Boehm, J., 2020. Cost-benefit analysis of rail tunnel inspection for photogrammetry and laser scanning. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLIII-B2-2020, 1137–1144.
- 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.
- Perfetti, L., Polari, C., Fassi, F., 2018. Fisheye multi-camera system calibration for surveying narrow and complex architectures. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2, 877–883.
- Perfetti, L., 2020. Multi-Camera Rig For 3D Reconstruction: Concept, Design and Accuracy Evaluation. *Luhmann, T., Schumacher, C., Photogrammetrie - Laserscanning – Optische 3D-Messtechnik, Beiträge der Oldenburger 3D-Tage 2020*, 124-131.
- Perfetti, L. and Fassi, F., 2022. Handheld fisheye multicamera system: surveying meandering architectonic spaces in open-loop mode – accuracy assessment. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVI-2/W1-2022, 435–442.
- Perfetti, L., Spettu, F., Achille, C., Fassi, F., Navillod, C., Cerutti, C., 2023. A multi-sensor approach to survey complex architectures supported by multi-camera photogrammetry. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-M-2-2023, 1209–1216.
- Teo, T., 2015. Video-based point cloud generation using multiple action cameras. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XL-4/W5, 55–60.
- Teppati Losè, L., Chiabrando, F., Giulio Tonolo, F., 2021: Documentation of Complex Environments Using 360° Cameras. The Santa Marta Belltower in Montanaro. *Remote Sens.* 13, 3633.
- Teruggi, S. and Fassi, F., 2022. Mixed reality content alignment in monumental environments, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLIII-B2-2022, 901–908.
- Torresani, A., Menna, F., Battisti, R. and Remondino, F., 2021. A v-slam guided and portable system for photogrammetric applications. *Remote Sensing* 13, no. 12.