3D MODELING WITH 5K 360° VIDEOS

L. Barazzetti[a] , M . Previtali[a], F. Roncoroni[b]

[a] Dept. of Architecture, Built environment and Construction engineering (ABC)
Politecnico di Milano, Piazza Leonardo da Vinci 32, Milan, Italy
(luigi.barazzetti, mattia.previtali)@polimi.it
[b] Polo territoriale di Lecco, via Previati 1/c, 23900, Lecco, Italy
fabio.roncoroni@polimi.it

Commission II

KEY WORDS: 360°, 5k, Accuracy, Automation, Low-cost sensor, Orientation, Video

ABSTRACT:

Video acquisition with 360° (spherical) cameras is becoming increasingly popular for the opportunity to capture the entire scene around the user in a relatively short time. The method can also be attractive for photogrammetric applications. As the overlap between consecutive frames is undoubtedly guaranteed, 3D models can be generated with an automated processing workflow. The paper illustrates the results achieved with 5k 360° videos captured with different Insta360 cameras. As the number of frames can become large, two complementary solutions are proposed to provide approximate initial exterior orientation parameters: the integration of the trajectory captured through GNSS, and the creation of an acquisition plan with a GIS-based application. The availability of approximated EO parameters provides a visibility map between the frames and reduces the computational cost during image matching. Experimental results demonstrate that such preliminary information is necessary for large datasets. Indeed, the photogrammetric processing of the entire dataset without the proposed preliminary EO parameters resulted in unreliable or incomplete orientation results.

1. INTRODUCTION

The 360° camera model (also called spherical or equirectangular) is supported by commercial software based on the Structure from Motion (SfM) / photogrammetric image processing workflow for 3D modeling. For instance, Agisoft Metashape and Pix4Dmapper can process images and videos captured with low-cost 360° commercial cameras.

In recent years, rapid technological advances in 360° cameras and the use of such sensors for virtual reality applications have reduced hardware costs and enhanced image/video quality. Different papers by different authors have discussed the results achievable using images acquired with low-cost 360° cameras, showing different case studies, and testing metric accuracy and model completeness with laser scanning or traditional photogrammetry. Some examples are illustrated and discussed in Aghayaria et al. (2017), Abate et al. (2017), Barazzetti et al. (2015;2017), Fangi, (2009; 2017), Fassi et al. (2019)., Kwiatek et al. (2014-2015), Matzen et al. (2017), Pisa et al., 2010, Strecha et al. (2015).

A comprehensive discussion of the spherical bundle adjustment theory and different practical case studies are described in Fangi (2017). The adjustment method is an extension of methods for adjusting geodetic networks based on angular measurements. Pixel coordinates (x, y) and angles expressed as complement of latitude (ϕ) and longitude (λ) are related by the following equations: $x = r\lambda$ and $y = r\phi$. The radius of the sphere r (in pixels) can be estimated as $r = h/\pi = w/(2\pi)$, where (w, h) are image width and height in pixels. The radius corresponds to the focal length of the camera.

As an example, the spherical camera model used in Metashape has the form:

$$u = \frac{1}{2}w + \frac{w}{2\pi}\frac{X}{Z} \qquad v = \frac{1}{2}w + \frac{w}{2\pi}\frac{Y}{\sqrt{X^2 + Y^2}} \tag{1}$$

where (X, Y, Z) are point coordinates in the local camera coordinate system. This implementation does not support distortion correction.

The development of high-resolution panoramic cameras able to provide equirectangular projections in a fully automated way is making the use of spherical photogrammetry increasingly interesting for digital documentation projects. An example is the Weiss AG Civetta camera, featuring 230 megapixels. More details can be found in Zhao (2021). This paper distinguishes from previous work for the direct use of videos instead of "static" images. Although the use of frames extracted from a video is conceptually similar, new problems arise and require processing strategies able to reduce CPU time and create reliable datasets of matched image points for bundle adjustment. Videos can provide many frames that could increase CPU cost, making data processing impossible even with powerful hardware configurations. At the same time, when multiple long-duration videos are taken in the same area, the experimental results show that matching could not identify some tie points between frames of different videos, leading to significant errors and deformations in the reconstruction.

The used videos have 5k resolution (5760 × 2880) and were acquired using two different Insta360 cameras. The advantage of using a video is the very rapid acquisition. The user has only to walk with the camera on a pole (such as a selfie stick). Overlap between consecutive frames is automatically guaranteed. The 360° field of view provides complete coverage of the entire scene, which can be fully reconstructed notwithstanding ground sampling distance (GSD) can be highly variable depending on object geometry. Particular attention must also be paid to those geometries aligned with camera locations, for which the intersection of rays in 3D space cannot be achieved.

This paper describes two solutions developed to overcome these limitations, describing pros and cons and possible future developments based on their integrated use. The first solution consists of using a GIS-based tool to generate approximate exterior orientation (EO) parameters, which are used as initial approximation during matching and bundle adjustment. Such a solution is advantageous when no information is available about camera trajec-

tory. The second solution provides approximated EO parameters captured with a mobile phone associated with the low-cost 360° camera. In this way, initial information can be used to document large outdoor spaces, such as narrow streets in city centers.
Examples and results for both methods are illustrated and discussed within the paper. Finally, the possible integration of the two methods is discussed.

2. METHOD N.1: APPROXIMATED TRAJECTORY USING GIS

The first solution is based on an approximated trajectory created using existing 2D drawings (cartographies, maps, plans, etc.). The aim is to generate an approximated acquisition plan depending on the videos. The user must trace the trajectory of the video on a map, then approximated camera locations are automatically generated depending on the duration of the video. Instead of acquiring a single long video, splitting video acquisition into multiple parts is recommended to manually trace the trajectory. The proposed solution was implemented in QGIS and used the frame rate chosen to subsample the video, assuming a constant speed motion.
Figure 1 shows the trajectory for a project in the historic city center of Sondrio (Italy). Fourteen videos with variable length were acquired with an Insta360 One R. The extracted frames form a dataset of 3,337 spherical images. The total path is about 1,650 m, and the area is about 31,500 m^2.

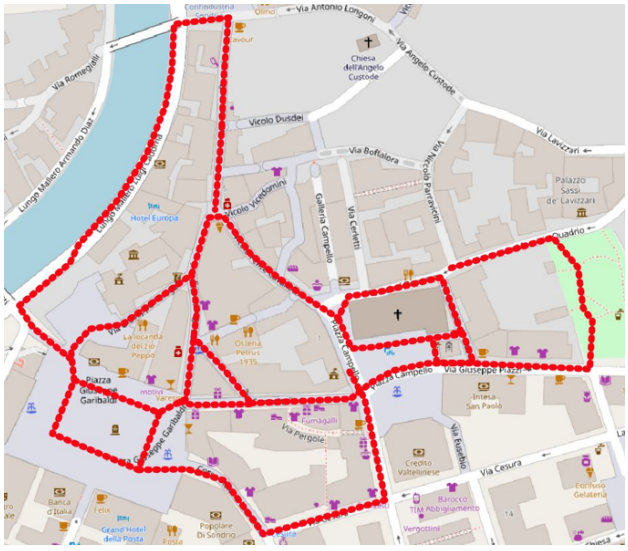
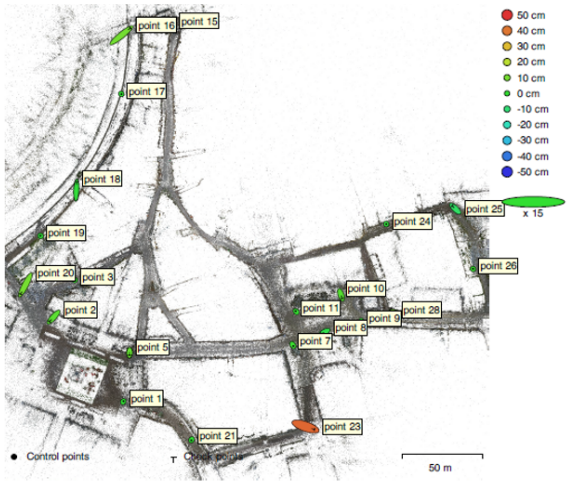


Figure 1: The computed approximated EO parameters using the GIS-based tool.

Data processing was carried out in Metashape, importing the approximated EO parameters to initialize image matching and orientation. It is essential to mention that the software did not successfully process the same dataset without using such an initial approximation. After orienting the frames, additional control points (measured with an Emlid Reach RS2 in RTK) were integrated into the adjustment using the optimization tool of the software. The regional service of GNSS permanent stations (SPIN3 GNSS) was used with an Internet connection. The closest station (Sondrio) was very close (less than 300 m) to the survey area. The reference system used for the project is UTM32-WGS84 ETRF2000-RDN (2008.0).
The resulting sparse cloud was imported on a digital orthophoto

Figure 2: The points used for image orientation visualized on a satellite orthophoto.



Count	X error (cm)	Y error (cm)	Z error (cm)	XY error (cm)	Total (cm)
11	2.26397	3.75757	1.04463	4.3869	4.50956

Control point RMSE

Count	X error (cm)	Y error (cm)	Z error (cm)	XY error (cm)	Total (cm)
10	39.6188	43.1581	15.1474	58.5856	60.5121

Check point RMSE

Figure 3: Statistics (in cm) after adding precise control points in the adjustment.

(Figure 2) of the city center. Finally, the results for a set of control and check points are illustrated in Figure 3. The discrepancy with control points is consistent with the expected precision of GNSS measurements (centimeter level: 2.3 cm, 3.7 cm, 1.0 cm for X, Y, Z, respectively), whereas the error on check point is more significant (39.6 cm, 43.1 cm, 15.1 cm) notwithstanding some check points are relatively close to control points.
Such a result is still not clear and requires additional analysis. In fact, if all GNSS points are integrated as control points in the adjustment, RMSE values of control points become 2.7 cm, 3.2 cm, and 1.9 cm for X, Y, Z, respectively. On the other hand, this last result cannot be accepted as no check points remain available.
A second set of GNSS points will be measured in the area to clarify such a large discrepancy between control points and check points.

3. METHOD N.2: APPROXIMATED TRAJECTORY VIA MOBILE PHONE

3.1 Overview

Most 360° cameras can be connected to a mobile phone using a wireless connection and specific applications. The idea behind the second method proposed in the paper is to exploit the GNSS trajectory and obtain initial exterior orientation parameters for the different frames, which can then be used to determine the visibility between images and reduce the number of combinations during image matching.
Figure 4 shows the trajectory acquired in the city center of Bassano del Grappa (Italy). The camera used is an Insta360 One X2, which can capture 5k videos. The total distance is about 950 m. As can be seen, the trajectory provided by the GNSS receiver inside the mobile phone is not regular mainly for two reasons: (i) the low precision of the sensor itself, and (ii) the narrow spaces inside the city center, which provide several occlusions and have an impact on the acquisition of GNSS signals.

Figure 4: Trajectory provided by mobile phone (top), after bundle adjustment (middle), and comparison (bottom).

The approximated exterior orientation parameters are therefore used (i) to limit the number of combinations during image matching and (ii) to georeference the photogrammetric project using a rigid registration (scale, rotation, translation), obtaining the following errors:

- mean longitude error = 335 cm
- mean latitude error = 260 cm
- mean altimetric error = 297 cm

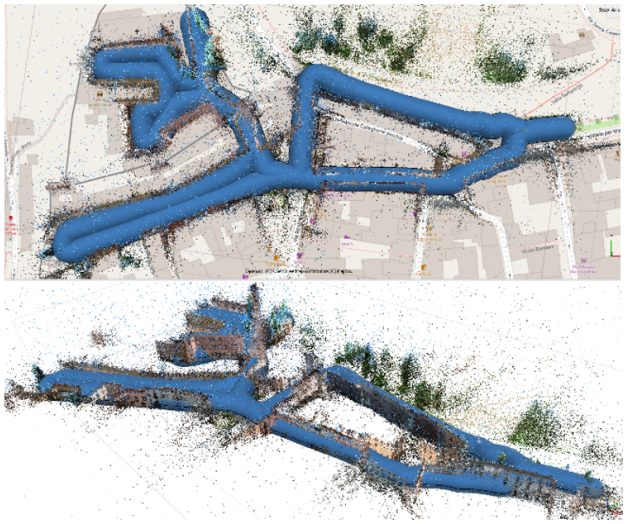


Figure 5: Orientation results of spherical 5k frames with the first proposed strategy. The project has been rigidly georeferenced using mobile phone camera parameters.

Such errors are consistent with the expected precision of a GNSS receiver inside a mobile phone. The software used for processing is Agisoft Metashape (Figure 5), which processed about 1800 frames acquired in 20 minutes.
The following section deals with the integration of precise GNSS points to compensate for the lack of reliable measurements for overall georeferencing.

3.2 Accuracy evaluation integrating precise GNSS points

3.2.1 Dataset n. 1. A second dataset was acquired around Lecco Campus of Politecnico di Milano (Italy), always using an Insta360 ONE X2 and a mobile phone to record the trajectory. The same procedure was applied to reduce the number of combinations during image matching. A comparison between mobile

Figure 6: Results for the Lecco dataset: mobile phone recorded path (yellow) and trajectory after bundle adjustment (red).

phone results and the trajectory after bundle adjustment is shown in Figure 6.
After the rigid registration, the following results were achieved:

- mean longitude error = 368 cm
- mean latitude error = 440 cm
- mean altimetric error = 63 cm

The dataset features 2450 frames acquired in about 30 minutes. Processing was carried out with Metashape obtaining the results in Figure 7.

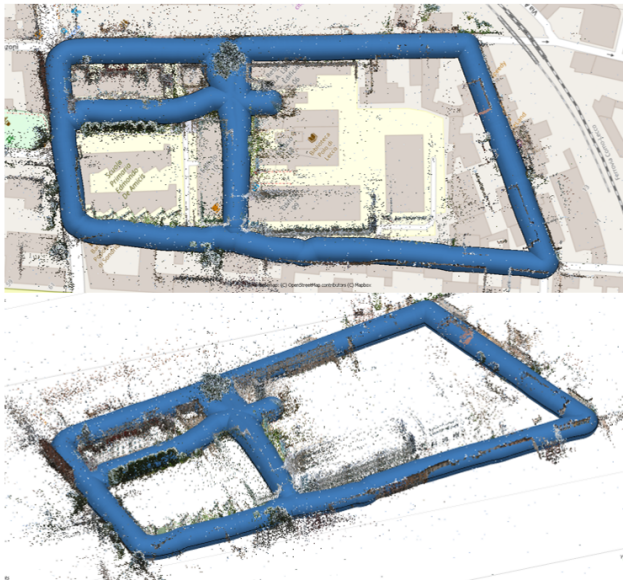
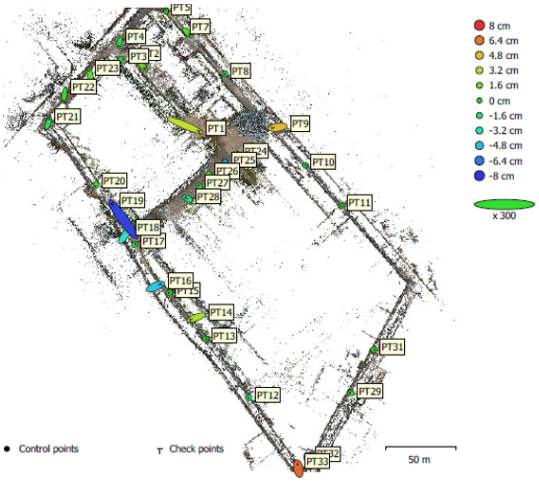


Figure 7: Results after bundle adjustment with a rigid registration on mobile phone GNSS coordinates.



Count	X error (cm)	Y error (cm)	Z error (cm)	XY error (cm)	Total (cm)
20	0.394304	1.17949	2.09314	1.24366	2.43473

Control point RMSE

Count	X error (cm)	Y error (cm)	Z error (cm)	XY error (cm)	Total (cm)
11	3.22706	3.18549	3.92587	4.53445	5.99781

Check point RMSE

Figure 8: Statistics (in cm) after adding precise control points in the adjustment.

After image orientation, additional points (31) were measured

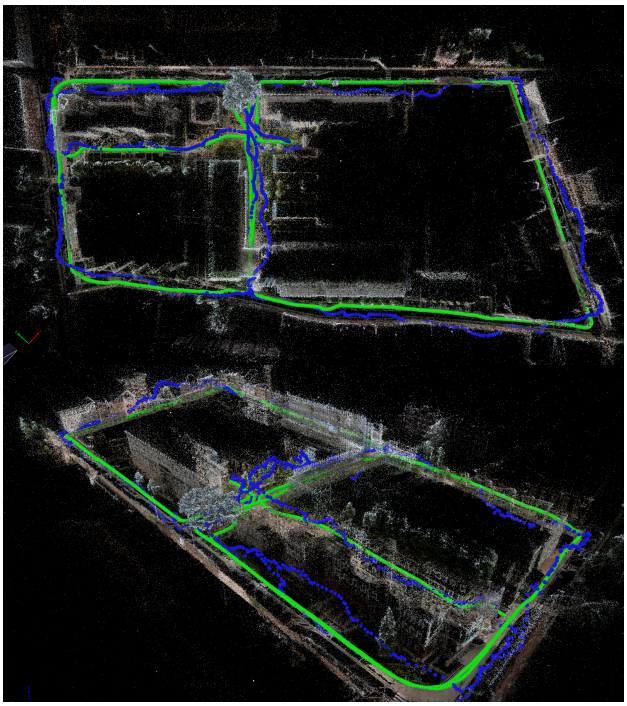


Figure 9: Orientation results obtained with Pix4Dmapper.

with an Emlid Reach RS2 GNSS antenna, featuring a better metric precision (±2-4 cm). The same method used for the Sondrio dataset was replicated. The closest permanent GNSS station of the SPIN3 GNSS network was located on a building of the Lecco Campus, making the baseline very short (less than 200 m) and providing good metric precision.
Some control points (20) were manually measured in the images and a set of check points (11). The optimization of the project with such additional constraints allowed one to reach better metric results with a discrepancy of a few centimeters, as shown in Figure 8.
The same dataset was processed with Pix4Dmapper, which can process blocks and sequences of equirectangular projections. The same approximated parameters were imported at the beginning of the project and used as initial approximation during matching and orientation. The same control (20) and check (11) points were used to evaluate metric accuracy after bundle adjustment.
RMSE values of control points were 1.8 cm, 2.3 cm, and 4.1 cm for X, Y, Z, respectively. RMSE values on check points were 1.6 cm, 3.5 cm, and 4.3 cm.
Such statistics are consistent with the results obtained using Agisoft Metashape and confirm the need for a precise set of control points to compensate for network deformation errors.

3.2.2 Dataset n. 2. Another dataset was acquired in a small mountain village Erve/Nesolio, in the surrounding of Lecco (Italy). The dataset was acquired with an Insta360 ONE X2. It covers the entire village, approximately a surface of 0.56 ha. The acquisition of this dataset was organized into 4 main loops (Figure 10), resulting in a few minutes of video recording. The videos were then sampled at the frequency of one frame per second, obtaining 1382 frames. The GNSS track was acquired by a mobile phone. The data processing was carried out using the same workflow previously presented. As it can be observed in Figure 10 one of the main issues of this dataset is the low accuracy of the GNSS track, which is probably due to the weak visibility of GNSS satellites in a mountainous region. In particular, the camera position errors

are showing the following mean discrepancies:
- mean longitude error = 917 cm
- mean latitude error = 876 cm
- mean altimetric error = 1530 cm

Maximum discrepancies reach 40 m (horizontal) and 35 m (height) and are mainly located in the path marked in light blue in Figure 10.

Figure 10: Trajectory provided by mobile phone (top) and after bundle adjustment (centre) and GCPs/CPs position (bottom).

A set of control points and check points were manually measured. Different configurations of GCPs and CPs were tested in Agisoft Metashape (Figure 11). Then, the same dataset was processed with Pix4Dmapper (Figure 12). The same control (41) and check (7) points were used in both software. Statistics are summarized in Figure 13 and show a good correspondence.

Once the image block is oriented, it is also possible to carry out dense matching. Figure 14 shows an example of dense matching output carried out with Agisoft Metashape, obtaining about 215 million points. However, this point will be addressed more in detail in future work, as described in the next section.

Number of Ground Control Points	Point type	X – East Mean Error [cm]	Y – North Mean Error [cm]	Z – Altitude Mean Error [cm]	X – East Max. Error [cm]	Y – North Max. Error [cm]	Z – Altitude Max. Error [cm]
#8 (4, 6, 7, 8, 9, A1, A3, A4)	GCP (#8)	1.6	0.9	3.2	2.6	1.1	4.9
	CP (#40)	4.9	4.4	4.6	11.5	11.9	5.8
#13 (4, 6, 7, 8, 9, A1, A3, A4, A5, A20, A28, A34, A44)	GCP (#13)	2.2	1.6	2.9	6.1	4.1	4.9
	CP (#35)	4.0	3.9	4.6	7.0	12.9	8.3
#23 (4, 6, 7, 8, 9, A1, A3, A4, A5, A8, A10, A14, A16, A19, A20, A22, A25, A28, A31, A34, A38, A41, A44)	GCP (#23)	1.9	1.6	2.5	6.4	4.6	6.3
	CP (#25)	3.5	3.8	3.6	7.8	11.0	4.8
#23 (4, 6, 7, 8, 9, A1, A3, A4, A5, A6, A7, A8, A10, A11, A13, A14, A16, A18, A19, A20, A22, A23, A25, A26, A28, A29, A31, A32, A34, A37, A38, A40, A41, A44)	GCP (#34)	1.8	1.7	2.2	6.1	4.9	5.8
	CP (#14)	2.9	3.1	3.5	7.6	9.7	4.8
#48 (ALL)	GCP (#48)	2.0	1.7	2.0	6.3	5.2	5.4

Figure 11: Orientation statistics with different GCPs configurations.

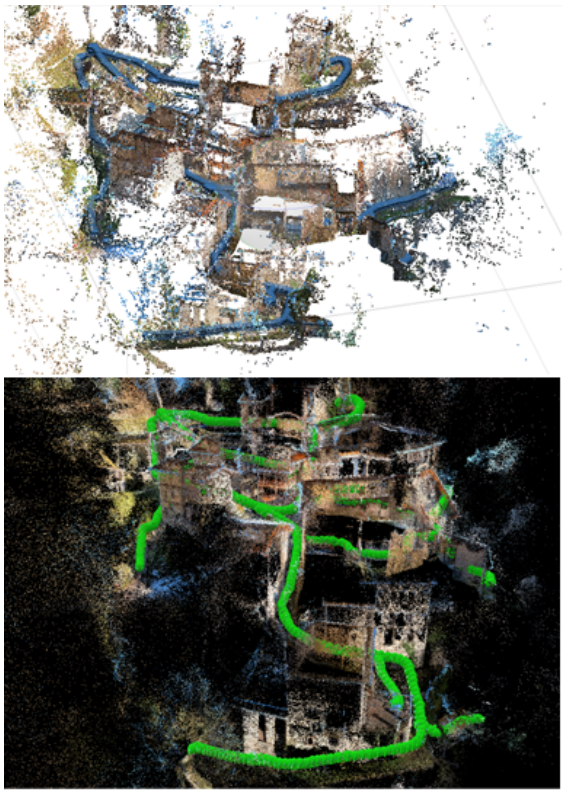


Figure 12: Orientation results obtained with Agisott Metashape (top) and Pix4Dmapper (bottom).

Number of Ground Control Points	Point type	X – East RMSE [cm]	Y – North RMSE [cm]	Z – Altitude RMSE [cm]
Agisoft Metashape	GCP	2.0	1.7	2.1
	CP	3.7	4.1	6.2
Pix4Dmapper	GCP	2.2	1.9	2.9
	CP	5.1	3.5	3.3

Figure 13: Statistics on GCPs and CPs with Agisoft Metshape and Pix4Dmapper.)

Figure 14: Dense cloud reconstructed for the Erve/Nesolio dataset.

4. CONSIDERATIONS AND OUTLOOKS

The use of high-resolution 360° videos is an attractive opportunity to reconstruct complex scenes. The method has pros and cons, which are summarized in this section along with future work. The manuscript has faced only some of the possible issues when such data are used in a photogrammetric workflow. For instance, illumination conditions play a fundamental role among practical issues. The need for uniform and homogenous illumination is a requirement in most photogrammetric projects, including those carried out with traditional images based on the central perspective camera models. In the case of 360° images, it is impossible to compensate for significant differences in lighting conditions. Although most low-cost cameras today acquire HDR images, the transition from bright and dark spaces remains problematic.

As the overlap between consecutive frames is guaranteed, the user can walk into the scene to capture the entire area. Moreover, the camera can be pointed in any direction, making the method very attractive for operators who are not specialists in digital photogrammetry. Data acquisition is very rapid, and the examples proposed in this paper showed that the survey of very long and narrow spaces could be carried out in a few minutes. On the other hand, several other aspects must be considered and were not exploited in the paper.

The first obvious consideration is the completion of the whole photogrammetric processing workflow. The paper only concentrated on the first step of data processing (image orientation). In contrast, the photogrammetric pipeline for 3D modeling would require the generation of dense point clouds and other deliverables such as meshes or digital orthophotos. These additional outputs were not tested and still require future work. Preliminary results revealed that the metric quality of a point cloud is questionable due to multiple problems, such as the use of short baselines and the highly variable camera-object distance, that could result in weak intersections of rays in 3D space. However, further tests and numerical results on dense point cloud quality are beyond the scope of this paper.

An example of results from the Sondrio dataset is shown in Fig. 15. One of the city hall facades was reconstructed using a point cloud of about 5.5 million points extracted from about 100 spherical images. Then, a mesh was generated, featuring 1.1 million faces. Finally, a digital orthophoto was produced (pixel = 1 cm). A detail of the bottom-left corner is shown in the figure. No metric evaluation was carried out on such deliverables, which are here presented as future extensions of the proposed approach.

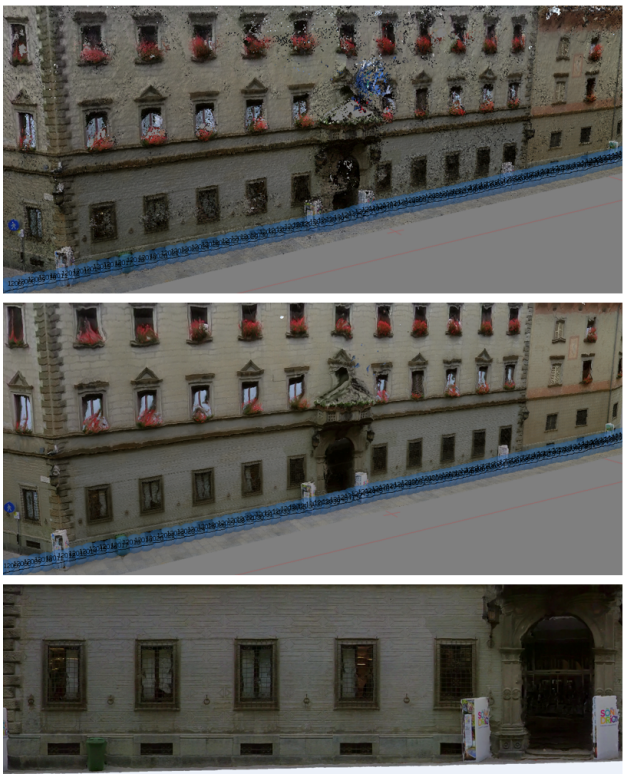


Figure 15: Dense cloud, mesh and orthophoto from about 100 spherical images in the Sondrio dataset.

An important consideration is also related to the 5k resolution used in this paper. The GSD of a project carried out with 360° images can be highly variable. It depends on both object geometry and trajectory of the camera. The proposed examples showed variable GSD values, ranging from a few millimeters (in narrow streets) to several centimeters (in squares or other “more open” spaces). Overall, metric quality is not constant for the entire project, resulting in some areas with a very high level of detail, whereas others can reveal inferior geometric detail.

5. CONCLUSIONS

The proposed work focused on 5k videos acquired with low-cost cameras (under 500 euros). Frames are automatically extracted from the video and processed using a photogrammetric workflow for image orientation. Starting from a set of tie points extracted with SIFT-like algorithms, bundle adjustment based on the equirectangular camera model provided EO parameters. The paper presented two methods to simplify both matching and orientation stages with the initial creation of approximated EO parameters. Such methods allowed one to orient large sequences

of frames (more than 2000-3000 images) that could not be successfully oriented without including the proposed approximated initial parameters. From this point of view, the method is not only a solution able to speed up processing, but it was also necessary to complete the image orientation phase.
Precise georeferencing can be achieved only when precise GNSS points are integrated into processing. If a mobile phone connected to the camera is used to measure approximated EO parameters, the survey can also be georeferenced using a 7-parameters rigid transformation (translations (3), rotation (3), scale (1)). Results showed that low metric accuracy (about ±100-300 cm) could be achieved in this way, as expected. Precise georeferencing requires a set of well-distributed control points measured with a more precise GNSS receiver using differential techniques. Using an Emlid Reach RS2 connected to a network of permanent GNSS available in the area provided centimeter-level precision control points. The integration of such points in the adjustment (beyond the use of a rigid transformation) allows one to partially control network deformations, which cannot be neglected for surveys in large spaces carried out with long frame sequences.
Future work will also consider a combination of the two proposed methods. The use of approximated GNSS parameters using a mobile phone can be used only in external areas. In contrast, the method based on the manually traced trajectory (in GIS) can work for closed spaces, notwithstanding a part of the work remains completely manual. Digital documentation projects connecting external and internal areas (i.e., the interior of a building) can be exploited using a mixed approach. The GIS-based method should also be extended to deal with buildings with several floors, whereas only a trajectory in a plane is available in the actual implementation. A preliminary study of the area can also simplify data acquisition operations. The user can plan a series of closed loops using multiple videos, breaking the acquisition work into more organized datasets, which are then connected in a single photogrammetric project.

REFERENCES

Abate, D., Toschi, I., Sturdy-Colls, C., Remondino, F., 2017. A low-cost panoramic camera for the 3d documentation of contaminated crime scenes. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W8, pp. 1–8.

Aghayaria, S., Saadatsereshta, M., Omidalizarandi, M., Neumannb, I., 2017. Geometric Calibration of Full Spherical Panoramic Ricoh Theta Camera. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 4(1/W1), pp. 237-245.

Barazzetti, L., Previtali, M., Roncoroni, F., 2017. 3D Modelling with the Samsung Gear 360. In: The Int. Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 42(2/W3), pp. 85-90

D’Annibale, E. and Fangi, G., 2009. Interactive modelling by projection of oriented spherical panorama. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 38(5/W1). 6 pages (on CD-ROM).

Fangi, G., 2007. The multi-image spherical panoramas as a tool for architectural survey. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 36(5/C53): 311–316.

Fangi, G., 2009. Further developments of the spherical photogrammetry for cultural heritage. XXII International Committee for Cultural Heritage (CIPA), Kyoto, Japan. 6 pages (on CD-ROM).

Fangi, G., 2017. The book of spherical photogrammetry: Theory and experiences. Edizioni Accademiche Italiane, 300 pages.

Fassi F., Troisi S., Baiocchi V., Del Pizzo S., Giannone F., Barazzetti L., Previtali M., Polari C., Perfetti L., Roncoroni F., 2018. Fisheye Photogrammetry to Survey Narrow Spaces in Architecture and a Hypogea Environment. In Latest Developments in Reality-Based 3D Surveying and Modelling, MDPI Books.

Kwiatek K., Tokarczyk R., 2014. Photogrammetric Applications of Immersive Video Cameras. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 1, pp. 211–218.

Kwiatek, K., Tokarczyk, R., 2015. Immersive Photogrammetry in 3D Modelling. Geomatics and Environmental Engineering, Volume 9, Number 2, pp. 51-62.

Pisa, C., Zeppa, F. and Fangi, G., 2010. Spherical photogrammetry for cultural heritage. Proceeding of the Second Workshop on eHeritage and Digital Art Preservation, Florence, Italy. 3–6.

Matzen, K., Cohen, M. F., Evans, B., Kopf, J., Szeliski, R., 2017. Low-Cost 360 Stereo Photography and Video Capture. In: Journal ACM Transactions on Graphics, Vol.36(4), pp. 148.

Mandelli, A., Fassi, F., Perfetti, L., and Polari, C., 2017. Testing Different Survey Techniques To Model Architectonic Narrow Spaces, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W5, 505-511.

Strecha, C., Zoller, R., Rutishauser, S., Brot, B., Schneider-Zapp, K., Chovancova, V., and Glassey, L., 2015. Quality assessment of 3D reconstruction using fisheye and perspective sensors. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2(3), 215.

Zhao, C. (2021). Creating point clouds, textured meshed models and orthophotos with Weiss AG Civetta - 230 magapixels 360° HDR camera. Technical Report available on the Internet, https://docplayer.net/209971609-Civetta-agisoft-metashape.html, last accessed December 2021, 5 pages.