

# INDOOR MAPPING WITH AN OMNIDIRECTIONAL CAMERA SYSTEM: PERFORMANCE ANALYSIS

Ladai<sup>1,\*</sup>, A., Toth<sup>2</sup>, C., Toth<sup>3</sup>, Z.

<sup>1</sup>Department of Manufacturing and Civil Engineering, Faculty of Engineering, NTNU – Norwegian University of Science and Technology, Gjøvik, 2815, Building B, Teknologiveien 22, Norway - [andras.ladai@ntnu.no](mailto:andras.ladai@ntnu.no)

<sup>2</sup>Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, [toth.2@osu.edu](mailto:toth.2@osu.edu)

<sup>3</sup>Alba Regia Technical Faculty, Óbuda University, [toth.zoltan@amk.uni-obuda.hu](mailto:toth.zoltan@amk.uni-obuda.hu)

## Commission I, WG I/10

**KEY WORDS:** Fisheye Sensors, Sensor Calibration, Point Cloud Generation, Surface Fitting, Performance Evaluation

### ABSTRACT:

The use of BIM (Building Information Modeling), a component of the Digital Twin concept, is on the rise, and the need for indoor data is rapidly growing. BIM information is not only used for management purposes, but it is essential to support navigation indoors. Observing building interiors by optical sensors, such as cameras and laser scanners, has challenges as the image scale changes over a broad range in rooms and floors, and then complete coverage is required, requiring images taken from several locations with various camera orientations. Using 360° imaging sensors partially addresses the need for efficient wide FOV observations. In this study, we investigate the feasibility of using a 6-sensor omnidirectional/fisheye camera system and report about its performance.

## 1. INTRODUCTION

Recently, the proliferation of various sensing technologies has dramatically increased the remote sensing capabilities at large. In particular, inexpensive consumer-grade sensors are deployed in huge and further increasing numbers, providing unprecedented observation capabilities of our environment. Smartphones are used by approximately half of the world population and represent the most powerful mapping device, equipped with a variety of sensors that can support positioning and optical sensing of nearby areas. The combined use of the acquired data provides powerful observation capabilities, and by now Crowdsourcing or VGI (Volunteered Geographic Information) data have become a growing part of geospatial data used professionally as well as in consumer applications. The need for Digital Twin is not new, what is new is that recently technology has reached to the point that such systems are affordable.

The past decade has seen phenomenal developments in sensor technologies, and by now our environment is continuously observed by an ever-growing network of navigation, imaging, mapping and a variety of other sensors (Toth and Jozkow, 2015). Furthermore, sensor integration has become the standard, providing additional benefits, such as increased redundancy, larger coverage, complementarity of various sensor data, georeferencing, etc.

Fisheye lens-based camera systems are used in professional photography and, most recently, increasingly in consumer applications. The ultrawide-angle lens provides wide panoramic or even hemispheric images at the price of significant image warping/distortion introduced during the projection. The excellent coverage makes these cameras attractive in surveillance applications, such as stores, hotels, airports, home security, traffic flow, etc., where object monitoring is the prime objective; see, for example, (Deng et al., 2017) or (Wang et al., 2015).

Fisheye cameras have been used in robotics for a while, see (Courbon et al., 2007) and are getting some attention in the mapping community too; see, for example, (Schneider et al., 2009) and (Alessandri et al., 2019). Nevertheless, since the use of these omnidirectional or 360°-FOV sensors is relatively new, there is interest in the mapping community to further investigate the mapping potential of these inexpensive and easily deployable sensors. In particular, the mobile mapping field can benefit from increased coverage, and the main question is what geometrical accuracy can be achieved under normal operating conditions.

Indoor mapping as well as indoor navigation are popular research topics as the need for these technologies is rapidly growing in, for example, BIM and personal navigation applications. Consumer applications are already pursuing similar tasks, such as robot vacuum cleaners map the floor for internal use as they roam through the rooms of a building. However, in these applications, metric accuracy is not required, as course localization is enough in these operations. Improving the sensor FOV and the georeferencing of a robotic platform can potentially offer an effective indoor mapping capability.

In this study, the feasibility of using the Insta360Pro system is investigated to create point cloud of the environment, which could be subject of reverse engineering of the object space to obtain, for example, the 3D footprint of rooms or offices, and to evaluate its performance limit for indoor mapping application.

Field tests were done in a typical office at NTNU to acquire highly redundant imagery. A strong control network, including 36 accurately surveyed control point was installed in the test site where multiple data acquisition sessions were executed. In addition, the area was surveyed by a terrestrial laser scanner, and that dataset served as a reference for the evaluation. To support this effort the camera system calibration and point cloud generation were primarily done by using commercially available geospatial software tools in order to judge to what extent these tools can be used for fisheye sensor-based workflows.

---

\* Corresponding author

## 2. DATA ACQUISITION

The Insta360Pro system, shown in Fig. 1 has six 200° F2.4 fisheye lenses, each paired with a 4,000x3,000-pixel sensor. Based on our experiences, the actual HFOV is most likely around 150°, as it can be seen on the sample image of Fig. 2, taken by one of the 6 fisheye sensors of the Insta360Pro camera.



**Figure 1.** Insta360Pro system.

The test area was a typical office with furniture. During the data acquisition, 6 fisheye format images were taken at 6 locations with 3 different vertical positions with approximately 30 cm height spacing at each location, totaling in  $6 \times 6 \times 3 = 108$  images. The acquisition points were distributed in skewed regular  $3 \times 2$  grid with an approximately 1 meter spacing. The average image overlap is about 9 or higher; clearly, adequate for multiview image photogrammetric processing.

To provide control points for photogrammetric processing, targets were evenly placed on the four walls of the room, ensuring full vertical distribution, see Fig. 2. The surveying accuracy of the target points is estimated to be better than 1mm; note that this level accuracy is not really needed for this study.

A reference point cloud of the office was obtained by using a Leica P40 TLS. From a single location about 163 million points were acquired, representing an average point density of 1 million  $1/m^2$ . The point cloud of a corner is shown in Fig. 3, clearly demonstrating the high point density. The accuracy of the point cloud is better than 3mm based on manufacturer's specification. The reason of acquiring the reference data from a single station was to ensure that the reference data accuracy was determined only by the instrument's capabilities; i.e., no need for adjustments for registering multiple point clouds.

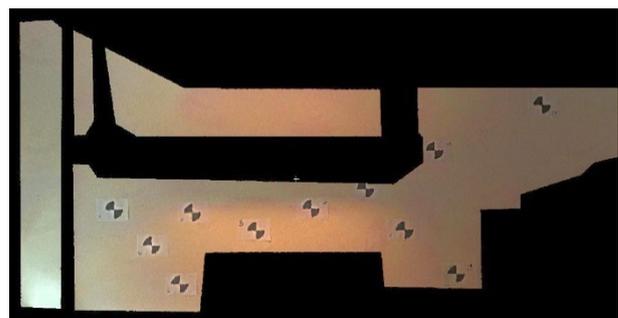
For the subsequent analysis, the non-wall objects are filtered out. Fig. 4 shows one wall with void areas where the non-wall points are removed. The cropping of the upper part of the wall data was determined by the shape of the ceiling.



**Figure 2.** A fisheye image taken at the test site.



**Figure 3.** Reference LiDAR point cloud of a corner area.



**Figure 4.** Non-wall points removed for one wall segment.

### 3. DATA PROCESSING

For point cloud generation, the AgiSoft software was selected, as in the practitioner community, it is known to be quite efficient to handle complex and challenging, even somewhat extreme image configurations. The Insta360Pro system can provide stitched panoramic images, formed from the six camera images. In our study, however, the original camera images were used since generating the panoramic images may introduce unknown distortions in the pictures. Furthermore, there was no system calibration in the sense that the relative orientation of the cameras was neither estimated nor used during the processing.

36 target points were manually measured to achieve the highest accuracy potentially available to support the thorough analysis for object space reconstruction.

There are many calibration tools offering various parametrization of the projection system of fisheye camera. (Mundhenk et al., 2001) provides a simple method for basic calibration adequate for robotics applications. Then, OpenCV and Matlab Toolbox are widely used for more accurate sensor modeling. The second tool implements the fisheye lens calibration method by (Scaramuzza et al., 2006). Finally, photogrammetric calibration represents the potentially most efficient calibration solution. However, it should be noted that calibrating a consumer grade fisheye camera that lacks stability has limitations, and thus, investing into more sophisticated calibration likely provides no real benefits. Therefore, the cameras in this study were calibrated during the triangulation process by allowing for self-calibration.

Since the 6 imaging sensors were individually used in the experiments, using the traditional camera calibration model was sufficient (Remondino and Fraser, 2006). The individual camera calibrations results showed noticeable variations, as expected for mass-produced inexpensive consumer cameras; note that besides the three linear parameters,  $F$ ,  $C_x$ ,  $C_y$ , the  $K1-K3$  and  $P1$ ,  $P2$  parameters were also estimated. The average positioning error estimates in the local mapping frame are listed in Table 1. Note that Z component has the lowest error, which is due to image geometry; the image plane is parallel to the walls.

X [mm]	Y [mm]	Z [mm]	XY [mm]	3D [mm]
5.7	2.2	1.6	6.1	6.3

Table 1. Average camera location error.

The vertical projection of the point cloud is shown in Fig. 5; the camera positions are overlaid, clearly showing the Insta360Pro sensor arrangement.

Using the 36 targets, the complete orientation and point cloud generation process was executed; the first one resulting in a 3-pixel reprojection error reported. Next, the entire processing using less ground control was repeated to assess the impact of ground control on the accuracy of the object space reconstruction, and subsequently, of the point cloud. Unfortunately, the reduction of the controls dramatically decreased the performance of the AgiSoft tool and was not pursued further; also, it is a different problem to address.

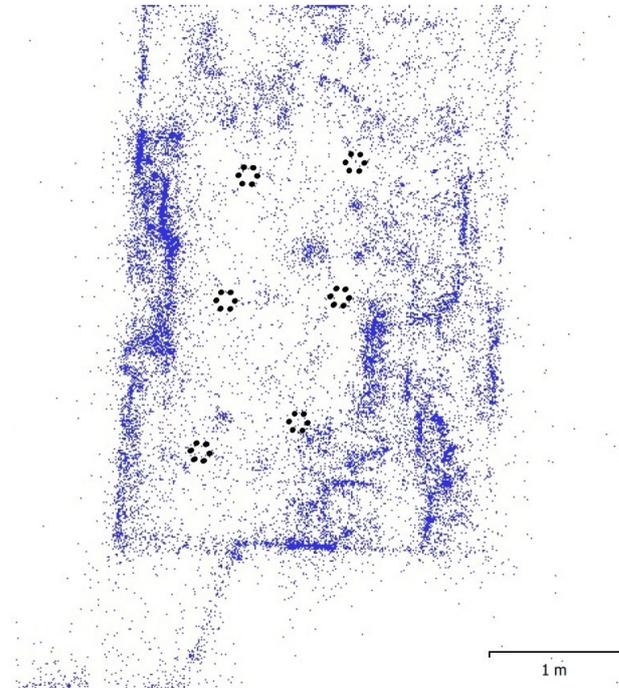


Figure 5. Point cloud in vertical view with camera locations.

### 4. PERFORMANCE EVALUATION

Since the primary objective of this effort is to assess the feasibility of fisheye lens-based camera system for indoor mapping, mainly surveying rooms, only the ceiling, walls and floor areas were used in the performance evaluation. As shown in Fig. 4, points not falling on these surfaces were filtered out. First, a point cloud to point cloud comparison was done, and then the smaller surface segments that could be modeled by a plane were analyzed for fitting.

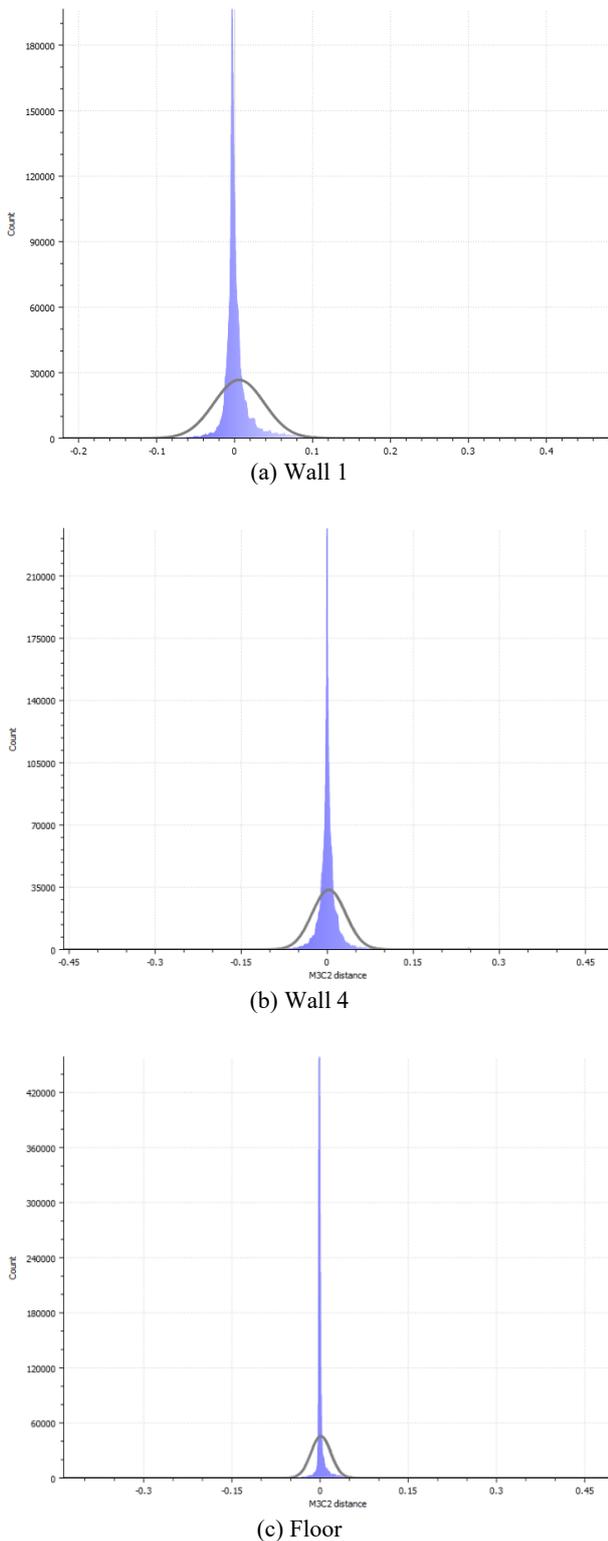
#### 4.1 Comparing Point Clouds

The popular CloudCompare software was chosen for basic point cloud comparison; more specifically, the M3C2 plugin (Lague et al., 2013). The photogrammetrically derived point cloud was compared to the reference point cloud obtained by TLS. Note that there was no validation of the laser data, accuracy was assumed to be provided by the manufacturer's specification. Results of comparison performed on two walls, Wall 1 and Wall 4 as well as the Floor point clouds are reported here.

Figure 6 shows the distribution of differences measured in M3C2 distanced for the three tested areas and Tables 2 list the numerical values.

	Mean [m]	STD [m]
Wall 1	0.01	0.03
Wall 4	0.01	0.03
Floor	0.01	0.02

Table 2. Statistical results of surface matching.



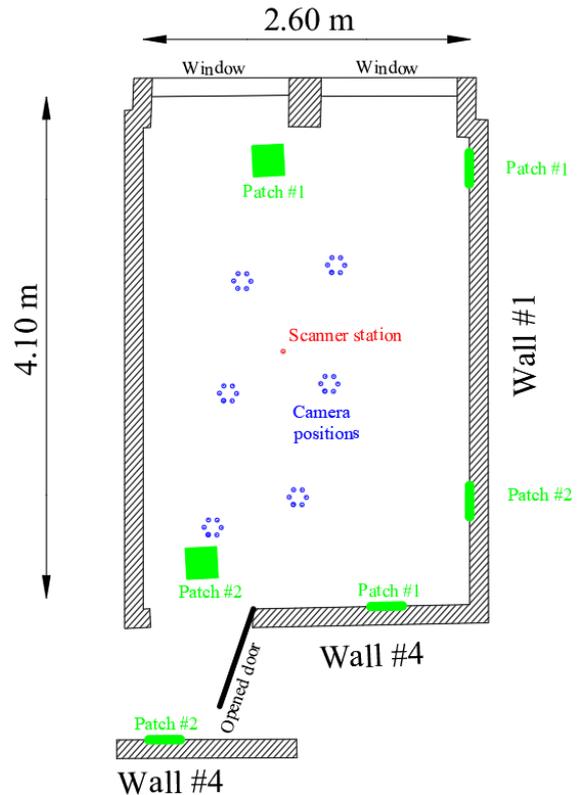
**Figure 6.** Distribution of differences in M3C2 distances (x is given in meters).

The results indicate close matches between the fisheye imagery created point cloud and TLS reference data. Given the different characteristics of two point clouds due to the different sensor types, these results represent clearly excellent performance. This level of accuracy is quite adequate for structural modeling an office interior.

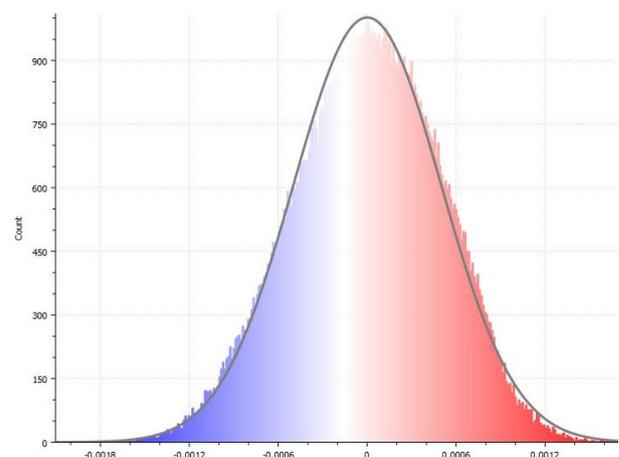
#### 4.2 Comparing to Targets and Reference Surfaces

The targets are easily identifiable in the point clouds and thus can be checked by manual measurements on a 3D photogrammetric workstation. Based on measuring a few points, the 3D measurement accuracy was estimated to be 1-2 cm.

Since man-made objects/structures are typically formed from geometrical primitives, such as planar, cylindrical, spherical, etc., surfaces, plane fitting was done for some selected smaller segments of the point clouds. Areas of approximately 0.25m x 0.25m size were extracted from the Wall 1, Wall 4 and Floor point clouds from both Insta360Pro imagery derived and TLS created point clouds, see Fig. 7.



**Figure 7.** Layout of the surface patches used for evaluation.



**Figure 8.** Typical distribution of residuals of plane fitting to TLS data; Patch 1 of Wall 1 (x is given in meters).

First, we did plane fitting to the reference data, using 2 surface patches from Wall 1, Wall 2 and Floor, respectively. The statistical parameters for the 6 areas showed similar distribution of the residuals; Fig. 8 shows a typical histogram. The 1mm RMSE of the plane fitting residuals indicate that both the planarity of the wall surfaces is quite good and then the TLS data is also very accurate. Note that the size of wall patches is relatively small, and thus there is barely noticeable warping effect.

The statistical analysis showed similar consistent results and as expected larger residuals. Fig. 9 shows a representative histogram for residuals. The statistical results of the plane fitting are listed in Tables 3, 4 and 5. The average 3-4 cm RMSE of residuals are realistic, given the scale, the object space complexity and the quality of the consumer grade imaging sensor. The 10 cm results for Patch 2 of Wall 4 are clearly significantly worse, which can be explained with the longer camera-object distance and the darkness of the area; the patch is situated on the visible part of the corridor outside of the office, which lacks natural lighting. Another interesting area is the Floor Patch 2, which shows significantly lower level of noise than the other areas, see Fig. 10. This can be explained with the good texture and strong contrast of the floor area, ideal for tie point matching. In addition, the area has good lighting conditions compared to the Floor Patch 1, which is in a darker area under the desk at window.

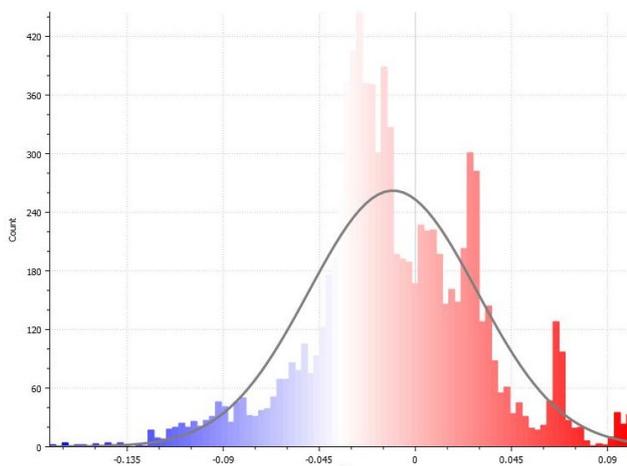


Figure 9. Distribution of residuals for Patch 1 of Wall 4 (x is given in meters).

	Wall #1			
	Patch #1		Patch #2	
	Reference	Insta360P	Reference	Insta360P
No. of points	102981	12795	260388	27171
Mean	0.0000	-0.0007	0.0000	-0.0170
1σ	0.0005	0.0132	0.0006	0.0375
3σ	0.0015	0.0395	0.0017	0.1124
Min	-0.0021	-0.0670	-0.0022	-0.1589
Max	0.0018	0.0662	0.0024	0.1498

Table 3. Statistical results for surface patches on Wall 1.

	Wall #4			
	Patch #1		Patch #2	
	Reference	Insta360P	Reference	Insta360P
No. of points	141104	8910	45041	4189
Mean	0.0000	-0.0108	0.0000	0.0352
1σ	0.0004	0.0390	0.0004	0.0966
3σ	0.0012	0.1170	0.0012	0.2898
Min	-0.0016	-0.1712	-0.0015	-0.1925
Max	0.0016	0.1018	0.0015	0.4755

Table 4. Statistical results for surface patches on Wall 4.

	Floor			
	Patch #1		Patch #2	
	Reference	Insta360P	Reference	Insta360P
No. of points	127971	3263	136278	32251
Mean	0.0000	0.0045	0.0000	0.0012
1σ	0.0004	0.0199	0.0004	0.0022
3σ	0.0012	0.0598	0.0012	0.0066
Min	-0.0023	-0.1221	-0.0016	-0.0027
Max	0.0019	0.0767	0.0018	0.0378

Table 5. Statistical results for surface patches on Floor.



Figure 10. Floor pattern; Floor Patch 2.

## 5. SUMMARY AND CONCLUSION

This paper reports about the experiments conducted with an Insta360Pro omnidirectional/fisheye camera system. 108 images of size of 3K by 4K were acquired at six locations at three different heights. The images covered the entire interior of a typical office, and the average overlap was about nine. For reference, 36 controls points were installed on walls with good spatial distribution; the target points in the image domain were measured at 1-5 mm level object space accuracy.

Using the AgiSoft software, the point cloud of the area was created and then segmented into walls, floor and ceiling datasets. In addition, the non-room-surface points were filtered out and were not subject to statistical evaluation. The analysis included two components: comparing the point clouds to the reference and then investigating the how well the point cloud segments can be fitted to a plane; note that small surface patches were selected for the latter purpose, where the planarity of the object surface could be assumed.

The point cloud comparisons were performed by using the CloudCompare software. The results showed fairly good matches between the Insta360Pro imagery derived point clouds and the reference data. The 2-3 cm average RMSE of the differences in the surface normal direction can be considered good; obviously, as expected, there is a variation of the performance with respect to object space conditions, such as texture, lighting conditions, occlusions, etc.

The plane fitting evaluation included six 0.25m x 0.25m surface patches extracted from two walls and the floor. The fitting results obtained compared well to the point cloud performance. In numbers, most patches producing 2-3 cm RMSE of the residuals. The most extreme result was 10 cm, which is clearly significant, where the imagery acquired by Insta360Pro was of poor quality due to the object space conditions. Obviously, it is not specific and applies to any image-based processing.

In summary, the Insta360Pro system is able to acquire imagery that is adequate for point cloud generation that is suitable for engineering scale mapping. Installing a moving platform, such as indoors robot can form a simple yet effective mobile mapping system. Since robots have navigation capabilities, they can provide approximate georeferencing for the acquired imagery that is essential for processing, as the triangulation process needs good initial approximations. Note that most mapping software is unable to handle a large number of images if no spatial relationship is provided.

## REFERENCES

- AgiSoft <https://www.agisoft.com/>
- Alessandri, L. and Baiocchi, V., Del Pizzo, S. and Rolfo, M. F. and Troisi, S., 2019: 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*, Vol. XLII-2/W9, pp. 25-32.
- CloudCompare <https://www.danielgm.net/cc/>
- Courbon, J., Mezouar, Y., Eckert, L., Martinet, P., 2017: "A generic fisheye camera model for robotic applications," *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2007, pp. 1683-1688, doi: 10.1109/IROS.2007.4399233.
- Deng, L., Yang, M., Qian, Y., Wang C., B. Wang B., 2017: "CNN based semantic segmentation for urban traffic scenes using fisheye camera," *2017 IEEE Intelligent Vehicles Symposium (IV)*, 2017, pp. 231-236, doi: 10.1109/IVS.2017.7995725.
- Ladai, A., Miller, J., 2014: Point Cloud Generation from sUAS-Mounted iPhone Imagery: Performance Analysis. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-1, pp. 201-205, 10.5194/isprsarchives-XL-1-201-2014.
- Lague, D., Brodu, N., Leroux, J., 2013: Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (N-Z), *ISPRS Journal of Photogrammetry and Remote Sensing*, Volume 82, pp. 10-26, ISSN 0924-2716.
- Matlab Computer Vision Toolbox™ <https://www.mathworks.com/products/computer-vision.html>
- Mundhenk, T. & Rivett, Michael & Liao, Xiaoqun & Hall, Ernest., 2001. Techniques for Fisheye Lens Calibration using a Minimal Number of Measurements. *Proceedings of SPIE*.
- OpenCV <https://opencv.org/>
- Remondino, F., Fraser, C., 2006: Digital camera calibration methods: considerations and comparisons, *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XXXVI Part 5, pp. 266-272.
- Scaramuzza, D., Martinelli, A., Siegwart, R., 2006: "A Toolbox for Easy Calibrating Omnidirectional Cameras." *Proceedings to IEEE International Conference on Intelligent Robots and Systems, (IROS)*. Beijing, China, October 7–15, 2006.
- Schneider, D., Schwalbe, E., Maas, H.-G., 2009: Validation of geometric models for fisheye lenses, *ISPRS Journal of Photogrammetry and Remote Sensing*, Volume 64, Issue 3, 2009, Pages 259-266,
- Toth, C., Jozkow, G., 2015: Remote Sensing Platforms and Sensors: A Survey, *ISPRS Journal of Photogrammetry & Remote Sensing*, 115 (2016), pp. 22-36.
- Wang, W., Gee, T., Price J., Qi, H., 2015: "Real Time Multi-vehicle Tracking and Counting at Intersections from a Fisheye Camera," *2015 IEEE Winter Conference on Applications of Computer Vision*, 2015, pp. 17-24, doi: 10.1109/WACV.2015.10.