ASSESSMENT OF A PORTABLE TOF CAMERA AND COMPARISON WITH SMARTPHONE STEREO VISION

A. Masiero[a,*], A. Guarnieri[a], A. Vettore[a]

[a] Interdepartmental Research Center of Geomatics (CIRGEO), University of Padova,
Viale dell'Università 16, Legnaro (PD) 35020, Italy -
masiero@dei.unipd.it
(alberto.guarnieri, antonio.vettore)@unipd.it

Commission II,

KEY WORDS: TOF camera, Depth sensor, Stereo vision, Smartphone

ABSTRACT:

Nowadays time-of-flight (ToF) cameras and multiple RGB cameras are being embedded in an increasing number of high-end smartphones: despite their integration in mobile devices is mostly motivated by photographic applications, their availability can be exploited to enable 3D reconstructions directly on smartphones. Furthermore, even when a ToF camera is not embedded in a smartphone, low cost solutions are available on the market in order to easily provide standard mobile devices with a lightweight and extremely portable ToF camera. This work deals with the assessment of a low cost ToF camera, namely Pico Zense DCAM710, which perfectly fits with the above description. According to the results obtained in the considered tests, the ranging error (precision) of the DCAM710 camera increases linearly approximately up to the nominal maximum range in the considered working mode, up to approximately 1 cm. Despite the device allows to acquire measurements also at larger ranges, the measurement quality significantly worsen. After assessing the main characteristics of such ToF camera, this paper aims at comparing its 3D reconstruction ability with that of a smartphone stereo vision system. In particular, the comparison of a 3D reconstruction obtained with stereo vision from images acquired with an LG G6 shows that the stereo reconstruction leads to a much larger point cloud. However, points generated by the ToF camera are more homogeneously distributed, and they seem to slightly better describe the real geometry of the reconstructed object. The combination of such two technologies, which will be investigated in our future work, can potentially lead to a denser cloud with respect to the ToF camera, while preserving a reasonable accuracy.

1. INTRODUCTION

The availability of low cost depth sensors (such as RGB-D cameras), which has been developed during the last decade mostly motivated by gaming purposes (e.g. Microsoft Kinect), led to the realization of alternative 3D reconstruction methods, in particular for small objects and indoor environments (Jóźków et al., 2014). Differently from traditional laser scanners used in surveying applications, the maximum range of such low cost RGB-D cameras is usually limited to few meters, e.g. 5-10 m. Nevertheless, their usage in indoor environments became quite popular, also eased by the development of simultaneous localization and mapping (SLAM) methods in the robotics and computer vision communities (Whelan et al., 2016, Zollhöfer et al., 2018).

Nowadays, depth sensors, in particular Time-Of-Flight (ToF) cameras, are embedded on several high-end smartphones, hence opening the possibility of using them for smartphone-based 3D reconstructions (which have already been investigated for example by means of "standard" single RGB camera, e.g. (Poiesi et al., 2017, Schöps et al., 2014, Al Hamad, El Sheimy, 2014, Masiero et al., 2016, Fissore et al., 2018)). Furthermore, most of such devices are also provided with multiple cameras, hence stereo vision can also be implemented.

Despite several smartphones are provided with such sensors, currently most of the producers do not allow their complete access to developers. Given such restriction, this work considers the use of a portable ToF camera that can be used as external sensor with a large variety of smartphones. To be more specific, a Pico Zense DCAM710 depth camera is used in this work (Fig. 1). Pico Zense DCAM710 camera is a low cost, small and lightweight device that can simultaneously acquire RGB, IR images and depth information at a 30 frames-per-second frequency, with the maximum resolutions reported in the following table:

information type	resolution
RGB	1920 × 1080
IR	640 × 480
Depth	640 × 480

Table 1. Resolution of the information provided as outputs by Pico Zense DCAM710 camera.

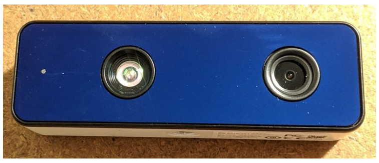


Figure 1. Pico Zense DCAM710.

The first aim of this paper is that of assessing the characteristics (e.g. ranging quality, statistical behavior) of the depth (and 3D) information provided by the Pico Zense DCAM710 camera. Such kind of assessment, which is similar in terms of tests and results to that of the Microsoft Kinect v2 provided in (Lachat et al., 2015a, Lachat et al., 2015b), will be presented in

[*] Corresponding author.

https://doi.org/10.5194/isprs-archives-XLII-2-W17-187-2019

Section 2.

Then, Pico Zense DCAM710 3D reconstruction results will be compared in Section 3 with those provided by smartphone stereo vision in a case study. To such aim, a specific Android application has been implemented ad hoc in order to (almost) simultaneously acquire images from the two rear cameras of a smartphone LG G6 (Fig. 2). The characteristics of such two rear cameras (a "standard" and a "wide-angle" camera) of the LG G6 are reported in Table 2. The reader is referred to (Masiero et al., 2019) for a more detailed description of the approach used for stereo vision reconstruction with the images acquired with the dual camera of smartphone LG G6.

Figure 2. LG G6.

sensor resolution	4160 pix × 3120 pix
pixel physical side size	1.12 μm
standard camera focal length	4.03 mm
wide-angle camera focal length	2.01 mm
baseline between cameras	≈ 1.8 mm

Table 2. LG G6 characteristics)

It is worth to notice that, despite the smartphone LG G6 provides images from both the cameras at the same resolution (12 Mpixels), only part of such pixels can be used for stereo vision. Indeed, since the two cameras have quite different focal lengths, the overlapping area is just one fourth of the image provided by the wide-angle camera, approximately.

Despite such loss of resolution, smartphone stereo vision still typically ensures a higher resolution with respect to the considered low cost ToF camera. Hence, the rationale driving the considered comparison is that of a potential future integration between such two 3D reconstruction technologies in order to obtain reconstructions at higher resolution with respect to the ToF camera, but with a geometric accuracy comparable with that ensured by such camera. Consequently, the ToF camera assessment and the comparison are done in the range of depths that can be interest for such combination. In practice, given the short baseline between the two cameras of the LG G6, reasonable 3D reconstruction results are possible only at quite low distances, e.g less than 1.5 m, approximately (Masiero et al., 2019).

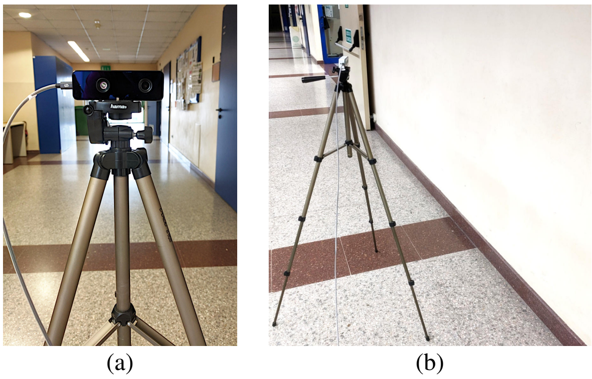
2. CHARACTERIZATION OF THE DEPTH CAMERA PICOZENSE DCAM710

Several range setting modes can be distinguished in the Pico Zense DCAM710, depending on the current interval of ranges of interest to be measured. Since this work aims at comparing the Pico Zense DCAM710 results with those of smartphone stereo vision, with the future goal of integrating them in the future, the ToF camera characterization presented in this section performed using the "Range0" setting in the Pico Zense DCAM710, which corresponds to setting the depth interval of interest between 35 cm and 150 cm.

It is worth to notice that this section presents an assessment of the Pico Zense characteristics when working in static conditions: the ToF camera was mounted on a tripod (Fig. 3(a)), acquiring mostly RGB-D images of a flat wall (Fig. 3(b)).

Furthermore, image undistortion and point cloud generation utilities provided in the Pico Zense SDK were used in order to generate the results shown in this paper.

Detailed investigations of the ToF camera characteristics are provided in the next subsections.



(a) (b)

Figure 3. Acquisitions with Pico Zense DCAM710 on a tripod.

2.1 Consistency over long time intervals

First, the behavior of the Pico Zense DCAM710 is investigated over a quite long time interval (one hour). To this aim, the ToF camera was positioned on a tripod at approximately 120 cm from a internal wall, which is assumed to well approximate a planar surface, of a building of the Uniersity of Padua. Data collection lasted for one hour from the starting time (i.e. few instants after the device was activated). In particular, a sequence of eleven RGB-D images was acquired every five minutes. Images in the same sequence were acquired at a sample frequency of 0.2 s.

Data were acquired without applying any kind of time or spatial filtering, whose effect are instead investigated in subsection 2.2.

First, the standard deviation σ_{depth} is computed (over time) for each pixel in the sensor and for each of the considered time sequences. Then, for each time sequence, the values obtained on all the sensor are averaged, obtaining the average σ_{depth} shown in Fig. 4. To be more specific, Fig. 4 shows that the computed sample standard deviation has very small variations among the considered time interval (at sub-millimeter level).

Then, Fig. 5(a) shows the depth values (dashed line) on the sensor center collected during the above described procedure. Values related to different time sequences are plotted with different colors. Furthermore, the average depth value evaluated for each of such time sequences is reported as a bold solid line.

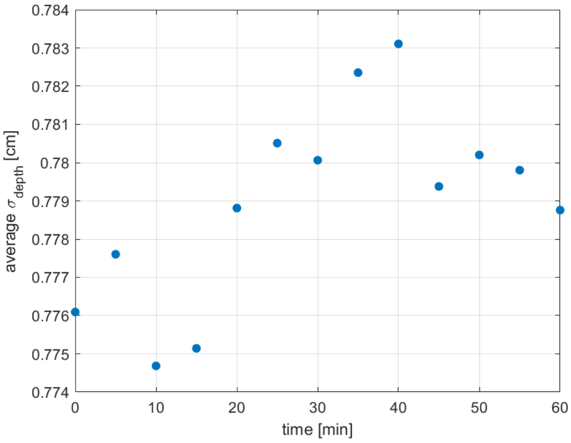
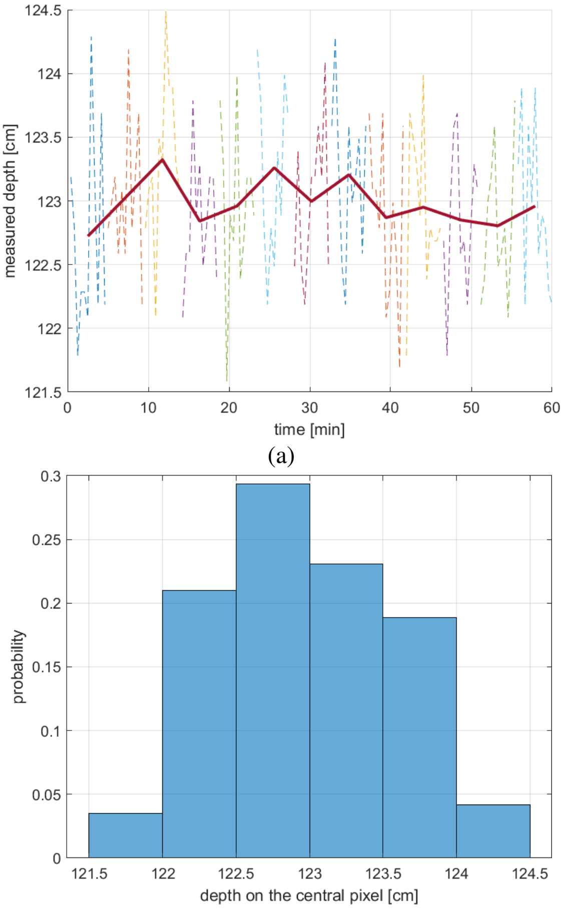


Figure 4. Values of the average σ_{depth} separately evaluated on different time sequences collected during a one-hour acquisition.



(a)

(b)

Figure 5. (a) Values of the depths on the sensor center over one hour (dashed line). Average of the depths collected in different sequences is shown in bold solid line. (b) Distribution of the values of the depths on the sensor center shown in (a).

The variation among such averaged values apparently tends to converge approximately after 40 minutes.

Fig. 5(b) shows the distribution of the depth values for the sensor center (already depicted in Fig. 5(a) as well). The obtained distribution is quite symmetric with an only quite rough Gaussian aspect.

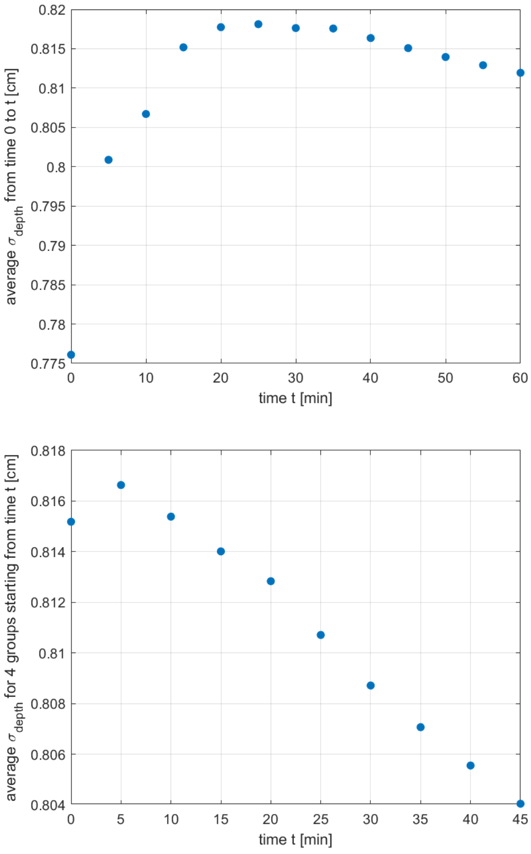


Figure 6. Average σ_{depth} obtained averaging acquired depths on consecutive time instants. (a) Standard deviation computed considering measurements acquired from time 0 to t. (b) Standard deviation computed considering measurements from a 16-minute fixed-horizon starting from each of the considered time t.

Then, the effect of long time averaging is assessed in Fig. 6. Fig. 6(a) shows the average σ_{depth}, where in this case the σ_{depth} standard deviation was computed by averaging samples corresponding to the same pixel over the data collected from instant at time 0 to the considered time t.

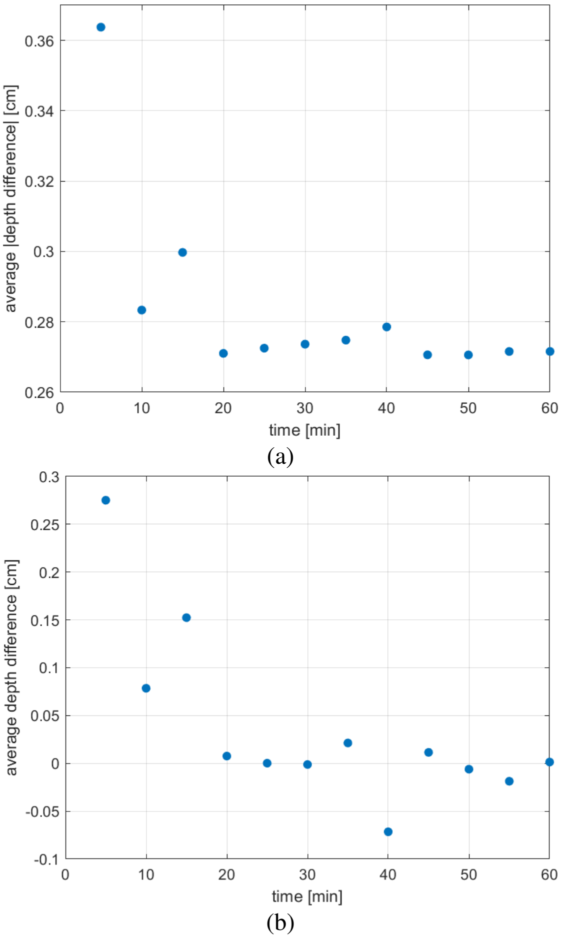
As shown in Fig. 6(a) long time averaging oif the depths do not lead to any advantage in terms of quality of the obtained values: a systematic depth variation shall influence the outcomes of such computation implying that the lowest standard deviation can be obtained without averaging on successive time sequences. Nevertheless, the slight decrease in the last part at the right of Fig. 6(a) shows that such systematic variation shall become smaller (and/or more stable) in the last part of the acquired dataset.

The latter is also confirmed by Fig. 6(b): in this case depths were averaged on the images acquired in four consecutive time sequences (e.g. during the 16 minutes started after each of the time instants shown in the figure). It is quite clear that such “fixed-horizon” standard deviation starts decreasing after the 5 minutes sample. This observation confirms that the systematic variation shall become smaller with the increase of the working time of the ToF device.

Finally, Fig. 7(a) shows the average of the absolute value of the differences between depths measured on the same pixel but on two consecutive time sequences. This graph confirms that after a while the behavior of the sensor tends to stabilize.

Fig. 7(b) confirms also that after an initial time interval, the depth variations between successive time sequences tends to be approximately zero-mean.

Given the presented results, it is quite apparent that averaging the acquired depths over quite long intervals do not enhance the quality of the obtained results because of a systematic behavior of the considered sensor.



(a)

(b)

Figure 7. (a) Average of the absolute values of depth differences computed on successive time sequences. (b) Average of the values of depth differences computed on successive time sequences.

2.2 Time and spatial filtering

This subsection aims at testing the performance obtained with the options available by default in the considered devices: time and spatial averaging. It is worth to notice that, differently from the previous section, time averaging here is applied only on samples collected very close (in time) to each other.

Similarly, spatial averaging is applied on a small neighborhood of each considered point.

Table 3 compares the results obtained while applying time, spatial and both time and spatial filtering to the acquired data with those obtained without using any filter.

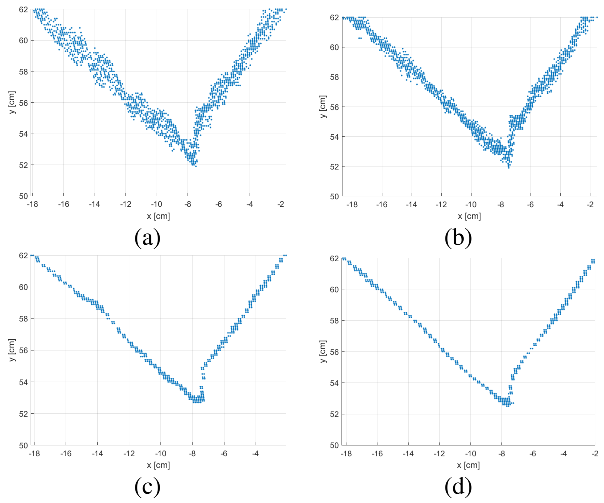
In particular, Table 3 presents the results with respect to three different criteria: average σ_{depth} (defined as previously), the average root mean square error obtained fitting the acquired point cloud with a planar surface, and, finally, the average sample standard deviation of the depth values on 3×3 spatial neighborhoods. The latter clearly aims at evaluating the spatial regularity of the measured depth values, which is should to be good because of the high regularity of the measured object (i.e. a planar surface).

All such statistics were computed considering 20 different RGB-D acquisitions and with the ToF camera positioned at approximately 120 cm from a wall.

Table 3. Performance comparison using time and spatial filtering

filtering	**avg** σ_{depth}	**avg RMSE**	**avg** σ_{neigh}
none	7.8 mm	9.4 mm	4.9 mm
time	3.4 mm	7.0 mm	3.6 mm
spatial	2.9 mm	5.4 mm	1.1 mm
time and spatial	1.1 mm	4.9 mm	1.0 mm

The effect of time and spatial filtering can also be seen in Fig. 8, where the points on a 2 cm-height slice acquired using the different acquisition settings are compared.



(a)

(b)

(c)

(d)

Figure 8. Top views of 2 cm-height slices of a point cloud acquired with PicoZense DCAM710 on the corner between two sides of a closet. Comparison between the different operative modes of the TOF camera: (a) standard, (b) time averaging, (c) spatial averaging, (d) time and spatial averaging.

2.3 Assessment of the precision varying the measurement distance

This subsection deals with the assessment of the effect of the distance to the measured object on the quality of the obtained data.

20 RGB-D samples were collected, at varying distances from the object, for each of the cases considered in this subsection.

Fig. 9 compares the top views of three point clouds acquired from different distances to the wall. It is worth to notice that the measurement uncertainty (which can be evaluated by checking the "thickness" of the sets obtained by projecting the points along the vertical direction) increases with the distance. Furthermore, an apparent distortion effect and presence of certain

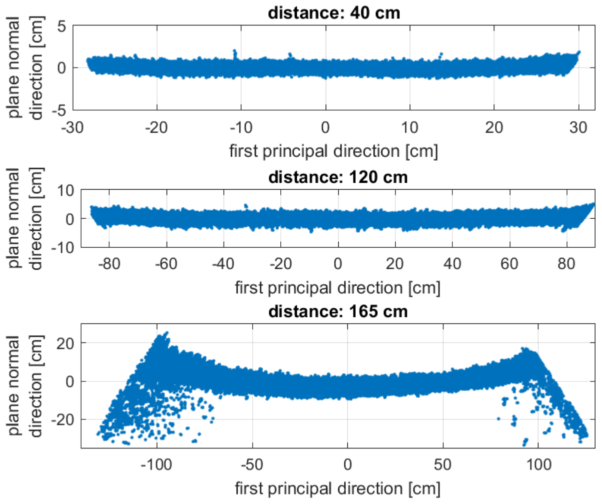


Figure 9. Fitting errors of PicoZense DCAM710 point acquisitions of a planar surface: comparison varying the distance from the plane.

outliers at the borders of the point cloud are apparent on the bottom graph of Fig. 9 (distance≈165 cm).

Fig 10 shows that both the average σ_{depth} and the RMSE obtained fitting a plane on the obtained point cloud increase linearly with the distance when working within the maximum range (150 cm) of the used Pico Zense operative mode.

Differently, when working at distances larger than the maximum range (150 cm) of the used Pico Zense (e.g. distance≈165 cm in Fig.10) the errors increase at a much higher rate (two points on the right of Fig.10).

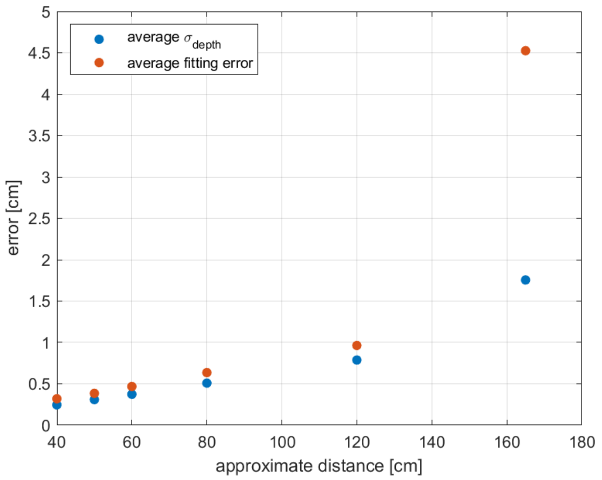


Figure 10. Fitting error and average standard deviation of the depth measurements for acquisitions at different distances from the plane.

2.4 Assessment of the precision varying the incident angle

Finally, the effect of the incident angle value on the sample standard deviation of the measured depth is evaluated on Fig. 11. The precision is evaluated just on the sensor center and while averaging the results from 20 time samples. Despite the error is quite independent of the incident angle value, for relatively small values, it becomes significantly larger for high values of the angle.

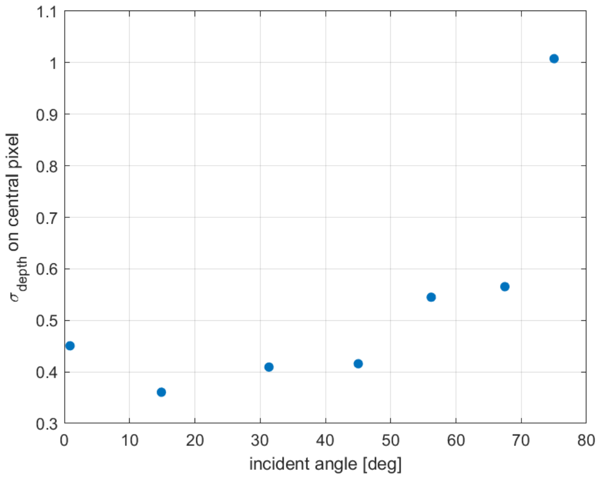
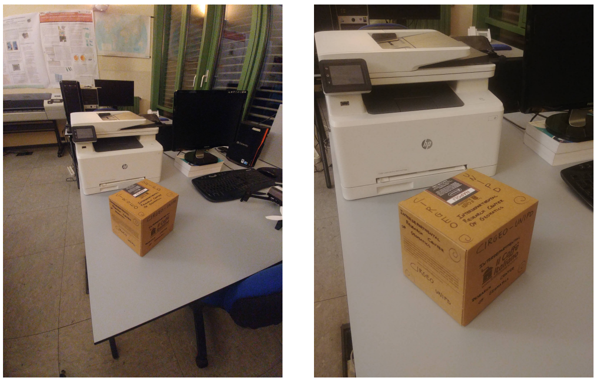


Figure 11. Standard deviation of the depth measurements on the center of the depth sensor varying the value of the incident angle.

3. COMPARISON ON 3D RECONSTRUCTION

In this section the 3D reconstruction results obtained with the Pico Zense ToF camera and with the LG G6 stereo vision are compared. In particular, the reconstruction of a (17 cm× 15 cm × 16 cm) box on a table is used as case study here. Both the ToF and the RGB cameras acquired from a distance of approximately 50 cm from the object.



(a) (b)

Figure 12. Images acquired with the LG G6 dual camera.

Fig. 12 shows two images acquired with the smartphone LG G6 of the object of interest.

Fig. 13 shows for instance the 3D information concerning the box provided by Pico Zense.

Finally, the number of points in the obtained point cloud, the RMSE of the fitting plane, and the angles between the three detected box planes are reported in Table 3.

4. DISCUSSION

Tests for evaluating the measurement consistency on long time intervals proved the presence of a certain systematic effect causing the variation of the measurements with time. Similarly to the Microsoft Kinect v2 case (Lachat et al., 2015a), such variation can be a direct consequence of the device heating process.

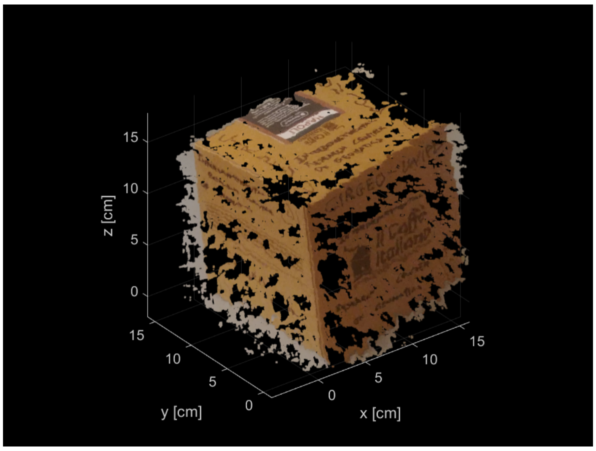


Figure 13. Box reconstructed with stereo vision from images acquired with LG G6 dual camera.

Table 4. Comparison between TOF camera and smartphone stereo vision on the reconstruction of a box

	DCAM710 ToF cam	LG G6 Stereo
number of points	31.7 k	387.7k
RMSE fit plane x	0.41 cm	0.46 cm
RMSE fit plane y	0.41 cm	0.36 cm
RMSE fit plane z	0.29 cm	0.46 cm
angle planes x-y	87.4 deg	81.1 deg
angle planes x-z	95.1 deg	86.4 deg
angle planes y-z	89.9 deg	89.1 deg

Such variation, which makes useless the averaging of measurements over a long time period, decreases after some tens of minutes, leading to a more stable performance of the system.

Table 3 and Fig. 8 compared the effects of the different settings for what concerns the available working modes on Pico Zense DCAM710. On the one hand, time averaging over a short time period can be useful to slightly reduce the measurement error, however this can probably make sense only for static devices, which in certain cases can be a too stringent operative requirement (e.g. mobile mapping). On the other hand, spatial averaging significantly reduces the noise, and it can be directly applied to a single RGB-D image, hence not requiring time filtering. Despite spatial filtering is expected to potentially reduce the spatial resolution of the acquired object, this is not significantly visible in the acquired dataset.

Instead, it is worth to notice that the corner depicted in Fig. 8 corresponds to a $\pi/2$ corner, whereas some artifacts are present in Fig. 8.

Furthermore, the obtained results shows a linear dependence of the error with respect to the distance from the object (up to the maximum tolerable distance according to the selected operative mode), and a significant increase of the error when the incident angle is quite large.

The 3D reconstructions of the object of interest obtained with the two considered methods led to the generation of point clouds, characterized by a quite different cardinality: the stereo vision point cloud typically has a quite higher cardinality. However, such point cloud is often prone to gaps in the reconstruction (13). Instead the point distribution in the ToF camera case is much more homogeneous.

The statistics reported in Table 4 show that the Pico Zense ToF camera allowed to obtain, in the considered case study, a 3D reconstruction slightly more geometrically consistent with respect to the real 3D object.

5. CONCLUSIONS

This paper assessed the characteristics of a low cost, lightweight and easily portable ToF camera, the Pico Zense DCAM710. Such camera can interestingly be integrated with devices running most of the operative system of major interest nowadays, e.g. Android, Windows, Linux. Furthermore, given the exceptional portability, such ToF camera can be an effective imaging/mapping solution also for drones aiming at flying at low heights.

The conducted tests showed an error linearly increasing with the distance from the camera, and a significant increase of the error when the incident angle is quite large.

Differently from the stereo vision point cloud, whose cardinality was larger, but the lack of texture in the measured object caused the presence of gaps clearly visible in the reconstruction, the cloud generated by the low cost ToF camera seems to be geometrically more reliable.

Finally, since the resolution of the depth image provided by the considered low cost ToF camera is quite lower than the potential resolution of the smartphone stereo vision reconstruction, the integration between such two technologies, which will be considered in our future work, can potentially lead to an increase of both reliability and resolution of the produced 3D point cloud.

ACKNOWLEDGMENTS

The authors acknowledge Marco Santoni and GRS s.r.l. for their financial support during the development of this project.

REFERENCES

Al Hamad, A., El Sheimy, N., 2014. Smartphone based mobile mapping systems. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XL-5, 29–34. doi.org/10.5194/isprsarchives-XL-5-29-2014, 2014.

Fissore, F., Masiero, A., Piragnolo, M., Pirotti, F., Guarnieri, A., Vettore, A., 2018. Towards surveying with a smartphone. *New Advanced GNSS and 3D Spatial Techniques*, Springer, 167–176.

Jóźków, G., Toth, C., Koppanyi, Z., Grejner-Brzezinska, D., 2014. Combined matching of 2d and 3d kinect data to support indoor mapping and navigation. *Proceedings of Annual Conference of American Society for Photogrammetry and Remote Sensing*.

Lachat, E., Macher, H., Landes, T., Grussenmeyer, P., 2015a. Assessment and calibration of a RGB-D camera (Kinect v2 Sensor) towards a potential use for close-range 3D modeling. *Remote Sensing*, 7(10), 13070–13097.

Lachat, E., Macher, H., Mittet, M., Landes, T., Grussenmeyer, P., 2015b. First experiences with Kinect v2 sensor for close range 3D modelling. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XL-5/W4, 93–100. doi.org/10.5194/isprsarchives-XL-5-W4-93-2015, 2015.

Masiero, A., Fissore, F., Pirotti, F., Guarnieri, A., Vettore, A., 2016. Toward the use of smartphones for mobile mapping. *Geo-Spatial Information Science*, 19(3), 210–221.

Masiero, A., Tucci, G., Conti, A., Fiorini, L., Vettore, A., 2019. Initial evaluation of the potential of smartphone stereovision in museum visits. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W11, 837–842. doi.org/10.5194/isprs-archives-XLII-2-W11-837-2019.

Poiesi, F., Locher, A., Chippendale, P., Nocerino, E., Remondino, F., Van Gool, L., 2017. Cloud-based collaborative 3d reconstruction using smartphones. *Proceedings of the 14th European Conference on Visual Media Production (CVMP 2017)*, ACM, 1.

Schöps, T., Engel, J., Cremers, D., 2014. Semi-dense visual odometry for ar on a smartphone. *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on*, IEEE, 145–150.

Whelan, T., Salas-Moreno, R. F., Glocker, B., Davison, A. J., Leutenegger, S., 2016. ElasticFusion: Real-time dense SLAM and light source estimation. *The International Journal of Robotics Research*, 35(14), 1697–1716.

Zollhöfer, M., Stotko, P., Görlitz, A., Theobalt, C., Nießner, M., Klein, R., Kolb, A., 2018. State of the art on 3d reconstruction with rgb-d cameras. *Computer Graphics Forum*, 37number 2, Wiley Online Library, 625–652.