INITIAL EVALUATION OF 3D RECONSTRUCTION OF CLOSE OBJECTS WITH SMARTPHONE STEREO VISION

A. Masiero[a, *], F. Fissore[a], M. Piragnolo[a], A. Guarnieri[a], F. Pirotti[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
(francesca.fissore, marco.piragnolo, alberto.guarnieri, francesco.pirotti, antonio.vettore)@unipd.it

Commission I, WG I/7

KEY WORDS: Stereo-Vision, Smartphone, Depth sensor, Indoor Mapping, Mobile Mapping, Photogrammetry

ABSTRACT:

The Worldwide spread of relatively low cost mobile devices embedded with dual rear cameras enables the possibility of exploiting smartphone stereo vision for producing 3D models. Despite such idea is quite attractive, the small baseline between the two cameras restricts the depth discrimination ability of this kind of stereo vision systems. This paper presents the results obtained with a smartphone stereo vision system by using two rear cameras with different focal length: this operating condition clearly reduces the matchable area. Nevertheless, 3D reconstruction is still possible and the obtained results are evaluated for several camera-object distances.

1. INTRODUCTION

The fast development of imaging and positioning sensors enabled the realization of several Mobile Mapping Systems (MMSs), i.e. systems able to efficiently acquire three dimensional data using moving sensors (Schwarz and El-Sheimy, 2004, Puente et al., 2013). Despite most of the mobile mapping systems were based on the use of terrestrial vehicles, actually nowadays a number of different solutions are available, e.g. aerial, marine platforms, and human-carried equipment (Nex and Remondino, 2014, Lo et al., 2015).

Thanks to the availability of such systems and to the spread of several location based services, the worldwide quest for accurate and ubiquitous 3D models of the reality is continuously increasing, making the fast collection of geospatial data a must for a wide spread of applications based on the use of geographical information. Thanks to their ability to quickly collect huge amount of data and efficiently surveying quite large areas (Tao and Li, 2007, Piras et al., 2008), mobile mapping systems represent an ideal solution to such quest for accurate 3D models generation.

Actually, the large demand of location based services related to places inside of buildings and human constructions is significantly increasing the request for indoor mapping and navigation. Differently from the outdoor case, the typical dimensions of rooms and corridors in indoor environments are quite narrow, hence this makes indoor mapping and navigation very difficult with the previously mentioned solutions developed for the outdoors case.

The necessity for very portable devices and the increasing computational power and availability of sensors already embedded in the device are making smartphones very attractive solutions for such kind of applications. Nowadays, standard smartphones are provided of several sensors such as accelerometer, gyroscope, magnetometer, barometer, WiFi and Bluetooth receivers. Several techniques of information fusion of measurements provided by such sensors have already been investigated by a number of authors (Zhuang et al., 2016, Yu et al., 2017), showing that their combination can enable indoor positioning with a reasonable accuracy (even if the error is typically higher with respect to outdoors performance of high grade GNSS receivers).

Furthermore, as already considered in several works in the literature (Nocerino et al., 2017), the standard smartphone camera can be conveniently used as an imaging sensor for obtaining 3D photogrammetric models of the area of interest. However, certain external information should be added in order to obtain a metric reconstruction (Alsubaie et al., 2017, Masiero et al., 2018). Actually, a double rear camera is embedded in most of the recent smartphones, thus potentially enabling stereo-vision 3D reconstruction. Despite stereo-vision might potentially be implemented on a wide variety of smartphones, in practice most of the producers have limited the dual camera access to developers, hence stereo-vision in most of the cases can be implemented only by the producers themselves.

Given the above limitations, this paper aims at assessing the results of stereo-vision reconstruction obtained by using an LG G6 smartphone. Indeed, to the best of our knowledge, LG is the only smartphone producer that is currently allowing developers to fully access to the dual camera. In fact, since the two cameras are very close to each other, 3D reconstruction can be done only for quite close objects. Furthermore, since the two cameras embedded in the LG G6 have a quite different focal length, pixel matching between the two images cannot be done exploiting all the sensor pixels. A clear benefit of using such dual camera stereo-vision instead of single camera photogrammetry is clearly the possibility of 3D metric reconstruction without the need of any external information.

During the recent years several depth sensors have been develeped, which can be used for indoor mapping and object 3D modelling. Among them, a quite famous example is the Microsoft Kinect depth sensor, which has been already widely used in literature for indoor mapping (Khoshelham and Elberink, 2012).

This aim of this paper is that of assessing the stereo-vision 3D reconstruction ability of a smartphone, in particular taking into account also of the expected performance of other relatively low cost systems such as the Kinect sensor (Toth et al., 2012).

[*] Corresponding author.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLII-1-289-2018 | © Authors 2018. CC BY 4.0 License.

2. SMARTPHONE BASED PHOTOGRAMMETRY

Photogrammetry has been widely used in the recent past for producing 3D representations of reality. In particular, it has become extremely popular after the development of several commercial software based on the implementation of the Structure from Motion (SfM) approach. Indeed, the typical easiness of use of such kind software is enabling the production of photogrammetric 3D models to not so specialized persons as well.

Given the large spread of smartphones, tablets and similar mobile devices embedded with cameras, it is quite clear that the usage of such kind of devices for producing photo-based 3D models (Al Hamad and El Sheimy, 2014, Micheletti et al., 2015) and for

However, photo-based 3D models produced by single camera systems typically requires external information in order to be properly scaled, i.e. to become a metric reconstruction. Control points and GNSS measurements are often used for such purpose (and for georeferencing the 3D model as well). Despite this procedure is widely used and it is well known to ensure reliable and accurate results, the use of GNSS and control points is not always possible, e.g. indoors, in GNSS denied environments (Tucci et al., 2018, Dabove et al., 2018). Scale estimation for smartphone-based photogrammetric reconstructions has been investigated also by exploiting inertial sensor information (Mustaniemi et al., 2017, Ham et al., 2014, Alsubaie et al., 2017)), and, actually, the availability of a reliable navigation system may enable also direct georeferencing-like solutions (Chiang et al., 2017, Pfeifer et al., 2012, Lo et al., 2015, Masiero et al., 2017)).

The usage of stereo-vision is often considered in order to develop vision based stand-alone metric reconstruction systems. Given the recent spread of a quite large amount of dual rear camera smartphones it is quite clear that such solution shall be considered as well (Konrad and Padmanaban, 2017). However, to the best of our knowledge, until now most of the smarphone producers did not provided complete access to the dual cameras of their devices, hence making hard the development of smartphone-based stereo vision systems.

The only exception to the above consideration is LG, which provided the access to certain of their dual camera phones. This motivated the usage of the smartphone LG G6 in this work.

3. SMARTPHONE LG G6

The stereo-vision system developed in this work is based on the use of smartphone LG G6, shown in Fig. 1. The main characteristics of interest for this work of such smartphone are summarized in the following table:

Table 1. LG G6 characteristics)

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	$\approx$ 1.8 mm

LG G6 is provided of a “standard” and a wide-angle camera, with quite different focal lengths, as shown in Table 1 . Since the overlapping area of the two images obtained by such two cameras is quite smaller than the original image size, it is clear that having such different focal lengths have the practical consequence of reducing the number of pixels available for the reconstruction: pixel matching is practically possible only in the central area of the wide-angle image (see Fig. 2 and 3).

(a) (b)

Figure 1. Smartphone LG G6: front (a) and rear view (b).

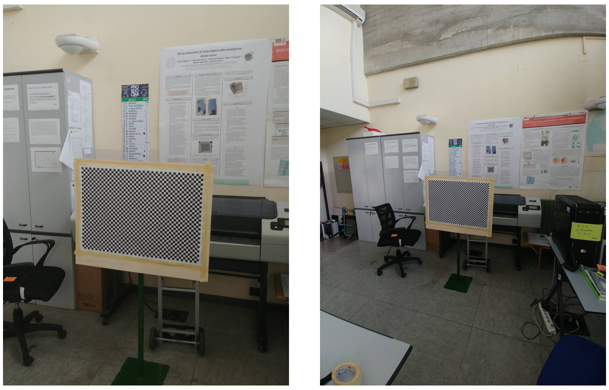
(a) (b)

Figure 2. Sample image acquired by main LG G6 camera (a) and wide-angle one (b).

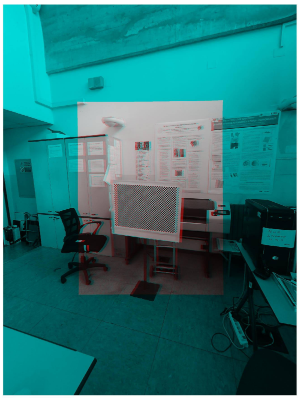

Figure 3. Overlapping of images in Fig. 2 after rectification and normalization.

Furthermore, focus distance should be constant in order to obtain best reconstruction results. Unfortunately, the limited control on camera focus distance on smartphones, and the presence of objects close to the cameras but at relatively quite different distances clearly reduces the system potential reconstruction accuracy.

4. STEREO-VISION

The following standard steps have been considered in order to obtain stereo-vision based 3D reconstruction.

- *Stereo-camera calibration.* First, the stereo camera system has been calibrated in order to estimate internal camera parameters (and in particular lens distortion) and

- *Image acquisition.* An ad hoc application has been developed in order to approximately simultaneously acquire an image from each of the two rear cameras. Comparing time stamps obtained by the Android operating system just after each image acquisition, the acquired images are usually synchronized up to few centiseconds. During such time interval the smartphone is assumed to be rigidly held, i.e. approximately at a constant position. In practice, it is clear that this assumption holds only approximatively when the smartphone is hand-held by a human operator. The reconstruction performance is obviously negatively affected by any camera movement during the two image acquisitions: an in-depth investigation on the effects of the stereo-camera time synchronization accuracy and of the hand stability on the obtained reconstruction results will be object of our future investigation.

- *Image rectification.* Images acquired by the cameras are rectified and normalized in order to (i) correct lens distortion, (ii) obtain similar image views (e.g. compensate for the different focal lengths and orientation views), (iii) ease pixel matching between the two images (thanks to the generation of normalized images) (Förstner and Wrobel, 2016).

- *Disparity map.*: Given two normalized images, a disparity map is obtained by means of pixel matching between the two images. In this work pixel matching is considered only for areas characterized by a high intensity gradient in order to ensure quite reliable matching results. Points characterized by a high intensity gradient are obtained by determining zero-crossings of the image values after filtering them with a Laplacian of Gaussian (LoG) filter (Lindeberg, 2015). Once such points have been computed, matches are obtained by means of local template matching. In this work template matching is performed at pixel level, however sub-pixel template matching (e.g. (Brunelli, 2009, Korman et al., 2013)) should be considered in order to improve the system performance for what concerns the 3D reconstruction accuracy. Different matching strategies might be considered in order to provide reliable dense matches (Kolmogorov and Zabih, 2002). Semi-global matching can also be considered for the computation of dense pixel matching (Hirschmuller, 2005, Hirschmuller, 2008).

- *Depth estimation.* Depth estimation from normalized image pairs with already matched pixels is a standard problem. Depth is estimated as usual by means of triangulation, which, in the stereo-vision scenario with normalized images, leads to a particularly easy formulation: let M be a real point and m_1, m_2 its coordinates on the normalized images. Furthermore, let also f be the focal length and b the baseline distance between the two optical centers, then the depth z_M of point M can be easily computed as a function of the disparity $d_m = m_1 - m_2$:

$$z_M = \frac{f\,b}{d_m} \tag{1}$$

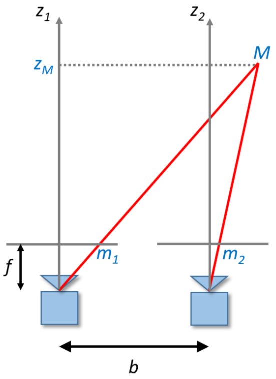

Figure 4. Stereo-vision with normalized images.

Given the depth computation expressed as in (1) and the point matching resolution, done at pixel level in this work, it is clear that the quite small baseline between the two images is a severe limiting factor in the depth estimation. This aspect is analyzed in Fig. 5 , where depth is shown as a function of the disparity: being the matching done at pixel level, only integer values of the disparity have been considered.

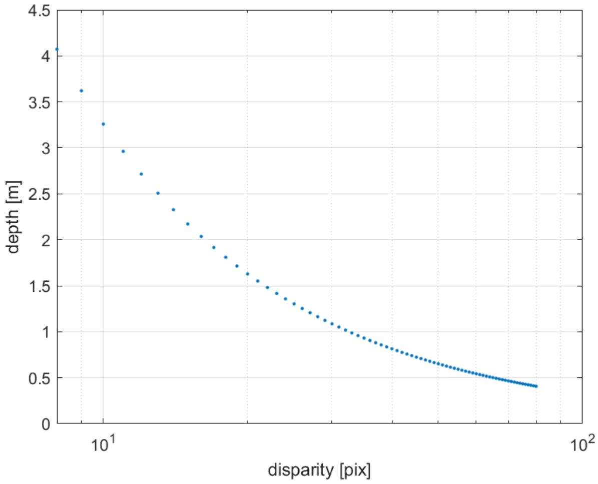

Figure 5. Estimated depth as a function of the disparity computed at pixel resolution level.

As shown in Fig. 5 , the level of resolution of the depth estimation decreases significantly for objects not so close to the camera: this imposes a stringent closeness requirement on the distance of the object to be reconstructed.

5. RESULTS

Similarly to (Khoshelham and Elberink, 2012) the reconstruction of a planar surface is used as a test for providing a first evaluation of the reconstruction ability of the system.

In particular, a set of seven stereo images of the surface have been collected, varying distance and observation angle. Then a planar surface has been fitted to the points representing the reconstructed plane.

For each reconstructed point the distance from the fitting plane has been computed and the following statistics have been evaluated for each image: root mean square (RMS) distance from the plane, maximum and average error.

Fig. 6 shows the computed statistics as functions of the depth value (average depth value of the planar surface has been computed for each image).

Similarly, Fig. 7 shows the computed statistics as functions of the angle between the normal of the planar surface and the plane orthogonal to the camera observation direction.

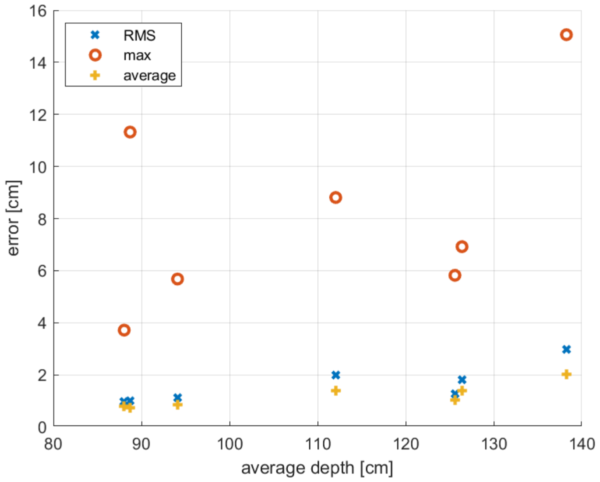

Figure 6. Depth error on a planar surface varying the average depth from the cameras.

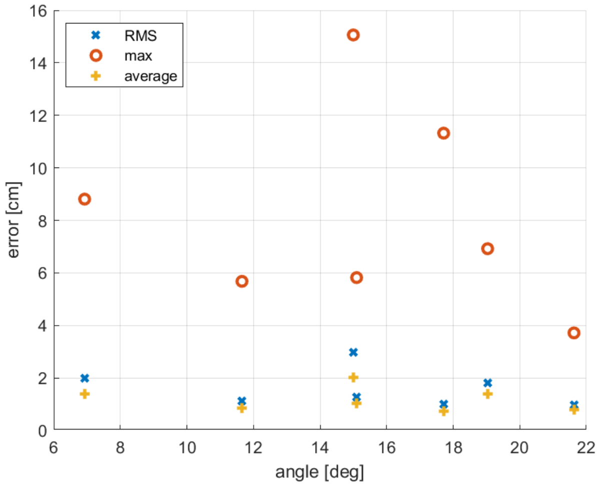

Figure 7. Depth error on a planar surface varying the angle between the normal of the planar surface and the plane orthogonal to the camera observation direction.

6. DISCUSSION

Thanks to the presence of a rear dual camera, smartphones can be considered as stereo vision-based depth sensors and hence compared with range cameras. Furthermore, since nowadays dual cameras are mounted also to quite low cost smartphones, and since this kind of devices are largely used by the worldwide population, smartphone based stereo-vision can be considered as very attractive and potentially usable in a wide range of cases.

However, as shown in Fig. 5, the closeness between the two rear cameras severely reduces the potentiality of the system in properly estimating depths: in practice, the distance to the object to be reconstructed should be limited to approximately 2 m to limit the depth resolution error to 10 cm.

As expected, Fig. 6 shows that the error when reconstructing a planar surface, evaluate in this case as RMS with respect to the fitting plane, increases for larger values of the depth. It is worth to notice that the RMS shown in Fig. 6 is typically smaller with respect to theoretical depth resolution limit shown in Fig.5: this is because the plane fitted to compute the error shown in Fig. 6 does not necessarily coincide with the real one, e.g. due to the quantization of the possible depth levels (shown in Fig.5 and caused by the matching at pixel resolution level) a bias with respect to the real one might be present and hence not taken into account in the errors shown in Fig. 6.

Differently, Fig. 7 shows that, at least for the considered angles, there isn't a relevant dependence of the depth estimation error with respect to the angle formed by the surface and the observation direction.

7. CONCLUSIONS

This paper presented some preliminary results on the 3D reconstruction error with smartphone-based stereo-vision. The interest in such kind of system is clearly motivated by the potential use of smartphones as stand alone photogrammetric surveying devices.

Despite such potential, the typical short baseline between the two rear cameras of a smartphone practically restricts the usage of this kind of system to the reconstruction of close objects. Despite this can be a severe limitation in certain applications, this is not that different from the typical case of certain commercial low cost range cameras (e.g. despite being designed for gaming purposes, Kinect sensor is actually a consumer-grade range camera).

Alternatively to the considered method the integration of information from the navigation system can be used to estimate the scale factor ((Ham et al., 2014, Widyawan et al., 2012, Saeedi et al., 2014, Masiero et al., 2014)). Future investigation foresees the combination of stereo-vision with such information provided by the navigation system, in particular when exploiting accurate IMU error modelling (Radi et al., 2018), and the improvement of the integration between single and dual camera visual information and reconstruction.

A more detailed evaluation of the reconstruction error obtained with the considered system will also be considered in our future investigations.

REFERENCES

Al Hamad, A. and El Sheimy, N., 2014. Smartphone based mobile mapping systems. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-5, pp. 29–34.

Alsubaie, N. M., Youssef, A. A. and El-Sheimy, N., 2017. Improving the accuracy of direct geo-referencing of smartphone-based mobile mapping systems using relative orientation and scene geometric constraints. Sensors 17(10), pp. 2237.

Brunelli, R., 2009. Template matching techniques in computer vision: theory and practice. John Wiley & Sons.

Chiang, K.-W., Liao, J.-K., Huang, S.-H., Chang, H.-W. and Chu, C.-H., 2017. The performance analysis of space resection-aided pedestrian dead reckoning for smartphone navigation in a mapped indoor environment. ISPRS International Journal of Geo-Information 6(2), pp. 43.

Dabove, P., Di Pietra, V. and Lingua, A. M., 2018. Close range photogrammetry with tablet technology in post-earthquake scenario: Sant'Agostino church in Amatrice. GeoInformatica pp. 1–15.

Förstner, W. and Wrobel, B. P., 2016. Photogrammetric Computer Vision. Springer.

Ham, C., Lucey, S. and Singh, S., 2014. Hand waving away scale. In: European Conference on Computer Vision, Springer, pp. 279–293.

Hirschmuller, H., 2005. Accurate and efficient stereo processing by semi-global matching and mutual information. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, Vol. 2, IEEE, pp. 807–814.

Hirschmuller, H., 2008. Stereo processing by semiglobal matching and mutual information. IEEE Transactions on pattern analysis and machine intelligence 30(2), pp. 328–341.

Khoshelham, K. and Elberink, S. O., 2012. Accuracy and resolution of kinect depth data for indoor mapping applications. Sensors 12(2), pp. 1437–1454.

Kolmogorov, V. and Zabih, R., 2002. Multi-camera scene reconstruction via graph cuts. In: European conference on computer vision, Springer, pp. 82–96.

Konrad, R. and Padmanaban, N., 2017. Stereophonic: Depth from stereo on phones. In: CS231 Course report, Stanford University.

Korman, S., Reichman, D., Tsur, G. and Avidan, S., 2013. Fast-match: Fast affine template matching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2331–2338.

Lindeberg, T., 2015. Image matching using generalized scale-space interest points. Journal of Mathematical Imaging and Vision 52(1), pp. 3–36.

Lo, C., Tsai, M., Chiang, K., Chu, C., Tsai, G., Cheng, C., El-Sheimy, N. and Habib, A., 2015. The direct georeferencing application and performance analysis of UAV helicopter in GCP-free area. ISPRS - International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 40(1), pp. 151.

Masiero, A., Fissore, F. and Vettore, A., 2017. A low cost UWB based solution for direct georeferencing UAV photogrammetry. Remote Sensing 9(5), pp. 414.

Masiero, A., Fissore, F., Guarnieri, A., Pirotti, F., Visintini, D. and Vettore, A., 2018. Performance evaluation of two indoor mapping systems: Low-cost uwb-aided photogrammetry and backpack laser scanning. Applied Sciences 8(3), pp. 416.

Masiero, A., Guarnieri, A., Pirotti, F. and Vettore, A., 2014. A particle filter for smartphone-based indoor pedestrian navigation. Micromachines 5(4), pp. 1012–1033.

Micheletti, N., Chandler, J. H. and Lane, S. N., 2015. Investigating the geomorphological potential of freely available and accessible structure-from-motion photogrammetry using a smartphone. Earth Surface Processes and Landforms 40(4), pp. 473–486.

Mustaniemi, J., Kannala, J., Särkkä, S., Matas, J. and Heikkilä, J., 2017. Inertial-based scale estimation for structure from motion on mobile devices. In: Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International Conference on, IEEE, pp. 4394–4401.

Nex, F. and Remondino, F., 2014. Uav for 3d mapping applications: a review. Applied Geomatics 6(1), pp. 1–15.

Nocerino, E., Poiesi, F., Locher, A., Tefera, Y. T., Remondino, F., Chippendale, P. and Van Gool, L., 2017. 3d reconstruction with a collaborative approach based on smartphones and a cloud-based server. In: International archives of photogrammetry, remote sensing and spatial information sciences: Proceedings of the ISPRS Commission V Mid-Term Symposium'Close Range Image Measurement Techniques', Vol. 42number W8, Copernicus, pp. 187–194.

Pfeifer, N., Glira, P. and Briese, C., 2012. Direct georeferencing with on board navigation components of light weight UAV platforms. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.

Piras, M., Cina, A. and Lingua, A., 2008. Low cost mobile mapping system: an italian experience. PLANS 2008 proceedings, Monterey (CA), 6-8 May 2008.

Puente, I., González-Jorge, H., Martínez-Sánchez, J. and Arias, P., 2013. Review of mobile mapping and surveying technologies. Measurement 46(7), pp. 2127–2145.

Radi, A., Nassar, S. and El-Sheimy, N., 2018. Stochastic error modeling of smartphone inertial sensors for navigation in varying dynamic conditions. Gyroscopy and Navigation 9(1), pp. 76–95.

Saeedi, S., Moussa, A. and El-Sheimy, N., 2014. Context-aware personal navigation using embedded sensor fusion in smartphones. Sensors 14(4), pp. 5742–5767.

Schwarz, K. P. and El-Sheimy, N., 2004. Mobile mapping systems–state of the art and future trends. ISPRS - International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 35(Part B), pp. 10.

Tao, C. V. and Li, J., 2007. Advances in mobile mapping technology. Vol. 4, CRC Press.

Toth, K., Molnar, B., Zaydak, A. and Grejner-Brzezinska, D., 2012. Calibrating the ms kinect sensor. In: ASPRS Annual Conference.

Tucci, G., Visintini, D., Bonora, V. and Parisi, E. I., 2018. Examination of indoor mobile mapping systems in a diversified internal/external test field. Applied Sciences 8(3), pp. 401.

Widyawan, Pirkl, G., Munaretto, D., Fischer, C., An, C., Lukowicz, P., Klepal, M., Timm-Giel, A., Widmer, J., Pesch, D. and Gellersen, H., 2012. Virtual lifeline: Multimodal sensor data fusion for robust navigation in unknown environments. Pervasive and Mobile Computing 8(3), pp. 388–401.

Yu, C., El-Sheimy, N., Lan, H. and Liu, Z., 2017. Map-based indoor pedestrian navigation using an auxiliary particle filter. Micromachines 8(7), pp. 225.

Zhuang, Y., Yang, J., Li, Y., Qi, L. and El-Sheimy, N., 2016. Smartphone-based indoor localization with bluetooth low energy beacons. Sensors 16(5), pp. 596.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLII-1-289-2018 | © Authors 2018. CC BY 4.0 License.