

DESIGN AND IMPLEMENTATION OF A NOVEL PORTABLE 360° STEREO CAMERA SYSTEM WITH LOW-COST ACTION CAMERAS

D. Holdener *, S. Nebiker, S. Blaser

FHNW University of Applied Sciences and Arts Northwestern Switzerland, Institute of Geomatics Engineering, Gründenstrasse 40,
4132 Muttenz, Switzerland – david.holdener@hotmail.com, (stephan.nebiker, stefan.blaser)@fhnw.ch

Commission II

KEY WORDS: Action Camera, Calibration, Fisheye Stereo, Indoor, Low-Cost, Mobile Mapping, Panorama

ABSTRACT:

The demand for capturing indoor spaces is rising with the digitalization trend in the construction industry. An efficient solution for measuring challenging indoor environments is mobile mapping. Image-based systems with 360° panoramic coverage allow a rapid data acquisition and can be processed to georeferenced 3D images hosted in cloud-based 3D geoinformation services. For the multi-view stereo camera system presented in this paper, a 360° coverage is achieved with a layout consisting of five horizontal stereo image pairs in a circular arrangement. The design is implemented as a low-cost solution based on a 3D printed camera rig and action cameras with fisheye lenses. The fisheye stereo system is successfully calibrated with accuracies sufficient for the applied measurement task. A comparison of 3D distances with reference data delivers maximal deviations of 3 cm on typical distances in indoor space of 2-8 m. Also the automatic computation of coloured point clouds from the stereo pairs is demonstrated.

1. INTRODUCTION

Image-based mobile mapping is a well-established method for capturing the outdoor environments such as road or rail infrastructures in 3D. Such systems can be equipped with stereoscopic panoramic cameras either in a vertical (Earthmine, 2014) or horizontal arrangement (Blaser et al., 2017) to enable full coverage of complex environments. With the trend in architecture and construction towards digital building design and construction progress control, the need for accurate and rapid mapping of indoor scenes in 3D is also rising. Existing indoor mobile mapping systems such as presented by Bergsli & Schroth (2017) are hybrid systems obtaining their 3D information from LiDAR sensors and subsequently combining those observations with images.

Such georeferenced images with color and depth information are an important help to enable exploration and interaction in LiDAR point clouds. This is especially true for users unfamiliar with point clouds and for complex scanning objects such as indoor environments. In contrast to depth values captured with LiDAR, the depth from dense image matching ensures the spatial and temporal coherence of radiometric and depth data of the 3D imagery (Nebiker et al., 2015). The 3D information can hereby be derived directly from the images. This allows a presentation in a cloud-based imaging service without the need for accurately co-registering the cameras and LiDAR sensors.

The indoor environment brings new challenges to multi-view stereo mobile mapping. These are, for example, the need of a compact sensor frame or potentially short distances between cameras and objects. Both factors limit the possible base length. In order to maximise the base length and still enable enough overlap between the images even on short object distances, action cameras with fisheye lenses were used.

The subsequent investigations were part of a project with the final goal of designing a portable stereo camera system for an indoor mobile mapping system (IMMS). The IMMS should have a horizontal 360° coverage without any occlusion in the images by the operator or the system itself. A sufficient measurement accuracy for typical applications in architecture and construction should be reached. The focus of the work lies on the elaboration and evaluation of multiple system designs and a subsequent low-cost implementation of a prototype system. This also includes the system calibration and evaluation of the relative accuracy.

In the following section 2, the related work to this project is discussed. Section 3 shows multiple system designs and their evaluation regarding the coverage and possible accuracy. The implementation, including the calibration and applied processing workflow, is described in section 4, followed by the system evaluation in section 5.

2. RELATED WORK

The use of multi-view stereo for outdoor mobile mapping applications – without a need for LiDAR scanners – is already well established (for example Burkhard et al., 2012). The same mobile mapping system was upgraded by Blaser et al. (2017) with two stereoscopic multi-head panoramic cameras. The stereo base is horizontally aligned in driving direction which enables large stereo baselines. The configuration also complies with Swiss privacy laws which prohibit street-level image acquisition from more than 2 m above ground. The authors also present a calibration and processing workflow for the stereo fisheye images. The obtained accuracy with fisheye cameras and the equidistant camera model by Abraham & Förstner (2005) does not differ significantly from the results with

* Corresponding author

pinhole stereo images. Another mobile mapping system with two stereoscopic multi-head panoramic cameras in a configuration with a vertical base is the Mars Collection System (Earthmine, 2014). A system that uses a single panoramic image with a virtual stereo base from Structure from Motion (SfM) is presented by Van den Heuvel et al. (2006). Even with a high positioning accuracy of the system, the 3D accuracy of such a SfM-based system cannot reach the high accuracy of a system with a calibrated physical base. Rau et al. (2016) present a SfM-based mobile mapping solution with a multi-head panoramic system mounted on a backpack for indoor and outdoor applications. They show the necessity of many tie points as the positioning sensors are not capable of computing the virtual bases for dense image matching with an adequate accuracy. More popular on IMMS than passive image sensors are active sensors such as LiDAR (Nüchter et al., 2015) or hybrid systems combining LiDAR and cameras (Leica Geosystems, 2017). Further active sensors integrated in IMMS are time-of-flight cameras (Khoshelham & Elberink, 2012) or regular cameras with structured light projection (Zhu et al., 2007).

Another research area with increasing interest in 360° stereo solutions is the field of robotics. An example for such a system is shown by Meilland et al. (2015) with different configurations of the cameras either vertical or horizontal. With the increase of virtual reality (VR) applications, also consumer grade systems for taking 360° stereo videos and for watching them on VR-glasses are getting more popular. An example for such a system is the Project Beyond from Samsung Electronics (2017).

A different approach for 360° stereo coverage without the need for a multi-head camera system is based on two catadioptric optics placed vertically on top of each other (Lui & Jarvis, 2010; Ragot et al., 2008). A more sophisticated double lobed catadioptric mirror is even capable of taking two stereo panoramic images with a single camera in one shot (Cabral et al., 2004). The disadvantage of these systems however are the large distortions. They result in an inhomogeneous distribution of the resolution across the image as well as a lower resolution in general compared to a multi-head system.

A potential approach to generate depth images from fisheye images is presented by Schneider et al. (2016). Prior to the dense image matching, the images are converted into an epipolar equidistant image pair according to Abraham & Förstner (2005). A good overview for different but obsolete algorithms for stereo matching are presented by Scharstein & Szeliski (2002). Newer algorithm with focus on real-time application and fisheye lenses can be found in Krombach et al. (2015). Well-known algorithms especially for fisheye stereo matching are Semiglobal Matching (SGM) by Hirschmüller (2008) and Efficient Large-Scale (ELAS) by Geiger et al. (2010).

3. SYSTEM DESIGNS

Holdener (2017) created and evaluated different designs for a 360° multi-headed stereo panoramic camera arrangement with respect to their coverage, to the handling in a portable mobile mapping system and the possible base length of stereo image pairs (see Figure 1). All concepts have five stereo bases to ensure a good overlap between the stereo pairs.

The first conceptual design is a horizontal circular arrangement of the stereo bases (Fig. 1, left). The images of such a system do

not have occlusions by the operator if the rig is placed on a backpack and carried above or around the body. The possible base length is limited by the width of doors and narrow hallways but can be up to approximately 50-60 cm. The stitching of a single panoramic image from multiple cameras will not be appropriate due to the large distances between the projection centers. A vertical multi-head arrangement is the second solution (Fig. 1, middle). The base length constraint of such a system is less strict and it is possible to generate a panoramic image with small stitching errors for each of the two multi-head systems. The major drawback of this design is the significant occlusion by the operator as the rig has to be carried in front or behind the body. The third concept (Fig. 1, right) combines the advantages of the previous two. With the shifted design, the stereo base is vertically in line but has an offset horizontally. This enables a 360° coverage around the operator without occlusion. In addition, the base can be large and a panoramic image can be stitched from the top cameras. However, the stereo images cannot be processed in a standard dense image matching process due to the horizontal offset and the resulting scale differences in the image pairs.

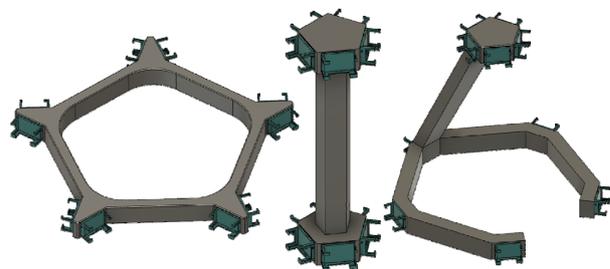


Figure 1. Conceptual designs for a possible IMMS: circular (left), vertical (middle) and shifted (right)

The realised system is a horizontal ring as shown in Figure 1, left. This layout enables a full horizontal 360° stereo image acquisition even for objects at short distance. A base length of 50-60 cm should enable a centimeter level 3D measurement accuracy even with low-cost cameras.

4. IMPLEMENTATION AND CALIBRATION

The implementation of the system is based on the low-cost action cameras Git2 from GitUp. The cameras have a resolution of 4608 x 3456 pixels with a focal length of 3 mm and a pixel size of 1.34 μm . The fisheye lenses provide a large field-of-view of 120° x 90°. The camera mounts and the frame of the system were printed with a 3D printer. The camera rig in this current version (see Figure 2) has a total diameter of 70 cm and a base length of 60 cm.

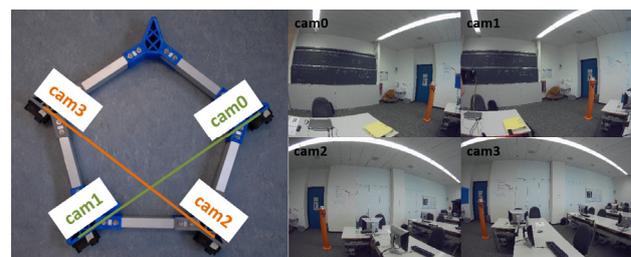


Figure 2. Realised camera rig (left) with captured images from the same location for two of the five stereo image pairs (right)

A Raspberry Pi 3 single board computer together with an associated touchscreen was used to trigger the cameras. The triggering signal for the GitUp cameras is a specific pulse-width modulation (PWM). The synchronisation of the four cameras was tested by taking images of a watch with a 10 ms resolution. The mean deviation is 20 ms with a maximum of 40 ms. The following calibration and evaluation were conducted with four cameras and therefore only two of the five stereo systems (Holdener, 2017).

4.1 Calibration

The calibration software was adopted from the mobile mapping calibration system by Blaser et al. (2017). It uses a test field calibration with bundle adjustment according to Ellum & El-Sheimy (2002) and supports the equidistant camera model for fisheye cameras (Abraham & Förstner, 2005). The resulting parameters are

- the interior orientation parameters (IOP) for each camera;
- the relative orientation parameters (ROP) between left and right camera for each stereo system;
- the boresight alignment (BA) with the lever arm and misalignment between the left cameras of each stereo system and the master camera cam0.

The exterior orientation parameters (EOP) for the whole system and for each frame are also computed using the bundle adjustment but are not of interest in our calibration. However, their standard deviation gives an impression of the resulting accuracy. All parameters could be determined in a single calibration in an indoor calibration field (see Figure 3) as there were only imaging and no positioning sensors included. The calibration was conducted with 8 epochs (32 images) and a total of 3952 observations excluding 61 outliers. The circular targets were measured with a least-squares ellipse fit.



Figure 3. Calibration field

The resulting standard deviations for the ROP, boresight parameters and EOP are all below 0.7 mm or 30 mdeg (Holdener, 2017). A complete list of the results can be found in Table 4. But since all parameters are correlated with each other and dependent on the measurement accuracy, a separate analysis for each parameter is not feasible.

The average reprojection error is 0.7 pixels. The results show some systematic behaviour – possibly due to an imprecise triggering or the rolling-shutter-effect. Furthermore, the size and number of outliers increase with the radial distance in the

image plane where the fisheye distortions grow larger and do no longer match the mathematical model in case of these low-cost cameras. A comparison with previous calibrations with the action camera GoPro from Hastedt et al. (2016) with a reprojection error of 1.2 pixels or Balletti et al. (2014) with 3.9 pixels is not possible since they both calibrated only single cameras and not a stereo system. They both report problems in regions with large radial distance and question the stability of the action camera's IOPs.

Parameter	Cam	Std. dev. of position [mm]			Std. dev. of orientation [mdeg]		
		X	Y	Z	ω	ϕ	κ
ROP base 1	0-1	0.3	0.3	0.6	16	24	3
ROP base 2	2-3	0.4	0.3	0.7	12	19	2
BA	0-2	0.5	0.3	0.5	31	20	28
EOP	0	0.4	0.4	0.4	16	12	11

Table 4. Standard deviations of the calibration parameters ROPs, BA and EOPs. The EOPs are the mean values from the 8 epochs

4.2 Processing Workflow

The stereo image processing workflow was also adopted from the solution by Blaser et al. (2017). The goal was to obtain geospatial 3D images which consist of the corrected equidistant RGB images and the corresponding depth values, ideally for every pixel. First, the IOP and ROP are corrected and the images are converted to the epipolar geometry using the epipolar equidistant projection model as presented by Abraham & Förstner (2005). These steps are required for the dense image matching conducted with SURE, based on tSGM (Rothermel et al., 2012). The disparity map and its values are then reconverted to the equidistant projection model to match the IOP and ROP corrected RGB images. The formulas for this reversion are as well presented by Abraham & Förstner (2005). A colored point cloud can subsequently be derived from the disparity map and the RGB image.

5. SYSTEM EVALUATION

To evaluate the system, the relative accuracy has been determined with manual point measurements in the corrected images. Additionally, a comparison of the derived point cloud with a reference scan from a terrestrial laser scanner (TLS) has been conducted.

5.1 Manual point measurements

Images from the calibration field (see Figure 3) have been used to evaluate the relative accuracy in the stereo image pairs with manual point measurements. The images were corrected with the IOP and ROP and the target observations were obtained with a least-squares ellipse fit. All 3D coordinates were computed using the same parameters as for the processing workflow for obtaining disparity maps. Sets of two tie points of the calibration field form the reference 3D distances. By comparing the photogrammetric 3D distances with the reference, only the IOP and ROP uncertainties from the calibration and the measurement accuracy affect the results. The reference distance can be regarded as error-free as the reference point coordinates are known with sub-millimeter accuracy.

A total of 78 3D distances from 3 epochs of each stereo base were defined. The measurements were done on typical object distances between 2-8 m, with the 60 cm base length, different radial distances in the image plane and different orientations in 3D space.

The resulting mean deviation between the photogrammetrically measured and the reference 3D distance is 5.7 mm with a standard deviation (1σ) of the differences of 5.6 mm (see Figure 5). These results exclude two outliers with differences higher than 70 mm. Without these outliers, the remaining maximal deviation is at 33 mm.

The influence of the object distance as well as the radial distance in the image plane on the resulting 3D measurement accuracy has been further investigated. The expected measurement accuracy declines with increasing object distance due to a weaker intersection geometry (Luhmann et al., 2006). The conducted experiment with the stereo system shows only a slight and insignificant decrease (see Figure 5). A declining measurement accuracy with increasing radial distance can also be expected for fisheye cameras, especially for low-cost models. This is due to the increasing distortion where the mathematical model does not fit well with the actual lens properties. The observed differences show a clear decrease in accuracy as expected (see Figure 6).

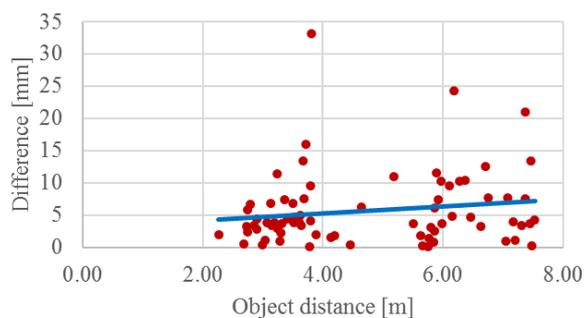


Figure 5. Differences between photogrammetrically measured distances and reference distances according to their 3D distance in object space (without two outliers > 70 mm)

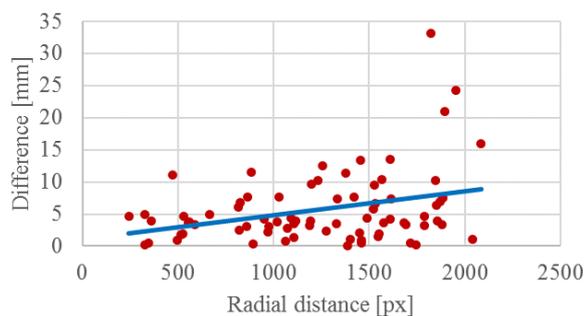


Figure 6. Differences between photogrammetrically measured distances and reference distances according to their radial distance in the image plane (without two outliers > 70 mm)

5.2 Point cloud comparison

A comparison between the 3D data obtained from the stereo system and a TLS evaluates the resulting disparity map or point cloud of the applied processing workflow. The TLS reference

scan was performed with a Leica P20 from the same measurement location as the images. This scanner has a specified measurement accuracy of <1 mm on short distances (Leica Geosystems, 2013). The two point clouds are referenced with the iterative closest point (ICP) algorithm. The tested scene is a typical indoor space with textured and also large untextured areas, furniture and with measurement distances between 1.5 and 7 m. The comparison shows big discrepancies in regions with planar, untextured walls or ceilings as this is a typical problem in dense image matching (see Figure 7).

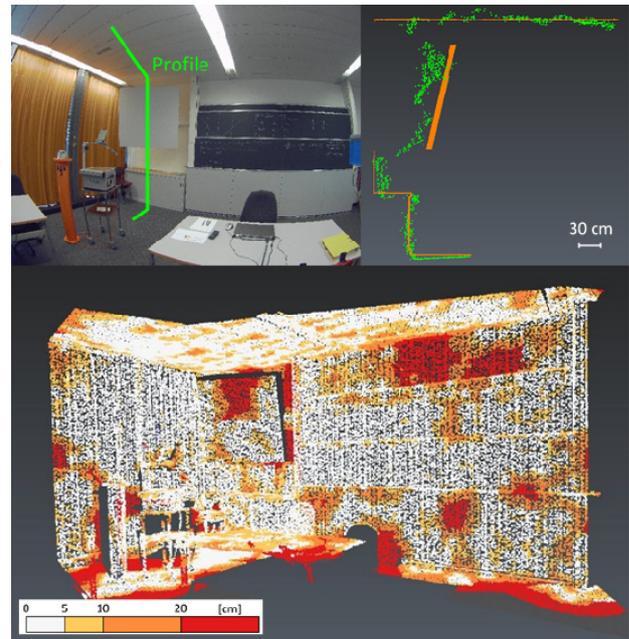


Figure 7. The testing area for a point cloud evaluation with a profile and an areal comparison

Also, a systematic drift towards the edges of the point cloud was observed. This is due to the degrading agreement between the actual distortion and the mathematic model with increasing radial distance. Thanks to the wide opening angles of the fisheye cameras, the overlapping areas between the stereo bases are large enough so that the drift parts can be replaced by the point cloud of the next stereo images.

6. CONCLUSIONS AND OUTLOOK

This work shows multiple conceptual designs for a possible 360° stereo imaging system. The calibration and processing of 3D imaging data and point clouds could be implemented in a low-cost prototype. Comparisons between 3D distances and point clouds with reference data evaluate the system.

The presented realisation of the imaging system with a horizontal ring of 5 stereo systems can map indoor spaces in 3D and with 360° coverage. Action cameras (GitUp Git2) with fisheye lenses, a single-board computer with an associated touch screen display (Raspberry Pi 3) and a 3D printed camera rig were used for the low-cost implementation. The system realised with two of the five stereo pairs was calibrated using a rigid bundle adjustment-based approach with resulting standard deviations of the ROP below 0.7 mm or 30 mdeg. A processing workflow for fisheye stereo image pairs based on Abraham &

Förstner (2005) was successfully applied to the imagery of the prototype camera system.

The disadvantages of such low-cost action cameras with regard to kinematic mobile mapping applications are the rolling shutter and the PWM-based camera triggering. The latter showed differences in the camera synchronization of up to 40 ms. Nevertheless, an accuracy evaluation in indoor space on typical object distances in the range of 2-8 m showed the potential of the system with a mean 3D measurement accuracy obtained at a sub-centimeter level. This demonstrates the feasibility of the low-cost approach for practical applications. The disparity map and point cloud showed some typical drifts for action camera based fisheye stereo applications in the outer region of the image plane. However, these regions of the point cloud can be ignored thanks to a sufficiently large overlap between the stereo pairs.

A solution for the many untextured areas in indoor spaces is yet open for further research. By implementing positioning sensors to the camera rig, the presented system can be evaluated as a complete mobile mapping solution. In addition, a stability test according to Habib et al. (2014) would be very interesting in order to evaluate the stability of the cameras IOPs and the ROPs of the mount.

REFERENCES

- Abraham, S. and Förstner, W., 2005. Fish-eye-stereo calibration and epipolar rectification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 59(5), pp.278-288.
- Balletti, C., Guerra, F., Tsioukas, V. and Vernier, P., 2014. Calibration of Action Cameras for Photogrammetric Purposes. *Sensors*, 14(9), pp.17471-17490.
- Bergsli, M. and Schroth, G., 2017. NavVis - Enabling Digital Value Creation Indoors. In: *Surveying the world of tomorrow - From digitalisation to augmented reality*. FIG Working Week. Helsinki, Finland.
- Blaser, S., Nebiker, S. and Cavegn, S., 2017. System Design, Calibration and Performance Analysis of a Novel 360° Stereo Panoramic Mobile Mapping System. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. ISPRS Hannover Workshop. Hannover, Germany, pp.207-213.
- Burkhard, J., Cavegn, S., Barmettler, A. and Nebiker, S., 2012. Stereovision mobile mapping: system design and performance evaluation. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. XXII ISPRS Congress. Melbourne, Australia.
- Cabral, E.L.L., de Souza, J.C. and Hunold, M.C., 2004. Omnidirectional stereo vision with a hyperbolic double lobed mirror. In: *Proceedings of the 17th International Conference on Pattern Recognition*. Cambridge, UK.
- Earthmine, 2014. *earthmine Mars Collection System*. http://www.earthmine.com/html/products_mobile.html (Accessed 10.10.2017).
- Ellum, C. and El-Sheimy, N., 2002. The calibration of image-based mobile mapping systems. In: *2nd Symposium on Geodesy for Geotechnical and Structural Engineering*. The International Association of Geodesy (IAG). Berlin, Germany.
- Geiger, A., Roser, M. and Urtasun, R., 2010. Efficient Large-Scale Stereo Matching. In: *Computer Vision – ACCV*, Lecture Notes in Computer Science. Asian Conference on Computer Vision. Springer Berlin Heidelberg, pp.25-38.
- Habib, A., Datchev, I. and Kwak, E., 2014. Stability Analysis for a Multi-Camera Photogrammetric System. *Sensors*, 14(8), pp.15084-15112.
- Hastedt, H., Ekkel, T. and Luhmann, T., 2016. Evaluation of the quality of action cameras with wide-angle lenses in UAV photogrammetry. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. XXIII ISPRS Congress. Prague, Czech Republic, pp.851-859.
- Hirschmüller, H., 2008. Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2), pp.328-341.
- Holdener, D., 2017. Konzeption und Realisierung eines neuen portablen 360°-Stereokamerasystems. Master Thesis. University of Applied Sciences and Arts Northwestern Switzerland, Muttenz, Switzerland (unpublished).
- Khoshelham, K. and Elberink, S.O., 2012. Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications. *Sensors*, 12(2), pp.1437-1454.
- Krombach, N., Droschel, D. and Behnke, S., 2015. Evaluation of stereo algorithms for obstacle detection with fisheye lenses. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. II-1/W1.
- Leica Geosystems, 2013. *Leica ScanStation P20 Product Specifications*. http://w3.leica-geosystems.com/downloads123/hds/hds/ScanStation_P20/brochures-datasheet/Leica_ScanStation_P20_DAT_en.pdf (Accessed 10.10.2017).
- Leica Geosystems, 2017. *Leica Pegasus:Backpack*. http://www.leica-geosystems.ch/de/Leica-PegasusBackpack_106730.htm (Accessed 10.10.2017).
- Luhmann, T., Robson, S., Kyle, S. and Harley, I., 2006. *Close Range Photogrammetry: Principles, techniques and applications*. Dunbeath: Whittles Publishing.
- Lui, W.L.D. and Jarvis, R., 2010. Eye-Full Tower: A GPU-based variable multibaseline omnidirectional stereovision system with automatic baseline selection for outdoor mobile robot navigation. *Robotics and Autonomous Systems*, 58(6), pp.747-761.
- Meilland, M., Comport, A.I. and Rives, P., 2015. Dense Omnidirectional RGB-D Mapping of Large-scale Outdoor Environments for Real-time Localization and Autonomous Navigation. *Journal of Field Robotics*, 32(4), pp.474-503.
- Nebiker, S., Cavegn, S. and Loesch, B., 2015. Cloud-Based Geospatial 3D Image Spaces—A Powerful Urban Model for the Smart City. *ISPRS International Journal of Geo-Information*, 4(4), pp.2267-2291.
- Nüchter, A., Borrmann, D., Koch, P., Kühn, M. and May, S., 2015. A man-portable, IMU-free mobile mapping system. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and*

- Spatial Information Sciences*. ISPRS Geospatial Week. La Grande Motte, France.
- Ragot, N., Rossi, R., Savatier, X., Ertaud, J.Y. and Mazari, B., 2008. 3D volumetric reconstruction with a catadioptric stereovision sensor. In: *Industrial Electronics*. IEEE International Symposium. pp. 1306-1311.
- Rau, J.Y., Su, B.W., Hsiao, K.W. and Jhan, J.P., 2016. Systematic calibration for a backpacked spherical photogrammetry imaging system. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. XXIII ISPRS Congress. Prague, Czech Republic.
- Rothermel, M., Wenzel, K., Fritsch, D. and Haala, N., 2012. SURE: Photogrammetric Surface Reconstruction from Imagery. In: *Proceedings LC3D Workshop*. Berlin, Germany.
- Samsung Electronics, 2017. *Project Beyond*. <http://thinktankteam.info/beyond/> (Accessed 10.10.2017).
- Scharstein, D. and Szeliski, R., 2002. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision*, 47(1/2/3), pp. 7-42.
- Schneider, J., Stachniss, C. and Förstner, W., 2016. On the Accuracy of Dense Fisheye Stereo. *IEEE Robotics and Automation Letters*, 1(1), pp. 227-234.
- Van den Heuvel, F.A., Verwaal, R. and Beers, B., 2006. Calibration of fisheye camera systems and the reduction of chromatic aberration. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(5).
- Zhu, J., Humphreys, G., Koller, D., Steuart, S. and Wang, R., 2007. Fast Omnidirectional 3D Scene Acquisition with an Array of Stereo Cameras. In: *Sixth International Conference on 3-D Digital Imaging and Modeling*. Montreal, Canada, pp. 217-224.