

APPLICATION OF STEREO CAMERAS WITH WIDE-ANGLE LENSES FOR THE INDOOR MAPPING

D. Wierzbicki ^{1*}, P. Stogowski ¹

¹Department of Imagery Intelligence, Faculty of Civil Engineering and Geodesy, Military University of Technology, 2 Gen. S. Kaliskiego st. Warsaw, Poland - damian.wierzbicki@wat.edu.pl

Commission II, WG II/4

KEYWORDS: Photogrammetry, Point clouds, Dense matching, Disparity Map, Indoor 3D modelling, Indoor localization, Accuracy

ABSTRACT:

Recently, there has been an increase in interest in the use of wide-angle cameras in multi-image matching for the indoor 3D mapping and indoor localization. The demand for rapid 3D models of spaces in unknown environments is increasingly observed. That is particularly important when modelling unknown objects to conduct reconnaissance or building intervention after a disaster. In this case, developing a 3D model using a robot equipped with a system of synchronized stereo cameras with a short length longitudinal base is extremely desirable. In these studies, we present the approach to indoor location based on a 3D model developed from a dense point cloud with multi-image matching technique. As part of the research, an imaging system was developed, and an algorithm that converts images of selected objects to 3D model was implemented. The research presents the method of determining the object position based on the calculation of reference points' disparity based on the Sum of Absolute Differences (SAD). Next, a dense point cloud was generated based on the method of mutual image matching using Structure from Motion (SfM) algorithms. The resulting dense cloud of points had a resolution of 0.05 m. Based on the developed algorithm, a method for generating a quick model of the environment based on multi-image matching and disparity maps was presented. The obtained test results confirmed the possibilities of using the developed methodology for the needs of rapid reconnaissance of the environment to determine the distance, location and size of objects of interest. The mapping accuracy is at a decimeter level, and the possibility of geolocation of objects can be performed with an accuracy of ± 0.15 m. Based on the obtained test results, the potential of using miniature, portable mobile image-based mapping systems has been demonstrated to identify and model inaccessible rooms. Further work will be focused on the improvement of the geometric image quality and to increase the accuracy of the calibration.

* Corresponding author

1. INTRODUCTION

Recently, there has been an increase in interest in the development of methods for rapid spatial data acquisition to develop 3D models of interiors while maintaining good geometric accuracy of the mapped areas (Hermans et al., 2014). Furthermore, taking into account the needs of imagery reconnaissance in inaccessible objects and the increasing popularity of location-based services (LBS), the development of methods for quick interior mapping is an important research issue. Currently, multicamera systems (Biber et al., 2004) and multi-module systems consisting of RGB-D depth cameras and laser scanners (Chen et al., 2018), robotics-based solutions or navigation of SLAM-type solutions (Haber et al., 2012, Whelan et al., 2016) are used to obtain this type of data. Those methods are complemented by the use of low-cost cameras and the processing of images obtained by them using the Structure from Motion (SfM) algorithms based on which it is possible to extract dense point clouds (Chen et al., 2014). The use of cameras equipped with fish-eye lenses, which allow obtaining images of small spaces, has become particularly popular. The popularity of fish-eye lenses is caused by their wide focus range and field of view, which allows the more efficient acquisition of scenes with a limited number of shots. With the simultaneous evolution of calibration of this type of camera (Westboy et al., 2012), it is possible to obtain dense and accurate point clouds (Strecha et al., 2015). The use of this type of solutions related to close-range photogrammetry is now more accessible and interdisciplinary while reducing the cost of data acquisition (León-Vega, Rodríguez-Laitón, 2019).

The approach proposed in this paper was to develop a stereo imaging system using cameras with wide-angle lenses for the purposes of mapping of the environment. As part of the work, the possibilities of the application of OpenCV library in the processes of image analysis, and object detection systems were also shown. Furthermore, algorithms for depth maps and modeling were developed. Next, various possibilities for depth map creation were analyzed due to a number of problems that limit the use of this type of results. The practical part of the research also includes a description of the development of 3D model in the form of a dense, metric point cloud.

The contents of the paper are organized as follows. Section 1 gives Introduction and in Section 2 a review of related works is presented. Methodology was presented in Section 3. Section 4 contains test results from individual processing stages. Section 5 contains Discussion, and Section 6- Conclusions and also, an outlook to future developments.

2. RELATED WORKS

Stereovision is a breakthrough in many areas of the economy and industry, but along with it, many problems regarding the acquisition and processing of images obtained in stereo arose. One of them is the development of a dense 3D map for robot navigation (Yeang et al., 2018). Most of the work related to 3D imaging from stereo systems are associated with solutions based on a pair of synchronized cameras including epipolar matching for the needs of 3D point triangulation (Camunas-Mesa et al., 2014). Some papers also deal with the problem of instantaneous - developing depth maps based on short time intervals using two non-simultaneous cameras, (Schraml et al., 2015) (Zhou et al., 2018).

In terrestrial systems using two-camera imaging, integration with LIDAR systems can be found. Undoubted progress is the use of this type of technology in newly designed cars adapted for semi-autonomous driving. The generated point clouds are constantly analysed and compared, which helps the vehicle user in collision prevention. Hybrid of stereo and LIDAR

technologies are also used in SLAM systems (González-Fraga et al., 2017). Other solutions use SfM algorithms, which allow determining the location and orientation of the camera as well as sparse features. Solutions based on SLAM and SfM have many similarities, i.e. nonlinear optimization. In addition, the use of multi-core processing significantly speeds up the calculation. Until now, multi-image matching using SfM has been applied in close-range photogrammetry for outdoor mapping, e.g. building facades (Chen et al., 2017). However, concerning 3D indoor mapping, the use of SfM faces many challenges related to the occurrence of objects, shadowing, but also the occurrence of various internal construction elements (Kim et al., 2012), (Holdener et al., 2017). Therefore, it is necessary to develop new solutions that will recreate the structure of internal scene structures, but also the correct geometry of objects located in mapped (modelled) spaces (Ding et al., 2018).

3. MATERIALS AND METHODS

In this section, the characteristics of the developed stereo system to create 3D maps is presented. The experiments were conducted based on images obtained by a system of two cameras with wide-angle lenses installed onboard a radio-controlled self-propelled vehicle. A set of two low-cost commercial cameras with fisheye lenses of the GoPro type was used for the experiments. As part of the research, a camera set was calibrated to determine the interior orientation parameters (IOPs). Mutual orientation of cameras included in the stereo vision system was also made to determine the parameters of the relationship between the primary system, the system associated with a single camera and the stereo system. The OpenCV library was used to develop the stereoscopic image recording and post-processing algorithm.

3.1 Configuration and implementation of the stereo system

The most important element of the system were two cameras with wide-angle lenses. The camera was chosen because of its small size and relatively low weight (83g). In the research, two GoPro Hero 4 Black cameras were used. GoPro 4 camera can work in camera and video modes. In these possible arrangements, is also possible the use of different dividedness and the speeds of the recording of the sequence of video with the different FOV. Record videos in 4K / 30fps modes in ultra-wide FOV combination up to 170 °, 2.7K / 50fps and Full HD / 120fps. [GoPro, 2019]. Table 1 shows the camera's technical specification.

Item	Description
Size [mm]	41 x 59 x30
Weight [g]	88
Optical sensors type	CMOS
Digital Video Format	H.264
Nominal focal length [mm]	3
Image Recording Format	JPEG
Max Video Resolution	3840 x 2160
Effective Photo Resolution	12.0 MP
Sensor size [mm]	6.16 x 4.62
Pixel pitch [µm]	1.55

Table 1. Technical specification of GoPro 4 hero black

In the system, the short length of the longitudinal base and the horizontal position of the objectives relative to each other has been taken into account- see Figure 1. The symmetrical setting of cameras about the shortest longitudinal base was applied.

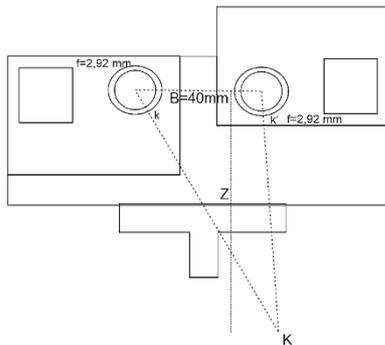


Figure 1. The layout of two imaging cameras (B- Longitudinal base, f-focal length, Z-distance from the point, K- point in space, k, k'- the projection of point K

The next stage in the construction process of the system was the adaptation of an appropriate tripod and balancing it in order to connect with the running platform. The radio-controlled vehicle in the 2.4 GHz range was used as the drive.

3.2 Stereo Camera Calibration

The purpose of calibration of the stereo vision system is to determine the parameters of the relationship between the basic system, the system associated with a single camera and the stereo system in order to match images (rectification). Based on multiple images of the test field (Fig. 2), the program determines the elements of internal orientation by the simultaneous assessment of linear epipolar geometry.

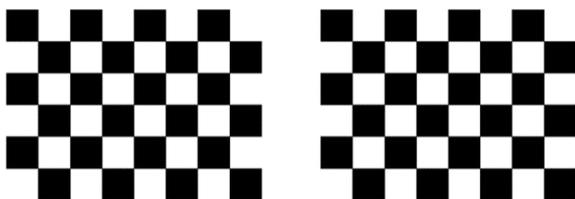


Figure 2. Tests used for the calibration process

Calibration is based on the detection of characteristic points used for calibration purposes. At least two calibration tests are needed to solve the equations and obtain results of calibration parameters. The following exposure parameters were used in the calibration process: ISO-106, exposure time - 1/120 s, aperture value $f / 2.8$. In order to determine elements of internal orientation, six series of photos of the calibration tests are required and then the results should be averaged (Table 2.). The system was calibrated in Agisoft Photoscan software and a script written based on the OpenCV library (for fish-eye camera models).

Mode HD	Agisoft Photoscan	OpenCV script
c [mm]	2.7166	2.7252
x0 [mm]	0.1422	0.1565

y0 [mm]	-0.0084	-0.0043
A1 [1/mm ²]	-0.0351	-0.0365
A2 [1/mm ⁴]	0.0015	0.0018
A3 [1/mm ⁶]	0.0000	0.0000
B2 [1/mm ²]	0.0000	0.0000
B3 [1/mm ³]	0.0000	0.0000

Table 2. Calibration results for stereo camera

The determined value of focal length differs by 0.3 mm from the value provided by the manufacturer. Whereas the focal length and coordinates of the main point are comparable with other results (Hastedt et al., 2016).

3.3 Data acquisition

As the study area a room with an area of approx. 25 m², located in a multi-storey residential building, was used. To simplify and speed up the process of image matching and creating a point cloud, calibration disks were used located at different heights throughout the entire tested room according to the following diagram (Figure 3).

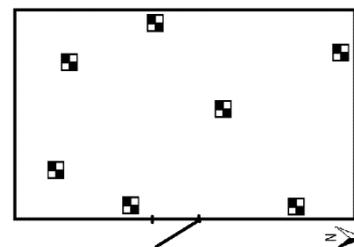


Figure 3. Plan of the location of calibration disks in measured room

The use of the created system to object detection is one of the the main role of the 3D map. For this purpose, in addition to fixed elements, a dummy model was placed in the room. Thanks to this solution, various options of the map and point cloud application to people detection could be simulated. For the research purpose, two video sequences were acquired with stereo cameras. Next, they were processed into individual frames. In this way, 1800 frames were obtained from each camera.



Figure 5. Assessment of the importance of the smear in the acquired image sequences

Due to the high speed of the vehicle with a mounted imaging system and artificial lighting, the vast majority of the sequences

were blurred. Smear values were determined based on the formula (1):

$$\Delta s = \frac{t \cdot c_k \cdot V_{kam}}{d} \quad (1)$$

where Δs = smear,
 V_{kam} = platform speed
 d = distance to the object,
 c_k = focal length,
 t = acquisition time.

The smear is caused by the vehicle's progressive movement and the too-long acquisition time. Image sequences with smear value lower than 2 mm were selected for further investigation. After data selection, the best 40 frames from each film were used.

3.4 Determination of disparity map

To accurately determine the position of objects of interest, the disparity values for all points of the reference image were calculated. It is difficult or even impossible to match individual points in normal conditions. Therefore the area or block matching techniques are used. To measure similarities between individual rectangular image fragments, the sum of absolute differences (SAD) was used, which can be calculated by the equation (Miyajima, Maruyama, 2003):

$$SAD = \sum_{c \in C} \sum_{i=-n}^n \sum_{j=-n}^n |I_L^c(x+i, y+j) - I_R^c(x+i+d, y+j)| \quad (2)$$

where
 I_L^c, I_R^c – the left and right image respectively,
 c – selected colour component (RGB or greyscale),
 n – dimension of the analysed area,
 d – disparity (between 0 and the maximum search distance).

First, the basic version of the block matching algorithm for similar regions in the first and second image is presented. According to the diagram (see Figure 6), the search window is moving to the left (assuming the created system where the cameras are at the same height). There are two parameters used in block matching algorithm - block size and its maximum difference (the maximum number of pixels that will be searched). First, the end nodes first should be initialized, i.e. input data, i.e. left and right image, and depth map as output. Next, in order to implement the matching block, nodes that will merge the images should be created.

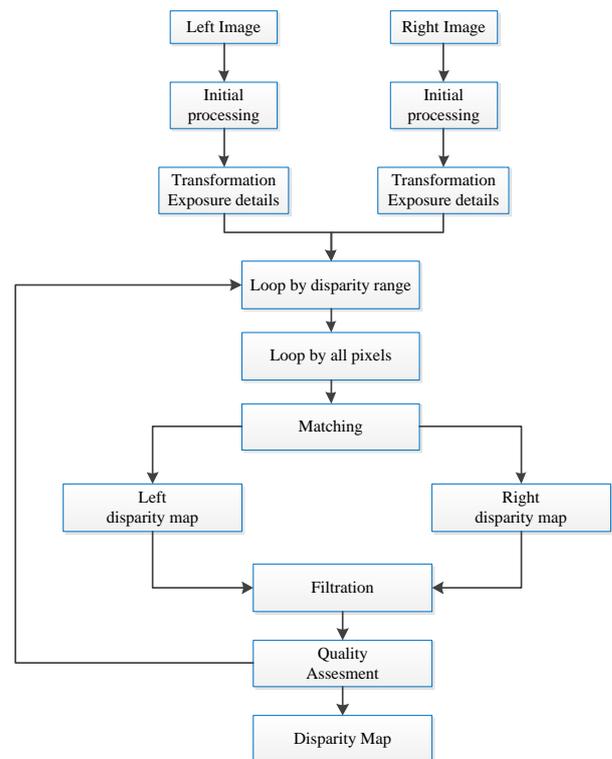


Figure 6. Diagram of the stereovision algorithm matching by areas

The next step is to create a node that will help create a depth map. *Neighbor_operation* processing has four parameters (up, down, right, left). In order to implement the block matching algorithm, auxiliary functions were created - the sum of absolute differences (SAD) for better matching and *getBlock*, which retrieves the block from the current pixel position. The block matching algorithm takes the current block from the first image and calculates SAD from the blocks from the second image, thus obtaining the position of the second image with the lowest SAD. The result is the discrepancy values, i.e. the differences between the current pixel and the best SAD location. After calculating the depth map, the distance values should be scaled to better visualization of the result.



Figure 7. The depth map obtained from the stereogram with optimal parameters

The resulting depth map (Figure 7) was determined using a block matching algorithm. The brighter the colour of the object, the closer to the camera during the acquisition. In the centre of the scene, a dummy very close to the camera can be distinguished. Moreover, the noise caused by different lighting conditions during data acquisition can be observed.

3.5 Reconstruction of the 3D scene

A dense point cloud was generated based on the acquired stereo images. To determine the metricity of the created space model, point cloud coordinates will be needed. They can be determined based on disparity maps, internal orientation parameters of standardised images, and external orientation of the model (and having elements of the internal orientation of standardized images and external orientation of the model (written in M_R , M_L transformation matrices). The calibration performed earlier allows determining elements of the external orientation of both cameras by registration of calibration tests. The rotation matrix is created based on known orientation elements:

$$R_{LR} = R_R R_L^T \quad T_{LR} = T_L - R_{LR}^T T_R \quad (3)$$

gdzie:

R_{LR} – rotation matrix,

T_{LR} – translation vector,

R_R , T_R – external orientation parameters of the right camera,

R_L , T_L – external orientation parameters of the left camera.

For a specific point P in the space of a stereo scene, the relationship between its coordinates in both cameras is as follows:

$$P_R = R_{LR}(P_L - T_{LR}) \quad (4)$$

The main issue of stereovision is the creation of a canonical system of two cameras. It is correct when epipolar lines are parallel with lines of digital images. The creation of such a system must be preceded by calibration (i.e. determining the essential elements of internal orientation and distortion parameters) and the normalization of images pair. Using orientation elements and a pair of rectification matrices (M_R , M_L), the following images can be normalized with the following formulas:

$$\begin{aligned} a_L \quad x_{pL} \quad a_R \quad x_{pR} \\ b_L = M_L \times y_{pL} \quad b_R = M_R \times y_{pR} \\ c_L \quad 1 \quad c_R \quad 1 \end{aligned} \quad (5)$$

$$\begin{aligned} x_{rectL} = a_L / c_L \quad x_{rectR} = a_R / c_R \\ y_{rectL} = b_L / c_L \quad y_{rectR} = b_R / c_R \end{aligned}$$

where

x_{rectL} , y_{rectL} – coordinates of the image points from the left camera after normalization

x_{rectR} , y_{rectR} – coordinates of the image points from the right camera after normalization

The last element to calculate the spatial coordinates of a point of interest is the solution of the linear equation:

$$ap_R - bRp_L + c(p_R \times Rp_L) = T \quad (6)$$

If the solution of equation (6) would be $[a_0, b_0, c_0]$, then the coordinates of the searched point in space are determined based on the formula:

$$P = a_0 p_R + 0,5 * (b_0 (T + Rp_L) - a_0 p_R) \quad (7)$$

Based on the implementation of the above equations and acquired image sequences, dense point clouds of the imaged room were obtained (Figure 8).

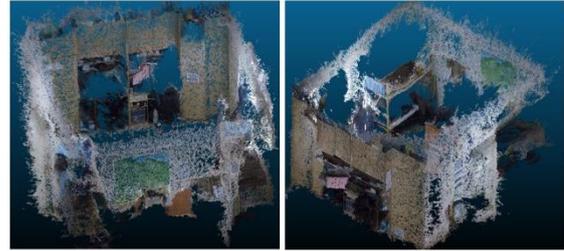


Figure 8. A spatial model of the mapped area

The obtained dense point cloud had a resolution of 0.05 m, and consist of more than 0.5 million points.

4. RESULTS

The results are presented graphically and in tables in the following sections. The algorithm described earlier has a specific reference image. Based on it, the version of this code assumes a left search. The block matching method consists in taking the current block from the first image and calculating the sum of the absolute differences from the second image, i.e. the differences between the current pixel and its best location in the second image. In creating the disparity map, it was important to select the appropriate parameters. Otherwise, the result would be useless (Table 2).

Configuration number	Resolution	Block Size	Window Size	bandwidth filter	Lambda	Brightness	Contrast
1	4	5	5	60	80	100	10
2	1	1	5	60	80	120	13
3	3	1	5	60	80	120	13
4	3	1	7	52	72	120	13
5	6	1	7	52	72	120	13
6	2	8	7	52	72	120	13
7	7	8	7	52	72	120	13
8	7	12	5	52	72	120	13
9	19	1	4	52	72	120	13
10	4	19	5	52	72	120	13
11	9	6	8	66	61	90	9

Table 2. Parameters for individual configurations of disparity map configurations

The first configuration was the default configuration after loading images into the program. The visual analysis provides enough information so that it can be considered an acceptable disparity map. In subsequent configurations, the influence of individual parameters on the accuracy of the resulting products was checked in order to select the optimal setting. In the second configuration, the brightness and contrast parameters were changed first so that the visual effect was better for the recipient. The next step was to change the value of the block size and resolution to a minimum. That gave a disparity map without any information. In the next sitting, the size of the

matching block remained the smallest, while the resolution was increased to three. That gave a blurry image with the delicate outlines of the examined objects. In the fourth configuration, the resolution, and block size were the same as in the previous settings, but the window size has increased, while the lambda value and filter capacity have been reduced. That did not affect the image. Then all parameters were left unchanged except for the resolution, which was increased to six. The obtained result gave much information about the position of objects in the stereogram; the preliminary assessment showed that it was too bright and illegible. In the sixth configuration, the resolution was reduced to two, and the block size increased to eight. It gave a dark image with distorted object edges, so the results were not satisfactory. In subsequent settings, both block size (eight) and resolution (seven) were increased. The result turned out to be much better than the previous ones. Unfortunately, that configuration could not remain due to the doubling of some objects and the occurrence of so-called double edges. The last three configurations were the minimum and maximum size settings alternately for resolution and block size, i.e. the values that are most important when generating the dispatch map. Finally, it was decided to choose the optimal settings. Changing settings several times gave the results shown in table 2. The eleventh configuration had a block size of 6. The resolution turned out to be the best at the medium level (nine). The brightness was reduced to 90 and the contrast to nine. The filter throughput and lambda value were not significantly changed. The resolution and size of the fit block had the most significant impact of all parameters.

4.1 Accuracy analysis of 3D mapping

This section presents the results of the analysis of the accuracy of detail reproduction. It allowed evaluating the metricity of the resulting model. To perform a correct assessment of accuracy in the stereo imaging systems, a five-degree scale was selected, as shown in the table below (Table 3).

Level	Group	Accuracy
I	Very accurate	< 0.10 m
II	Precise	0.10- 0.20 m
III	Acceptable	0.20- 0.40 m
IV	Low	0.40- 0.60 m
V	Insufficient	> 0.60 m

Table 3. Details' accuracy classes

Ten different measures on the model were made. The values have been rounded to 0.05 m due to the point cloud resolution. The measure on the model was defined as the distance measured between cloud points, and the real measure was defined as the result of direct measurements taken in the test room- below are some examples of restored objects from the terrain on the model, whose measures have been checked and analyzed (Figure 9).



Figure 9. Examined details (from left: bed and doors)

Based on the visual assessment and the measurements, it was concluded that the test objects were deformed. Many factors,

i.e., matching algorithm, imaging system instability, noise associated with uneven lighting affected the results. The real dimensions differed from those measured on the model by 0.091 m. The highest distortions were detected on objects furthest from the imaging system. As the distance increases, the field size of the pixel decreases, which causes objects to lose their detail. The objects of equal dimensions, in fact, depending on the place of measurement (top, middle, bottom) were not uniform, which means that they were deformed.

4.2 Geolocation accuracy analysis

One of the most sought after parameters of spatial products is the correct determination of the location of objects of interest. That can also be considered as a measure of the product's metric and correctness. In order to examine the resulting model, a geolocation analysis was performed (see Figure 10). It consisted of the selection of points in the examined object, which are clear to identified on the model and distributed evenly throughout the point cloud and measurements of their coordinates.



Figure 10. Distribution of checkpoints in the analysed space

The model has a local coordinate system with the conventionally assumed origin point, i.e. the place where the imaging system was located. In order to determine the accuracy and reliability of the study, it was necessary to measure the coordinates of the points of the tested object in reality and on the model and compare their differences. Verification of 10 points was the basis for calculating the average position error, which for X was ± 0.11 m, and for Y ± 0.09 m. When analyzing the results of the differences in theoretical and measured coordinates values, the highest X coordinate deviation of 0.15 m for points No. 6 and 8 and the lowest Y coordinate deviation equal to 0.00 for point 5 (see Figure 11).

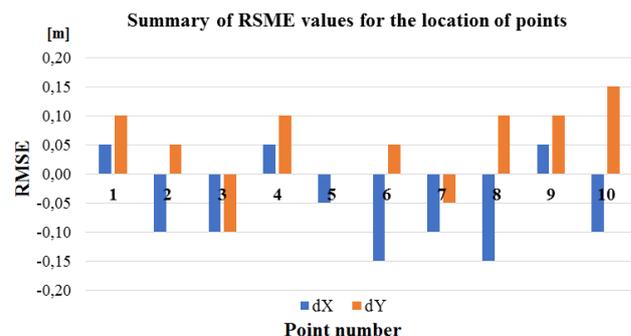


Figure 11. Summary of RMSE values for the location of points

A large amount of noise and shifts were probably caused by the movement of the vehicle and uneven lighting of the room. That made it challenging to select measuring points, which were often obscured by erroneous points. It should be borne in mind that the origin of the used coordinate system is only a theoretical place because the camera system changed position; therefore the possibility of immobilizing the system would reduce the final error results.

5. DISCUSSION

Based on the performed experiments, the accuracy of the geolocation of objects was ± 0.15 m. The results are similar to those obtained in other studies (Bassie et al., 2017), where the geolocation accuracy was 0.18 m. Other authors (Ma et al., 2016) presents products with accuracy at a similar level - 0.12 m. In the case of research (Holdener et al., 2017) where camera rig was used in a similar solution, the accuracy of 0.03 m was achieved. If the results are compared to the results of the authors' research (Yuda et al., 2016) and (Pei and Rui, 2015), a significant degree of similarity as a result of pixel matching of images when generating the point cloud can be observed. The obtained accuracy is at a very high level, considering the low budget of the project and the use of only imaging cameras in the visible range.

The use of the OpenCV library makes the presented project a starting point for the issues raised. The algorithm, due to its complexity, has many possibilities for development. It could be expanded in the future by adding an element of people or objects location and tracking. Future works should be started with the process of autonomous frame selection, and even an attempt to transform the platform to real-time mapping and modelling system. Such topics are already taken up but based on SLAM technologies.

6. CONCLUSIONS

As part of the research, a low-budget imaging system with an algorithm to process images to the environmental model was developed. The article deals with problems related to image processing and object detection on digital images. Based on the literature, a review of systems using stereovision was carried out. The OpenCV library gives many possibilities and ready solutions in the field of image processing. Based on the available literature, the block matching method was selected and implemented into the code. The developed software is not coupled with the system in real-time. Therefore the model is developed during post-processing. The developed map has the character of a metric point cloud. On its basis, the use of camera systems on a mobile platform to obtain information on the location and dimensions of objects inside rooms has been proven.

After applying the appropriate systems, creating a mini-computer and mounting it additionally on the platform, the algorithm presented in the future could be modified to the on-line version. The obstacle to using the created imaging system may be the lack of sufficient lighting, so it should be equipped with imaging cameras in the thermal range. The results of this research can be a starting point for location and people tracking and building an environment model by several systems simultaneously.

ACKNOWLEDGEMENTS

This paper has been supported by the Military University of Technology, the Faculty of Civil Engineering and Geodesy, Department of Imagery Intelligence.

REFERENCES

- Bassie, A. L., Meacham, S., Young, D., Turnage, G., & Moorhead, R. J., 2017. 3D reconstruction optimization using imagery captured by unmanned aerial vehicles. In *Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping II* (Vol. 10218, p. 102180I). International Society for Optics and Photonics, doi: 10.1117/12.2254852.
- Biber, P., Andreasson, H., Duckett, T., Schilling, A., 2004. 3D modeling of indoor environments by a mobile robot with a laser scanner and panoramic camera. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*(IEEE Cat. No. 04CH37566) (Vol. 4, pp. 3430-3435). IEEE.
- Camunas-Mesa, L. A., Serrano-Gotarredona, T., Jeng, S. H., Benosman, R. B., Linares-Barranco, B., 2014. On the use of orientation filters for 3D reconstruction in event-driven stereo vision. *Frontiers in neuroscience*, 8, 48. doi: 10.3389/fnins.2014.00048
- Chen, C., Yang, B., Song, S., Tian, M., Li, J., Dai, W., & Fang, L., 2018. Calibrate multiple consumer RGB-D cameras for low-cost and efficient 3d indoor mapping. *Remote Sensing*, 10(2), 328.
- Chen, S., Li, M., Ren, K., 2014. The power of indoor crowd: Indoor 3D maps from the crowd. In *2014 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)* (pp. 217-218). IEEE.
- Chen, Y., Chan, A. B., Lin, Z., Suzuki, K., Wang, G. 2017. Efficient tree-structured SfM by RANSAC generalized Procrustes analysis. *Computer Vision and Image Understanding*, 157, 179-189. doi: 10.1016/j.cviu.2017.02.005.
- Ding, Y., Zheng, X., Zhou, Y., Xiong, H., Gong, J., 2019. Low-Cost and Efficient Indoor 3D Reconstruction through Annotated Hierarchical Structure-from-Motion. *Remote Sensing*, 11(1), 58, doi: 10.3390/rs11010058.
- González-Fraga, J. A., Kober, V., Diaz-Ramirez, V. H., Gutierrez, E., Alvarez-Xochihua, O., 2017. Accurate generation of the 3D map of environment with a RGB-D camera. In *Applications of Digital Image Processing XL* (Vol. 10396, p. 103962A). International Society for Optics and Photonics. doi: 10.1117/12.2273074.
- GoPro 2019, Gopro Hero 4 Black user manual, https://gopro.com/content/dam/help/hero4-black/music-bundle-manuals/UM_H4Black-Music_ENG_REVB_WEB.pdf. (accessed on 10 December 2019).
- Haber, A., McGill, M., Sammut, C., 2012. Jmesim: An open source, multi-platform robotics simulator. In *Proceedings of Australasian Conference on Robotics and Automation* (December 2012).
- Hastedt, H., Ekkel, T., Luhmann, T. 2016. Evaluation of the quality of action cameras with wide-angle lenses in UAV photogrammetry, *Int. Arch. Photogramm. Remote Sens. Spatial*

- Inf. Sci.*, XLI-B1, 851–859, <https://doi.org/10.5194/isprs-archives-XLI-B1-851-2016>, 2016.
- Hermans, A.; Floros, G.; Leibe, B., 2014. Dense 3d semantic mapping of indoor scenes from RGB-D images. In *Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA)*, Hong Kong, China, 31 May–5 June 2014; pp. 2631–2638.
- Holdener, D., Nebiker, S. Blaser, S., 2017. Design and Implementation of a Novel Portable 360° Stereo Camera System with Low-Cost Action Cameras. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Hamburg, Germany, Vol. XLII-2/W8, pp. 105-110, doi: 10.5194/isprs-archives-XLII-2-W8-105-2017.
- Kedzierski, M., Fryskowska, A., Wierzbicki, D., Dabrowska, M., and Grochala, A., 2015. Impact of the method of registering Terrestrial Laser Scanning data on the quality of documenting cultural heritage structures, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XL-5/W7, 245–248, <https://doi.org/10.5194/isprsarchives-XL-5-W7-245-2015>, 2015.
- Kim, Y. M., Mitra, N. J., Yan, D. M., Guibas, L., 2012. Acquiring 3D indoor environments with variability and repetition. *ACM Transactions on Graphics (TOG)*, 31(6), 138.
- León-Vega, H. A., & Rodríguez-Laitón, M. I., 2019. Fisheye lens image capture analysis for indoor 3d reconstruction and evaluation. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 179-186. Revised May 2019.
- Ma, Y., Guo, Y., Lu, M., Zhao, J., Zhang, J., 2016. Global localization in 3D maps for structured environment. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 6680-6683). IEEE.
- Miyajima, Y., Maruyama, T., 2003. A Real-Time Stereo Vision System with FPGA, *Lecture Notes in Computer Science, Field Programmable Logic and Application*, vol. 2778, pp. 448-457, doi: 10.1007/978-3-540-45234-8_44.
- Pei, L., Rui, Z., 2015. The Analysis of Stereo Vision 3D Point Cloud Data of Autonomous Vehicle Obstacle Recognition. In *2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics* (Vol. 2, pp. 207-210). IEEE. doi: 10.1109/IHMSC.2015.192.
- Schraml, S., Belbachir, A. N., & Bischof, H., 2015. An event-driven stereo system for real-time 3-D 360 panoramic vision. *IEEE Transactions on Industrial Electronics*, 63(1), 418-428.
- Strecha, C., Zoller, R., Rutishauser, S., Brot, B., Schneider-Zapp, K., Chovancova, V., Krull, M., and Glassey, L., 2015. Quality assessment of 3d reconstruction using fisheye and perspective sensors, *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, II-3/W4, 215-222, <https://doi.org/10.5194/isprsannals-II-3-W4-215-2015>, 2015.
- Westoby, M. J., Brasington, J., Glasser, N. F., Hambrey, M. J., Reynolds, J. M., 2012. ‘Structure-from-Motion’ photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology*, 179, 300-314.
- 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.
- Yang, Y., Qiu, F., Li, H., Zhang, L., Wang, M. L., Fu, M. Y., 2018. Large-scale 3D Semantic Mapping Using Stereo Vision. *International Journal of Automation and Computing*, 15(2), 194-206.
- Yuda, M., Xiangjun, Z., Weiming, S., Shaofeng, L., 2016. Target accurate positioning based on the point cloud created by stereo vision. In *2016 23rd International Conference on Mechatronics and Machine Vision in Practice (M2VIP)* (pp. 1-5). IEEE. doi: 10.1109/M2VIP.2016.7827268.
- Zhou, Y., Gallego, G., Rebecq, H., Kneip, L., Li, H., Scaramuzza, D., 2018. Semi-dense 3D reconstruction with a stereo event camera. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 235-251).