

USING AN LOW-COST STEREO CAMERA FOR AUTONOMOUS NAVIGATION OF MOBILE ROBOT

A.G. Chibunichev*, A.P. Makarov, E.V. Poliakova

Moscow State University of geodesy and cartography (MIIGAiK), 4, Gorokhovskiy pereulok, Moscow, RUSSIA, agchib@mail.ru,
mapdir@yandex.ru, ekat.polykova@yandex.ru

Commission II

KEY WORDS: low-cost stereo camera, mobile robot, navigation, copter, change and charge the battery of the copter, homologue points.

ABSTRACT:

The paper considers the possibility of using low-cost stereo cameras for autonomous robot navigation. An low-cost stereo camera with a focal length of 5 mm and a photo base of 6 cm was chosen for the research. Experimental studies have shown that the accuracy of determining the coordinates of object points from a pair of images obtained by such a stereo camera is sufficient for organizing autonomous navigation of the robot. In order to improve the reliability and accuracy of determining the trajectory of the robot, this paper proposes to use two stereo cameras. One is directed forward by the robot's movement, and the other is directed at the nadir. Thus, the trajectory is determined twice, independently of each other. Moreover, each case has its own algorithm for finding the homologue points. In the first case, a sparse point cloud is constructed for each stereo pair based on the selection of interesting points and their identification based on the comparison of descriptors. In addition, blunder detection of points identification are realized based on the analysis of the values of the relative orientation equations using the fundamental matrix. In the second case, when the stereo camera is pointed at the nadir, the usual method of correlation is used in the nodes of the grid specified at one image. Experimental studies have shown sufficient efficiency of autonomous navigation of mobile robot based on the use of two stereo cameras.

1. INTRODUCTION

At the photogrammetry department of MIIGAiK is developing an "Automated information generation system" that is designed to perform aerial photography of an object using an unmanned aerial vehicle and obtain documents about the area (orthophoto, DTM, etc.) in automatic mode. A mobile robot is an integral part of this system (Fig. 1) designed to automatically change and charge the battery of the copter.



Figure 1. Two versions of the mobile robot

The system for changing and charging the battery operates as follows. There is a battery charging point (point A) and a battery change point (point B), it is also the point of takeoff and landing of the copter and they are marked on the ground (Fig.2). Distance between these points can be from 10 meters to 1 kilometer. The mobile robot must travel this distance, change the battery, launch the copter for the next aerial survey, and return to point B to charge the battery. All this is carry out automatically.

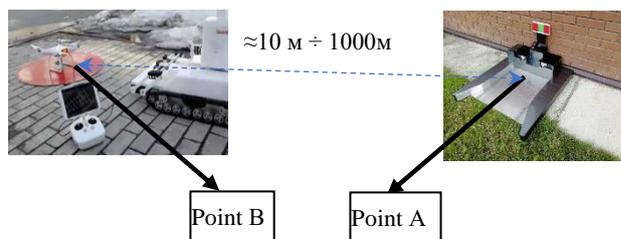


Figure 2. Battery changing and charging system.

Approximately the coordinates of points A and B are known from the navigation GPS. However, the mobile robot must approach points A and B with an accuracy of a few centimeters and stop in front of the point at a distance of approximately 50 cm. In this case, when the robot moves, it may encounter obstacles that should be avoided.

To solve this problem, it is necessary first of all to develop algorithms for autonomous tracking of the robot's trajectory (autonomous navigation), based on low-cost cameras. Today, there are some solutions based on stereo cameras on the market, for example, the Artisense system (Artisense Corporation, 2019), ZED stereo camera (StereoLabs, 2020) or the Parrot system (Parrot 2020). However, these systems are designed mainly for drones, cars, etc., where the distance from the camera to the object is calculated in meters. In our case, the distances can vary

* Corresponding author

from centimeters to meters. In addition, these systems are quite expensive. So we decided to use low-cost stereo camera, that costs about \$ 70.

2. STEREO CAMERA

Figure 3 shows the stereo camera that was used to implement autonomous robot navigation (Chibunichev A. G., et al, 2019). Table 1 shows its characteristics.



Figure 3. Stereo camera with a photo base of 6 cm and a focal length of 5 mm

Frame size (pixel)	1280x960		
Focal length (pixel)	855,9	853,4	
The coordinates of the principal point (pixel)	695.1, 449.8	703.7, 459.3	
Distortion parameters	k1	6,65e-07	6,8e-07
	k2	-1,03e-13	-1,6e-13
	k3	1,89e-18	1,99e-18
	p1	6,93e-07	1,69e-06
	p2	-2,07e-07	-4,81e-07
photo base (mm)	58		
Frame rate (frames / s)	56		

Table 1 Main parameters of the camera

Distance (mm)	number	control points			number	check point		
		RMS (mm)				RMS (mm)		
		X	Y	Z		X	Y	Z
4000	5	6.0	3.5	14.4	94	8.8	3.0	19.4
3000	5	4.2	2.5	6.4	81	9.8	2.8	10.6
1000	5	2.0	2.0	4.6	31	6.4	2.8	6.1
300	5	0.1	.05	0.2	10	0.1	0.1	0.3

Table 2 evaluating the accuracy of determining the coordinates of object points

Table 2 shows the results of evaluating the accuracy of determining the coordinates of object points from a pair of images, taken by this camera after its calibration. This table shows that this camera can be used for autonomous robot navigation in terms of accuracy.

The navigation itself is carried out using two stereo cameras, one pointing forward in the direction of travel, and the second pointing down. The first camera is used to track obstacles and orient the robot relative to points A and B. Both cameras are used for navigation

3. AUTONOMOUS NAVIGATION

Currently, quite a lot of algorithms for autonomous navigation of robotic systems based on the use of single or stereo cameras have been developed. The problem of determining the orientation and position of images in a spatial coordinate system is a standard photogrammetric problem (Wang et al., 2019), which is often solved using the method of "structure from motion" (SfM), "simultaneous localization and mapping" (SLAM), or "visual odometry" (VO). SfM allows you to simultaneously build a three-dimensional model of an object and determine the camera positions from multiple images of the object. Usually, the final position of cameras and object models is obtained as a result of bundle adjustment. This process takes quite a long time and grows with the increasing number of images (Frahm et al., 2010). And vice versa, visual odometry (VO) is mainly aimed at obtaining a three-dimensional trajectory of the camera's movement in real time, analyzing images sequentially. Photo-triangulation can be used for local trajectory correction when mapping the surrounding area and locating the camera on this plan, as implemented in SLAM (Scaramuzza et al., 2011). At the same time, this method aims to implement everything in real time. The SLAM method can be implemented on the basis of a single camera (Davison et al., 2007; Engel et al., 2014; Mur-Artal et al., 2015, V. V. Kniaz, 2015) and on the basis of a stereo camera (Engel et al., 2015; Mur-Artal et al 2017; Smolyansky et al 2017). The accuracy of determining the trajectory is low and largely depends on many factors, primarily the focal length of the camera, field of view and the ratio of the basis of photography and the distance to the object (Paz et al., 2008; Strasdat et al., 2011; Mur-Artal et al., 2015). The accuracy of determining the trajectory by a stereo camera is naturally higher than that by a single camera. To improve positioning accuracy using a stereo camera, it is proposed to increase the angle between the optical axes of the cameras (Shtain, et al. 2019).

In fact, all these methods for determining the robot's trajectory (VO, SLAM, SfM) are based on the same principles. First, the corresponding points on neighboring images are found in one of the known ways, and then direct intersection, resection and/or aerial triangulation are solved to find the spatial position of the images.

The main stage that determines the accuracy of the problem solution is to find the homologue points. Currently, a lot of algorithms for solving this problem have been developed, that are invariant to scale changes in images, their orientation and illumination (Lowe, 2004; Rublee et al., 2011; Leutenegger et al., 2011). However, research shows that these methods do not always give a good result, in the sense of identifying the homologue points, and as a result, the trajectory of the robot is determined with errors.

In order to improve the reliability and accuracy of determining the robot trajectory, this paper proposes to use two stereo cameras. One is directed forward by the robot's movement, and the other is directed at the nadir. Thus, the trajectory is determined twice, independently of each other. Moreover, each case has its own algorithm for finding the homologue points.

Figure 4 shows a schematic diagram for solving the problem of determining the robot's trajectory using two stereo cameras (forward, in nadir).

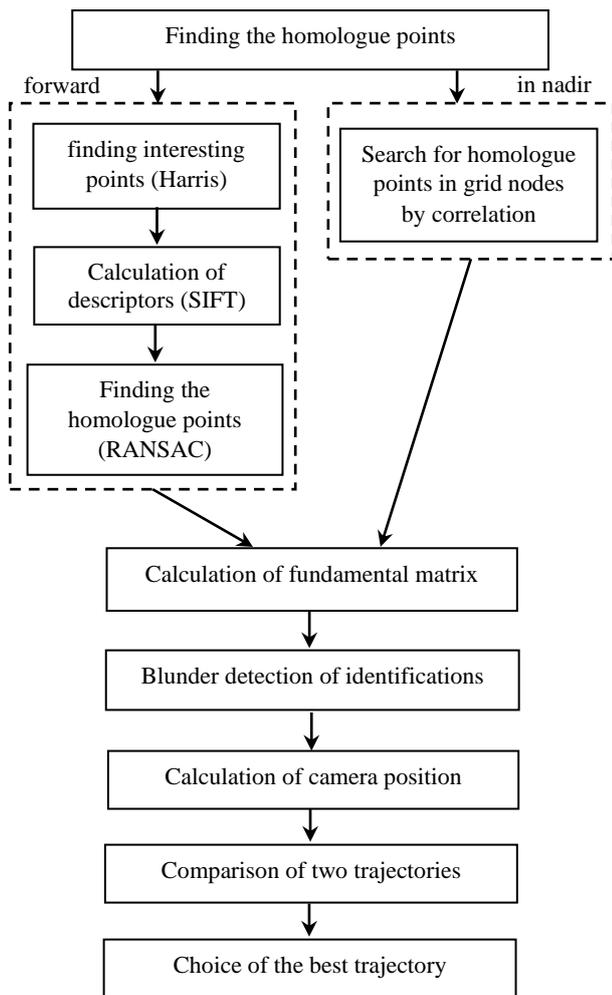


Figure. 4. Schematic diagram of solving the problem of determining the robot trajectory using two stereo cameras.

Finding the homologue points on a pair of images is performed by different methods, depending on the direction of imaging. If the stereo camera is directed forward in the direction of the robot's movement, then the homologue points are found by selecting interesting points on each image, using the Harris operator, calculating the descriptor for each point using the SIFT algorithm, and finally using the RANSAC algorithm to find the homologue points. For a stereo camera aimed at the nadir, first a regular grid is set on one of the images, and then the corresponding points on the other image are found in the nodes of this one using the correlation method. The use of the correlation method in this case is because there are usually few clear contours (the robot moves on asphalt, ground, etc.).

Regardless of the method used to find the homologue points, there is always the possibility of identifications with an error. Therefore, the next stage involves the blunder detection of identifications. To do this, it is enough to make a relative orientation of a pair of images and reject the points based on the value of the residual transverse parallax. Traditional relative orientation takes quite a long time, since there is an iterative process. It is much faster to make relative orientation through the fundamental matrix. The original equation is known to be written as follows:

$$r_1^T \cdot F \cdot r_2 = 0 \quad (1)$$

or in coordinate form

$$(x_1, y_1, z_1) \begin{pmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = 0 \quad (2)$$

Here r_1, r_2 are vectors that determine the position of the corresponding points on the first and second images in the image coordinate systems; F is the fundamental matrix. Equations (2) are linear with respect to unknown elements of the fundamental matrix. To find the elements of the fundamental matrix, it is sufficient to have at least 8 homologue points. However, it is advisable to find the fundamental matrix using the RANSAC method to avoid finding an erroneous matrix. Having found the fundamental matrix, you can detect blunder of identifications by checking each pair of points for compliance with condition (2). As you know, in the right part of expressions (1) and (2) is the volume of the figure built on the vectors r_1, r_2 and the photo base. To say in advance what amount should not exceed this volume for a good identification is quite difficult. Therefore, in the paper (Khrush R.M, et al. 2020), it is proposed to calculate the y-parallax in the model space and use it as a criterion for blunder detection of identifications. However, this requires additional calculations, which is not always acceptable when calculating the robot's trajectory in real time. Therefore, it is proposed in equations (2) to use homogeneous coordinates of the image points and normalize the fundamental matrix to f_{33} , then (2) will be written as:

$$(x_1, y_1, 1) \begin{pmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & 1 \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \\ 1 \end{pmatrix} = 0 \quad (3)$$

In this case, the right side of expressions (3) will contain a value comparable to the residual y-parallax q , which can be used to detect blunder of identifications. To do this, we calculate the root mean square error of the residual y-parallaxes:

$$m_q = \sqrt{\frac{\sum q^2}{n}}, \quad (4)$$

where n is the number of homologue points in the pair of images.

Calculating the position of the cameras (the robot's trajectory) is performed by solving space intersection by the previous stereo pair for corresponding points and resection for images of the subsequent stereo pair. To reduce accumulation errors, bundle adjustment is performed for every 5 stereo pairs, and the entire block is adjusted at the end.

When execute resection and bundle adjustment is also performed blunder detection of identifications. For this purpose, known robust method is used, in which each equation of corrections from the collinearity equations is multiplied by the weight, which is calculated using the following formula:

$$P_i = \begin{cases} 1, & \text{if } |v_i| \leq 2\mu \\ e^{-0.1\left(\frac{|v_i|}{\mu}\right)^4}, & \text{if } |v_i| \leq 2\mu \text{ and } N \leq 3 \\ e^{-0.1\left(\frac{|v_i|}{\mu}\right)^3}, & \text{if } |v_i| \leq 2\mu \text{ and } N > 3 \end{cases} \quad (5)$$

where v_i is the discrepancy in the i correction equation; μ is standard error calculated by v_i ; N is the iteration number.

This way the robot's trajectory is obtained. Moreover, the trajectory is obtained twice for each of the stereo cameras. During imaging, the time when the camera shutter is triggered is recorded. This allows us to compare the two paths, other word to get $\Delta X, \Delta Y, \Delta Z$ between the cameras looking forward and into the nadir. These values must be constant and equal to the values determined as a result of system calibration. If $\Delta X, \Delta Y, \Delta Z$ differ from the calibrated values by more than the specified value, then μ for stereo pairs from different cameras are analyzed. The final trajectory is the one that μ is smaller.

A sparse point cloud obtained from a forward-facing stereo camera describes the surrounding space at inflection points (obstacles). This information is also used to detect obstacles. If the distance to the nearest points in the point cloud is less than the allowed distance, the robot's trajectory changes in order to avoid the obstacle.

When point A or B (figure 2) is in the field of view of the first camera, the algorithm for tracking the distance to the point is enabled. It works in the following way. A point A or B is allocated on each stereo pair, then their relative position on the stereo pair is specified by correlation and the space intersection is solved. Since points A and B (figure 2) are well marked (different colors are used), it is not difficult to find them automatically in the images. If the distance corresponds to the specified distance, the robot stops.

4. RESULT

The algorithm described above is implemented as software in the "Python" language, using standard image processing functions from the OpenCV library. The software is installed on the StereoPic microcomputer by the basic raspberry Pi CM 3 Lite microprocessor, which controls a robot, implemented by a crawler scheme. To evaluate the accuracy and determine the robot's trajectory, several closed paths and paths between two points with a known distance between them were set. A trajectory with arbitrary rotation angles. The minimum trajectory length is 5 m, the maximum is 50 m. The estimation was based on discrepancies between the coordinates of the start and end points of the trajectory. In addition, the distances between two points measured using a tape measure and a stereo camera were compared when the robot moved in a straight line. The relative position error in all experiments was 0.1% - 1% of the trajectory length.

In addition, the accuracy of stopping the robot before points A and B (figure 2) was evaluated. To do this, the robot started moving in the direction of points A and B from a distance of 1 to 10 meters (a total of 15 times). After stopping the robot at a distance of 50 cm from points A and B, this distance was

measured using a tape measure. In all cases, the discrepancy did not exceed 2 cm, with an average value of 1.2 cm.

5. CONCLUSION

The paper considers the possibility of using low-cost stereo cameras for autonomous robot navigation. Experimental studies have shown that the accuracy of determining the coordinates of object points from a pair of images, obtained by such a stereo camera is sufficient for organizing autonomous navigation of the robot. The use of two stereo cameras increases the accuracy and reliability of autonomous robot navigation. However, the results of experimental studies should be considered preliminary. These studies should be continued for various operating conditions of the robot (indoor and outdoor, driving on asphalt, soil, sand, grass, etc.). In addition, the algorithms should be investigated for various implementations of the robot's mechanics (tracks, wheels), its ability to accurately perform specified turns and rectilinear movements.

ACKNOWLEDGEMENTS

The work was performed with the support by Grant 17-29-04410 of Russian Foundation for Basic Research (RFBR).

REFERENCES

- Artisense Corporation, 2020. Visual-Inertial Navigation System (VINS) DevKit. <https://www.artisense.ai/localization/>. Last Accessed: March 2020
- Chibunichev A.G., Makarov S.B., Poliakova E.V. 2019. Study of the possibility of using low-cost stereo cameras for robotic systems navigation. *Izvestiya vuzov «Geodeziya i aerofotosyemka»*. *Izvestia vuzov «Geodesy and Aerophotosurveying»*. 63 (6): 645–649. DOI: 10.30533/0536-101X-2019-63-6-645-649
- Davison, A. J., Reid, I. D., Molton, N. D., Stasse, O., 2007. MonoSLAM: Real-time single camera SLAM. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 29(6), 1052–1067.
- Engel, J., Schops, T., Cremers, D., 2014. LSD-SLAM: Large-scale direct monocular SLAM. *European Conference on Computer Vision (ECCV)*, Springer, 834–849.
- Engel, J., Stuckler, J., Cremers, D., 2015. Large-scale direct SLAM with stereo cameras. *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 1935–1942.
- Frahm, J.-M., Fite-Georgel, P., Gallup, D., Johnson, T., Raguram, R., Wu, C., Jen, Y.-H., Dunn, E., Clipp, B., Lazebnik, S. et al., 2010. Building Rome on a cloudless day. *European Conference on Computer Vision (ECCV)*, Springer, 368–381.
- Khrush R.M., Grin' A.N., Solovev A.V., 2020. The choice of access for the reliable corresponding points identification of the overlapping images. *Izvestiya vuzov «Geodeziya i aerofotosyemka»*. *Izvestia vuzov «Geodesy and Aerophotosurveying»*. 2020, 2: [In Russian].

V. V. Kniaz, 2015. "Fast instantaneous center of rotation estimation algorithm for a skied-steered robot", Proc. SPIE 9528, *Videometrics, Range Imaging, and Applications XIII*, 95280L (2015); <https://doi.org/10.1117/12.2184834>

Leutenegger, S., Chli, M., Siegwart, R., 2011. BRISK: Binary robust invariant scalable keypoints. *IEEE international conference on computer vision (ICCV)*, IEEE, 2548–2555.

Lowe, D. G., 2004. Distinctive image features from scaleinvariant keypoints. *International journal of computer vision*, 60(2), 91–110.

Mur-Artal, R., Montiel, J. M. M., Tardos, J. D., 2015. ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics*, 31(5), 1147– 1163.

Mur-Artal, R., Tard'os, J. D., 2017. ORB-SLAM2: An open-source slam system for monocular, stereo, and rgbd cameras. *IEEE Transactions on Robotics*, 33(5), 1255– 1262.

Paz, L. M., Pini'es, P., Tard'os, J. D., Neira, J., 2008. Largescale 6-DOF SLAM with stereo-in-hand. *IEEE Transactions on Robotics*, 24(5), 946–957

Parrot, 2020. <https://www.parrot.com/global/drones/>. Last Accessed: March 2020

Rublee,E.,Rabaud,V.,Konolige,K.,Bradski,G.R.,2011. ORB: An efficient alternative to SIFT or SURF. *International Conferenceon Computer Vision (ICCV)*, IEEE, 2564–2571.

Scaramuzza, D., Fraundorfer, F., 2011. Visual Odometry [Tutorial]. *IEEE Robotics Automation Magazine*, 18(4), 80-92.

Shtain, Z. and Filin, S. 2019. TOWARDS AN ACCURATE LOW-COST STEREO-BASED NAVIGATION OF UNMANNED PLATFORMS IN GNSS-DENIED AREAS, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W17, 323–329, <https://doi.org/10.5194/isprs-archives-XLII-2-W17-323-2019>.

Smolyanskiy, N., Gonzalez-Franco, M., 2017. Stereoscopic First Person View system for Drone navigation. *Frontiers in Robotics and AI*, 4, 11

Strasdat, H., Davison, A. J., Montiel, J. M. M., Konolige, K., 2011. Double window optimisation for constant time visual SLAM. *2011 International Conference on Computer Vision (ICCV)*, 2352–2359

Wang, X., Rottensteiner, F., Heipke, C., 2019. Structure from motion for ordered and unordered image sets based on random kd forests and global pose estimation. *ISPRS Journal of Photogrammetry and RemoteSensing*,147,19– 41.

ZED stereo camera. StereoLabs, 2020. <https://www.stereolabs.com/solutions/robotics/>. Last Accessed: March 2020