This is a post-peer-review, pre-copyedit version of an article published in **2017 Baltic Geodetic Congress (BGC Geomatics)**. The final authenticated version is available online at: https://doi.org/10.1109/BGC.Geomatics.2017.21

Towards Precise Visual Navigation and Direct Georeferencing for MAV Using ORB-SLAM2

Pawel Burdziakowski
Faculty of Civil and Environmental Engineering,
Gdansk University of Technology,
Gdansk, Poland
pawel.burdziakowski@wilis.pg.gda.pl

***Abstract*—A low accuracy of positioning using Global Navigation Satellite System (GNSS) are not meet geodetic requirements for direct images georeferencing for Unmanned Aerial Vehicle (UAV) photogrammetry. A majority of UAVs are equipped with a monocular or stereo non-metric cameras for either visual data gathering or live video feed for operator. A cheap positioning techniques used on board commercial UAVs are not that precise as geodetic community requires. Moreover, a traditional satellite navigation suffers from a multipath Global Positioning System (GPS) signal in an industrial environment and cities or during operation in poor GPS space segment geometry or in indoor environment. In that cases, the UAV position can be enhanced by a visual navigation method. The paper presents a results of implementation of recently developed robust simultaneous localization and mapping method for a low cost hexacopter platforms navigation system augmentation, in order to support a precise and robust direct images georeferencing.**

Keywords—unmanned aerial vehicles; navigation; simultaneous localization and mapping.

I. INTRODUCTION

A direct georeferencing, a new method in photogrammetry, is referred to a Direct Platform Orientation (DPO). The method is based on two navigation devices, Global Navigation Satellite System (GNSS) for X_p, Y_p and Z_p platform coordinates and Inertial Measurement Unit (IMU) for roll, pitch and yaw platform orientation. With known relationship between navigation devices location and orientation on board the platform and imaging sensor location and orientation it is possible to calculate Exterior Orientation Parameters (EOP), including the three camera position coordinates (X_c, Y_c, Z_c) and the three attitude angles of the camera (ω, ϕ, κ). In traditional photogrammetry EOP are derived from Aerial Triangulation (AT), where Ground Control Points (GCP) are prerequisite [1]. Theoretically, when platform is able to record EOP directly form navigation and orientation system, GCP are not required. This is a perfect solution for areas or objects where measuring GCP is unmanageable, or areas are unreachable and traditional geodetic measurement methods are unavailable (sea cliffs, high buildings, etc). Due to the fact, that in the direct georeferencing method, navigation module precision is crucial for overall process accuracy, sensors became fully integrated and navigation data fusion is a fundamental step [2]. In fact, as for the airborne photogrammetry, where platform is able to carry high accuracy and substantial weight GNSS and INS it is able to possess precise platform position and orientation [3]. On the other hand, there is a recently brought into general use the UAV photogrammetry. This technique of carrying sensors requires low weight and small size navigation solutions, which delivers significantly lower accuracy. Small IMU sensors, used in Micro Air Vehicles (MAV) or small UAVs are based on microelectromechanical system (MEMS) technology. MEMS IMU accuracy and performance varies and cannot be refereed as a standalone navigational system [4]. Also, a low accuracy of positioning using small GNSS modules on MAV are not meet geodetic requirements for direct images georeferencing, and cannot be applicable for photogrammetry with a reliable precision [5].

In order to, increase accuracy of navigation module for low-cost MAV [6], in this paper it is proposed to integrate GNSS and INS data with trajectory and orientation calculated using images obtained by onboard stereo camera. A process of navigation which uses images for localization is referred to comparative navigation methods, as described in [7]. In this research the trajectory calculation is performed using the recently developed, real-time simultaneous localization and mapping (SLAM) algorithms based on Oriented FAST (Features from Accelerated Segment Test) and Rotated BRIEF (Binary Robust Independent Elementary Features) features [8]. This algorithm, called ORB features is composed with FAST keypoint detector [9] and BRIEF descriptor [10]. Both techniques present good performance in commuter vision (CV) technologies. SLAM algorithm based on ORB features and implemented for stereo images is called ORB-SLAM2 [11].

In the previous work [12], where SIFT (scale-invariant feature transform) algorithm were used with monocular camera, it was assumed that, this approach not guarantee own trajectory and orientation calculation in the real time. The research [8] shows that, ORB is an order of magnitude faster

than other very popular algorithm called SURF (Speeded Up Robust Features), and over two orders faster than SIFT.

In the research [11] it is proven that ORB-SLAM2 is able to calculate MAVs coordinates and orientation based on stereo images in real time, and it delivers the best trajectory precision among other SLAM algorithms in a different environments. Evaluation of ORB-SLAM2 was performed on a three most popular data sets: KITTI dataset [13] for a car and urban environments, EuRoC dataset [14] for MAV, and TUM RGB-D dataset for indoor environment [15]. Among all navigation and reference data, those datasets delivers images recorded during tests. Images for two datasets are rectified, and for one are not rectified. Cameras used in stereo systems used in those projects presents a relatively low distortions parameters, as compared to fisheye cameras. ORB-SLAM2 originally was not tested on low-cost fisheye cameras. Researchers community recently started to implement ORB-SLAM2 algorithms based on fisheye cameras in multi camera configuration [16], however there is still lack of reliable work in this area.

A. *System Setup*

Based on research [12], in order to recover own track referenced to a word coordination system, a right scale of own trajectory are to be calculated during additional process. Using stereo camera it is possible to recover track scale without additional measurements and processes. Due to that reason, it has been decided to use the same GoPro Hero 3 Black Edition camera like in [12], but in a stereo configuration. Both cameras has been mounted on a stereo rig. Two stereo rigs has been prepared, with base line 60 mm and 200 mm. Both cameras are synchronized and started at the same time via remote WiFi module. The stereo rig had been mounted on tested hexacopter platform [6]. All navigation data for referencing are logged on board autopilot using two GNSS modules and INS. Synchronization method between referenced position and images were conducted as in [5].

Fig. 1. Stereo rig with GoPro cameras mounted on tested MAV

B. *Camera Calibration*

During the process of nonmetric camera calibration, the intrinsic parameters of the camera are determined, as well as distortion parameters. The intrinsic parameters describe the metric characteristics of the camera needed for photogrammetric processes. It is assumed, that calibration process is crucial for the presented work and camera configuration will play important role, as for the accuracy results. The calibration was conducted using methods [17], implemented in C++ language using OpenCV library [18] on Ubuntu 16.04 LTS (Long Term Support) operation system. The fisheye camera calibration model was used for calibrating process, as in [19]. The GoPro cameras generate images that are significantly distorted and calibration approach as in [20], which is designed for low distorted images, is unacceptable (Fig. 2).

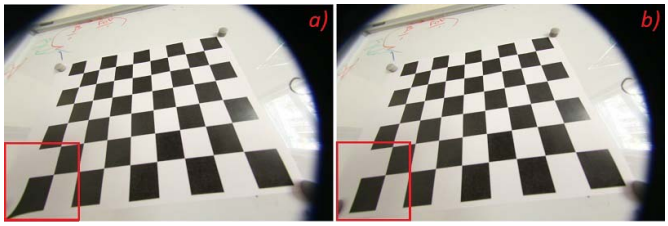

Fig. 2. Differences between rectyfication of fisheye camera using model for low distorted cameras (a) and fish-eye (b) [18]

For the stereo camera configuration, it is also obligatory to recover the extrinsic parameters [21]. During the calibration process extrinsic and intrinsic parameters for presented camera configuration was calculated.

II. IMPEMENTATION

ORB-SLAM2 was originally developed and tested on Ubuntu operation system using C++ 11 thread and chrono functionalities, in order to run multithread calculation. The same approach and implementation was used in the presented research. The calculations have been run on mobile computer with 8 GB RAM and Intel Core I5-2410M CPU at 2.3 GHz. The Ubuntu operation system is the base operation system for robust embedded systems designed for graphic and commuter vision operations (e.g. NVIDIA Jetson TX2). For the presented work, the Ubuntu operation system was chosen due to the reason that, it can be simply run on embedded system, without significant work.

III. RESULTS

A. *Calibration*

Due to the reason, that the calibration process in crucial in photogrammetry tasks and will affect final result, three different camera calibration patterns were used (Table I). Patterns used for the calibration are based on chessboard corners, sizes as in Table I.

TABLE I. CALLIBRATION PATTERNS DETAILS

No.	Pattern details		
	Width (squares)	*Height (squares)*	*Square size (mm)*
1	9	5	25
2	13	9	25
3	8	5	40

ORB-SLAM2 implementation uses a configuration file to pass all arguments to algorithm. First part of the algorithm parameters list are detailed intrinsic and extrinsic camera parameters including distortion coefficients for each camera. The algorithm is able to use the calibrated and stereo rectified image pairs or it can process the uncalibrated images. In both ways obligatory input data are: left and right camera matrix (A), left and right camera distortion coefficients (radial coefficients k_1, k_2, and tangential p_1, p_2) (if using rectified images for algorithm, distortions coefficients are assumed zero), size of the images in pixels, rectification transform (rotation matrix R_1) for the left camera, rectification transform (rotation matrix R_2) for the right camera, projection matrix (P_1) in the new (rectified) coordinate systems for the left camera and projection matrix (P_2) in the new (rectified) coordinate systems for the second camera, the stereo rig b_f parameter (calculated by multiply camera *fx (*from projection matrix P_1, resulted in the stereo calibration*)* and stereo baseline in meters). All above compulsory matrixes are defined in accordance with [18]. What is important, ORB-SLAM2 algorithm uses above data for rectification only detected feature points, and not for entire image points.

The radial distortion coefficients were obtained in calibration process using fish-eye camera model [19] up to k_4 radial coefficient, tangential coefficients are assumed zero. Detailed cameras intrinsic parameters and distortions coefficients are presented in Table II.

TABLE II. DETAILLED CAMERAS PARRAMETERS

Left Cam.	**Calibration pattern no. 1**	**Calibration pattern no. 2**	**Calibration pattern no. 3**
fx	348.10082114039	376.47547629978	347.58072962486
fy	358.24495827426	376.23007127923	351.04413716858
cx	386.56743589748	385.08410093667	384.51404377562
cy	206.91858917630	198.60617928093	199.81581667963
k1	-0.01562142665	-0.16627353302	-0.00636529234
k2	0.05509976923	0.59787783285	0.04654113072
k3	0.06301262495	-0.87274640148	0.03959455985
k4	-0.06298664098	0.43886169373	-0.03148326848
Err	**0.43165342947**	2.45300763052	0.71865006544
Right Cam.	**Calibration pattern no. 1**	**Calibration pattern no. 2**	**Calibration pattern no. 3**
fx	348.00082114015	397.17794883441	348.02907936745
fy	357.94495827426	397.76774781281	351.83762703559
cx	382.26741238749	381.30932160935	376.53481715598
cy	202.91858917233	198.12907253341	198.30320603185
k1	-0.01542142665	-0.18344175595	-0.02982282388
k2	0.05499976923	0.46636102951	0.11781117613
k3	0.06281262493	-0.55003195040	-0.05115326827
k4	-0.06308664098	0.21333727605	0.00761778501
Err	**0.43365342947**	1.93228887627	0.71090927164

As the calibration results show, average reprojection error (*Err*) has the smallest value for the calibration performed with set of images with calibration pattern no. 1. The reprojection error is the distance between a pattern keypoint detected in a calibration image, and a corresponding world point projected into the same image. The average reprojection error provides a qualitative measure of calibration accuracy [18]. All images used for processing were undistorted with parameters obtained with calibration pattern no. 1.

The stereo calibration was conducted using undistorted images form cameras calibration process. In this step P_1, P_2, R_1, R_2 matrixes were calculated. The stereo calibration process for rig 60 mm resulted with success, algorithm were able to estimate the relative position and orientation of both stereo cameras and compute the rectification transformation (first stereo pair on Fig. 3). The stereo rectification was performed for using all calibration patterns.

For the 200 mm rig the stereo calibration algorithm were unable to correctly calculate respective matrixes. The rectified stereo pairs were distorted and respective points were not laid on the same epipolar lines (second and third stereo pairs on Fig. 3). It is assumed. that patterns for 200 mm rig stereo calibration were placed to close distance from the cameras. For this stereo rig, calibration process should be performed using a large size calibration pattern and placed distant from the rig. In that case the algorithm would be able to correctly calculate matrixes.

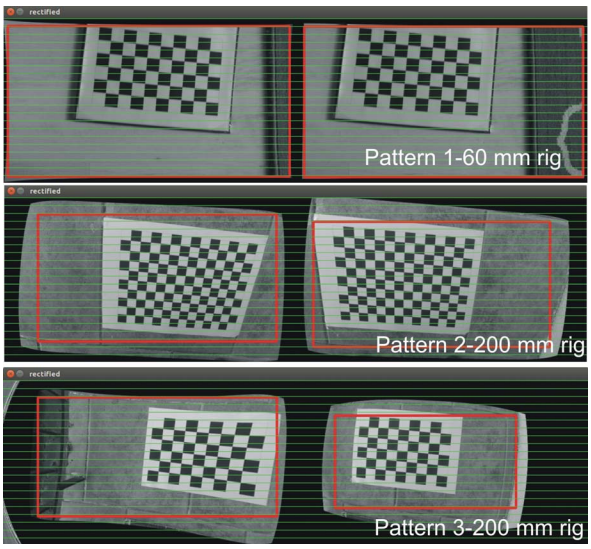

Fig. 3. Results of stereo rectification (first pair correct, socond and third incorrect stereo rectyfication)

B. *Trajectories*

The testing trajectories was located in the urban environment where GPS suffers from multipath or poor segment geometry [22] (Fig. 4). The flight altitude was planned to *1.5 m* above the ground, in order to enhance bad segment geometry. The test flight location stimulates main

problems for MAVs navigation based on GNSS caused by urban environment [5]. The number of satellites fixed during test flight are less than 10 for GPS1 and HDOP for GPS1 in between *1* and *1.5*. GPS2 suffers severe interferences and was able to fix only *4* satellites at few moments during the flight (Fig. 5). The flight was carried out in altitude hold mode (autopilot controls only altitude), the flight was controlled by MAV pilot (man controlled). The realization of flight in mode based on GPS stabilization (loiter mode) in this particular environment was too dangerous, due to the high GPS HDOP. Normally, the safe flight in the loiter mode can be performed when HDOP stable and less than *1*. In that case, if the MAV is stabilized in loiter mode, any GPS position deviation would result in the rapid MAVs maneuver, what is unacceptable.

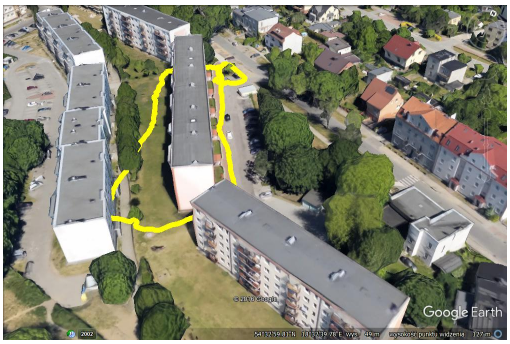

Fig. 4. The urban environment - location of the flight track (yellow lines-GPS possitions recorded during flight).

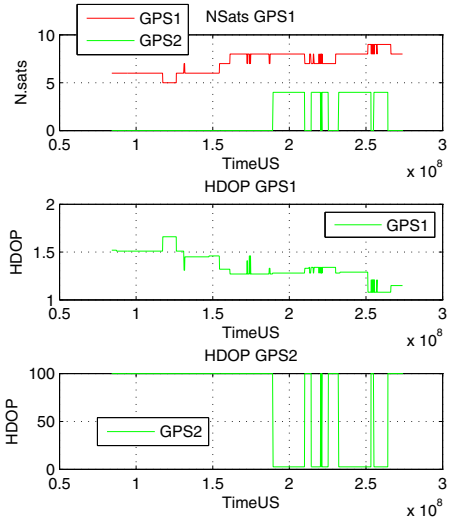

Fig. 5. GNSS HDOP and number of sats fixed (NSATS)

The ORB-SLAM2 uses ORB features [8] for tracking, mapping and place recognition tasks. For the recorded track number of ORB features (Table III) was set to *3000*. If this parameter was set less than *2000* tracking was interrupt, and trajectory was lost. Despite the fact that, the ORB features are robust to the rotation and scale to less number of features, during rapid rotations track was lost while MAV was changing the course (heading). The ORB features presents good invariance to camera auto-gain and auto-exposure, and illumination changes and was also resistant to "jello effect". That effect appears when the camera is vibrating when shooting from the MAV. In this case cameras was mounted without active stabilization system (gimbal), and "jello effect" affected images significantly. The ORB features are fast to extract and match allowing for real-time operation, faster than used in previous work [12] SIFT features, and allowed to calculate tracks in real-time for 15 frames per second (FPS).

TABLE III. ORB-SLAM2 INPUT PARAMETERS FOR CALCULATIONS

Trajectory	**Pattern details**			
	Calibration pattern	***FPS***	***ORB features***	***b_f***
D1	1	15	3000	22.651538
D2	2	15	3000	30.423335
D3	3	15	3000	21.385282
D4	3	15	4000	21.385282
D5	1	15	3500	22.781588

The results, visual tracks and GPS1 track are presented in Fig. 6. The relations between GPS and ORB-SLAM2 trajectories (Vision D1, Vision D2, Vision D3, Vision D4, Vision D5 on Fig. 6) were calculated using unit quaternions [23], which describes the relationship between two coordinate systems using pairs of measurements of the coordinates of a number of points in both systems.

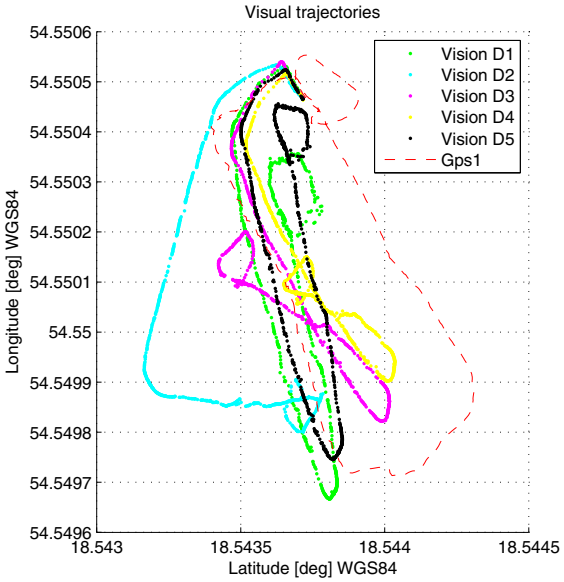

Fig. 6. The visual and GPS trajectories.

IV. CONCLUSIONS

The paper presents the results of preliminary research on visual navigation for MAV using fisheye stereo camera system in order to support direct images georeferencing. The visual navigation is based on recently developed and presented ORB-SLAM2 algorithm, which allows precisely calculate

own trajectory in real time. This attribute allows to implement algorithm directly on MAV, using either lightweight companion computers or robust embedded systems.

The results are not as precise as presented on the original work [11], mainly due to different camera system. The cameras used in the presented research (Gopro Hero 3 Black) mounted in the designed stereo rig present a low-cost solution, as for stereo vision. The calibration process, and stability of extrinsic parameters is low, in comparison to approach presented in data sets [13]–[15]. The stereo rig was moulded in plexiglas material and was vulnerable to MAV vibrations. Nevertheless, the low-cost solution showed in the paper proves, that real time calculation and precise visual navigation are affordable for a MAV, in order to support direct georeferencing for UAV photogrammetry.

The calibration process in presented solution turned out to be a crucial part of the research. The calibration process of the same camera resulted in slightly deferent results when using deferent calibration pattern, what resulted in the different visual trajectory calculation. It is expected, that an improvement in this part of the process will result in the accuracy progression. It is planned to use new stereo calibration method, as described in [24], for improve the fisheye camera calibration process. Moreover, used algorithm for the 200 mm rig stereo calibration was not able to correctly compute the parameters. It was expected, that longer stereo base will effect in accuracy improvement, nevertheless this thesis has not been confirmed in this research.

The ORB features proved to be invariance to camera auto-gain and auto-exposure, and illumination changes and was also resistant to "jello effect". That effects are common on MAV used low cost camera usually set to automatic mode. The trajectory calculations were performed without significant disturbances.

Future work will focus on real-time integration of data streamed from stereo camera. The new stereo camera will be use, camera which guarantee the exterior and interior parameters stability and on the fly image and GPS data synchronization.

References

[1] A. Rizaldy and W. Firdaus, "Direct Georeferencing: a New Standard in Photogrammetry for High Accuracy Mapping," ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci., vol. 39, no. September, pp. 5–9, 2012.

[2] A. Stateczny and I. Bodus-Olkowska, "Sensor data fusion techniques for environment modelling," in 2015 16th International Radar Symposium (IRS), 2015, pp. 1123–1128.

[3] D. Grejner-Brzezinska, "Direct exterior orientation of airborne imagery with GPS/INS system: performance analysis," Navig. J. Inst. Navig., vol. 46, no. 4, pp. 261–270, 2000.

[4] A. Kealy, G. Retscher, and D. Grejner-Brzezinska, "Evaluating the Performance of Mems Based Inertial Navigation Sensors for Land Mobile Applications," Arch. Fotogram. Kartogr. i Teledetekcji, vol. 22, pp. 237–248, 2011.

[5] P. Burdziakowski and K. Bobkowska, "Accuracy of a low-cost autonomous hexacopter platforms navigation module for a photogrammetric and environmental measurements," in Environmental Engineering 10th International Conference, 2017.

[6] P. Burdziakowski, "Low cost hexacopter autonomous platform for testing and developing photogrammetry technologies and intelligent navigation systems," in Environmental Engineering 10th International Conference, 2017.

[7] A. Stateczny, The methods of the comparative navigation. Gdansk: Gdanskie Towarzystwo Naukowe, 2004.

[8] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in Proceedings of the IEEE International Conference on Computer Vision, 2011, pp. 2564–2571.

[9] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2006, vol. 3951 LNCS, pp. 430–443.

[10] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary robust independent elementary features," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2010, vol. 6314 LNCS, no. PART 4, pp. 778–792.

[11] R. Mur-Artal and J. D. Tardos, "ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras," arXiv, no. October, 2016.

[12] P. Burdziakowski, A. Janowski, M. Przyborski, and J. Szulwic, "A modern approach to an unmanned vehicle navigation," in 16th International Multidisciplinary Scientific GeoConference SGEM 2016, Book 2, 2016, vol. 2, no. SGEM2016 Conference Proceedings, ISBN 978-619-7105-59-9 / ISSN 1314-2704, pp. 747–758.

[13] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," Int. J. Rob. Res., vol. 32, no. 11, pp. 1231–1237, 2013.

[14] M. Burri et al., "The EuRoC micro aerial vehicle datasets," Int. J. Rob. Res., vol. 35, no. 10, pp. 1–7, 2016.

[15] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of RGB-D SLAM systems," in IEEE International Conference on Intelligent Robots and Systems, 2012, pp. 573–580.

[16] S. Urban and S. Hinz, "{MultiCol-SLAM} - A Modular Real-Time Multi-Camera SLAM System," arXiv Prepr. arXiv1610.07336, 2016.

[17] J.-Y. Bouguet, "Complete Camera Calibration Toolbox for Matlab," Jean-Yves Bouguet's Homepage. 1999.

[18] O. S. Project, "OpenCV (Open source Computer Vision)," Internet, 2012. [Online]. Available: http://opencv.org/.

[19] J. Kannala and S. S. Brandt, "A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses," IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 8, pp. 1335–1340, 2006.

[20] Z. Zhang, "A flexible new technique for camera calibration," IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 11, pp. 1330–1334, 2000.

[21] R. I. Hartley, "Theory and practice of projective rectification," Int. J. Comput. Vis., vol. 35, no. 2, pp. 115–127, 1999.

[22] A. Nowak, "The Proposal to 'Snapshot' Raim Method for Gnss Vessel Receivers Working in Poor Space Segment Geometry," Polish Marit. Res., vol. 22, no. 4, pp. 3–8, 2015.

[23] B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," J. Opt. Soc. Am. A, vol. 4, no. 4, p. 629, 1987.

[24] S. Urban, J. Leitloff, and S. Hinz, "Improved wide-angle, fisheye and omnidirectional camera calibration," ISPRS J. Photogramm. Remote Sens., vol. 108, pp. 72–79, 2015.