2018 24th International Conference on Pattern Recognition (ICPR)
Beijing, China, August 20-24, 2018

Real-time Vehicle Localization and Tracking Using Monocular Panomorph Panoramic Vision

B. Akdemir, A.N. Belbachir, *IEEE Member,* and L.M. Svendsen
Smart Instrumentation Department
Teknova AS
Grimstad, Norway
{bak ; anb ; lms} @ teknova.no

***Abstract*—This paper presents a feasibility analysis of the ORB-SLAM [1] for real-time vehicle localization and tracking using a monocular visual camera providing 360˚ panoramic views. This method described in [1] was initially designed and developed for conventional cameras, making use of a method for detection and tracking visual features and estimating the camera trajectory while reconstructing the environment. The accuracy of the tracking depends on the ability of this method to robustly detect and match sufficient visual features. This work aims to extend this method to large monocular round views using fish-eye-like cameras allowing an increase of visual features with the aim of improving localization robustness. The main challenge in using a standard fish-eye camera for generating panoramic views is the reduction of visual performance due to a potential higher distortion and lower spatial resolution compared to that using a standard camera lens. The objective of this research is to perform a feasibility analysis of a method combining a camera equipped with a panomorph lens to generate real-time panoramic views at minimal distortion and ORB-SLAM to robustly detect and track visual features for real-time camera localization and tracking. A quantitative evaluation is performed on a vehicle driving in an outdoor natural scene with the monocular panomorph camera mounted on-front and without any other additional sensors. The results with analysis and a concluding summary are included as well.**

***Keywords*— panomorph; panoramic vision; visual SLAM; localization and tracking.**

I. INTRODUCTION (*HEADING 1*)

In the past decade, the real-time visual Simultaneous Localization and Mapping (SLAM) [2][3][1][4][5][6] has gained higher attention and larger contributions from the scientific community, driven by the large interest and needs within the self-driving autonomous cars sector. Visual SLAM methods make use of images as input data with the objective of estimating the camera trajectory while reconstructing the visual environment, by applying Bundle Adjustment (BA) for camera localization and sparse geometric reconstruction. This reconstruction is very relevant for application in robotics and autonomous systems when cameras are mounted on-board mobile platforms (e.g. vehicles, aircrafts and drones). Two categories of visual SLAM methods are typically used with monocular cameras. The first category, called feature-based visual SLAM methods [1][2] that typically detect and match a set of visual features across a sequence of images using descriptors extracted from every image frame. As most of the information contained in each image is discarded, matching sparse keypoints leads to efficient BA and computation at the cost of reduced accuracy and robustness. The second category, called direct methods [6], makes use of the full image to directly minimize the photometric error between an image and a map for increased accuracy, dense reconstructions and some robustness to viewpoint change and blur. The main limitation for these methods is the implicit assumption of static scene illumination required for the photometric error metric, which is only valid in controlled environments (e.g. indoor) and not in natural dynamically changing environments (e.g. outdoor). The recent feature-based ORB-SLAM method [1], selected for this research, has proven to provide real-time camera localization in natural outdoor environments. This method was developed and evaluated on standard images with limited field of view and thus limited visual features.

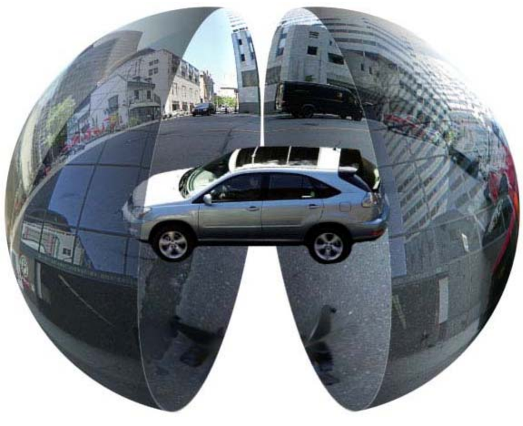


Fig. 1. Illustration of panoramic vision around a vehicle (credit [15]).

Camera systems providing wide panoramic views (i.e. up to 360˚ views) enable enhanced capabilities for wide area imaging [8] and extraction of a larger set of visual features, which is very useful in autonomous navigation, localization and mapping (see illustration in Fig. 1). Majority of these camera systems make use of at least one of the following four principles to record full panoramic views: i) systems using wide angle lenses e.g. fish-eye lenses, ii) systems using mirror(s), e.g. catadioptric cameras, iii) systems using multiple cameras, e.g. poly-dioptric cameras and ii) those using scanning platforms, e.g. multi-perspective cameras. A detailed classification of panoramic cameras and a comparison of existing cameras is given in [10][9]. The categories i) and ii) face a challenge with undesirable distortion that limits the optical resolution. On the other hand, the categories iii) and iv) should deal with bulky systems and a complex mechanics limiting compact deployment and the system lifetime.

Within this research, we are interested in involving a camera system equipped with a panomorph lens [12][14] to keep simplicity offered by category i), providing real-time panoramic views at minimal distortion. We evaluate the combination of this camera system for generating a larger set of features and ORB-SLAM to robustly detect and track visual features for real-time camera localization and tracking. The paper is structured as follows: In section II, a brief review of the monocular ORB-SLAM is given. Section III presents imaging capabilities using a camera equipped with the panomorph lens and its evaluation with respect to the camera projection models. The feasibility analysis of the ORB-SLAM on images generated by a panomorph camera for real-time deployment given in section IV including the calibration approach. Section V provides a short concluding summary.

II. REVIEW OF THE MONOCULAR ORB-SLAM

ORB-SLAM algorithm [1] performs detection, matching and tracking of visual features in real-time by making use of a sequence of images generated from a monocular camera as input. It incorporates the tasks of tracking, local mapping, re-localization after tracking failure and loop closing, which occurs when the trajectory of the camera is on its own previous path/trajectory. Fig. 2 depicts flowchart of the algorithm [1]. It aims to estimate the trajectory of the camera while at the same time reconstruct the surroundings around the camera. By visual it means using images or/and video. Features in the images are being detected by using the orientated FAST corner detection method. FAST is a feature detector method which looks at the pixel intensities in an image without consideration of the orientation. ORB descriptor is a combination of an orientated FAST keypoints added to a rotated BRIEF feature descriptor. BA is used to provide accurate estimates of the camera localization. ORB-SLAM provides BA with corresponding observation of features among a subset of selected frames.

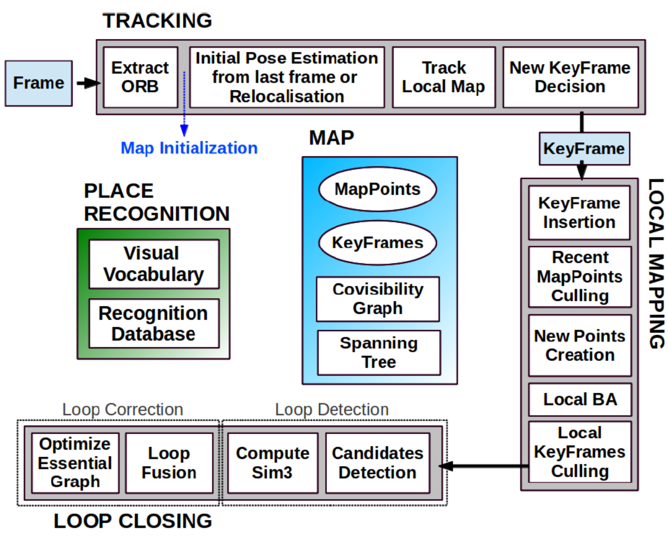


Fig. 2. Flowchart of the monocular ORB-SLAM method (credit[1]).

The algorithm (Fig. 2) consists of three threads which run in parallel; tracking, local mapping and loop closing. In the tracking thread, the camera is being localized within every frame, using the features that are being detected. This is also the thread that deals with initialization of the system when starting, and also re-localization in case the tracking is lost. The local mapping process new images and performs calculations (BA) to get an optimal reconstruction of the surroundings of the camera. In this thread new features are being matched with unmatched features from previous images to see if they correspond. The loop closing thread is looking for loops in every image, and if a loop is detected calculations are performed to adjust the trajectory if drifting has occurred.

III. PANOMORPH PANORAMIC VISION

This section provides a short description of panoramic vision using a camera equipped with a panoramic lens and the projection models related to.

A. Panoramic Vision Using a Panomorph Lens

The panomorph lenses [15] developed by ImmerVision are advanced hemispheric wide-angle lenses and are designed with patented anamorphosis (optimal sensor coverage) and/or magnification (targeted distortion), to enable covering more pixels and magnify desired zone of interest by controlling the distortion through optical design. The optical distortion is used to magnify zones of interest and to increase resolution, which enables to distribute the pixel density along the detector photosensitive surface. The concept anamorphosis provides optimal pixel coverage where the image is being stretched to optimize coverage on a rectangular sensor. The comparative analysis of the panomorph lens and fish-eye lens of the resolution gain shows the following[13]:

1. Using a standard fisheye lens on a VGA imager (640×480 pixels), the resulting resolution is a constant over the entire field-of-view and can be calculated as follows:

$$R_{fisheye} = 480 \text{ pixels} / 190 \text{ deg} = 2.52 \text{ pixels} / \text{deg} \tag{1}$$

2. With a panomorph lens, the image mapping can be managed in a manner that the resolution for each zone will be the same, and we can decrease the resolution in the inter-zone (between the forward and side views). This approach is claimed to provide a better control over the resolution. The panomorph lens includes an anamorphic correction, hence the 190 degree field-of-view will be spread not only over 480 pixels but on the longer axis, and will use 640 pixels (30% more). If we consider that the inter-zone also covers 38 degrees and that the resolution can be reduced by a factor two in this zone, the new resolution of the panomorph lens is:

$$R_{panomorph} = 160 \text{ pixels} / 38 \text{ deg} = 4.21 \text{ pixels} / \text{deg} \tag{2}$$

3. In this calculation, it is considered that 160 pixels are required to image each zone of interest and 80 pixels are required for each inter-zone for a total of 640 pixels.

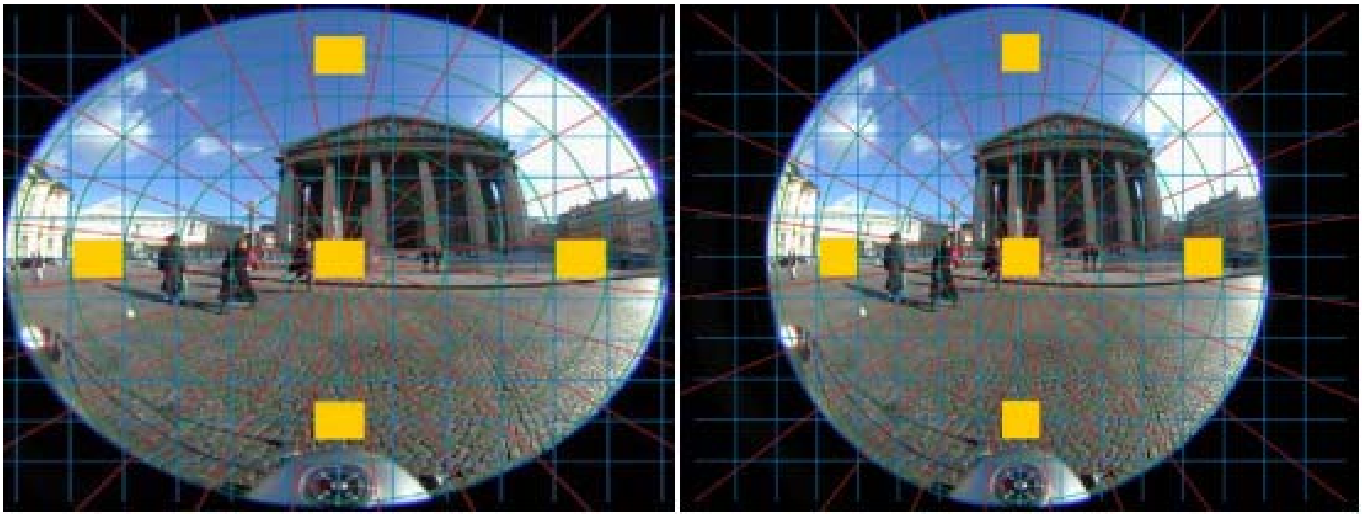


Fig. 3. Images acquired with a camera equipped with panomorph lens (left) and a conventional fisheye lens (right). Yellow boxes represent equivalent areas. (credit [13]).

The resolution provided by the panomorph lens seems to be 170% higher than the resolution provided by a conventional fisheye lens in the zones of interest. For the inter-zones, the resolution is only 10% smaller. This example shows how a custom-designed panomorph lens can provide a higher resolution where required, while still providing full hemispheric views. It is also possible to have a lower resolution in the inter-zones. Fig. 4 shows a front view image, taken by a fisheye lens and a panomorph lens with equidistance projection. The yellow boxes show the relative dimension of an object in the field. By performing an anamorphic correction (elliptical footprint), the panomorph lens also provides a gain in resolution along the long axis of the sensor.

B. Camera Projection Models

In [11], the different projection models of camera configurations are provided, which are summarized in this section. The perspective projection of a pinhole camera can be described by the following formula:

$r = f \tan \Theta$ (i. perspective projection) (3)

where Θ is the angle between the principal axis and the incoming ray, r is the distance between the image point and the principal point and f is the focal length.

Conventional fish-eye lenses instead are usually designed to obey one of the following projections:

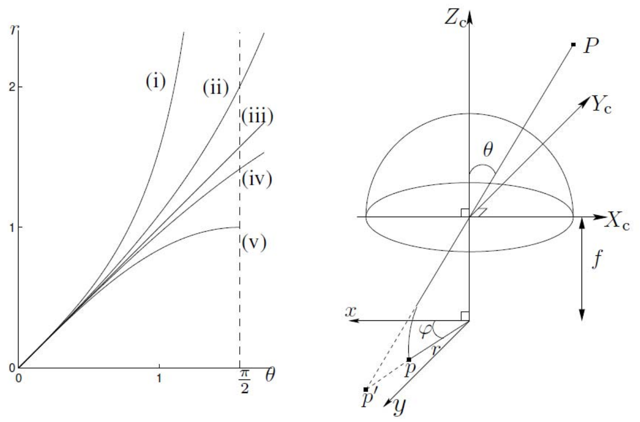
$r = 2f \tan \Theta\ (\Theta/2)$ (ii. stereographic projection) (4)

$r = f \Theta$ (iii. equidistance projection) (5)

$r = 2f \sin(\Theta/2)$ (iv. equisolid angle projection) (6)

$r = f \sin(\Theta)$ (v. orthogonal projection) (7)

Perhaps the most common model is the equidistance projection. The behavior of the different projections is illustrated in Fig. 4(a) and the difference between a pinhole camera and a fish-eye camera is shown in Fig. 4(b).



(a) (b)

Fig. 4. (a) Projections (1)-(5) with f = 1. (b) Fish-eye camera model. The image of the point P is p whereas it would be p0 by a pinhole camera. (credit [11]).

Panomorph lenses do not, however, exactly follow the designed
projection model. From the viewpoint of automatic calibration, it would also be useful to have only one model suitable for different types of lenses. Therefore, projections can be considered in the general form of:

$$r(\Theta) = k_1\Theta + k_2\Theta^3 + k_3\Theta^5 + k_4\Theta^7 + k_5\Theta^9 + \ldots, \quad (8)$$

where, without any loss of generality, even powers have been dropped. This is since r may be extended onto the negative side as an odd function while the odd powers span the set of continuous odd functions. For facilitating computations, the number of terms in (8) should be fixed. It is assumed that the first five terms, up to the ninth power of Θ, give enough degrees of freedom for good approximation of different projection curves.

IV. Feasibility Analysis of the ORB-SLAM Using Panomorph Panoramic Vision

In this section, the test setup, the calibration procedure and the evaluation results are provided and discussed.

A. *Test Set-up*

For the feasibility analysis and evaluation of this approach, two test runs were performed. First, a data collection run was performed in the city of Kristiansand in Norway (see the vehicle trajectory in Fig. 5) where more than 30 minutes of video data were collected for a trajectory of 6 km.

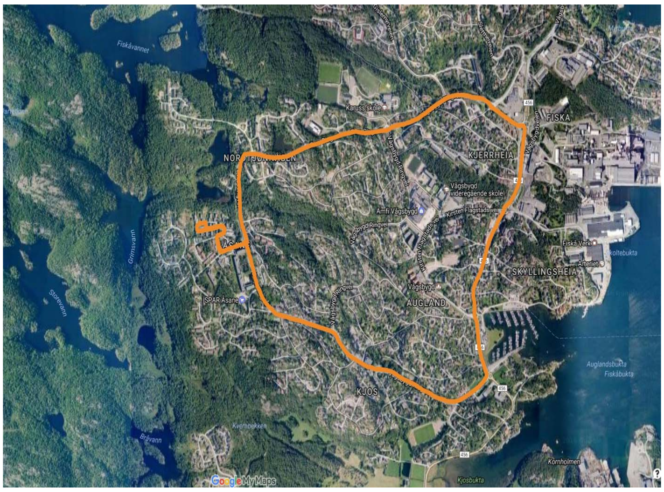


Fig. 5. Overview of the trajectory for data collection.

Fig. 6 depicts the illustration of the on-vehicle test set-up system for real-time localization and tracking. The development hardware consists of a visual camera equipped with a panomorph lens for 360° imaging. It is mounted on the front of the car in order to visually sense the surroundings in 360° panoramic views. A high performance mini PC is attached to the camera to register and process data and estimate the vehicle location on the image views. The camera has a 6 megapixels resolution generating images at 25 frames per second. The mini PC is a Zotac Magnus PC with a quad core of 3.6 GHz

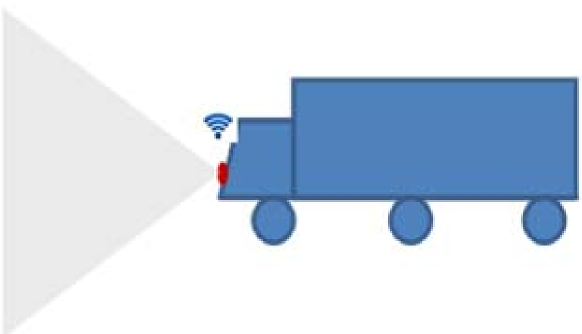


Fig. 6. Illustration of the camera mounting position on-board a vehicle.

B. *Calibration*

The camera calibration [7] is one of the key tasks in this research for providing accurate feature matches and robust localization. It estimates the distortion parameters, the intrinsic and the extrinsic parameters of a lens. The distortion parameters describe how much the lens distorts an object shape from the world 3D-points to its view as 2D image points. The intrinsic parameters describe the focal length and the optical center of the camera, whereas the extrinsic parameters describe the location of the camera within the world scene (3D).

These camera parameters will enable for correction of lens distortion and to undistort images, to get the size of objects in world units and also determine the location of the camera within a scene. The camera calibration is paramount to be able to undistort the images for accurately implementing ORB-SLAM, track and match features. In other words, to accurately track features from an image to the next image, we need to undistort each image. This is especially important when dealing with cameras using non-conventional lens distortion models, which is the case of the lens used in this research.

The camera calibration is performed by taking multiple images from different angles of a known pattern, such as a checkerboard, to extract the 3-D world points and their corresponding 2-D image points. The camera parameters are then extracted by using these correspondences, and the given/calculated reprojection errors and parameter estimation errors are used to evaluate the accuracy of the estimated parameters. The main challenge is the calibration of images acquired with a panomorph lens due to its non-conventional distortion model.

Initially there was a challenge to perform camera calibration since there are few existing camera calibration toolboxes for fisheye/wide-angle lenses and none for the special type of panomorph lenses. To alleviate this issue, we integrated the calibration look-up table provided by the lens manufacturer for in-lab calibration and generation of a virtual 360˚ view with a tunable zoom and tilt function to project image patches for different viewing areas. The front image patches were used for feature detection and matching while side image patches were used for calibration update as depicted in Fig. 7 and Fig. 8 of the next subsection.

C. *Evaluation Resultss*

In the first run, a drive with the camera attached to a car was performed in Kristiansand (Norway) to collect images around the city (Fig. 5) and perform offline analysis. A second test run was performed in Grimstad (Norway), with a drive around the University of Grimstad (Fig. 7) to test the accuracy of the ORB-SLAM for localization. In addition to the images from the camera as inputs to the mini PC, the raw GPS position (converted into x and y image coordinates) was incorporated into the algorithm to compare the location deviation with respect to the visual SLAM location results. TABLE I depicts a quantitative analysis of accuracy of the visual ORB-SLAM localization where every 10 second the position given by the algorithm (and corrected for by GPS) is compared with the position given by the GPS. The average error in meter is 0.2676 meters.

TABLE I. QUANTITATIVE RESULTS OF THE VISUAL SLAM ACCURACY COMPARED TO THAT OF GPS

Time output SLAM	Time output GPS	X-position SLAM [m]	Y-position SLAM [m]	X-position GPS [m]	Y-position GPS [m]	Error [m]
10.56	10.55	64.68	29.2	64.57	28.94	0.28
22.04	22.05	62.88	95.64	62.82	95.62	0.06
30.72	30.7	-18.14	133.26	-18	133.81	0.56
41.2	41.2	-79.77	168.65	-79.77	168.65	0
52.88	52.9	-64.28	283.12	-64.16	283.31	0.22
64	64	-17.88	359.5	-18	360.12	0.63
76.48	76.5	96.16	296.1	96.25	296	0.13
84.68	84.7	196.76	231.99	196.76	231.99	0
92.8	92.8	294.66	169.36	294.47	169.65	0.35
104.64	104.65	377.37	98.63	377.51	98.96	0.35
118.04	118.05	368.17	-16.38	368.1	-16.48	0.13
126.04	126.05	358.16	-106.87	358.16	-106.87	0
134.28	134.3	324.74	-196.49	324.62	-196.26	0.27
145.8	145.8	274.36	-321.73	274.31	-322.27	0.54
155.2	155.2	262.63	-440.21	262.5	-440.38	0.21
172.04	172.05	262.91	-621.27	262.91	-621.27	0
180.2	180.2	230.82	-717.38	230.83	-717.01	0.37
190.16	190.15	179.4	-833.37	179.4	-833.56	0.19
198.88	198.9	156.53	-947.58	156.55	-947.66	0.09
207	207	133.76	-1055.31	133.76	-1055.31	0

Fig. 7 depicts the results of the vehicle localization using the combination of ORB-SLAM only using images from the panomorph camera. The top-left image shows an image patch from the 360˚ panoramic view including visual features detected by the method. The top-right image shows the resulting trajectory in green color on a Google map. The two bottom images are two additional patches extracted from the virtual view used for updating the calibration parameters.

Fig. 8 illustrates a visual representation of the selected three image patches used for this feasibility analysis based on the virtual 360˚ view generated by the panomorph camera and the calibration strategy.

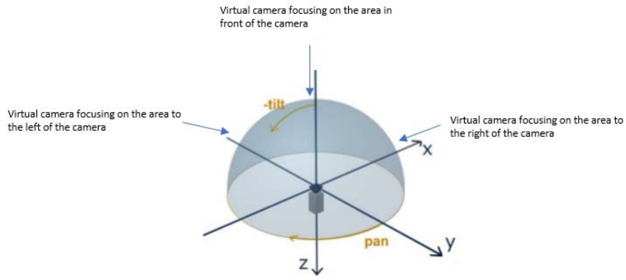


Fig. 8. Overview of three areas from which patches in Fig.7 were chosen.

V. CONCLUSIONS AND OUTLOOK

Within this paper, we aimed to transmit two main messages related to several findings. First, with this work, we showed that it is possible to perform real-time vehicle localization using 360° panoramic views based on a non-conventional wide-view lens (panomorph). Panomorph lenses are claimed to provide single camera panoramic views at higher optical

Fig. 7 A selection of detected features, resulting trajectory and selected image patches from the full 360˚ view.

resolution than that of fish-eye lenses using the same imager chip. However, a lens-specific calibration approach has to be performed for having control of the image distortion. Secondly, we showed that the ORB-SLAM method can be adapted to extract a larger set of features exploiting the wide camera field of view for real-time deployment in natural scenes. This system provides new localization capabilities, which can be exploited in robotics and autonomous driving applications at low-cost deployment. This work is a feasibility analysis, which shows the applicability of the combination of both panomorph cameras with ORB-SLAM but still the localization accuracy can be improved for longer trajectories due to the accumulation of the error. In a next step, we aim to add additional on-board sensors for this analysis such as low-cost accelerometers, gyroscopes and/or stereo panomorph cameras to evaluate the accuracy improvement vs. longer driving trajectories.

ACKNOWLEDGMENT

This work is supported by the project iTRACK running under the Horizon 2020 grant agreement number 700510 and funded by the European Commission. (http://cordis.europa.eu/project/rcn/202695_en.html)

REFERENCES

[1] R. Mur-Artal, J. M. M. Montiel and J. D. Tardós, “ORB-SLAM: A Versatile and Accurate Monocular SLAM System,” IEEE Transactions on Robotics, vol. 31, no. 5, pp. 1147-1163, October 2015.

[2] H. Lim, J. Lim, and H. J. Kim, “Real-time 6-DOF Monocular Visual SLAM in a Large-scale Environment,” in IEEE International Conference on Robotics and Automation (ICRA), pp. 1532–1539, Hong Kong, China, June 2014.

[3] E. Eade and T. Drummond, “Scalable monocular SLAM,” in IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 469–476, New York City, USA, June 2006.

[4] G. Pascoe, W. Maddern, M. Tanner, P. Pinies and P. Newman, “NID-SLAM: Robust Monocular SLAM Using Normalised Information Distance,” In Proc. of IEEE CVPR, pp. 1436-1444, USA, 2017.

[5] K. Tateno, F. Tombari, I. Laina and N. Navab, “CNN-SLAM: Real-Time Dense Monocular SLAM With Learned Depth Prediction,” In Proc. of IEEE CVPR, pp. 6243-6252, USA, 2017.

[6] G. Siveira, E. Malis and P. Rives, “An Efficient Direct Approach to Visual SLAM,” IEEE Transactions on Robotics, vol. 24, issue 5, pp. 969-979, October 2008.

[7] R. Hartley and A. Zisserman, “Multiple View Geometry,” 2nd edition, Cambridge, 2003.

[8] Y.C. Su and K. Grauman, “Making 360deg Video Watchable in 2D: Learning Videography for Click Free Viewing,” In Proc. of IEEE CVPR, pp. 6476-6484, USA, 2017.

[9] T.E. Boult, X. Gao, R. Micheals, and M. Eckmann, “Omni-directional visual surveillance”, Image and Vision Computing, 22(7), pp. 515-534, July 2004.

[10] T. Luhmann, “A Historical Review on Panorama Imagery,” ISPRS Workshop on Panorama Photogrammetry, Dresden, 2004.

[11] J. Kabala and S.S. Brandt, “A Generic Camera Model and Calibration Method for Conventional, Wide-Angle and Fish-Eye Lenses,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, issue 8 pp 1335 -1340, August 2006.

[12] S. Thibault, J. Parent, M. Larivière-Bastien, A.S. Poulin-Girard, A. Arfaoui and P. Desulniers, “Developments in modern panoramic lenses: lens design, controlled distortion, and characterization,” in SPIE Proceeding, vol. 8197, Beijing, China, 2011.

[13] S. Thibault, “360° Vision System: Opportunities in Transportation,” Proceedings of SPIE, the International Society for Optical Engineering, Society of Photo-Optical Instrumentation Engineers, 67580H. 1-67580H. 9., 2007.

[14] S. Thibault, A. Arfaoui and P. Deslauniers, “Cross-diffractive optical elements for wide angle geometric camera calibration,” In Optics letters, 36, pp. 4770-4772, 2011.

[15] S. Thibault, “Novel compact panomorph lens based vision system for monitoring around a vehicle,” in Optical Sensors, vol. 7003, 2008.