Journal of Korea Robotics Society (2022) 17(3):353-358
https://doi.org/10.7746/jkros.2022.17.3.353 ISSN: 1975-6291 / eISSN: 2287-3961

다중 카메라 시스템을 위한 전방위 Visual-LiDAR SLAM

Omni-directional Visual-LiDAR SLAM for Multi-Camera System

지샨 자비드[1] · 김 곤 우[†]
Zeeshan Javed[1], Gon-Woo Kim[†]

Abstract: Due to the limited field of view of the pinhole camera, there is a lack of stability and accuracy in camera pose estimation applications such as visual SLAM. Nowadays, multiple-camera setups and large field of cameras are used to solve such issues. However, a multiple-camera system increases the computation complexity of the algorithm. Therefore, in multiple camera-assisted visual simultaneous localization and mapping (vSLAM) the multi-view tracking algorithm is proposed that can be used to balance the budget of the features in tracking and local mapping. The proposed algorithm is based on PanoSLAM architecture with a panoramic camera model. To avoid the scale issue 3D LiDAR is fused with omnidirectional camera setup. The depth is directly estimated from 3D LiDAR and the remaining features are triangulated from pose information. To validate the method, we collected a dataset from the outdoor environment and performed extensive experiments. The accuracy was measured by the absolute trajectory error which shows comparable robustness in various environments.

Keywords: Omnidirectional, Simultaneous Localization, And Mapping, Multi-View, KLT

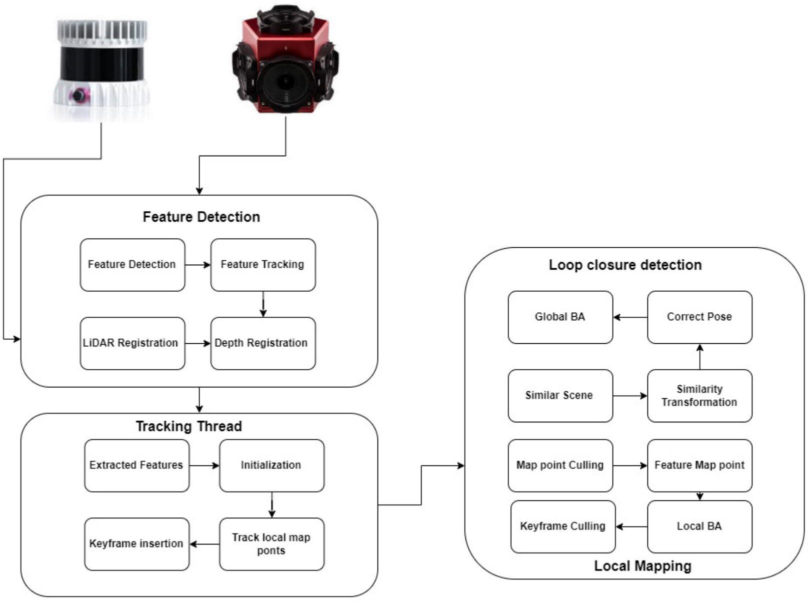
1. Introduction

SLAM enables mobile robots to move autonomously in an unknown environment. Visual SLAM is a technique used to estimate camera motion and generate a map from a sequence of images. It is a fundamental block for robot navigation, virtual reality, augmented reality like applications [1,2]. vSLAM can be divided into a mono-SLAM, stereo-SLAM and RGBD-SLAM. The many popular algorithms have been published in the literature and recognized widely such as ORB-SLAM2 [3], LSD-SLAM [4], PTAM [5], DSO [6], SVO [7] based on either monocular camera, stereo camera or RGB-D sensor.

In recent years, the panoramic vision system is widely used by many researchers to obtain more visual information for use in various environments. In literature, three types of panoramic vision systems are used such as the catadioptric vision system [8,9] is widely used for 2D motion estimation, fisheye cameras [10] are designed for multiple fisheye systems and multi-camera vision systems [11]. The most popular and widely used panoramic. environments, such as the problem of direct sunlight, lack of texture and rough terrain and sensor failure could complicate the environment for reliable motion estimation. All these factors lead to the lack of stable trackable points in complex environments which affect the overall performance of visual

Received : Mar. 10. 2022; Revised : Apr. 7. 2022; Accepted : Apr. 19. 2022
※ This research was supported in part by the MSIT (Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program (IITP-2022-2020-0-01462) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation), and in part by the Technology Innovation Program (or Industrial Strategic Technology Development Program-ATC+) (20009546, Development of service robot core technology that can provide advanced service in real life) funded By the Ministry of Trade, Industry & Energy (MOTIE, Korea)

1. PhD Student, Chungbuk National University, Cheongju, Korea (sszeshan667@gmail.com)
† Professor, Corresponding author: Chungbuk National University, Cheongju, Korea (gwkim@cbnu.ac.kr)



[Fig. 1] visual-LiDAR SLAM pipeline

simultaneous localization and mapping (SLAM).

Multi-camera and panoramic imaging systems have been widely used in the field of computer vision and robotics. Recently, spherical panoramic data have become increasingly popular for application in visual Odometry, and localization and mapping. Panoramic SLAM [11], Multi-camera SLAM [12], omniSLAM [13], Cubemap-SLAM [14], and Multicol-SLAM [15] are the complete feature-based SLAM system being used in the literature.

Multi-camera setup can solve the problems of visual SLAM in featureless region. However, in same time they increase the computational resources of algorithm. Therefore, our vSLAM algorithm balance the feature with appropriate number of feature selection from all camera. Instead of taking all features for tracking and mapping, only limited feature is selected from all views. The overall features from all views are managed based on complexity of environment, such as if the environment is feature rich the overall features are maintained to a suitable number good for tracking and mapping other features are discarded. Furthermore, the depth is added for each camera from 3D LiDAR.

The major contribution of this paper is as follows:

- The visual LiDAR SLAM for multi camera setup.
- The multi-view feature tracking and depth registration from 3D-LiDAR.
- The experimentation is performed for our campus dataset with GPS provided ground truth.

2. Visual LiDAR SLAM

The extrinsic calibration is performed between the camera and LiDAR with mutual information maximization [16,17]. The method is targetless without using any specific calibration target. The calibration parameters are obtained between the camera and LiDAR by maximizing the mutual information obtained from the surface intensities. The dataset is recorded from the outdoor environment with ROS-based recorded software. A detailed description of the dataset can be found at [18]. The data is time-stamped as it reaches the system. The manual delay is calculated between the camera and LiDAR as the camera used firewire. There is a transmission offset caused by the 800 Mb/s Firewire between the camera and Lidar.

The time synchronization is performed with provided timestamps for the camera, LiDAR, and IMU. The close timestamp is chosen with reference to the camera.

The multi-camera rig is used for visual odometry as utilized in [11]. The multi-camera consists of several fisheye cameras having a slight offset from each other and a panoramic center. First each point in the fisheye coordinate is translated into a panoramic point according to the rotation and translation between each individual fisheye camera and the panoramic camera by equation (1).

$$x' = mR_iK_i(x) + T_i \tag{1}$$

Where, R_i is the rotation matrix, $K_i(x)$ represents the intrinsic calibration parameters for each camera and the T_i is the translation vector for each camera. The detailed of camera model is presented in PanoSLAM [11].

This paper presents the omnidirectional Visual LiDAR SLAM algorithm for the multi-camera panoramic system, the module of the proposed algorithms is shown in [Fig. 1].

The features are detected from each view of the multi-camera panoramic rig. The fast feature detector is used to detect features from high resolutions images. To overcome the computational complexities, the features are detected in parallel threads. The detected features are then tracked based on KLT [10] tracking algorithm. The tracking is performed for each of the features of the individual camera in a parallel thread.

The multiple cameras have the advantages in case of the complex region to provide enough information for tracking, but in the case of a structured environment, the environment provides a lot number of features for tracking. These features increase the computational complexities of the system. Due to a large number of cameras (five in our case), the total number of extracted features increases the computation complexities of the algorithm. Therefore, features are budgeted for tracking and mapping. The adaptive strategies are used to budget overall features. The limited features are enough for tracking and mapping (say 500-600 features in the map for each key frame). If one camera provides enough features for tracking, other camera features are not added to the map. The overall budget is kept constant based on the number of extracted features.

only a few features are placed into the map for pose estimation. Firstly, the feature probability is checked for each camera, the more the probability there are high chances to get a good pose. Based on the probability of features one view is selected.

Once features are tracked, the depth is estimated from 3D LiDAR for each camera. The depth extraction is performed by selecting the neighborhood region across each feature. Firstly, foreground and background features are separated based on depth information. Then a plane is fitted to foreground features, the depth is extracted by finding the intersection of with feature. The map is initialized by extracting depth from 3D LiDAR.

The estimated pose is then refined based on pose-only optimization of triangulated map points. The initial pose between two adjacent spherical frames is used to triangulate the 3D landmarks of the remaining point for the sphere. The g2o [19] is utilized to solve the optimization problem based on the panoramic camera model similar to and only pose is optimized. The optimization problem is formulated as:

$$min\frac{1}{2}\sum_{k=1}^{n}\left\| X_S - \frac{r}{\| \exp(\xi).X_w \|}\exp(\xi).X_w \right\| \qquad (2)$$

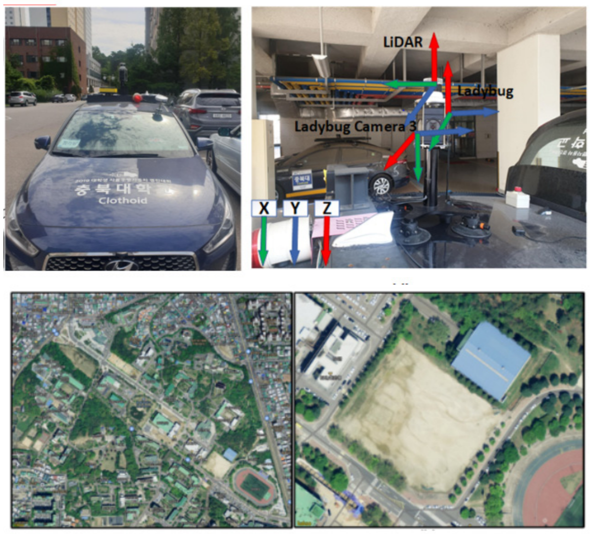
Where X_S and X_w are the 3D features of the sphere and world coordinate system and the r is the sphere radius. The equation is used to minimize the re-projection error of the sphere point.

The loop closure detection is based on bag of word (bow) [20].

3. Experiments and Results

In order to evaluate the proposed system, we conduct experiments with real-world datasets. The proposed framework is implemented and tested on the Hyundai i30 (Hyundai Motor Company, Seoul, South Korea), shown in [Fig. 1]. The platform is equipped with a Ladybug omnidirectional camera, mounted on the center top of the platform. The RTK Novatel GPS is on the left side of the platform with dual antennas setup. Ladybug is a high resolution spherical digital camera system with 360-degree coverage at a high-speed interface. The Ladybug3 has six 2-Megapixel cameras with five cameras in a circular rig and one camera is positioned at the top. This helps the system to cover more than 80 percent area of the full sphere, and all the cameras are pre-calibrated to enable high-quality spherical image stitching. The Ladybug3 allows to capture data at multiple resolutions and as well as different frame rates. Moreover, it also provides the hardware jpeg compression to support high frame rate. Novatel RTK GPS is compact, robust, high precision fully integrated global positioning system. The maximum data rate of GNSS is up to 100hz. In the proposed platform, GPS measurements are recorded at 100hz to provide the ground truth trajectories of a dataset that are used to calculate average trajectory error (ATE) for evaluation of visual odometry and SLAM algorithms.

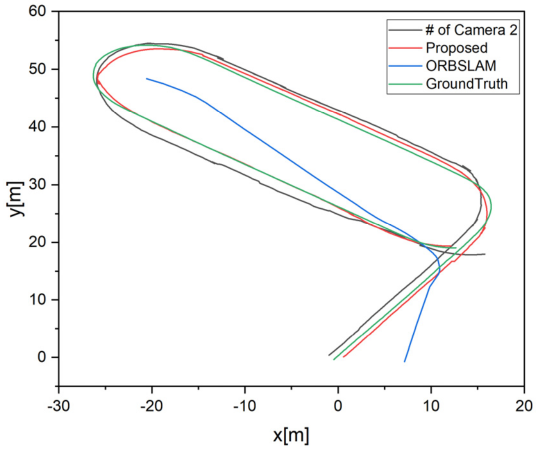
The dataset is recorded from the outdoor environment in 2 sequences including outdoor parking, campus main road as shown in [Fig. 2]. The RTK GPS is used as ground truth for testing and evaluation of the proposed method. In the experiments, the images are captured at full resolution (1616x1232).



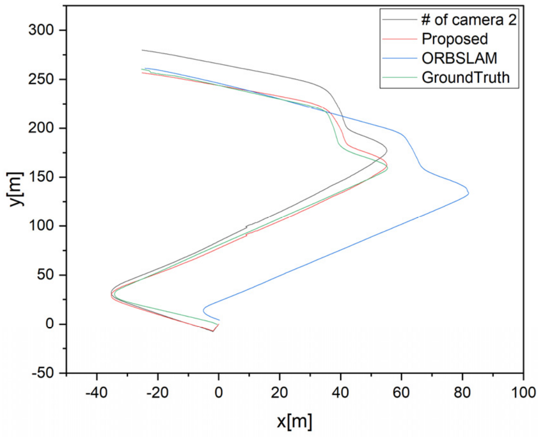
[Fig. 2] Experimental setup, Top images shows the platform used while recording. While bottom image show the data recorded region in campus

The parking sequence contains more than 400 images captured at 6 fps with total length of nearly 300 meters and contains the featureless region specially at turns. Similarly, the other two sequence have length nearly 200 meters and 0.7 km respectively.

The qualitative results are presented in [Fig. 2] with GPS-provided ground truth in the parking environment. The results show that our algorithm can recover complete trajectory in the complex, featureless outdoor environment. To evaluate the overall performance of the proposed method we conducted some experiments with different camera settings and with state of the art method. cameras are used to show the advantage of adding more cameras to pose estimation accuracy. Three output trajectories are shown namely cam2 (2 cameras), ORBSLAM-mono, and the proposed method. The ORBSLAM-mono is run only on front facing camera, because it provides obvious structural information.



[Fig. 3] Output plot of parking trajectory with GPS



[Fig. 4]. Output plot of sequence 2 with GPS trajecto

The dataset is recorded from fisheye camera therefore it contains distortion, the data is pre-rectified for and cropped for ORBSLAM.

For each experiment, 1000 features per frame are extracted. The resultant qualitative output with different combinations of a camera and ORBSLAM -mono is shown in [Fig. 3] and [Fig. 4]

[Table 1] Qunatitative Results (Rotation and Translation error) for proposed method

Sequence	Cameras	ATE (Translation)				ARE (Degree)	
		RMSE	Min	Median	RMSE	min	S.D
Parking	Cam2	1.3671	0.0056	1.1670	0.2372	0.0050	0.2805
	Cam3	1.1480	0.0041	1.0894	0.1231	0.0049	0.1610
	ORB-SLAM	2.2734	0.0912	1.8584	7.6301	0.0545	5.0784
	Proposed	0.9121	0.0021	0.6421	0.1308	0.0026	0.1138
Building	Cam2	5.7856	0.0288	4.4511	5.1542	0.0163	2.4732
	Cam3	5.1622	0.0163	4.1281	4.7273	0.0121	2.1245
	ORB-SLAM	4.8923	0.0253	4.6968	5.0123	0.0059	3.3245
	Proposed	4.1312	0.0217	3.3617	2.1706	0.0041	1.1723

respectively. The output shows that large field of view with different camera setup has better performance than other methods including limited field of view. The quantitative results are presented in [Table 1]. The [Table 1] shows the comparison between ORBSLAM, and the proposed method. The cam2 and cam3 are added for ablation study. The ORBSLAM is integrated to front facing camera only with pre-rectified and cropped images. The cam2 and cam3 show that only 2 and 3 cameras features are used for tracking and mapping. The Average Translation error (ATE) is used to compare the result with RMSE, min and S.D. The result shows that adding more views improves the accuracy of the system for a complex environment.

4. Conclusion

This research proposes visual LiDAR framework for simultaneous localization and mapping (vSLAM) for omnidirectional camera LiDAR setup. The algorithm is based on feature detection and tracking from multiple camera setup. The feature depth is estimated from 3D LiDAR and remaining feature are triangulated from previous motion. The local and global bundle adjustment is performed with loop closure detection. The algorithm is tested on real dataset with GPS provided ground truth. The overall results suggest the significant improvement with state of the art methods.

References

[1] G. Bresson, Z. Alsayed, L. Yu, and S. Glaser, “Simultaneous localization and mapping: A survey of current trends in autonomous driving,” *IEEE Transactions on Intelligent Vehicles*, vol. 2, no. 3, pp. 15-64, Sept., 1964, DOI: 10.1109/TIV.2017.2749181.

[2] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, “Past, present, and future of simultaneous localization and mapping: Toward the robust perception age,” *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1309-1332, Dec., 2016, DOI: 10.1109/TRO.2016.2624754.

[3] R. Mur-Artal and J. D. Tardos, “Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras,” *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 255-1262, Oct., 2017, DOI: 10.1109/TRO.2017.2705103.

[4] J. Engel, T. Schops, and D. Cremers, “LSD-SLAM: Large-scale direct monocular SLAM,” *European Conference on Computer Vision*, pp. 834-849, 2014, DOI: 10.1007/978-3-319-10605-2_54.

[5] G. Klein and D. Murray, “Parallel tracking and mapping for small AR workspaces,” *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, Nara, Japan, 2007, DOI: 10.1109/ISMAR.2007.4538852.

[6] J. Engel, V. Koltun, and D. Cremers, “Direct sparse odometry,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 1, pp. 611-625, Mar., 201 DOI: 10.1109/TPAMI.2017.2658 577.

[7] C. Forster, M. Pizzoli, and D. Scaramuzza, “SVO: Fast semi-direct monocular visual odometry,” *2014 IEEE International Conference on Robotics and Automation (ICRA)*, Hong Kong, China, 2014, DOI: 10.1109/ICRA.2014.6906584.

[8] H. Chen, K. Wang, W. Hu, K. Yang, R. Cheng, X. Huang, and J. Bai, “PALVO: visual odometry based on panoramic annular lens,” *Optics Express*, vol. 27, no. 17, pp. 24481-24497, 2019, DOI: 10.1364/OE.27.024481.

[9] D. Scaramuzza and R. Siegwart, “Monocular omnidirectional visual odometry for outdoor ground vehicles,” *International Conference on Computer Vision Systems*, pp. 206-215, 2008, DOI: 10.1007/978-3-540-79547-6_20.

[10] P. Liu, L. Heng, T. Sattler, A. Geiger, and M. Pollefeys, “Direct visual odometry for a fisheye-stereo camera,” *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Vancouver, BC, Canada, 2017, DOI: 10.1109/IROS.2017.8205988.

[11] S. Ji, Z. Qin, J. Shan, and M. Liu, “Panoramic SLAM from a multiple fisheye camera rig,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 159, pp. 169-183, Jan., 2020, DOI: 10.1016/j.isprsjprs.2019.11.014.

[12] Y. Yang, D. Tang, D. Wang, W. Song, J. Wang, and M. Fu, “Multi-camera visual SLAM for off-road navigation,” *Robotics and Autonomous Systems*, vol. 128, Jun., 2020, DOI: 10.1016/j.robot. 2020.103505.

[13] C. Won, H. Seok, Z. Cui, M. Pollefeys, and J. Lim, “OmniSLAM: Omnidirectional Localization and Dense Mapping for Wide-baseline Multi-camera Systems,” *2020 IEEE International Conference on Robotics and Automation (ICRA)*, Paris, France, DOI: 10.1109/ICRA40945.2020.9196695.

[14] Y. Wang, S. Cai, S.-J. Li, Y. Guo, T. Li, and M.-M. Cheng, “CubemapSLAM: A Piecewise-Pinhole Monocular Fisheye SLAM System,” *Asian Conference on Computer Vision*, pp. 34-49, 2018, DOI: 10.1007/978-3-030-20876-9_3.

[15] S. Urban and S. Hinz, “MultiCol-SLAM - A Modular Real-Time Multi-Camera SLAM System,” *arXiv:1610.07336*, 2016, DOI: 10.48550/arXiv. 1610.07336.

[16] G. Pandey, J. R. McBride, S. Savarese, and R. M. Eustice, “Automatic Extrinsic Calibration of a 3D Lidar and Camera by Maximizing Mutual Information” *Twenty-Sixth AAAI Conference on 2012*, [Online] https://www.aaai.org/ocs/index.php/AAAI/AAAI12/paper/view/5029/5371.

[17] G. Pandey, J. R. McBride, and R. M. Eustice, “Ford Campus vision and lidar data set,” *Sage Journals*, vol. 30, no. 13, 2011, DOI:10.1177/0278364911400640.

[18] Z. Javed and G. W. Kim, “PanoVILD: a challenging panoramic vision, inertial and LiDAR dataset for simultaneous localization and mapping,” *The Journal of Supercomputing*, vol. 78, 2022,

DOI: 10.1007/s11227-021-04198-1.

[19] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolig, and W. Burgard, “G^{2}o: A general framework for graph optimization,” *2011 IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, 2011, DOI: 10.1109/ICRA.2011. 5979949.

[20] A. Angeli, D. Filliat, S. Doncieux, and J.-A. Meyer, “Fast and Incremental Method for Loop-Closure Detection Using Bags of Visual Words,” *IEEE Transactions on Robotics*, vol. 24, no. 5, Oct., 2008, DOI: 10.1109/TRO.2008.2004514.

Zeeshan Javed

2014 Quaid-e-azam University, Islamabad, Pakistan (MSc)

2017 Department of Electrical Engineering, COMSATS University Islamabad, Pakistan (Master)

2019~Present Department of Control and Robot Engineering, Chungbuk National University, Korea (Ph.D.)

Interests: SLAM, Omnidirectional SLAM, Loop Closure Detection, Sensor Fusion

Gon-Woo Kim

2008 Assistant Professor, Electronics and Control Engineering, Wonkwang University

2012 Assistant Professor, School of Electronics Engineering, Chungbuk National University

2014 Associate Professor, School of Electronics Engineering, Chungbuk National University

2021~Present Professor, Department of Intelligent Systems and Robotics, Chunbuk National University

Interests: Navigation, Localization, SLAM