ACCURACY ANALYSIS OF PHOTOGRAMMETRIC STEREO VISUAL ODOMETRY ACCORDING TO IMAGING GEOMETRY

Sung-Joo Yoon (1), Taejung Kim (1)

[1] Inha University, 100 Inha-ro, Michuhol-gu, Incheon 22212, Korea
Email: 22181415@inha.edu; tezid@inha.ac.kr

KEY WORDS: Visual odometry, Photogrammetry, Imaging geometry, Feature optimization

ABSTRACT: Visual Odometry (VO) is a technique to estimate platform motion using an image sequence. This is one of promising image processing techniques because it uses only images at relatively low cost. However, for the same reason, this technique's accuracy becomes sensitive to imaging geometry. In this study, we analyze the accuracy with two conditions: Field of View (FOV) of images used for experiments, and baseline, the distance between two imaging locations. The Photogrammetric Stereo Visual Odometry (PSVO) developed in this paper performs feature extraction and matching, feature optimization, and photogrammetric motion estimation. For experiments, we used a dataset provided by Karlsruhe Institute of Technology and Toyota technological Institute (KITTI) community and a dataset acquired with our platform. Our dataset is called Inha dataset and has a smaller FOV and a longer moving distance per frame than KITTI dataset. We compared the results depending on imaging geometry and performed visual inspection of feature matching and accuracy verification. In the experimental results, as FOV decreases, the estimation errors tended to increase. Also, the longer moving distance per frame, the worse performance of outlier filtering, especially Random Sample Consensus (RANSAC). Through these experiments, we observed that not only the status of features, but also imaging geometry was critical factor and the general RANSAC filtering was not suitable for VO. This paper proposes the importance of imaging geometry and feature optimization in VO.

1. INTRODUCTION

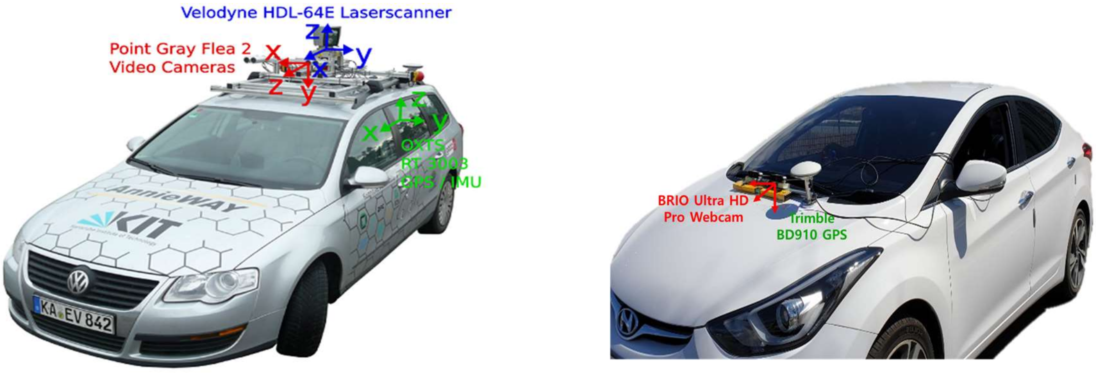
Stereo Visual Odometry (SVO) estimates platform motion from a stereo image sequence. It is one of interest in the various fields where positioning is necessary, such as autonomous driving and robotics. This technique quantifies platform translation like an odometer and rotation like an Inertial Measurement Unit (IMU) simultaneously. Thus, it has a great advantage that the both motion factor is estimated at low cost. But since it uses only images, its accuracy becomes sensitive to imaging geometry and condition. For example, even if this technique estimates the motion on a same path, the result may vary depending on the conditions, such as camera specifications, frame intervals and camera placement. Therefore, we need to analyze SVO accuracy for some situations.

Estimation of geometry with images, such as SVO and Aerial Triangulation (AT), generally uses features on an image as observed values (Jung et al., 2016). Therefore, it is important to extract a lot of features on image and then select precise ones. In Karlsruhe Institute of Technology and Toyota technological Institute (KITTI) community (Geiger et al., 2013), there have been many researches on Visual Odometry (VO) to extract correct features without any outliers on image. Stereo Odometry algorithm relying on Feature Tracking (SOFT2) (Cvišić et al., 2018) technique performed Simultaneous Localization and Mapping (SLAM). This technique selected correct features using prediction value with Inertial Measurement Unit (IMU) sensor, a model for velocity assumption, and Normalized Cross Correlation (NCC) algorithm. For this technique, there may be a restriction by additional cost for IMU and proper model selection. The RotRocc+ (Buczko and Willert, 2016) technique extracted inliers by a model for restrictive motion and decoupling the optical flow within image. Sometimes outliers may be included in inliers once the restriction model was improper to the imaging geometry. Circular Fast Retina Keypoint (FREAK) - Oriented Fast and Rotated Binary Robust Independent Elementary Feature (CFORB) (Mankowitz and Rivlin, 2015) technique matched features using the circular matching based Gauss-Newton Optimization and removed outliers by continuing to apply Random Sample Consensus (RANSAC) (Wu and Fang, 2007). For this technique, it was difficult to determine stop condition for iterative RANSAC filtering and the performance may vary by the model appropriacy for RANSAC filtering. Also, techniques recently listed in KITTI community carried out feature optimization by careful feature selection (Zhang, 2019), loop closure detection and optical flow check (Bultmann et al., 2019), and multiple sensor (Qin et al., 2019). As these techniques, Most of VO research focus on feature optimization for accuracy improvement. In this study, we focus to apply photogrammetric feature optimization and bundle adjustment to VO technique.

This study proposes Photogrammetric Stereo Visual Odometry (PSVO) technique and shows its accuracy according to Imaging geometry. This paper shows experiment datasets, a method with algorithm, experiment results, discussion and conclusions.

2. MATERIALS AND METHODS

We used the KITTI sequence 03 provided by the KITTI community and Inha dataset acquired by ourselves. The vehicle for KITTI dataset is shown in (a) of Figure 1, and for Inha dataset is shown in (b) of Figure 1.

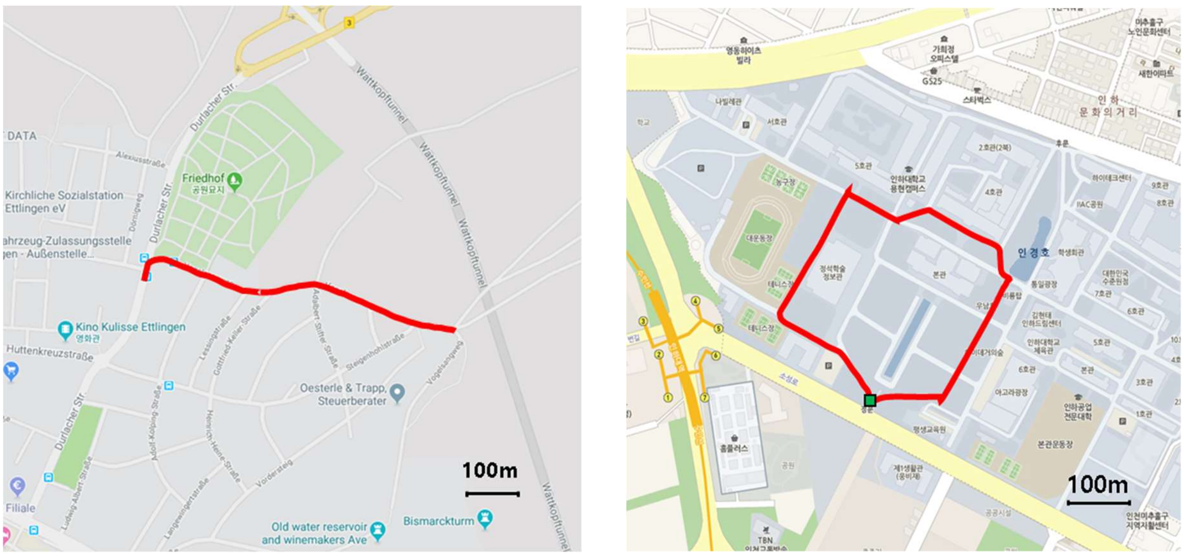


(a) (b)

Figure 1. (a) Vehicle for KITTI dataset; (b) Vehicle for Inha dataset

For the KITTI dataset, while maintaining about 0.5m intervals between frames, it was constructed in urban areas. It contains stereo image sequences with Point gray flea 2 video cameras, Interior Orientation Parameters (IOP) and truth for the poses with OXT RT 3003 as in (a) of Figure 2. The Field of View (FOV) is about 120 degrees and the baseline between stereo cameras is about 0.5m.

For the Inha dataset, while maintaining about 3.5m intervals between frames, we constructed it in Inha University in Korea using BRIO Ultra High Definition (HD) Pro Webcam and Trimble BD910 as in (b) of Figure 2. The FOV is about 80 degrees and the baseline is about 0.2m.



(a) (b)

Figure 2. (a) KITTI sequence 03 trajectory; (b) Inha dataset trajectory

The flowchart of the proposed technique is as in Figure 3. First, we extracted features on stereo image pairs and matched them. For the feature extraction and matching, there have been a lot of methods, whose processing time and result are different depending on image condition. Therefore, we compared the processing time and result according to several method combinations. Second, we optimized features using three methods: Photogrammetry-based Feature Optimization (PFO), Vision-based Feature Optimization (VFO), and Statistic-based Feature Optimization (SFO). As mentioned in Section 1, we focused to apply PFO for outlier elimination. For the PFO method, it checks the projection and reprojection errors based photogrammetry. For the VFO method, it eliminates the outlier using sequentially RANSAC filtering within multiple images. For the SFO method, it continues to detect the outlier through data snooping as statistical posteriori. Finally, we estimated the platform motion using photogrammetric bundle adjustment. The details are given in the subsections.

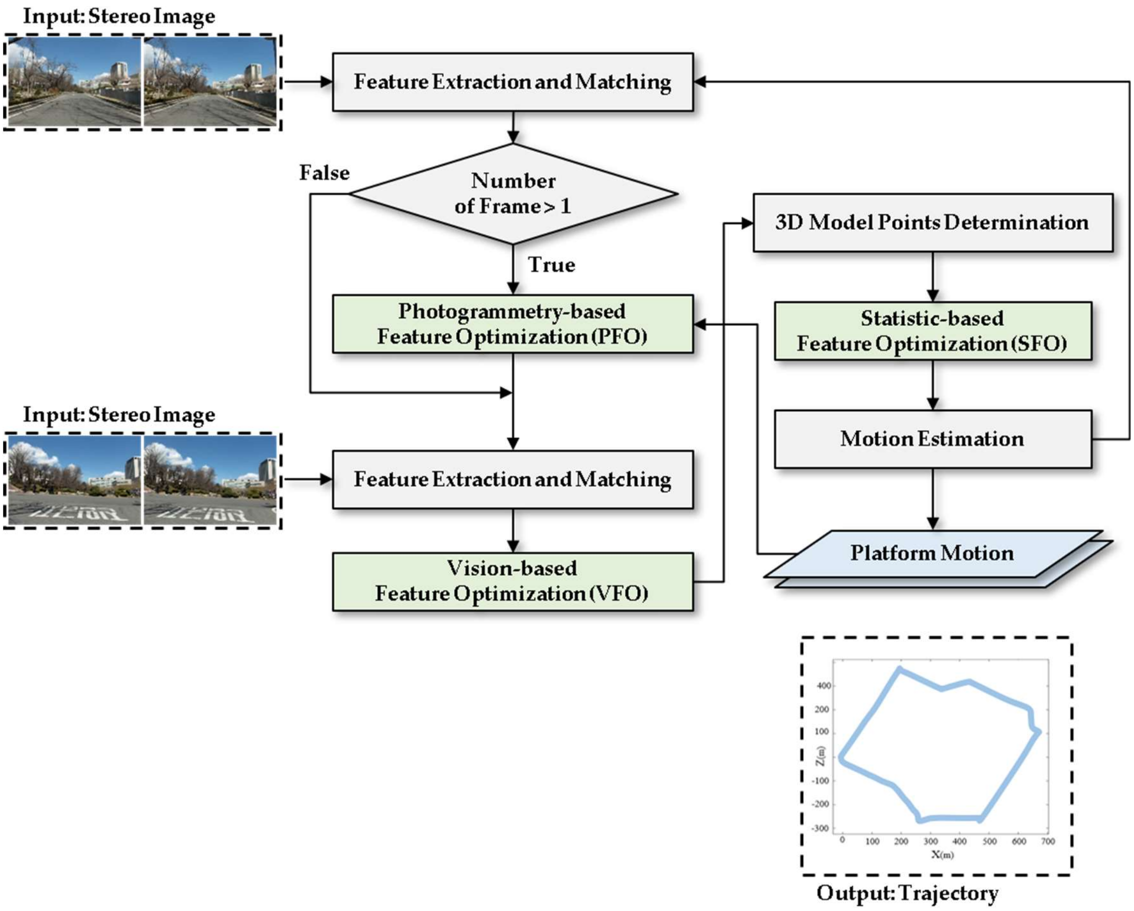


Figure 3. Overview of the proposed technique

2.1 Feature Extraction and Matching

The corresponding point is extracted through feature extraction and feature matching. For feature extraction method, it extracts corner points as features on the image. There are representative feature extractors provided by OpenCV and have different characteristics, such as scale invariability and optical flow. For feature matching method, it matches the features based pairwise matching or sequential tracking. Pairwise matching algorithm matches the features more accurately and sequential tracking algorithm tracks the corresponding points more quickly. We have studied the feature extraction and matching methods and chose the Shi-Thomasi corner-KLT tracker method because it speedily tracked a lot of corresponding points as in Table 1 (Yoon and Kim, 2019).

Table 1. Corresponding point extraction performance

Method	Algorithm combination	Number of points per time (num./ sec.)
Pairwise matching	SIFT-SIFT-BruteForce	1223.30
	SURF-SURF-BruteForce	2635.92
	FAST-BRISK-FLANN	15209.61
	FAST-ORB-FLANN	29101.80
	FAST-FREAK-FLANN	28143.88
Sequential tracking	FAST-KLT tracker	94782.52
	Shi-Thomasi corner-KLT tracker	180708.18

2.2 Feature Optimization and Motion Estimation

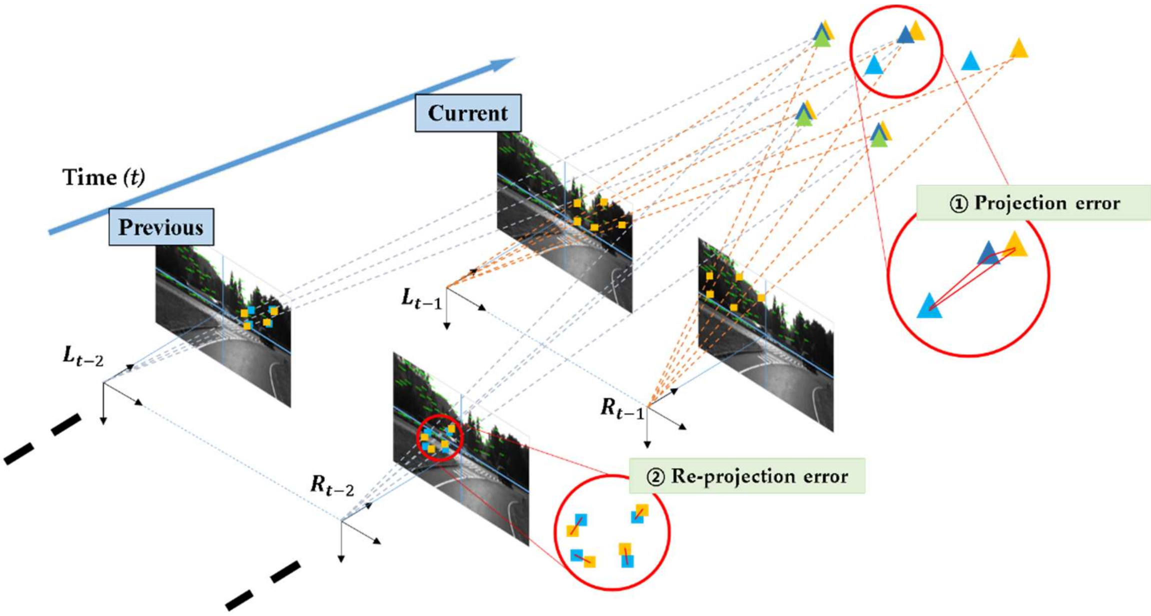


Figure 4. PFO method concept diagram

In the proposed technique, feature optimization performs three methods: PFO, VFO, and SFO. For only the first stereo pair, this method carries out feature optimization except PFO because it needs the previous motion.

For the VFO, it performs RANSAC filtering applying 5 point-based essential matrix assumption model. VFO method filters the features on the stereo pairs circularly and defines the detected inliers as the first candidate features. Also, the EOP is calculated, used in motion estimation as initial value.

As in Figure 4, the core concept of PFO is projection and re-projection error calculation based photogrammetry. Using the Exterior Orientation Parameters (EOP), PFO method calculates 3D model points within stereo image pairs. Then, it projects the model points from the previous to current model space using the estimated motion. At this time, the distance between original and projected model points is called projection error. Next, it re-projects the model point on the current model space onto the previous image space. At this time, the distance between original and re-projected image points is called re-projection error. It selects the correct features according to these sizes and defines these as the second candidate features.

For the SFO, it checks the features through data snooping of statistical posteriori method. SFO method proceeds with motion estimation method for statistical analysis. After motion estimation method, it calculates test statistic according to Equation (1) and (2), and detects outliers by one-tailed test. It removes the feature one by one with the greatest outlier iteratively until there is no outlier. Finally, this method defines inliers as the optimal features.

$$
\begin{aligned}
Q_{vv} &= W^{-1} - AQ_{xx}A^T \\
&= W^{-1} - A(A^TWA)^{-1}A^T
\end{aligned}
\tag{1}
$$

$$
t_i = \frac{\bar{v}_i}{S_0} = \frac{|v_i|}{S_0\sqrt{q_{ii}}} \tag{2}
$$

where t_i is test statistic of ith observation, Q_{vv} is cofactor matrix of residuals, Q_{xx} is cofactor matrix of unknown parameters, W is weight matrix, A is coefficient matrix, v_i is residual of ith observation, q_{ii} is ith diagonal element of Q_{vv}, and S_0 is unit standard deviation.

Motion estimation method uses absolute orientation using photogrammetric collinearity condition. To perform Least Squares Estimation (LSE), this method is structed as linear matrix form by linear approximation. It first utilizes the LSE with the second features until SFO is finished. Then, it re-estimates the platform motion using the optimal features and finally multiplies by the previous matrix to calculate the motion.

3. Experiment results

(a) (b) (c)

Figure 5. Feature matching results on Inha dataset

We first experimented PSVO with Inha dataset. In the experiment, the initial number of features was small on average. Additionally, there were problems that feature filtering did not work properly as in Figure 5. As a result, unlike our experiment with KITTI dataset, it showed gross errors on the motion estimation.
Then, we analyzed the PSVO performance according to FOV and the distance per frames using KITTI dataset. We set 7 experiment cases as Table 2 and checked translation error (%), rotation error (degree per meter), standard deviation, and processing time.

Table 2. Experiment case on this study

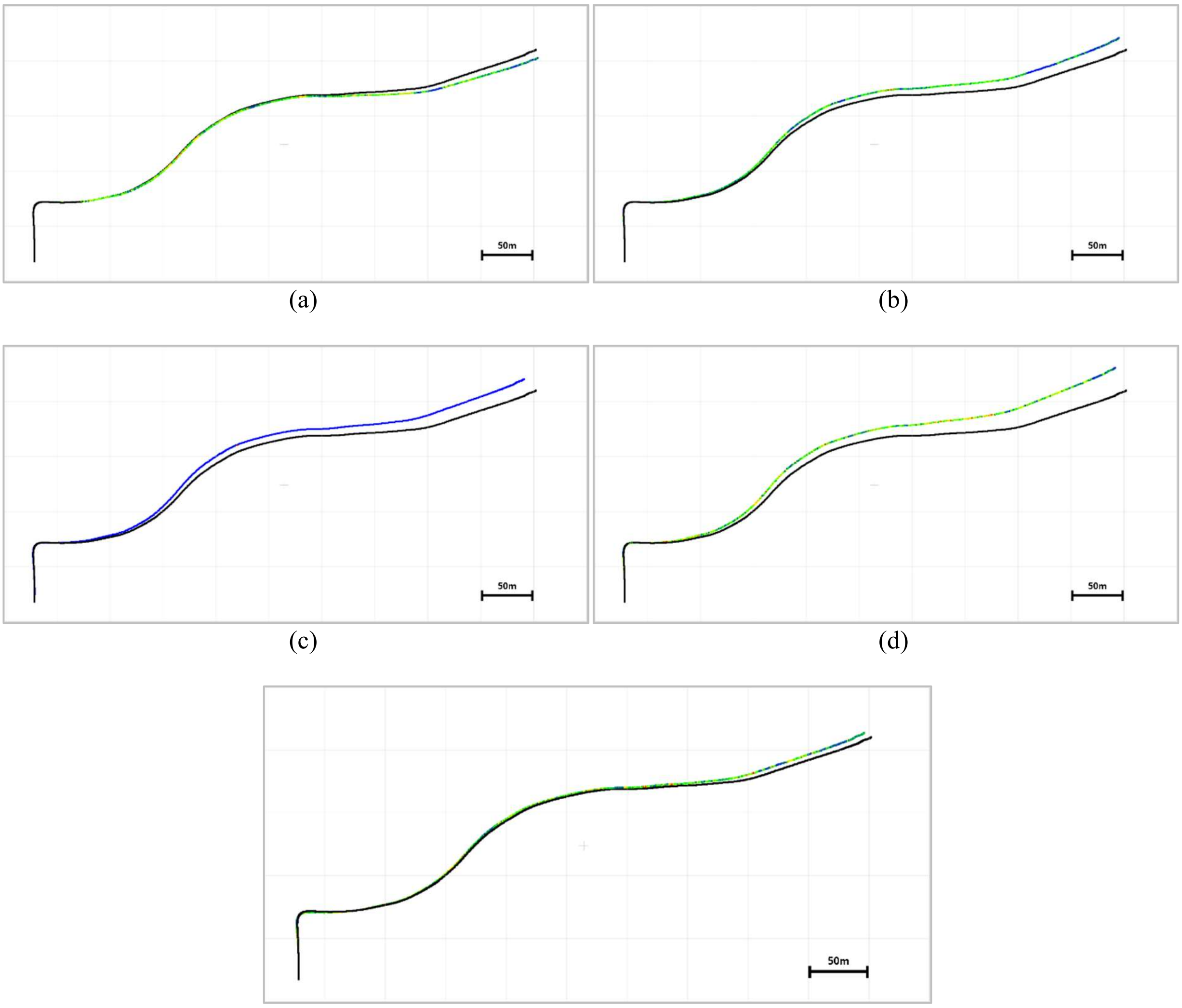
Experiment case	Purpose	Image size (width × height; pixel)	FOV (degree)	Distance per frame (m)
1	Original condition	1242×375	119.7	0.5
2	FOV experiment	966×375	106.5	0.5
3		828×375	97.9	0.5
4		690×375	87.4	0.5
5		414×375	59.7	0.5
6	Distance per frame experiment	1242×375	119.7	1.0
7		1242×375	119.7	1.5

3.1 Estimation Result According to FOV

For the FOV experiment, we set the FOV cases from 59.7 to 119.7 degrees by clipping the images. As in Table 3, PSVO showed the best performance at case 2 and the worst performance at case 1. While the processing time was shortened by FOV reduction, the accuracy was not significantly different. Figure 6 shows the trajectory estimation results at case 1 to 5. In this figure, the black line is truth and the color line is trajectory estimated. The color, which means standard deviation of each frame, showed that standard deviation increases as the road switches from straight to curved.

Table 3. Performance by FOV case

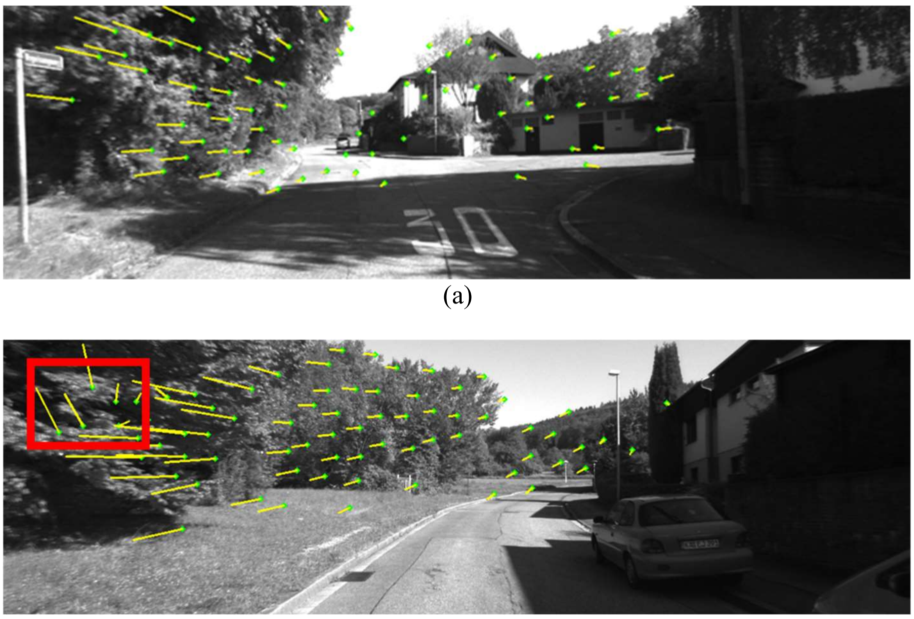
Experiment case	FOV (degree)	Translation error (%)	Rotation error (deg./m)	Standard deviation	Processing time (sec.)
1	119.7	1.7274	0.0203	0.3155	0.1336
2	106.5	1.9947	0.0083	0.2886	0.1342
3	97.9	2.6548	0.0110	0.2997	0.1170
4	87.4	2.6754	0.0076	0.2942	0.1038
5	59.7	1.6721	0.0124	0.2944	0.0670
Average		2.1449	0.0119	0.2985	0.1111



(a) (b) (c) (d) (e)

Figure 6. Estimated trajectory at Case 1 to 5: (a) to (e)

3.2 Estimation Result According to distance per frame



(a) (b)

Figure 7. (a) Feature matching result on straight road; (b) Feature matching result on curved road

For this experiment, we set the distance per frame from 0.5 to 1.5 meters by loading the images at regular intervals. When the distance per frame became longer, the number of outliers increased. This problem was found in curved rather than straight road as Figure 7. The experiment showed that the performance was sensitive to the distance as in Table 4. In particular, the translation error increased rapidly as the distance became longer.

Table 4. Performance by distance per frame

Experiment case	Distance per frame (m)	Translation error (%)	Rotation error (deg./m)	Standard deviation	Processing time (sec.)
1	0.5	1.7274	0.0203	0.3155	0.1336
6	1.0	24.5303	0.0991	2.4831	0.3550
7	1.5	48.3810	0.1319	8.1412	1.1118
Average		24.8795	0.0838	3.6466	0.5335

4. DISCUSSION

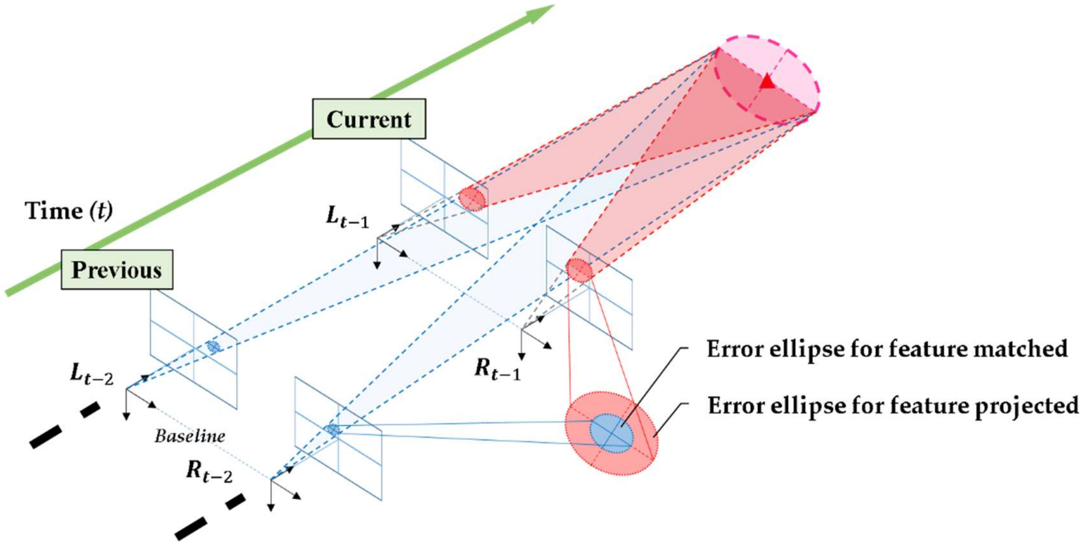


Figure 8. Error ellipse for feature matched and projected

In this study, we used the Inha dataset and sequence 03 of KITTI dataset. We tested PSVO developed according to imaging geometry by 7 cases and described the error per mileage. The experiment results showed that VO accuracy is more sensitive to the distance per frame than FOV.

The error ellipses for feature are as shown in Figure 8. When the feature is derived from previous to current image space, its error ellipse changes depending on the baseline and the distance per frame. To define two geometry factors, we calculated convergence angle between $\overline{L_{t-2}\,L_{t-1}}$ and $\overline{L_{t-1}\,R_{t-2}}$. The angle is proportional to the baseline and inversely proportional to the distance per frame. We found that the error ellipse is inversely proportional to the convergence angle.

Table 5 shows the convergence angle by cases. The error ellipses with Inha dataset are the biggest as the convergence angle is significantly small as about 3.4 degrees. This causes a large error propagation from image space to model space. Therefore, fitting the features to geometry model, such RANSAC and collinearity condition, becomes difficult, why feature filtering did not work properly.

Table 5. Convergence angle between the frames

Experiment case		Convergence angle (degree)
KITTI dataset	Case 1	47.0514
	Case 6	28.2425
	Case 7	19.7024
Inha dataset		3.3521

5. CONCLUSIONS

It is important for VO techniques to get rid of the outliers as possible. In this paper, we showed Photogrammetry-based VO focusing feature optimization. The PSVO developed carries out two main parts: Feature extraction and matching, and feature optimization and motion estimation. The PSVO performs the Shi-Thomasi corner-KLT tracker method in feature extraction and matching part, and PFO, VFO and SFO in feature optimization.

For the experiments, we set the 7 experimental cases using the KITTI dataset and Inha dataset with all different imaging geometries. We confirmed the PSVO performance checking feature matching result, trajectory estimation error, standard deviation, and processing time. In case 2 with FOV 106.5 degrees and distance per frame 0.5m, the performance was the best as translation error is about 1.99%, rotation error is about 0.01deg./m, standard deviation is about 0.29, and processing time is about 0.13 seconds. On the other hand, in case 7 with FOV 119.7 degrees and distance per frame 1.5m, the performance was the worst as translation error is about 48.38%, rotation error is about 0.13deg./m, standard deviation is about 8.14, and processing time is about 1.11 seconds. The experiment results described that VO accuracy is more affected by the distance per frame than the FOV and the translation accuracy is more sensitive than the rotation accuracy.

We analyzed the results using the convergence angle as a quality indicator of imaging geometry. Through the analysis, we were able to confirm the PSVO stability at side of various FOV and a correlation between its accuracy and imaging geometry. Based on this study, we expected to improve the PSVO in various environments and apply to other image-based estimation as the quality indicator proposed.

ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. NRF-2016R1A2B4013017).

REFERENCES

Buczko, M. and Willert, V., 2016. Flow-decoupled normalized reprojection error for visual odometry. In: Proceedings of the 2016 IEEE 19th International Conference on Intelligent Transportation Systems, Rio de Janeiro, Brazil, pp. 1161–1167.

Bultmann, S., Li, K. and Hanebeck, U. D., 2019. Stereo visual SLAM based on unscented dual quaternion filtering. In: Proceedings of the 22nd International Conference on Information Fusion (Fusion 2019), Ottawa, Canada.

Cvišić, I., Ćesić, J., Marković, I. and Petrović, I., 2018. SOFT-SLAM: Computationally efficient stereo visual simultaneous localization and mapping for autonomous unmanned aerial vehicles. Journal of Field Robotics, 35(4), pp. 578-595.

Geiger, A., Lenz, P., Stiller, C. and Urtasun, R., 2013. Vision meets robotics: The KITTI dataset. The International Journal of Robotics Research, 32(11), pp. 1231–1237.

Jung, J., Kim, T., Kim, J. and Rhee, S., 2016. Comparison of match candidate pair constitution methods for UAV images without orientation parameters. Korean Journal of Remote Sensing, 32(6), pp. 647-656.

Mankowitz, D.J. and Rivlin, E., 2015. CFORB: Circular FREAK-ORB visual odometry. arXiv 2015, arXiv:1506.05257

Qin, T., Pan, J., Cao, S. and Shen, S., 2019. A general optimization-based framework for local odometry estimation with multiple sensors. arXiv preprint arXiv:1901.03638.

Wu, F.L. and Fang, X.Y., 2007. An improved RANSAC homography algorithm for feature based image mosaic. In: Proceedings of the 7th WSEAS International Conference on Signal Processing, Computational Geometry & Artificial Vision, Athens, Greece, pp. 202–207.

Yoon, S.J., and Kim, T., 2019. Development of stereo visual odometry based on photogrammetric feature optimization. Remote Sensing, 11(1).

Zhang, S., 2019. Stereo visual Odometry based Optical Flow and Careful Feature Selection, Retrieved September 5, 2019, from http://www.cvlibs.net/datasets/kitti/eval_odometry_detail.php?&result=5c3bfe2a8a4aa1de737dd05 03498c4 a03e82b934.