2024 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)
July 15-19, 2024. Boston, MA, USA

Feature Point Selection Scheme of Stereo Visual Odometry for Planetary Exploration Rover

Masatoshi Motohashi, Takashi Kubota

***Abstract*—Planetary exploration rovers are required to estimate their position in GPS-denied environments accurately. Visual Odometry is one of the solutions in such environments. By tracking feature points in the image, it is possible to estimate the position with high accuracy, even in extreme environments. If all the feature points in the image are used for estimation, however, the computational cost increases. Especially in stereo camera methods, the calculation time required for stereo matching is highly extended. Therefore, this paper proposes a method for selecting feature points in the image before stereo matching. The accuracy will be diminished if the number of feature points is reduced. To address this issue, the feature points are selected separately for rotation and translation. As a result of verification using an actual rover, it is confirmed that the proposed method can reduce the computational cost by up to 33% compared to the conventional method.**

***Index Terms*—Visual Odometry, Stereo Camera, Feature Point Selection, Planetary Exploration Rover**

I. INTRODUCTION

In the last few decades, planetary exploration by rovers has been actively conducted. The rover brings detailed data by in-situ measurement on the planet's surface. Not only that but a wide range of data can also be obtained by moving, so it is possible to collect different data from various area. 6 rovers have landed on Mars so far, and 3 rovers are currently active. They gave much information about Mars, such as the existence of water in the past and the possibility of life [1]. In the near future, the role of rovers will be more critical in planetary exploration.

To safely traverse with communication latency between Earth and the Moon or the planet, the rover is required to have autonomous navigation capability. The navigation system for exploration rovers consists of perception, localization, path planning, and traverse [2]–[5]. Among them, localization significantly contributes because others require accurate rover position information. GPS is not available in the planetary environment. Wheel odometry is also ineffective because the error increases significantly on sandy soils and slopes. Therefore, visual odometry (VO) has been used for exploration rovers for the last few decades [6]–[8].

VO [9] is the image-based ego-motion estimation technique. For the planetary exploration rover, feature-based stereo VO is used. The feature points are extracted from the image taken at different locations, and the camera motion is estimated by matching the feature points between the images. The scale indeterminacy is solved by using the stereo camera. The procedure of stereo VO is as follows. First, feature points are extracted from 2 images taken at different locations. Next, the feature points are matched between the images. Then, stereo matching is performed to estimate the depth of the feature points and calculate the 3-dimensional position of the feature points. Finally, the camera motion is estimated by optimizing the reprojection error between two 3-dimensional feature points.

One of the issues in VO is the computational cost. Image processing is computationally demanding for the planetary exploration rover, in which computational resources are limited; in particular, stereo matching takes a lot of time. NASA/JPL's latest Mars rover, Perseverance, takes 4.9 seconds per frame [10]. This is the processing time when using an FPGA specialized for image processing, and it will take even longer if the power resource is severe. Reducing the number of feature points is one of the solutions to improve computational efficiency. However, the accuracy will be reduced if the feature points are reduced blindly. Therefore, it is necessary to select the appropriate feature points.

Bucketing Technique [11] is a typical method for feature selection. The image is divided into grids, and the feature points are extracted from each grid. This method reduces the number of feature points and distributes them uniformly over the entire image, making achieving both efficiency and accuracy possible. It is a simple yet highly useful method and is used not only in VO but also in various other methods, such as Visual SLAM [12] [13].

Zhao et al. [14], [15] proposed the feature points selection scheme based on the uncertainty model of least squares pose optimization. The combination of feature points that maximizes the determinant of the error matrix is searched. A stochastic-greedy search algorithm was also proposed to solve NP-hard combinatorial searches efficiently. A feature selection scheme in high-speed motion was proposed by Buczko [16]. However, these methods required detailed uncertainty information of the feature points, meaning stereo matching is required before selection.

Otsu et al.'s two-point algorithm [17] selects one nearby point and one far point from the robot perspective view and uses each point for translation and rotation estimation separately. This method reduces the minimum number of feature points required for the optimization from at least three to two. Similar methods were also proposed in [18], [19] which use infinite homography. In [20], the orthogonality-index is defined, representing the orthogonality of the coordinates of feature points, then preferentially selects those with a good index during optimization.

As mentioned above, the prior works have shortened the processing time of the optimization phase in VO. However, the computational cost of stereo matching, which requires the most processing time in stereo VO, is equivalent to the case where all feature points are used. There is room to achieve further reduction in computation time.

To this end, the authors propose a feature selection scheme that selects feature points before stereo matching. The main contributions of this article are as follows.

- Using the information on the camera ' s orientation attached to the rover, the rough positions of feature points are selected based on their positional relationships.
- Using singular value decomposition (SVD) in the optimization phase, separating the feature points for rotation and translation estimation is possible. Taking advantage of this, the proposed method separates rotation and translation and selects the appropriate one for each..
- The proposed method was evaluated using an actual rover. It was confirmed that the processing time per frame can be reduced by up to 33% compared to the conventional method. The validity in the natural environment is also demonstrated.

II. Preliminaries

A. Notations and Definitions

Let $\mathcal{F}_{(\cdot)}$ denote a 3-dimensional coordinate frame. $\mathcal{F}_c$ and $\mathcal{F}_r$ are camera-centered frame and rover-centered frame, respectively. $\boldsymbol{p}_{(\cdot)} \in \mathbb{R}^{3\times 1}$ denotes 3 dimensional coordinates in $\mathcal{F}_{(\cdot)}$, $\boldsymbol{P}_{(\cdot)} \in \mathbb{R}^{3\times N}$ is the point group consisting of N points in $\mathcal{F}_{(\cdot)}$, and $\bar{\boldsymbol{P}}$ is the centroid of the point group $\boldsymbol{P}$. $\Delta\boldsymbol{P}$ denotes the point group $\boldsymbol{P}$ subtracted by the centroid $\bar{\boldsymbol{P}}$. The rotation matrix and translation vector in $\mathcal{F}_d$ to $\mathcal{F}_s$ is expressed as ${}^s\boldsymbol{R}_d \in \mathrm{SO}(3)$, $\boldsymbol{t}^s_d \in \mathbb{R}^3$, respectively. The image's pixel coordinate is represented by $\boldsymbol{u} = [u, v]^{\mathrm{T}} \in \mathbb{R}^2$ where u and v are the horizontal and vertical coordinates, respectively.

B. Stereo Camera Model and Setups

This study assumes that the stereo camera mounted on the rover has two cameras attached at the same height, separated by b, and facing in the same direction. The camera is required to be calibrated in advance to obtain the intrinsic, extrinsic, and distortion parameters. The pinhole camera model [21] can be considered against well-calibrated cameras. Given corresponding points u_L, u_R in the left and right images, the 3-dimensional coordinate $\boldsymbol{p}_c$ is calculated by triangulation.

$$\boldsymbol{p}_c = \begin{bmatrix} x & y & z \end{bmatrix}^{\mathrm{T}} = \frac{b}{d}\begin{bmatrix} u-u_0 & v-v_0 & f \end{bmatrix}^{\mathrm{T}} \quad (1)$$

where f is the baseline and focal length of the stereo camera, and $d = u_L - u_R$ is the disparity. It is assumed that the stereo camera is mounted on the rover at a height h and a depression angle θ_{dep}.

III. Relative Pose Estimation of Stereo Visual Odometry

Let $\mathcal{F}_s$ and $\mathcal{F}_d$ be camera frames at different points, and consider the problem of finding the relative orientation from the images taken at those points. First, the feature points are extracted from these images, and find the corresponding points between them. Then, stereo matching is performed to estimate the depth of the feature points and calculate the set of two point groups $\boldsymbol{P}_s, \boldsymbol{P}_d$. The relative pose of the camera between two frames, $\mathcal{F}_s$ and $\mathcal{F}_d$, is equivalent to the pose of the 3-dimensional coordinates $\boldsymbol{P}_s$ and $\boldsymbol{P}_d$. This results in an optimization problem that finds ${}^s\boldsymbol{R}_d, \boldsymbol{t}^s_d$ that minimizes the reprojection error.

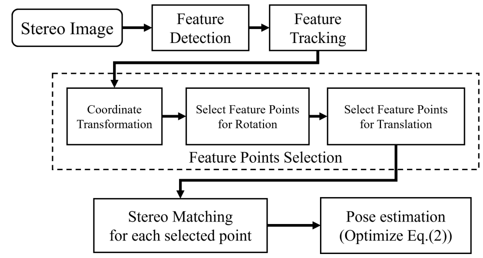


Fig. 1. VO algorithm with the proposed feature selection scheme.

$$\{{}^s\boldsymbol{R}_d, \boldsymbol{t}^s_d\} = \underset{\{{}^s\boldsymbol{R}_d, \boldsymbol{t}^s_d\}}{\arg\min} \sum_i ||\boldsymbol{p}^i_s - \boldsymbol{R}(\boldsymbol{p}^i_d - \boldsymbol{t})||^2 \quad (2)$$

The relative pose between $\mathcal{F}_s$ and $\mathcal{F}_d$ is recovered by solving the above optimization problem. The optimization problem is solved by methods such as [17]–[19], [22].

SVD is also the solution for the above optimization problem [23]. First, the ${}^s\boldsymbol{R}_d$ is obtained by the following equation.

$$\begin{aligned} {}^s\boldsymbol{R}_d &= \mathbf{V}\mathbf{U}^{\mathrm{T}} \\ \mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}} &= \mathrm{svd}(\Delta\boldsymbol{P}_s \Delta\boldsymbol{P}^{\mathrm{T}}_d) \end{aligned} \quad (3)$$

Next, the translation vector $\boldsymbol{t}^s_d$ is obtained by the following equation.

$$\boldsymbol{t}^s_d = \bar{\boldsymbol{P}}_s - {}^s\boldsymbol{R}_d \bar{\boldsymbol{P}}_d \quad (4)$$

As mentioned above, by using SVD, the rotational matrix and translational vector are separately estimated. So the feature points used for optimization can be separated into those for rotational estimation and translational estimation. This research aims to reduce the number of feature points used for optimization by selecting feature points suitable for rotation and translation estimation.

IV. Proposed Method

The VO algorithm using the proposed feature point selection method is shown in Fig. 1. First, as with general VO, feature point detection and tracking are performed. Next, the feature points are selected. Then, stereo matching is carried out on the selected feature points Finally, the relative pose is calculated using Eq. (3), (4). As described above, by incorporating the proposed method into VO, the number of times of stereo matching can be reduced, and the processing time can be decreased. The details of feature point selection are described below.

A. Coordinate Transformation

Before selecting feature points, the detected and tracked feature points are converted to the rover-centered frame $\mathcal{F}_r$. To calculate accurate 3-dimensional coordinates, it is necessary to perform stereo matching, which goes against the purpose of this research. Therefore, as shown in Fig. 2, a homography transformation is applied to the photographed image to convert it into a view from directly above. This conversion is based on the assumption that the mounting height and depression angle of the camera are easily predicted, and the approximate positional relationship with the ground is known. The following

Fig. 2. Homography transformation to convert the coordinate.

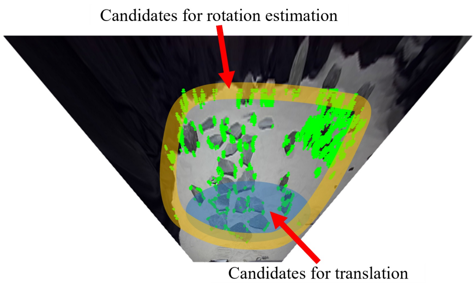


Fig. 3. Feature point candidates. The green points are all the feature points.

equation calculates the coordinate of each pixel $[x_r, y_r, z_r]$ in $\mathcal{F}_r$.

$$z_r \begin{bmatrix} x_r & y_r & 1 \end{bmatrix}^{\mathrm{T}} = h\,{}^r\boldsymbol{R}_c \begin{bmatrix} \frac{u-u_0}{f} & \frac{v-v_0}{f} & 1 \end{bmatrix}^{\mathrm{T}} \tag{5}$$

${}^r\boldsymbol{R}_c$ is derived from camera depression angle θ_{dep}.

The terrain is not flat and undulated on the planet's surface, so the coordinates are not strictly accurate. However, since the above transformation is used only for feature point selection, the effect of conversion error on position estimation is considered small. In the following, the coordinate system $\mathcal{F}_r$ is supposed to select feature points.

B. Feature Point Selection Scheme

This section selects N_R feature points for rotation estimation and N_t feature points for translation estimation. Figure 3 shows the outline of the feature point selection. In the proposed method, rotation and translation estimations are performed separately, so the feature points are also separately selected.

1) Selection for Rotation Estimation: As described in Eq. (3), SVD-based optimization considers the rotation around the centroid of the feature points. The spatial coordinates of each feature point include errors due to feature point matching and stereo matching, so when estimating rotation, the farther the center of gravity is, the more accurate estimation is possible. The proposed method preferentially selects from the outer edge of the feature point set. Given the set of all feature points $\boldsymbol{P}_r^{all}$, the vertices of the convex hull of $\boldsymbol{P}_r^{all}$ are selected as feature points for rotation estimation. Qucikhull [24] is used to find the convex hull. Then, the remaining feature points are randomly selected from the entire region to maintain robustness against outliers. In random sampling, the distance between the selected feature points is checked, and if the distance is less than ϵ_{dist}, the feature point is not selected.

2) Selection for Translation Estimation: The translation vector is calculated so that the centroids of two feature points overlap. To accurately estimate translation, it is desirable to select feature points with high stereo matching accuracy. Stereo matching can calculate depth with higher precision for points located in the rover's vicinity, so N_t points are selected in order from the bottom of the image, which is the area closest to the rover.

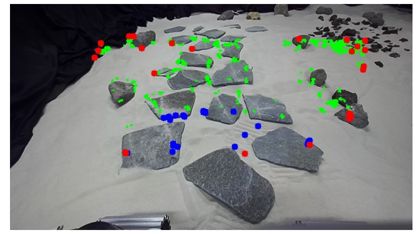


Fig. 4. Selected feature points. The red points are the feature points for rotation estimation, and the blue points are the feature points for translation estimation. The green points are the feature points that are not selected.

An example of the finally selected feature points is shown in Fig. 4.

After selection, stereo matching is performed on the selected feature points to calculate the 3-dimensional coordinates $\boldsymbol{P}_s^{rot}$, $\boldsymbol{P}_s^{tran}$, $\boldsymbol{P}_d^{rot}$, and $\boldsymbol{P}_d^{tran}$.

C. Optimization using selected feature points

Using the feature points $\boldsymbol{P}_s^{rot}$, $\boldsymbol{P}_s^{tran}$, $\boldsymbol{P}_d^{rot}$, and $\boldsymbol{P}_d^{tran}$, the relative pose is calculated using the following equations.

$$\begin{aligned} {}^s\boldsymbol{R}_d &= \mathbf{V}\mathbf{U}^{\mathrm{T}} \\ \mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}} &= \mathrm{svd}(\Delta\boldsymbol{P}_s^{rot}\Delta\boldsymbol{P}_d^{rot\,\mathrm{T}}) \\ \boldsymbol{t}_d^s &= \bar{\boldsymbol{P}}_s^{tran} - {}^s\boldsymbol{R}_d\bar{\boldsymbol{P}}_d^{tran} \end{aligned} \tag{6}$$

If there are feature points whose depth estimation by stereo matching or tracking between images is incorrect, the estimation accuracy may deteriorate significantly. Therefore, in order to increase the robustness against outliers, the proposed method adopts RANSAC (RANdom SAmple Consensus) [25] when solving the above equation.

V. Experiment

This section evaluates the proposed method using two types of data: a testbed rover in the laboratory environment and a dataset on the natural terrain.

A. Evaluation with Testbed Rover

This experiment aims to verify the effectiveness of the proposed method in terms of processing time and accuracy. The effect of the number of feature points on the accuracy is also demonstrated.

1) Setup: The experiment used a field that reproduced the surface of the Moon and Mars. The testbed rover developed by JAXA/ISAS which mounted ZED2i stereo camera was used in this experiment. The rover's position was also measured using the motion capture camera system OptiTrack. The rover drove one lap around the field at $0.05\,\mathrm{m/s}$, and images were taken at 3-second intervals. The overall trajectory length was $12.2\,\mathrm{m}$. Figure 5 shows images taken during the driving.

2) Evaluation method: The following three methods were used for comparison.

- SVD-ALL: Optimize using all feature points by Eq. (3) and (4)
- LM-ALL: Optimize Eq. (2) using the Levenberg-Marquardt method by all feature points

Fig. 5. Snapshots of the captured images

- SVD-BKT (N_s): Optimize using the Eq. (3), (4), and N_s feature points are selected by the Bucketing Technique [11].

All the methods used the Harris corner detector [26] for feature point detection and ORB [27] for feature descriptors. The input of the visual odometry is stereo images taken by ZED2i. ZED2i outputs depth images as well as stereo images, but this experiment does not use depth images to measure the processing time of stereo matching.

The accuracy and the processing time per frame evaluate the effectiveness of each method. Regarding accuracy, 3 types of metrics were evaluated: Absolute Trajectory Error (ATE), Relative Pose Error (RPE), and Relative Orientation Error (ROE) [28]. ATE is the error between the estimated trajectory and the ground truth trajectory. RPE and ROE are the errors between the estimated and ground truth trajectories at each frame. The CPU used for verification was Intel Core i9-11900K@3.5GHz.

The accuracy of SVD-BKT and the proposed method was evaluated while changing the number of feature points from 30 to 500. In the proposed method, this value is the sum of the number of feature points N_R for rotation estimation and the number N_t of feature points for translation estimation. In addition, N_R and N_t were set to be the same.

The evaluation was performed 10 times for each method.

3) Results and Discussion: The experimental results are summarized in Table I. Regarding the processing time, the proposed method improved the processing speed by 3.0 times compared to SVD-ALL. Further analysis of the processing time is shown in Fig. 6. As shown in the figure, the selected-based method (SVD-BKT and proposed method) significantly reduced the time required for stereo matching. The average number of tracked feature points in SVD-ALL and LM-ALL is 1768 ± 865 per frame. SVD-ALL and LM-ALL are required to execute stereo matching for all these feature points, while the proposed method and SVD-BKT are only required to execute stereo matching, in this case, for 50 points. The above demonstrated the effectiveness of narrowing down the feature points before stereo matching.

Although positive results were obtained in terms of processing time, performance in terms of ATE deteriorated. Since there is no significant difference in RPE and ROE between the proposed method and the method that uses all feature points, it is inferred that the error becomes large in some specific frames.

Figure 7 shows the relationship between the number of feature points and the accuracy. The more feature points used, the higher the estimation accuracy becomes, and the accuracy approaches that of a method that uses all feature points. Although the proposed method was superior in all cases in RPE, SVD-BKT showed better results in ROE. The proposed method reduced the rotation error by selecting points that move away from the center of gravity. It means the feature points for rotation estimation were preferentially selected from the outer edge. However, since the accuracy of stereo matching is low for points located far away, a dilemma arises in that the rotation accuracy deteriorates when such points are used. Further accuracy improvement can be expected by considering a selection method that considers the trade-off between the error caused by stereo matching and the error caused by the distance from the center of gravity. On the other hand, since the proposed method yields good results for RPE, it is concluded that it can obtain appropriate feature points for translation estimation.

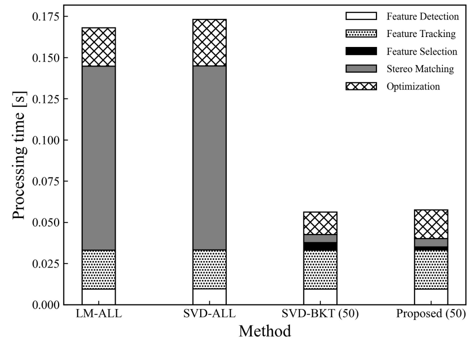


Fig. 6. Average processing time for each process of VO.

B. Evaluation on Natural Terrain

This experiment evaluates the proposed method in a natural terrain environment.

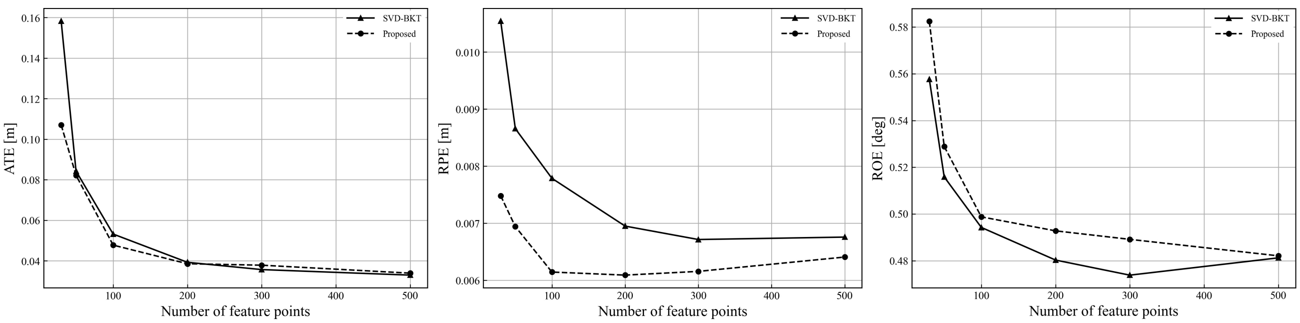
1) Setup: The MADMAX dataset [29] provides stereo image sequences on natural terrain. The dataset is taken in various environments, including flat, undulated, rocky, and sandy terrain. The image data is corresponded to the ground truth trajectory measured by RTK-GPS. In this experiment, sequences A-0 and H-0 were used. Sequence A-0 has flat terrain that is easy for the rover to drive on and is a feature-rich environment with many rocks and pebbles. Sequence H-0 is a challenging environment for the proposed method. The surface of the terrain is covered with sand, so it is difficult to extract valid feature points. In addition, the terrain is undulated, so the assumption that the posture relationship between the camera and the ground is known is violated. The distances of the parts used for evaluation are $41.5\,\mathrm{m}$ and $58.5\,\mathrm{m}$ for A-0 and H-0, respectively.

The MADMAX dataset uses hand-held modules, so the exact camera height and angle are unknown. But [29] reported that the approximate height and angle are $h = 1.20\,\mathrm{m}$ and $\theta_{dep} = 28\,\mathrm{deg}$, respectively. Therefore, the same parameters were used for the proposed method.

2) Evaluation: The proposed method was compared with SVD-ALL and SVD-BKT. SVD-BKT selects 200 feature points, and the proposed method selects $N_R = 100$ feature

TABLE I
AVERAGE VALUES OF PROCESSING TIME, ATE, RPE, AND ROE ON AKI ROVER

	Processing Time [s/frame]	ATE [m]		RPE [m]	ROE [deg]
SVD-ALL	0.173 ± 0.0882	0.0274 ± 0.00306	(0.22 %)	0.00632 ± 0.000147	0.466 ± 0.00409
LM-ALL	0.168 ± 0.0827	0.0298 ± 0.00435	(0.24 %)	0.00571 ± 0.000167	0.460 ± 0.00290
SVD-BKT (50)	0.0562 ± 0.0236	0.0843 ± 0.0442	(0.69 %)	0.0105 ± 0.00202	0.582 ± 0.0177
Proposed (50)	0.0576 ± 0.0220	0.0822 ± 0.0326	(0.67 %)	0.00694 ± 0.000486	0.489 ± 0.00452



(a) ATE (b) RPE (c) ROE

Fig. 7. The number of feature points vs. accuracy.

TABLE II
AVERAGE VALUES OF PROCESSING TIME, ATE ON MADMAX DATASET

	Sequence A-0			Sequence H-0		
	Processing Time [s/frame]	ATE [m]		Processing Time [s/frame]	ATE [m]	
SVD-ALL	0.357 ± 0.0444	0.744	(1.79 %)	0.295 ± 0.0757	1.77	(3.02%)
SVD-BKT	0.100 ± 0.00791	1.41	(4.41 %)	0.0873 ± 0.00912	2.18	(3.73%)
Proposed	0.112 ± 0.00777	1.00	(2.41 %)	0.100 ± 0.0107	1.88	(3.22%)

points for rotation estimation and $N_t = 100$ feature points for translation estimation. For all methods, feature detection and detector were performed by ORB. The evaluation was performed 1 time for each method. The processing time and ATE were evaluated.

3) Results and Discussion: Table II shows each sequence's processing time and ATE results. Similar to the experiment in the previous section, for both sequence A-0 and H-0, feature point selection-based methods such as SVD-BKT and the proposed method succeeded in shortening the calculation time. Regarding ATE, SVD-ALL scored the best results, followed by the proposed method, and SVD-BKT is the worst. Figure 8 shows the estimated trajectory for each method and each sequence. For Sequence A-0, all methods can estimate a generally correct trajectory. The error gradually spreads due to drift, and it is most noticeable in SVD-BKT. This result shows that the proposed method is effective even on natural terrain and that the error drift increases if feature points suitable for position estimation are not selected. In sequence H-0, SVD-ALL and the proposed method estimated the correct trajectory, but SVD-BKT failed to estimate it halfway. Sequence H-0 has severe ups and downs, and there is a risk that the results of co-ordinate transformation using Fig. 2 may deviate significantly from the actual coordinates. However, the results showed no direct negative effect on position estimation, confirming the effectiveness of the proposed method on natural terrain.

VI. CONCLUSION

This paper proposed a method to select feature points in VO based on their positional relationships to improve the processing speed of VO. By taking advantage of the fact that rotational and translational estimation can be separated by using SVD during optimization, the feature points for rotation and ones for translation were separately selected. For rotation, feature points are selected which are located at the outer edge of the feature point set, and for translation, are selected which feature points are located in the area close to the rover. The experiment result showed that the proposed method was highly effective in processing speed. It was also verified that the proposed method could improve the accuracy compared to Bucketing Technique. In addition, the experiment using the MADMAX dataset demonstrated the proposed method was valid in the natural environment. However, the accuracy of the proposed method was slightly lower than the VO using all feature points. To select important feature points for position estimation is under planning as a future work.

ACKNOWLEDGMENT

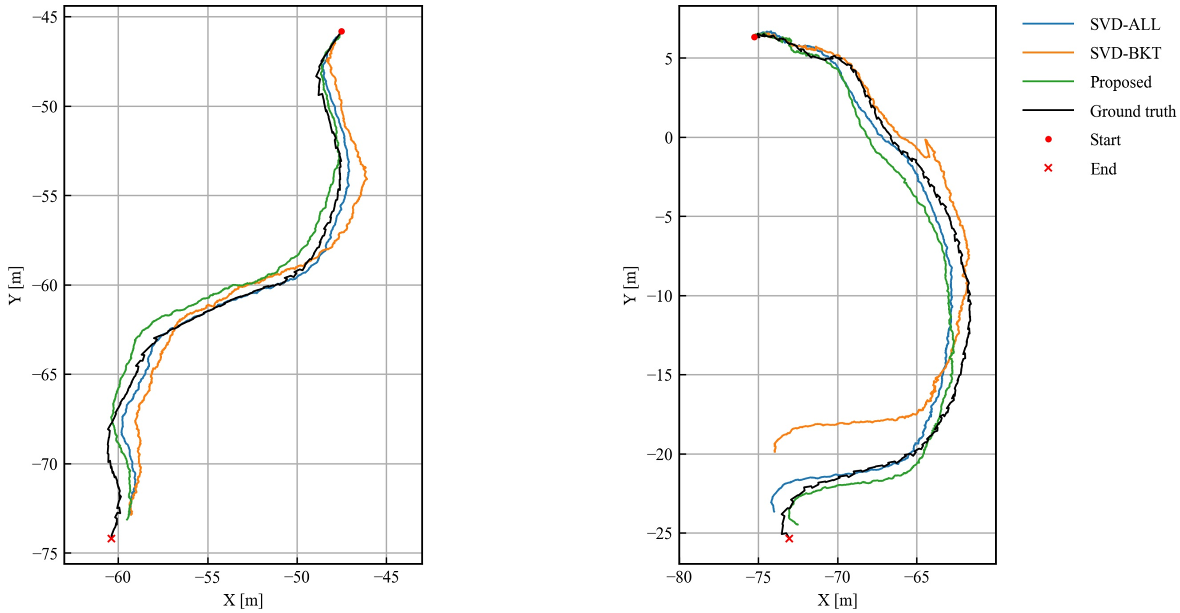
This work was supported by JST SPRING Grant Number JPMJSP2108 and JSPS KAKENHI Grant Number 23K26081.

REFERENCES

[1] J. Grotzinger, S. Gupta, M. Malin, D. Rubin, J. Schieber, K. Siebach, D. Sumner, K. Stack, A. Vasavada, R. Arvidson, *et al.*, "Deposition, exhumation, and paleoclimate of an ancient lake deposit, gale crater, mars," *Science*, vol. 350, no. 6257, p. aac7575, 2015.

[2] S. B. Goldberg, M. W. Maimone, and L. Matthies, "Stereo vision and rover navigation software for planetary exploration," in *Proc. IEEE Aerospace Conference (AERO)*, vol. 5, 2002, pp. 5–5.

[3] J. J. Biesiadecki, P. C. Leger, and M. W. Maimone, "Tradeoffs between directed and autonomous driving on the mars exploration rovers," *The International Journal of Robotics Research*, vol. 26, no. 1, pp. 91–104, 2007.



(a) Sequence A-0 (b) Sequence H-0

Fig. 8. Estimated trajectory for each method on MADMAX dataset.

[4] J. Carsten, A. Rankin, D. Ferguson, and A. Stentz, "Global planning on the mars exploration rovers: Software integration and surface testing," *Journal of Field Robotics*, vol. 26, no. 4, pp. 337–357, 2009.

[5] V. Verma, F. Hartman, A. Rankin, M. Maimone, T. Del Sesto, O. Toupet, E. Graser, S. Myint, K. Davis, D. Klein, J. Koch, S. Brooks, P. Bailey, H. Justice, M. Dolci, and H. Ono, "First 210 solar days of mars 2020 perseverance robotic operations - mobility, robotic arm, sampling, and helicopter," in *Proc. IEEE Aerospace Conference (AERO)*, 2022, pp. 1–20.

[6] M. Maimone, Y. Cheng, and L. Matthies, "Two years of visual odometry on the mars exploration rovers," *Journal of Field Robotics*, vol. 24, pp. 169–186, 3 2007.

[7] A. Rankin, M. Maimone, J. Biesiadecki, N. Patel, D. Levine, and O. Toupet, "Driving curiosity: Mars rover mobility trends during the first seven years," in *Proc. 2020 IEEE Aerospace Conference (AERO)*, 2020, pp. 1–19.

[8] S. Andolfo, F. Petricca, and A. Genova, "Precise pose estimation of the nasa mars 2020 perseverance rover through a stereo-vision-based approach," *Journal of Field Robotics*, vol. 40, no. 3, pp. 684–700, 2023.

[9] D. Scaramuzza and F. Fraundorfer, "Visual odometry [tutorial]," *IEEE Robotics & Automation Magazine*, vol. 18, no. 4, pp. 80–92, 2011.

[10] R. Rieber, "Autonavigation: intuitive autonomy on mars and at sea; the mobility system for mars2020," 2018. [Online]. Available: https://hdl.handle.net/2014/49767

[11] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," *Artificial Intelligence*, vol. 78, pp. 87–119, 10 1995.

[12] B. Kitt, F. Moosmann, and C. Stiller, "Moving on to dynamic environments: Visual odometry using feature classification," in *Proc. 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2010, pp. 5551–5556.

[13] A. Geiger, J. Ziegler, and C. Stiller, "Stereoscan: Dense 3d reconstruction in real-time," in *Proc. IEEE Intelligent Vehicles Symposium (IV)*, 2011, pp. 963–968.

[14] Y. Zhao and P. A. Vela, "Good feature selection for least squares pose optimization in vo/vslam," in *Proc. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 10 2018, pp. 1183–1189.

[15] Y. Zhao and P. A. Vela, "Good feature matching: Toward accurate, robust vo/vslam with low latency," *IEEE Transactions on Robotics*, vol. 36, pp. 657–675, 6 2020.

[16] M. Buczko and V. Willert, "How to distinguish inliers from outliers in visual odometry for high-speed automotive applications," in *Proc. 2016 IEEE Intelligent Vehicles Symposium (IV)*, 6 2016, pp. 478–483.

[17] K. Otsu and T. Kubota, "A two-point algorithm for stereo visual odometry in open outdoor environments," in *Proc. 2014 IEEE International Conference on Robotics and Automation (ICRA)*, 5 2014, pp. 1042–1047.

[18] K. Ni and F. Dellaert, "Stereo tracking and three-point/one-point algorithms - a robust approach in visual odometry," in *Proc. 2006 International Conference on Image Processing*, 2006, pp. 2777–2780.

[19] M. Kaess, K. Ni, and F. Dellaert, "Flow separation for fast and robust stereo odometry," in *Proc. 2009 IEEE International Conference on Robotics and Automation (ICRA)*, 2009, pp. 3539–3544.

[20] H. H. Nguyen and S. Lee, "Orthogonality index based optimal feature selection for visual odometry," *IEEE Access*, vol. 7, pp. 62 284–62 299, 2019.

[21] M. Young, "Pinhole optics," *Applied Optics*, vol. 10, p. 2763, 12 1971.

[22] Y. Cheng, M. Maimone, and L. Matthies, "Visual odometry on the mars exploration rovers," in *Proc. 2005 IEEE International Conference on Systems, Man and Cybernetics*, vol. 1. IEEE, 2005, pp. 903–910.

[23] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares fitting of two 3-d point sets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, no. 5, pp. 698–700, 1987.

[24] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, "The quickhull algorithm for convex hulls," *ACM Transactions on Mathematical Software*, vol. 22, pp. 469–483, 12 1996.

[25] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, p. 381–395, jun 1981.

[26] C. Harris, M. Stephens, *et al.*, "A combined corner and edge detector," in *Proc. Alvey vision conference*, vol. 15, no. 50. Citeseer, 1988, pp. 10–5244.

[27] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *Proc. 2011 International Conference on Computer Vision*, 2011, pp. 2564–2571.

[28] J. Sturm, W. Burgard, and D. Cremers, "Evaluating egomotion and structure-from-motion approaches using the tum rgb-d benchmark," in *Proc. of the Workshop on Color-Depth Camera Fusion in Robotics at the IEEE/RJS International Conference on Intelligent Robot Systems (IROS)*, vol. 13, 2012.

[29] L. Meyer, M. Smíšek, A. Fontan Villacampa, L. Oliva Maza, D. Medina, M. J. Schuster, F. Steidle, M. Vayugundla, M. G. Müller, B. Rebele, A. Wedler, and R. Triebel, "The madmax data set for visual-inertial rover navigation on mars," *Journal of Field Robotics*, vol. 38, no. 6, pp. 833–853, 2021.