A METHOD FOR ROTATION MATRIX ESTIMATION IN STEREO VISUAL ODOMETRY BASED ON LEFT AND RIGHT ROTATION AVERAGING

The Tien Nguyen[1], Cong Manh Tran[1], Quang Thi Nguyen[1], Xuan Phuc Nguyen[1], Huu Hung Nguyen[1,*]
[1]*Le Quy Don Technical University, Hanoi, Vietnam*

Abstract

For Stereo Visual Odometry (SVO), the rotation and translation of camera motion can be estimated simultaneously or separately in which rotation is extracted from an essential matrix. Most of state of art methods computing the parameters of rotation camera motion only use each pair of consecutive image frames on the left side. This leads to not leveraging the worth information of a pair of consecutive image frames on the right side. This paper presents an approach to leverage this information during computing rotation by averaging of value of rotations extracted from both left and right sides. The proposed method is evaluated on the KITTI dataset to verify the performance of this algorithm. The experimental result indicates that the proposed approach enhances about 10% accuracy compared to other methods in the same scenario.

Keywords: *Stereo Visual Odometry; Essential Matrix; Rotation Averaging; Robotics.*

1. Introduction

Visual Odometry (VO) is the process of estimating the pose (position and orientation) of an agent such as a vehicle, robot, etc., by analyzing the associated camera images. It has an essential role in major domains including robotics, automotive, wearable computing, and augmented reality. One of the fundamental challenges in an autonomous system is accurate localization. Maintaining knowledge of its position over time to achieve autonomous navigation is an essential duty of robot. Therefore, there are a lot of variety of sensors, techniques, and systems that are used for a mobile robot, such as wheel odometry, laser/ultrasonic odometry, global position system (GPS), global navigation satellite system (GNSS), inertial measurement units (IMUs) and VO. They have been studied and developed by scientists on over the world. However, each technique has its advantages and disadvantages. Compare to other approaches, VO is a low-cost technique that provides more accurate trajectory estimation as mentioned in [1, 2].

VO determines the pose of a vehicle by analyzing the image stream captured by camera attached on it. To work effectively, it needs to guarantee several necessary

[*] Email: hungnh.isi@lqdtu.edu.vn

conditions such as the sufficient illumination, the static scene with enough texture to allow apparent motion to be extracted. Furthermore, consecutive frames should be captured to ensure that they have sufficient scene overlap. Depending on the type of visual sensors, VO can be divided into two major types including Monocular Visual Odometry and SVO in which most of research done in VO has been produced using stereo cameras. Beside that, based on the input of pose estimation steps, there are several different methods to compute VO, and they can be divided into three categories as [1]: feature-based methods, appearance-based methods, and hybrid methods. Feature-based methods use salient and repeatable features to track over the frame; appearance-based methods use the intensity information of all pixels in the image or sub-regions of it; and hybrid methods combine of the two above methods.

The VO pipeline consists of five parts in sequence as Fig. 1: image sequence, feature detection, feature matching (or tracking), motion estimation and local optimization (bundle adjustment). *Motion estimation* is the core computation step performed for every image in VO system.

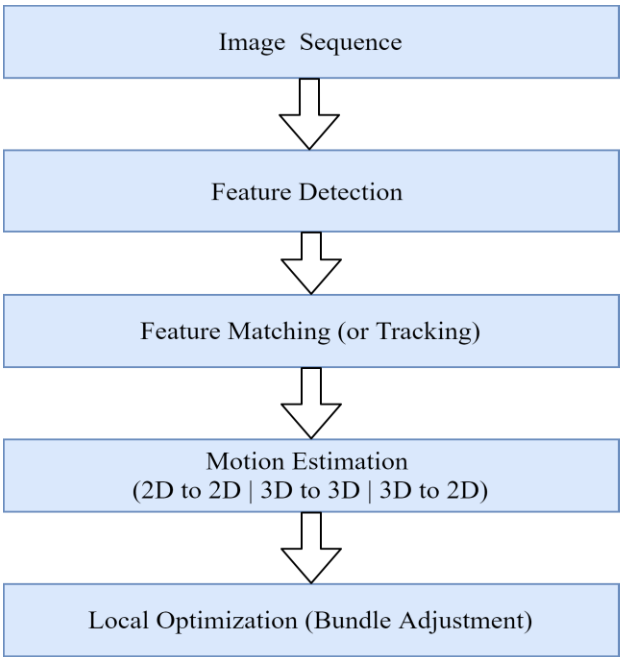

Fig. 1. The VO pipeline.

Depending on whether the feature correspondences f_{k-1}, f_k at time instants $(k-1)$ and k, respectively are specified in two or three dimensions, motion estimation can be divided into three methods: 1) 2-D to 2-D when f_{k-1}, f_k are in 2-D image coordinates; 2) 3-D to 3-D when f_{k-1}, f_k are in 3-D image coordinates; 3) 3-D to 2-D when f_{k-1} in 3-D coordinates and f_k is corresponding 2-D re-projection on the image coordinate. The survey in [3, 4, 5] concludes that 2-D to 2-D and 3-D to 2-D methods are more accurate than 3-D to 3-D methods because the motion estimation using 3-D to 3-D correspondences are more uncertainty then it may have a devastating effect in motion computing. To avoid the effect of uncertainty 3-D feature from stereo disparity, in [6] we presented a novel translation estimation for essential matrix based SVO. The rotation was extracted from the essential matrix of each pair of consecutive image frames on the left side; then the translation was rapidly estimated by solving a linear closed form only using 2D features as input with one-point RANSAC.

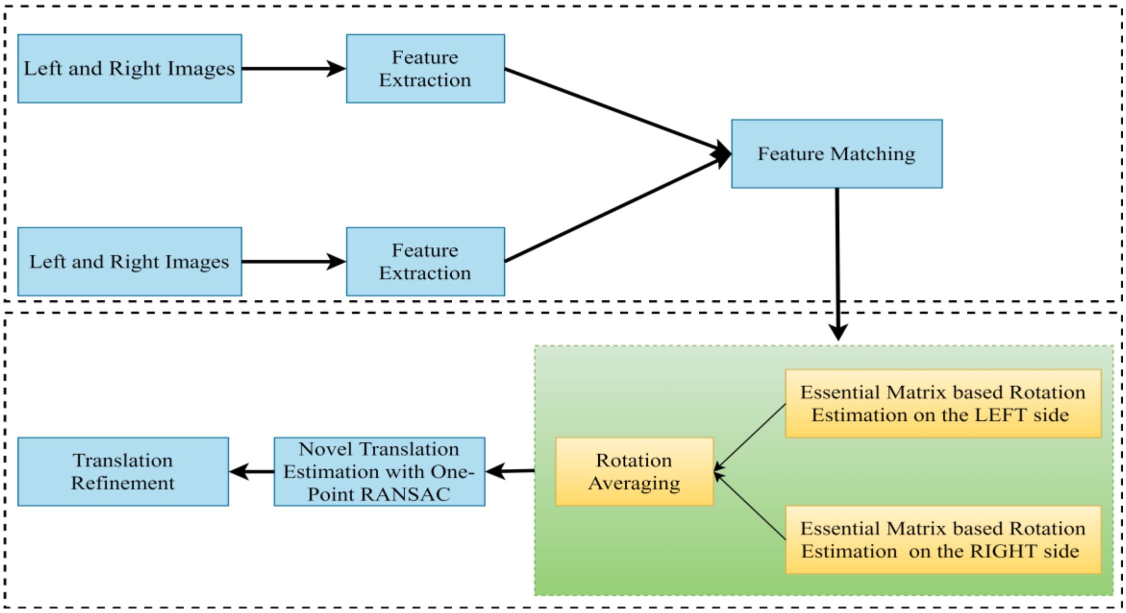

Fig. 2. The proposed VO pipeline with improvement of rotation averaging.

Carrying on improving the performance of algorithm, we realize that most of methods only use a pair consecutive of frames on the left side to compute the essential matrix, the worth information of a pair consecutive image frames on the right side is not leveraging. It is really waste of information because we are not sure whether the value of information extracted from each pair of image on the left side more accurate than that of the right side.

In this paper, we propose an idea computing the average of rotation using combination of each pair of consecutive image frames on the left and right sides. The algorithm was tested on the training KITTI dataset and achieved significant results. The approach enhances about 10% the accuracy of rotation and also improves a value of translation. Our algorithms is depicted as follows in Fig. 2. Like the traditional VO pipeline, the proposed VO pipeline consists of two main phases: 1) *feature extraction/matching* and 2) *pose estimation*. The contribution of paper is highlighted as green and yellow block where rotation estimation was determined by averaging rotation each image frames on the left and right side.

The rest of this paper is organized as follows. Section II summarizes the related works for essential matrix-based visual odometry. Section III deploys the improvement of rotation estimation by using each of pair of consecutive image frames on the left and right sides. Section IV gives several results and evaluates them via comparing to other approaches on KITTI dataset.

2. Related works

In this section, we briefly summarize conventional approaches to determine the value of rotation and translation. We divided the section into two parts including rotation estimation and translation estimation.

2.1. Rotation estimation

The essential matrix, E, is a 3×3 matrix expressing the geometric relation between two consecutive images I_k and I_{k-1}. It contains the camera motion parameters up to an unknown scale factor for the translation in the following form:

$$E = T^{\times}R \tag{1}$$

where matrix $T^{\times}$ is expressed as follows:

$$T^{\times} = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \tag{2}$$

The essential matrix can be computed from 2-D to 2-D feature correspondences. Each correspondence of two images satisfies the epipolar constraint as follows:

$$p^T E q = 0 \tag{3}$$

where p and q are a feature location in one image (e.g., I_k) and the location of its corresponding in another image (e.g., I_{k-1}), respectively. Note that, p and q are normalized image coordinate. The essential matrix E has two additional properties

$$\det\left(E\right) = 0 \tag{4}$$

and

$$2EE^T E - trace(EE^T)E = 0 \tag{5}$$

Note that, equation (3) can be rewritten in the linear formula as follows:

$$\hat{\upsilon}\hat{E} = 0 \tag{6}$$

where

$$\hat{\upsilon} = \left[p_1q_1, p_2q_1, p_3q_1, p_1q_2, p_2q_2, p_3q_2, p_1q_3, p_2q_3, p_3q_3 \right] \tag{7}$$

and

$$\hat{E} = \left[E_{11}, E_{12}, E_{13}, E_{21}, E_{22}, E_{23}, E_{31}, E_{32}, E_{33} \right]^T \tag{8}$$

Essential matrix can be solved by using five-point correspondences gives the linear equation (6) and by solving the system the parameters of E can be computed in [7]. Three equations (4), (5) and (6) are extended to 10 cubic constraints and transformed to a ten-degree polynomial. There are maximum 10 solutions for this polynomial resulting in maximum 10 candidates essential matrices. However, the solution yielding the highest number of inliers is known as a good representative. This five-point algorithm is applied in conjunction with RANSAC approach. A number of five-point sets are randomly chosen and evaluated the preemptive scores. The one with the best preemptive scoring together with the largest number of inliers is considered as the final solution.

2.2. Translation estimation

As above explained, we brief the method how to extract the value of rotation based an essential matrix by an efficient five-point algorithm proposed by Nister [7]. A RANSAC scheme was used to choose the smallest preemptive score from N set of five-point samples. From essential matrix E, we computed the rotation, however, there are three unknown parameters of the translation. A simple solution is that using a pair of 3-D feature correspondences (P, Q) with the RANSAC scheme.

$$P = RQ + t \tag{9}$$

where Q, P are 3-D corresponding feature of current and previous frames, respectively. This solution is similar to the 3-D to 3-D method, the result of this algorithm gets a high error of translation due to the high uncertainty of 3-D points. To avoid the effect of uncertainty 3-D points, in the paper [6], we propose a novel translation estimation. The process of the algorithm was summarized as follows. Firstly, the value was extracted from essential matrix using five-point Nister's algorithm [7] as mentioned above part. Secondly, the translation was determined via the proposed equations without 3-D input since projecting a 3-D point in the current left (world coordinate) to the pixel coordinate of two consecutive frames. Finally, we got an equation has a form as follows:

$$A_{8\times 6}\begin{bmatrix} X \\ Y \\ Z \\ t_x \\ t_y \\ t_z \end{bmatrix} = B_{8\times 1} \tag{10}$$

where X, Y, Z are 3-D points; t_x, t_y, t_z are the value of translation. Equation (10) includes eight linear equations with six unknown variables. It can be solved completely via Pseudo Inverse method to get M. Note that M is a matrix 6×1 is calculated by a following formula:

$$M = \begin{bmatrix} X & Y & Z & t_x & t_y & t_z \end{bmatrix}^T = \begin{bmatrix} A^T A \end{bmatrix}^{-1} A^T B \tag{11}$$

In the real situation, the noise of feature is always existing and they come from lot of resources such as light condition, imperfect camera calibration, etc. To guarantee the accuracy of translation estimation, we use 100 samples of closest 3-D features combining with RANSAC scheme to estimate candidate translation. Maximum inliers of best translation solution are used for refinement. Equation (10) is written for only one feature and it can be rewritten as the following equation:

$$\begin{bmatrix} A_{8\times 3}^{1XYZ} & A_{8\times 3}^{1T} \end{bmatrix}_{8\times 6}^{1} \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \\ t_x \\ t_y \\ t_z \end{bmatrix}_{6\times 1} = B_{8\times 1}^{1T} \tag{12}$$

where the matrix A is splited into 2 sub-matrices

$$A_{8\times 6}^{1} = \begin{bmatrix} A_{8\times 3}^{1XYZ} & A_{8\times 3}^{1T} \end{bmatrix}$$

For N features, equation (12) was generalized has following form:

$$A_n M_n = B_n \tag{13}$$

where

$$A_n = \begin{bmatrix} A_{8\times 3}^{1XYZ} & 0 & \cdots & \cdots & A_{8\times 3}^{1T} \\ 0 & A_{8\times 3}^{2XYZ} & \cdots & \cdots & A_{8\times 3}^{2T} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & A_{8\times 3}^{nXYZ} & A_{8\times 3}^{nT} \end{bmatrix}_{8n\times(3n+3)}$$

$$M_n = \begin{bmatrix} X_1 & Y_1 & Z_1 & \dots & X_n & Y_n & Z_n & t_x & t_y & t_z \end{bmatrix}^T_{1\times(3n+3)}$$

$$B_n = \begin{bmatrix} B_1 & B_2 & \dots & B_n \end{bmatrix}^T_{1\times 8n}$$

Equation (13) is solved by the Pseudo Inverse method to refine the initial estimation. Obviously, far distance features with small disparity do not only provide a good contribution to enhance translation accuracy but are also one cause increasing the amount of computation. To deal with this problem, the paper [6] only use 10 inliers closest 3-D features with top largest disparity for refinement.

3. Proposed rotation estimation

As mentioned above, the rotation was extracted from an essential matrix. Most of previous traditional methods only use each pair of consecutive image frames on the left side. This issue leads to not leverage lots of worth information of each pair of consecutive image frames on the right side. In this paper, we propose an approach to compute the average of rotation estimation by using information from each pair of consecutive image frames on the left and right sides. To be convenient for computing, using quaternion definition was introduced by Rowan Hamilton in 1843 in [8, 9] to express rotation matrix. Compared to other expression of rotation such as rotation matrices, Axis-Angle, Euler Angles, and Homogeneous Transformation Matrix, etc., quaternion is more compact, efficient, and numerically stable.

Our approach is implemented step by step as follows:

• *Firstly*, the value of rotation matrix R_1, R_2 are extracted separately from each pair of consecutive image frames on left and right side, respectively.

• *Secondly*, we convert the rotation matrices R_1, R_2 into quaternion forms q_1, q_2

• *Thirdly*, computing the average of rotation estimation by using Spherical linear interpolation (Slerp) function that expressed by following form by equation (14):

$$q_{avg} = Slerp(q_1, q_2, \alpha) \tag{14}$$

in which q_1, q_2 are two quaternions, α is coefficient range [0,1].

Slerp function can be described briefly as follows:

Quaternion spherical linear interpolation (Slerp) is an extension of linear interpolation along a plane to spherical interpolation in three dimensions. The algorithm was first proposed in [10]. Given two quaternions, q_1 and q_2, Slerp interpolates a new quaternion, q_{avg}, along the great circle that connects q_1 and q_2. The interpolation coefficient, α, determines how close the output quaternion is to either q_1 and q_2.

The Slerp algorithm can be described in terms of sinusoids:

$$q_0 = \frac{\sin(1-\alpha)\theta}{\sin(\theta)} q_1 + \frac{\sin(\alpha\theta)}{\sin(\theta)} q_2 \qquad (15)$$

where q_1 and q_2 are normalized quaternions, and θ is half the angular distance between q_1 and q_2.

• *Finally*, the average of rotation q_{avg} is converted back rotation matrix form R_{avg} that is input of getting translation estimation.

Algorithm:

Input: *Two rotation matrix on both left and right* $R_1 = R_{(k-1)l}$ *and* $R_2 = R_{(k-1)r}$ *computed from matrices from essential matrix using five-points of David Nister [7].*
Step 1: Compute q_1 *and* q_2*:* - $q_1 = converttoQua(R_1)$ - $q_2 = converttoQua(R_{2)}$ *Step 2: Compute average quaternion* q_{avg}*:* - $q_{avg} = Slerp(q_1, q_2, \alpha)$ *Step 3: Compute* R_{avg}*:* - $R_{avg} = converttoRot(q_{avg})$
Output: *Return* R_{avg} *is input to calculate* T *(translation)*

4. Experimental results

We assessed our approach using the KITTI dataset that be very popular with researchers in VO community. The dataset consists of different traffic scenarios that accommodate challenging aspects like lighting, shadow conditions, and dynamic object moving. The KITTI dataset contains 22 stereo sequences in total that was divided into two sub-sets including *Training and Testing dataset*. The training set provides 11 sequences (00-11) with the ground-truth trajectories for training evaluation and the testing set consist of 11 sequences (11-22) without the ground-truth for online evaluation. We will assess our approach on both two types of sub-sets to get its performance. The performance of the VO approaches is based on the RMSEs of measuring rotation/translation errors. These metrics are defined in [11] by computing the average errors from all possible sub-sequences of lengths (100, 200,..., 800 meters). We compared the result of our approach to other approaches such as *VISO2* [12], *MRPE* in [13], *Novel Translation* in [6] based on the training dataset to see the performance of algorithm.

A. Training dataset

Both of rotation and translation errors of 10 sequences (exception 1st sequence) of training KITTI dataset are shown in Tab. 1 that visualizes the average rotation error r_e in degree/100m, average translation in percentage (%) t_e and absolute error t_a in (meter) between the final frame of estimation and the ground-truth. Tab. 1 depicts the results of 4 methods including the popular VISO2 [12], MRPE [13], Novel Translation in [6] and our proposed method, respectively. The second sequence contains very fast moving frames (car moving up to 100 km/h), so the pose estimation was not stable. To compare fairly between methods, we considered ignoring its results.

Tab. 1. Offline performance evaluation on KITTI Dataset

Sec Num	**VISO2** [12]			**MRPE** [13]			**NOVEL TRANS** [6]			**OURS**		
	t_e (%)	r_e (deg/100 m)	t_{abs} (m)	t_e (%)	r_e (deg/100 m)	t_{abs} (m)	t_e (%)	r_e (deg/100 m)	t_{abs} (m)	t_e (%)	r_e (deg/100 m)	t_{abs} (m)
1	2.46	1.18	86.01	1.11	0.46	18.17	**1.08**	0.46	**11.06**	1.13	**0.45**	12.51
2	-	-	-	-	-	-	-	-	-	-	-	-
3	2.19	0.81	140.78	**0.95**	**0.36**	39.92	0.98	0.4	**21.10**	1.03	**0.36**	41.25
4	2.54	1.20	32.61	0.93	0.45	6.76	1.05	0.4	3.47	**0.85**	**0.37**	**3.08**
5	1.02	0.87	4.22	0.66	0.26	**2.64**	**0.56**	0.34	3.43	0.59	**0.17**	3.06
6	2.07	1.12	46.58	0.88	0.40	17.69	**0.83**	0.39	17.13	0.92	**0.38**	**14.90**
7	1.31	0.92	8.9	1.11	0.49	18.92	**0.85**	**0.40**	**8.07**	1.11	**0.40**	9.91
8	2.30	1.77	21.29	3.23	1.56	27.59	**1.44**	1.28	**13.30**	1.67	**1.19**	14.86
9	2.74	1.33	35.12	1.32	0.42	19.68	1.21	0.39	**9.55**	**1.11**	**0.35**	21.43
10	2.76	1.15	79.36	**0.91**	**0.28**	13.8	1.24	0.36	16.91	0.92	**0.28**	9.0
11	1.63	1.12	25.89	**1.06**	0.53	**8.71**	1.61	0.68	19.12	1.50	**0.61**	19.98
Avg	2.43	1.11	–	1.11	0.43	–	**1.08**	0.44	–	**1.08**	**0.40**	–

Look at Tab. 1, we can realize that the proposed method achieved lower errors for rotations r_e (deg/100m) in all sequences that was written in bold type. The average of rotation estimation of our approach is 0.40, MRPE is 0.43, NOVEL TRANS is 0.44, and it is the smallest value in these methods. This result indicates that the proposed method enhances approximately 10% the accuracy of rotation estimation by combining the orientation parameters of the left and right sides via quaternion presentation since comparing to the estimation in the left side only. Furthermore, MRPE and our approach are equal in the average of translation estimation and it is also lowest value translation estimation in four methods. This is understandable because the value of rotation after extraction is input of equation computing translation estimation. The value of rotation gets lower error lead to the value of translation also is improved.

Fig. 3 is an intuitive comparison between the performance of approaches. It is clear to see that camera tracks of proposed method in blue closer to the ground-truth than others like MRPE in black, Novel Translation in pink, VISO2 in green. This figure verifies the accuracy of our method compared to others shown in Tab. 1.

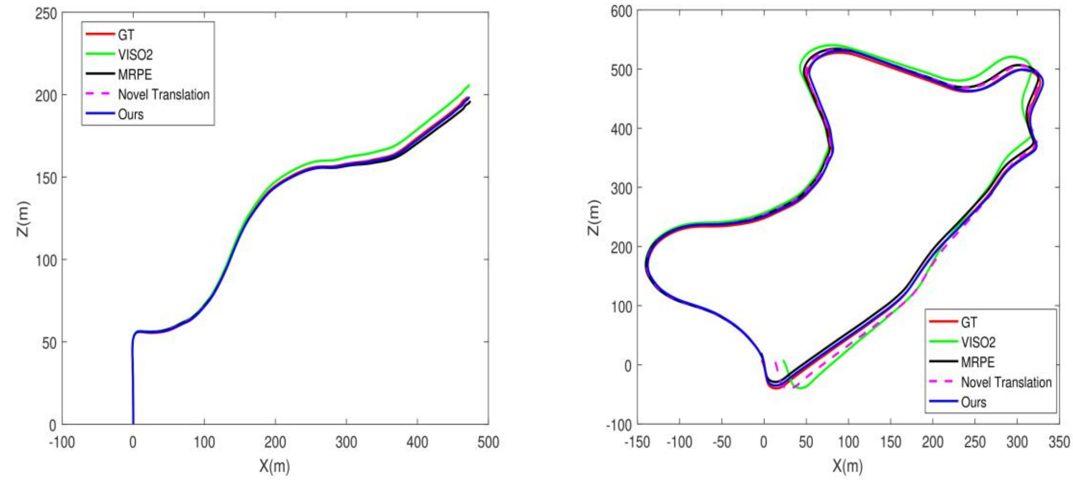

Fig. 3. Trajectory of sequence 3rd and 9th for four approaches compare to the ground-truth.

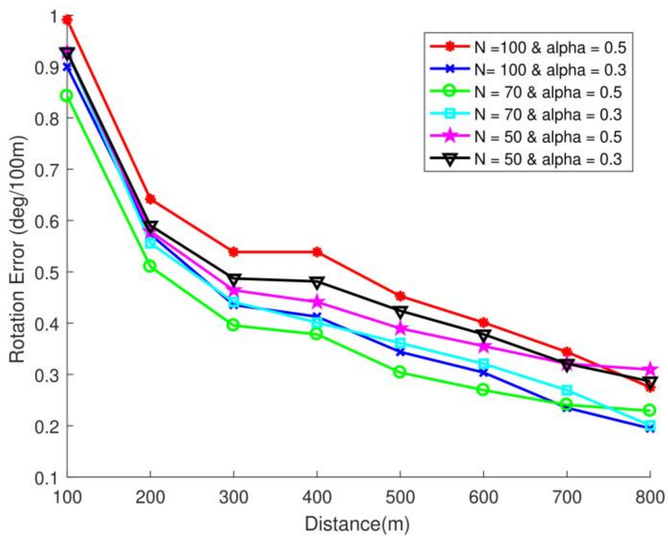

Fig. 4. Results of different set of five-point numbers and interpolation coefficient.

As mentioned above, Slerp function stores α variable. The interpolation coefficient α between quaternion rotations q_1, q_2 is one of factors affecting to the accuracy of rotation estimation. Furthermore, during the experiment, we also realize that the accuracy of rotation estimation depends on the number set of five-points, N. When N changes, the accuracy will be modified, too. These issues are illustrated in Fig. 4.

Fig. 4 shows that the accuracy of rotation estimation via quaternions q_1, q_2 is improved and be stable when we select least N = 70 and interpolation coefficient $\alpha = 0.3$. It is also an other evidence prove that the accuracy of rotation averaging is enhanced since combining the estimation from both the left and right sides. However, besides the advantages that were mentioned above, there is a drawback of our approach that is computing volume will increase at least two times when calculating rotation estimation using information a pair of frames on both the left and right stereo camera. Obviously, the computing time will rise, also in this situation. Therefore, depending on the computing capacity of hardware resources to consider to use or not. In this paper, we only consider how to improve the accuracy of rotation estimation and ignore the computing volume and time.

B. Testing dataset

In this sub-section, we will evaluate the performance of our method on the KITTI testing dataset that is publicly assessed on the web page. The result was compared to VISO2 [12] and Novel Trans [6] showed in Tab. 2. Look at Tab. 2, we can see that the rotation and translation error of our approach are 0.40 (deg/100m) and 1.24 (%), respectively and these are smallest value when comparing to other methods.

Tab. 2. Online performance evaluation on KITTI Dataset

	Rotation Error (deg/100m)	**Translation Error (%)**
VISO2 [12]	1.14	2.44
NOVEL TRANS [6]	0.48	1.42
Ours	**0.40**	**1.24**

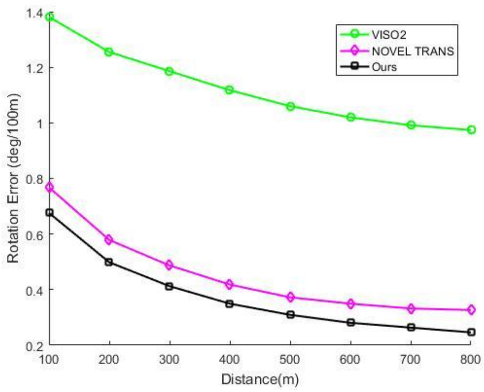

Fig. 5. Average rotation error along travel distance.

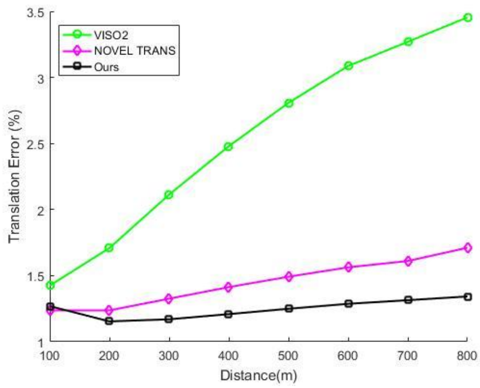

Fig. 6. Average translation error along travel distance.

To get an intuitive view point, we will visualize the translation and rotation errror along with travel distance in Fig. 5, Fig. 6. Based on graphs were plot on these figures, we can see that our results get better than VISO2 and NOVEL TRANS methods when run on the same dataset. The error of rotation and translation almost get lower comparing to other methods, it proves that our approach is correct because it enhances the accuracy of rotation and a little translation estimation.

5. Conclusions

The paper proposed an improvement the accuracy of rotation estimation for essential matrix based SVO. This approach fused the rotation estimation from both left and right sides via quaternion presentation. Evaluated the algorithm on KITTI dataset, the results prove that the proposed approach reinforces performance enhancement compared to other methods in the same scenarios. In future, we hope to use quaternion to express rotation and orientation instead of rotation matrix and translation vector due to the advantages of quaternion in computing.

References

[1] F. Friedrich and D. Scaramuzza, "Visual odometry: Part I: The First 30 Years and Fundamentals," *IEEE Robotics and Automation Magazine*, pp. 80-92, December, 2011.

[2] F. Friedrich and D. Scaramuzza, "Visual odometry: Part II: Matching, robustness, optimization, and applications," *IEEE Robotics and Automation Magazine*, 19.2, pp. 78-90, 2012.

[3] D. Nistér, N. Oleg and B. James, "Visual odometry," *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004* (CVPR 2004), Vol. 1, IEEE, pp. I-I, 2004.

[4] S. Poddar, R. Kottath, and V. Karar, "Motion Estimation Made Easy: Evolution and Trends in Visual Odometry," in *Recent Advances in Computer Vision*, Springer, Cham, pp. 305-331, 2019.

[5] D. Scaramuzza and F. Friedrich, "Visual odometry [tutorial]," *IEEE robotics and automation magazine*, 18.4, pp. 80-92, 2011.

[6] Huu-Hung Nguyen, The-Tien Nguyen, Cong-Manh Tran, Kim-Phuong Phung, Quang-Thi Nguyen, "A novel translation estimation for essential matrix based stereo visual odometry," *ICOM 2021*, IEEE, 2021, doi: 10.1109/IMCOM51814.2021.9377372.

[7] D. Nistér, "An efficient solution to the five-point relative pose problem," *IEEE transactions on pattern analysis and machine intelligence*, 26.6, pp. 756-770, 2004.

[8] On Quaternions; or on a new System of Imaginaries in Algebra, Letter to John T. Graves, 17 October 1843.

[9] Rozenfeld, Boris Abramovich, *The history of non-euclidean geometry: Evolution of the concept of a geometric space*, Springer, pp. 385, 1988, ISBN 9780387964584.

[10] Ken Shoemake, "Animating Rotation with Quaternion Curves," *ACM SIGGRAPH Computer Graphics*, Vol. 19, Iss. 3, pp. 345-354, 1985.

[11] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on.* IEEE, 2012.

[12] Geiger, J. Ziegler, and C. Stiller, "Stereoscan: Dense 3D reconstruction in real-time," *2011 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2011.

[13] H. H. Nguyen and S. Lee, "Orthogonality Index Based Optimal Feature Selection for Visual Odometry," in *IEEE Access*, Vol. 7, pp. 62284-62299, 2019, doi: 10.1109/ACCESS.2019.2916190.

MỘT PHƯƠNG PHÁP TÍNH TOÁN MA TRẬN QUAY KẾT HỢP MA TRẬN QUAY TRÁI VÀ PHẢI SỬ DỤNG STEREO CAMERA

Nguyễn Thế Tiến, Trần Công Mạnh, Nguyễn Quang Thi, Nguyễn Xuân Phục, Nguyễn Hữu Hùng

Tóm tắt: *Đối với các hệ thống Stereo Visual Odometry (SVO), chuyển động quay và chuyển động dịch của camera có thể được ước tính đồng thời hoặc riêng biệt, trong đó chuyển động quay được trích xuất từ ma trận thiết yếu. Cho đến nay, hầu hết các phương pháp tính toán các tham số của chuyển động quay camera chỉ sử dụng từng cặp khung hình ảnh liên tiếp ở phía bên trái. Điều này dẫn đến việc chưa tận dụng được thông tin giá trị của các khung ảnh liên tiếp ở phía bên phải của hệ thống camera. Bài báo trình bày một cách tiếp cận để tận dụng thông tin này trong quá trình tính toán chuyển động quay bằng cách tính giá trị trung bình của các phép quay được trích xuất từ cả hai phía bên trái và bên phải của hệ thống. Phương pháp đề xuất được đánh giá trên tập dữ liệu KITTI để xác định hiệu quả của thuật toán này. Kết quả thử nghiệm chỉ ra rằng phương pháp được đề xuất nâng cao độ chính xác khoảng 10% so với các phương pháp khác trong cùng một kịch bản đánh giá.*

Từ khóa: Rô bốt; định vị; phương tiện tự hành; stereo camera.

Received: 25/11/2021; *Revised:* 08/02/2022; *Accepted for publication:* 03/03/2022

❑