IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 9, 2019, accepted April 30, 2019, date of publication May 10, 2019, date of current version May 24, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2916190

Orthogonality Index Based Optimal Feature Selection for Visual Odometry

HUU HUNG NGUYEN AND SUKHAN LEE, (Fellow, IEEE)
Intelligent Systems Research Institute (ISRI), Sungkynkwan University, Suwon 16419, South Korea

Corresponding author: Sukhan Lee (lsh1@skku.edu)

This work was supported, in part, by ''Robocarechair: A Smart Transformable Robot for Multi-Functional Assistive Personal Care'' Project, KEIT P0006886, of the Korea Evaluation Institute of Industrial Technology (KEIT) and, in part, by the ''Space Initiative Program'' of National Research Foundation (NRF) of Korea (NRF-2017M1A3A9032518).

ABSTRACT The performance of visual odometry is dependent upon the quality of features selected for computing the frame-to-frame transformation. In order to ensure the quality of selected features, conventional approaches consider the spatial distribution of the selected features, in addition to their counts and matching scores, in which a small number of features are selected randomly from each of the uniformly distributed buckets. In this paper, we show that features can be selected optimally, rather than randomly, using a well-defined mathematical formalism. The proposed method of optimal feature selection minimizes the degree of uncertainty in estimating the essential, fundamental, or homography matrix involved in visual odometry by maximizing the orthogonality index of individual equations and constraints associated with computation. We found that, at a constant noise level, the mean of the residual error and the variance of an estimated essential, fundamental, or homography matrix decrease monotonically with increasing orthogonality index. The simulation validates the increased accuracy of the feature selection based on the proposed orthogonality index compared with the conventional random selection. For instance, it enhances accuracy by as much as 35% when a small number of feature sets, say, 20 feature sets, are used. The experiments using the KITTI and Devon Island datasets further reinforce the performance enhancement of simulations by 9% and 20%, respectively.

INDEX TERMS Visual odometry, ego-motion estimation, feature selection, orthogonality index.

I. INTRODUCTION

Visual odometry (VO) [1], which involves estimation of the relative motion based on a sequence of captured images, plays a crucial role in autonomous navigation. The ego-motion is estimated between current and previous images by resolving geometric constraints. Subsequently, the full camera trajectory is recovered via accumulation of these single movements. Therefore, VO is known as a dead-reckoning technique [2], indicating that it is subject to cumulative errors. Based on vision-based odometry, recent surveys [3]–[4] classified VO into feature-based, appearance-based, and a hybrid of feature- and appearance-based approaches. The feature-based VO is a geometric approach of extraction as well as matching of image features such as corners, lines, and curves in a sequential frame of images, and estimation of the motion. Appearance-basedVO is another geometric approach that utilizes information extracted from the pixel intensities within a whole image. The camera ego-motion can be estimated by changes in optical flow, which leads to robust pose estimation despite low-textured environments. Scaramuzza and Fraundorfer conducted a comprehensive review of feature-based VO [5], [6]. Accordingly, three major approaches were used to estimate the relative ego-motion between two frames:

- *2D-to-2D: Relative pose is estimated directly from 2D features.*
- *3D-to-3D: Relative pose is estimated directly from 3D features.*
- *3D-to-2D: Relative pose is estimated by re-projecting 3D features in one camera frame to another.*

The 3D-to-3D and 3D-to-2D methodologies require triangulation of 3D points via intersection of back-projected rays from 2D image correspondence of at least two image frames. However, uncertainties in feature measurement and imperfect calibration prevent intersection. Therefore, these two

The associate editor coordinating the review of this manuscript and approving it for publication was Mehul S. Raval.

approaches we're considered less accurate than the 2D-to-2D method in the VO survey of Scaramuzza and Friedrich [5]. Nistér [7] proposed an efficient state-of-the-art solution for 2D-to-2D relative pose estimation in the presence of outliers for the minimal case involving the five-point-problem. It is used to estimate the geometric relationship between two calibrated camera frames, which is designated as the essential matrix E, despite the five co-planar points. E accommodates the camera pose parameters up to an unknown scale factor for translation. In order to retrieve the full trajectory, the absolute translational scale for every single movement is calculated, and is possibly determined by the corresponding 3D points [1]. As it is impossible to observe 3D points with triangulation based on the two mono-camera views, at least three frames should be used. However, using multiple views does not completely solve the scale problem. Therefore, most studies recommend using stereo-cameras with well-defined baselines.

The performance of pose estimation not only depends upon the methods for computing ego-motion but also on feature selection. Current approaches for feature selection are summarized in [3], [4]. Three critical issues have been considered: feature detector, outlier removal, and feature distribution.

In the literature, a variety of feature extractions and matching methodologies have been used in VO, including corner detector [8]–[11], SIFT [12], SURF [13], and ORB [14]. Several studies reported the methodological robustness in image variation and performance of VO involving feature detector/descriptor pairs [15], [16]. In terms of VO performance, SURF-based VO yields the maximal accuracy while ORB-based VO is computationally inexpensive but results in decreased accuracy. However, fast detector/descriptor pairs are preferred for autonomous driving applications, which require high speed. Results of such features including corners [8]–[11], and ORB [14] in KITTI datasets still contribute to increased accuracy. Since the corresponding features of the two frames are detected via matching, the presence of outliers is still attributed to image noise, similar patterns, occlusion, and blurring in viewpoints. Because wrong matches lead to erroneous estimation, an outlier removal is a mandatory step. Random sample consensus (RANSAC) [17] is a well-known approach to eliminate outliers exploiting geometric constraints. Specifically, RANSAC finds the consensus set from random samples and re-estimate the model based on the consensus set. Kitt *et al.* [9] employed an Iterated Kalman Filter directly incorporating RANSAC-based outlier removal based on the trifocal geometry in image triples. For a stereo camera, where each frame comprises two left-right images, a robust loop chain matching scheme between two pairs of stereo images was proposed in VISO2-S [10], in order to extensively improve the matching performance. Fanfani *et al.* [11] further improved the loop chain quality using a more robust detector/descriptor with high temporal and spatial disparities.

Feature distribution has been considered in addition to outlier removal. Bucketing technique [10] is frequently used to reduce the computational complexity and to improve accuracy. Specifically, the image is divided into $M \times M$ grids, and only a limited number of features are selected for further processing. This process not only reduces the total number of features, but also distributes the selected features uniformly, resulting in improved efficiency and accuracy. In addition to using bucket technique, Cvišić and Petrović [8] proposed that careful selection and tracking for the classification of features into different categories and selection of longer age for pose estimation. The features tracked longer were considered more reliable, with a lower probability of being outliers. As a result, the older features should always be selected for subsequent steps. Buczko and Volker [18] showed that inclusion of the far and near features also enhanced the accuracy of estimating transformation. Furthermore, Badino *et al.* [19] refined the position of 2D features based on the overall history frames, and used the integrated features to improve ego-motion accuracy. The aforementioned approaches of feature selection focused on strong and stable features as well as on feature distribution to improve accuracy and efficiency. However, these approaches are inherently based on heuristics while the selection of minimum features needed for fitting model is rather random.

Instead of adopting traditional approaches for random selection of five-point feature sets, we present an optimal feature selection with a mathematic formalism based on orthogonality index to develop a stereo-visual odometry framework. The five-point algorithm [7] is used to obtain frame-to-frame relative orientation as well as translation vector, whereas the translational scale is based on resolving the linear closed-form equation, followed by inexpensive re-projection minimization. The stereo-visual odometry framework is summarized in Fig. 1.

The main contributions of this paper are summarized as follows:

- ➢ A mathematical formalism of optimal feature selection based on the orthogonality index is proposed.
- ➢ The proposed optimal feature selection results in a smaller mean residual error in VO estimation than random selection. The larger the noise level in feature measurement, the more the proposed method gains in error reduction over random selection with a minimal overhead in computation.

According to the publicly available KITTI leaderboard, the proposed method ranks among the top stereo VO filter-free methods such as bundle adjustment or Kalman Filter, with an average translational error of 1.13% and a rotational error of 0.003 deg/m. The experimental results of Devon Island show an average translational error of 0.9%.

The remainder of the paper is organized as follows. After an introduction, we discuss the problem definition and proposed an approach in Section II. Section III describes the key contribution of this paper, which is related to feature selection

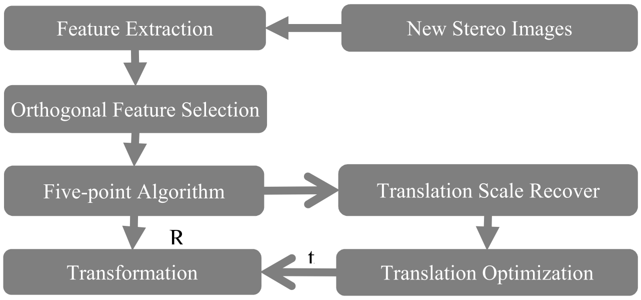


FIGURE 1. Visual odometry framework with feature selection based orthogonality index.

based on the proposed orthogonality index. We verified the proposed approaches via simulations in Section IV. Finally, the experimental results with KITTI and Devon Island are highlighted in Section V.

II. PROBLEM DEFINITION AND PROPOSED APPROACH

A. PROBLEM DEFINITION

Estimating ego-motion from two camera views presents a classical problem in computer vision, especially for autonomous navigation. Feature selection plays a critical role in the performance of VO. Conventional VO approaches have focused on qualitative subset features and their distribution from the full feature set to improve the accuracy of pose estimation as well as the computational time. Towards this end, the bucketing technique is preferred to ensure uniform distribution of strong and stable features. However, the random selection of samples, minimum features required to estimate the pose, presents an ad-hoc approach without a mathematical formalism for optimality, and therefore, does not always guarantee the minimum error of pose estimation. The accuracy of the estimated relative poses depends on the uncertainty of the selected points and their distribution. In order to obtain reliable results without resorting to a formal method, conventional approaches often rely on a large of number of randomly selected samples. A large number of samples improve accuracy at the cost of computational complexity. The goal of VO is to increase the precision while reducing the computational time. An optimal feature selection that is based on a mathematical formalism may require a smaller number of samples for the required accuracy.

B. PROPOSED APPROACH

The proposed approach for optimal feature selection is based on the selection of a set of features that minimizes the error in solving the equations associated with VO in the presence of feature noise. Specifically, the selected features are designed to determine the parameters associated with a set of equations,

$$E_i = f_{E_i}\left(p_i + \Delta\hat{p}_i, q_i\right), \quad i : 1 \sim m, \tag{1}$$

and constraints,

$$C_j = f_{C_i}\left(q_j\right), \quad j : 1 \sim l, \tag{2}$$

where p and q represent, respectively, the parameters determined by the selected features and the variables to be solved for VO. The solution represents the intersection of such sets of equations and constraints. Due to feature noise, the true parameters of a VO equation are probabilistically distributed or lie within a specific range around the nominal values, as represented by $\Delta\hat{p}_i$, such that the true solution resides in a hyper-volume, S, around the intersection of the VO equations and constraints with nominal parameters:

$$S = \bigcap_{\mathrm{i:1\sim m}}^{\mathrm{j:1\sim l}} (E_i, C_j) \tag{3}$$

Because the hyper-volume also represents the volume of error, the optimal feature selection proposed is designed for feature selection that minimizes this hyper-volume, S. The minimum hyper-volume can be achieved by selecting a set of features to ensure orthogonality of individual equations and constraints associated with VO. To this end, we defined the orthogonality index associated with a set of features to measure the degree of orthogonality of the individual VO equations and constraints with respect to each other for the selected set of features. In the follow section, we present the details of optimal feature selection based on the proposed orthogonality index for various cases of VO.

III. OPTIMAL FEATURE SELECTION BASED ON ORTHOGONALITY INDEX

A. ESSENTIAL MATRIX-BASED VISUAL ODOMETRY

The essential matrix-based VO is used to estimate the 3 × 3 essential matrix, E, from the epipolar constraints, $p^T Eq = 0$, to compute the translation and rotation between a pair of camera frames, where p and q represent the image coordinates of a feature in the respective camera pair. The epipolar constraint, $p^T Eq = 0$, can be rewritten as follows:

$$\hat{v}\,\hat{E} = 0, \tag{4}$$

where

$$\hat{v} = [p_1\mathrm{q}_1, p_2q_1, p_3q_1, p_1q_2, p_2q_2, p_3q_2, p_1q_3, p_2q_3, p_3q_3] \tag{5}$$

$$and\ \hat{E} = [E_{11}, E_{12}, E_{13}, E_{21}, E_{22}, E_{23}, E_{31}, E_{32}, E_{33}]^T \tag{6}$$

Note that $\hat{v}$ is determined by selected features while $\hat{E}$ is estimated using equation (4) together with the following additional constraints in E:

$$\det(E) = 0 \tag{7}$$

$$2EE^TE - tr\left(EE^T\right)E = 0 \tag{8}$$

It is known that E is determined by (4), (7) and (8) with five pairs of corresponding feature points [7] that define the following 5 × 9 matrix equation:

$$A\hat{E} = 0, \tag{9}$$

$$where\ \ A = \left[v_1v_2v_3v_4v_5\right]^T \tag{10}$$

Note that the increased error volume to be minimized is defined by the intersection of the two individual manifolds:

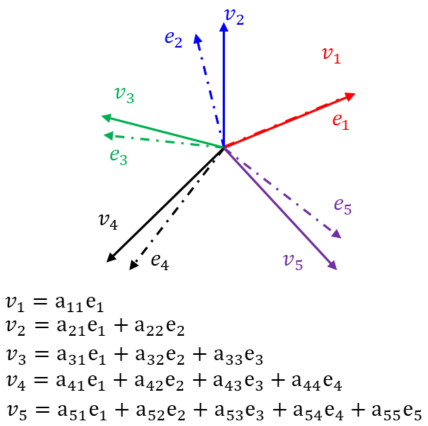

v_2
e_2
v_1
v_3
e_1
e_3
e_5
v_4
e_4
v_5
$v_1 = a_{11}e_1$
$v_2 = a_{21}e_1 + a_{22}e_2$
$v_3 = a_{31}e_1 + a_{32}e_2 + a_{33}e_3$
$v_4 = a_{41}e_1 + a_{42}e_2 + a_{43}e_3 + a_{44}e_4$
$v_5 = a_{51}e_1 + a_{52}e_2 + a_{53}e_3 + a_{54}e_4 + a_{55}e_5$

FIGURE 2. **An orthogonal basis of constraint vectors.**

one derived from (9) and the other from (8) and (7). Since the manifold obtained from (7) and (8) is fixed, the minimization of the intersection of two manifolds depends upon the manifold from (9) defined by the five pairs of corresponding features. To minimize the intersection, it is necessary to minimize the extent of uncertainty associated with the manifold from (9) while ensuring the correct intersection with the manifold from (7) and (8), where the extent of uncertainty in (9) is attributed to errors involved in selected features. First, to minimize the extent of uncertainty associated with the manifold derived from (9), we created five orthogonal vectors of (9),$v_1v_2v_3v_4v_5$, by selecting five pairs of corresponding features appropriately. Second, to ensure the accuracy of intersection with the manifold from (7) and (8), we searched for the minimum error solution among the multiple options of five orthogonal vectors obtained from different combinations of five feature pairs.

B. OPTIMAL SELECTION OF MULTIPLE SETS OF FIVE FEATURE PAIRS THAT MAKE $v_1v_2v_3v_4v_5$ AS ORTHOGONAL AS POSSIBLE

The orthogonality of the five vectors, $v_1v_2v_3v_4v_5$, can be measured based on their fit with the bases of an orthogonal frame, as shown in Fig 2. To reduce the ambiguity caused by vector length, the five vectors should be normalized. For mathematical representation, we defined the following orthogonality index:

$$Score = \sum\nolimits_{i=1}^{5} dot(e_i, v_i), \tag{11}$$

where an orthogonal frame $e_1e_2e_3e_4e_5$ is a result of QR factorization, such as the Gram Schmidt orthogonalization [20], with the first vector, $e_1 = v_1$.

Based on the foregoing analysis, our goal was to select sets of five corresponding pairs, which provide the highest orthogonality indices for estimation of the essential matrix. The feature selection based on the proposed orthogonality index is described as follows:

```
Algorithm 1 Orthognality Index Based Optimal Feature Selection
  Input: Correspondence pairs
  Output: $L$ five-point sets with the highest indices
  Step 1: Random selection of $K$ sets of five points.
  Step 2: For each set, an orthogonal frame was built using Gram-Schmidt orthogonalization followed by the calculation of the orthogonality index.
  Step 3: The $K$ sets were reordered using their orthogonality indices ranging from the maximum to the minimum value.
  Step 4: The $L$ sets were selected with the highest indices.
```

This orthogonal selection algorithm can be used for optimal selection of features in other methods of VO such as fundamental and homography matrix. The determination of the orthogonal score for all possible combinations is impossible due to the computational cost entailed. For a trade-off between efficiency and accuracy, we only selected a defined number of five-point $\boldsymbol{K}$ sets. The $\boldsymbol{L}$ sets with highest orthogonality indices are used to estimate the essential matrix. Optimal values of $\boldsymbol{K}$ and $\boldsymbol{L}$ are explained comprehensively using the results of simulations and real experiments.

C. FUNDAMENTAL AND HOMOGRAPHY MATRIX-BASED VISUAL ODOMETRY

Optimal feature selection based on the proposed orthogonality index is generally applicable to other methods of VO, including those based on fundamental and homography matrices.

1) FUNDAMENTAL MATRIX

Similar to the essential matrix, the 3 × 3 fundamental matrix, F, a pair of corresponding feature points (p, q) are related based on epipolar constraint:

$$p^T Fq = 0 \tag{12}$$

The constraint (12) can be rewritten in vector form as follows:

$$\hat{v}\,\hat{F} = 0, \tag{13}$$

where

$$\hat{v} = [p_1q_1, p_2q_1, p_3q_1, p_1q_2, p_2q_2, p_3q_2, p_1q_3, p_2q_3, p_3q_3] \tag{14}$$

$$and\ \ \hat{F} = [f_{11}, f_{12}, f_{13}, f_{21}, f_{22}, f_{23}, f_{31}, f_{32}, f_{33}]^T \tag{15}$$

It is known that the fundamental matrix can be estimated based on eight pairs of the corresponding feature points [21]:

$$\mathrm{A}\hat{F} = 0, \tag{16}$$

$$\text{where } A = [v_1v_2v_3v_4v_5v_6v_7v_8]^T \tag{17}$$

Note that the solution to the homogenous system (16) is the eigenvector corresponding to the smallest eigen-value of matrix A. It is adequate to minimize the extent of uncertainty associated with the solution of (16) based on the maximization of the orthogonality index associated with the eight vectors of (17). As indicated before, the orthogonality index of eight vectors can be measured using the following equation:

$$Score = \sum_{i=1}^{8} dot(e_i, v_i), \tag{18}$$

where $[e_1 e_2 e_3 e_4 e_5 e_6 e_7 e_8]$ represents an orthogonal frame derived from Gram-Schmidt orthogonalization. The optimal feature selection for the fundamental matrix-based VO can be performed using the same ***Algorithm I*** except for random selection of K sets involving eight-point features in ***Step 1*** and calculation of orthogonality index using equation (18).

2) HOMOGRAPHY MATRIX

In case feature pairs selected are placed on the corresponding planar surfaces, a pair of corresponding feature points, (p, q) satisfies the following 3 × 3 homography matrix equation:

$$q = Hp \tag{19}$$

(19) can be expressed as

$$\begin{bmatrix} p_1 & p_2 & 1 & 0 & 0 & 0 & -p_1 q_1 & -p_2 q_1 & -q_1 \\ 0 & 0 & 0 & p_1 & p_2 & 1 & -p_1 q_2 & -p_2 q_2 & -q_2 \end{bmatrix} \hat{H} = 0 \tag{20}$$

with $\hat{H}$ representing a nine-dimensional vector. It is known that four corresponding (p, q) pairs are sufficient to solve (20), resulting in

$$A\hat{H} = 0 \tag{21}$$

With $A = [v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8]^T$ representing an 8 × 9 matrix, where the solution corresponds to the smallest eigenvector of A [22]. Optimal feature selection to minimize the extent of uncertainty is based on maximizing the orthogonality index associated with $v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8$ vectors.

D. STRUCTURE FROM MOTION-BASED VISUAL ODOMETRY

The proposed optimal feature selection based on orthogonality index is applicable to the structure from motion-based VO. Note that SfM minimizes the re-projection error, RE:

$$\mathrm{RE} = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(x_{ij} - P_i\left(X_j\right)\right)^2, \tag{22}$$

where X_j is the jth 3D feature point, P_i represents projection matrix of the ith camera, and x_{ij} denotes the 2D image point of X_j in the ith camera. We conjecture that the optimal features selected based on the orthogonality index used in the fundamental matrix-based VO lead to minimize re-projection error in SfM, since the minimum extent of uncertainty associated with the fundamental matrix equation represents the minimum VO error. We validate this conjecture by simulation in the following section.

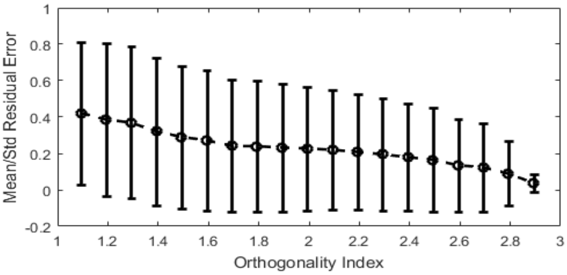


FIGURE 3. Mean/Std residual error with respect to orthogonality index.

IV. VALIDATION BY SIMULATION

This section validates the role of orthogonality index-based optimal feature selection proposed in the previous Section by comparing with the conventional random selection. The validation was performed for different methods of VO such as essential, fundamental, and homography matrices as well as SfM using the same configuration. With the known intrinsic parameters of the camera, we generated a transformation $\boldsymbol{C_1}^{\boldsymbol{C_0}}\boldsymbol{T}$ based on the camera frame,$\mathbf{C_0}$, to another, $\boldsymbol{C_1}$, as well as N random 3D feature points in $\boldsymbol{C_0}$. These 3D points were projected onto 2D camera image planes. All projections considered a noise level of 0.5 pixels.

A. ESSENTIAL MATRIX-BASED VISUAL ODOMETRY

1) ORTHOGONALITY INDEX VS. RESIDUAL ERROR

In order to demonstrate the benefit of the proposed orthogonality index, we conducted a statistical analysis to show the dependency of the residual error, number of inliers, and epipolar scores on this index. Multiple solutions are known for each five-point set. Since the essential matrix is defined only up to scale factor, the residual error was expressed as follows:

$$\underbrace{\dot{\min}}_{i}\ \dot{\min}\left(\left\|\frac{E_i}{\|E_i\|}-\frac{E_{GT}}{\|E_{GT}\|}\right\|,\left\|\frac{E_i}{\|E_i\|}+\frac{E_{GT}}{\|E_{GT}\|}\right\|\right) \tag{23}$$

The simulation can be described specifically as follows:

- ***Step 1.*** Random generation of M five-point sets;
- ***Step 2.*** Estimation of the essential matrices and calculation of the number of inliers, epipolar scores, residual errors, and orthogonal scores, and assignment of the set to the corresponding bin based on their orthogonal scores.
- ***Step 3.*** Measurement of the mean/standard of residual error, the number of inliers, and the epipolar score for each orthogonal score bin.

The above measurements related to the orthogonality index are shown below in Fig. 3, Fig. 4, and Fig. 5. The horizontal axis represents the increased orthogonality index. At each orthogonal score bin, the circle and black bar represent the means and standard deviations of measurements, respectively.

In Fig. 3, the mean of residual error was diminished almost linearly. Fig. 4 displays the increase in the inlier number.

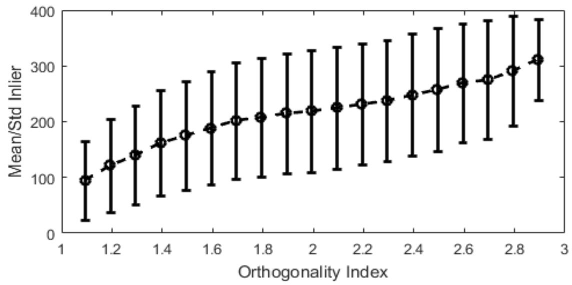


FIGURE 4. Mean/Std of inlier number with respect to orthogonality index.

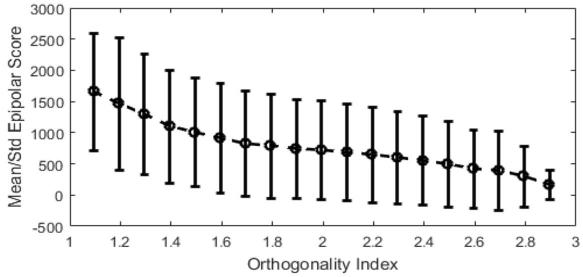


FIGURE 5. Mean/Std epipolar score with respect to orthogonality index.

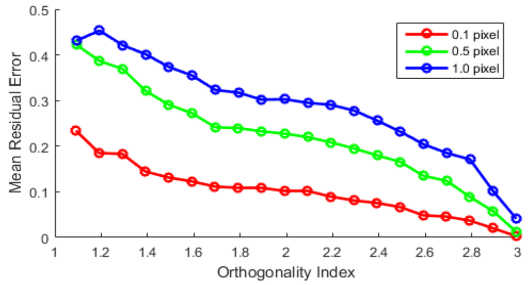


FIGURE 6. Mean residual error of the estimated essential matrix with respect to orthogonality index of the three fixed noise levels.

Fig. 5 represents a steady downward trend in the epipolar score. Addition of the standard deviation yielded similar trends. The standard deviation (std) of residual error and epipolar score was also reduced, which suggests that the five-point feature sets with higher orthogonal scores possibly provide higher reliability of estimation in the essential matrix.

We also measured the correlation between the mean/std residual error and orthogonality index of the different noise levels at 0.1, 0.5, and 1.0 pixels. The five-point feature set with the higher orthogonal score was minimally dependent on the noise levels. In particular, as shown in Fig. 6 and Fig. 7, both mean/std residual error decreased gradually with respect to the orthogonality index. Reduced noise level yielded lower mean/std residual error.

2) PERFORMANCE COMPARISON OF RANDOM- AND ORTHOGONAL-SELECTION

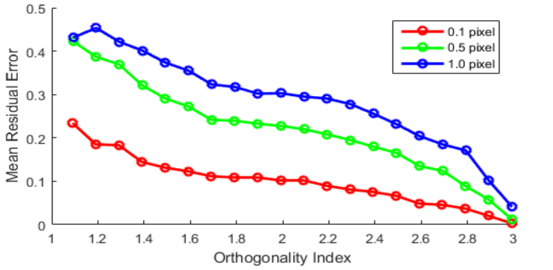


FIGURE 7. Mean residual error of the estimated essential matrix with respect to orthogonality index of the three fixed noise levels.

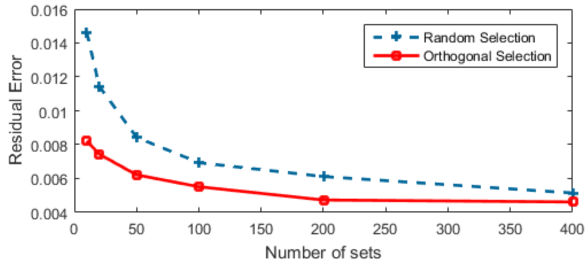


FIGURE 8. Mean residual error of orthogonal and random selections depending on the number of five-point feature sets.

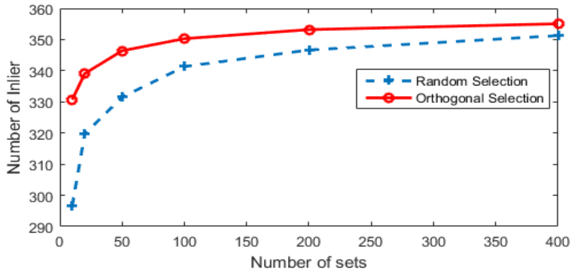


FIGURE 9. Mean number of inliers in orthogonal and random-selections depending on the number of five-point feature sets.

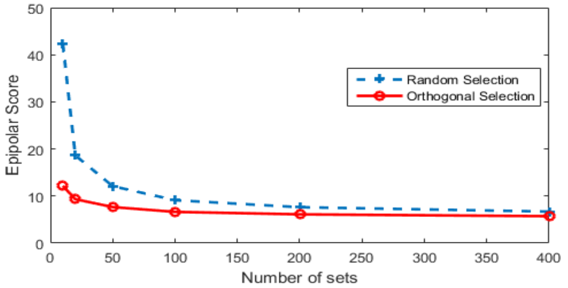
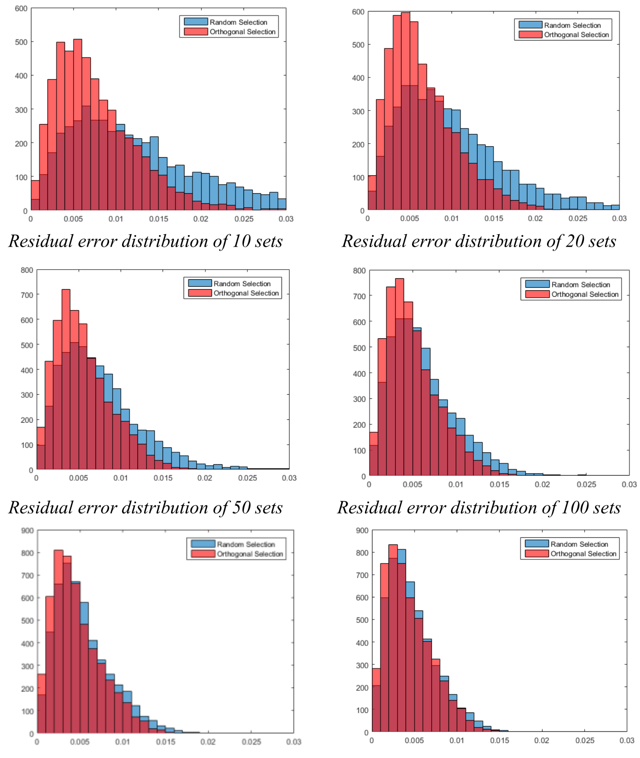


FIGURE 10. Mean epipolar score depending on the number of five-point feature sets in two approaches.

We compared the orthogonal- with the random-selection approaches for the selection of a varying number of five-point sets L, such as 10, 20, 50, 100, 200, and 400, to obtain the essential matrix. In order to eliminate bias, the selected L sets were repeated 5000 times. The average residual error, as well as the average inlier/epipolar scores were computed and shown in Fig. 8, Fig. 9, and Fig. 10. These figures indicate that, regardless of the use of random-selection or orthogonal-selection, the higher number of sets reduce the residual error and the epipolar score, as well as increase the number of inliers, on average.



Residual error distribution of 10 sets *Residual error distribution of 20 sets*

Residual error distribution of 50 sets *Residual error distribution of 100 sets*

Residual error distribution of 200 sets *Residual error distribution of 400 sets*

FIGURE 11. Residual error distribution of estimated essential matrix with respect to the number of five-point feature sets in two approaches.

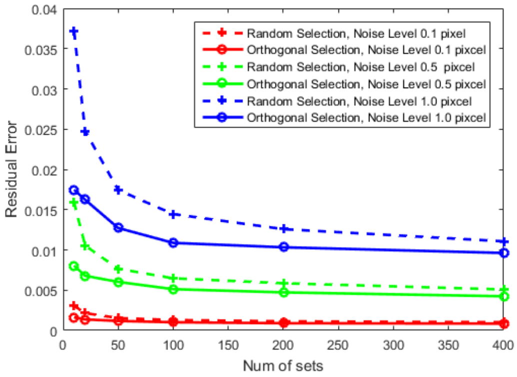


FIGURE 12. Mean residual errors of the orthogonal and random selections with respect to three noise levels.

TABLE 1. Computation time of orthogonal and randomselection with respect to varying number of sets.

Number of sets	10	20	50	100	200	400
Random-Selection (ms)	14	18	31	54	95	180
Orthogonal-Selection(ms)	23	27	40	64	104	190

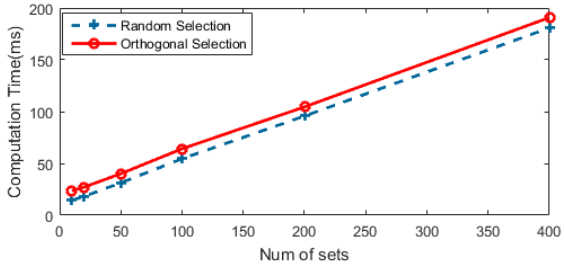


FIGURE 13. Computational time of the orthogonal– and random selection with respect to the different number of sets.

Specifically, Fig. 8 reveals a downward trend in residual errors between the orthogonal and random selections. The residual error in orthogonal selection was obviously lower than in the latter. The inlier number indicated in Fig. 9 shows a rising trend. The proposed method yields a higher number than the conventional one with all numbers of L sets. Conversely, the epipolar score, as shown in Fig.10, reveals a decreasing tendency. The epipolar score in our orthogonal selection method was always smaller than in random selection.

In general, the accuracy of orthogonal-selection is better than that of random-selection for the estimation of essential matrix. A small number of five-point sets (for instance, 20) enhances the accuracy to around 35%. We also found that orthogonal selection using 10 to 20 sets yielded similar inlier numbers and epipolar scores, as well as residual errors associated with random selection of 50 to 100 sets, respectively. The finding suggests that orthogonal selection guaranteed the success of essential matrix estimation using a small number of sets.

Additionally, Fig. 11 shows the distribution of residual errors when the selected L sets are repeated 5000 times. The horizontal axis represents the residual error and the vertical axis the count numbers. The residual error distributions of the proposed method and random-selection are shown in red and blue, respectively. The bin closer to zero represents smaller residual error, or highly accurate essential matrix. By increasing the number of sets, the distribution shifts gradually towards zero. With the same number of sets, the orthogonal selection bin closer to zero is higher than that of the random selection, while the orthogonal selection bin farther to zero is lower than that of the random one. Thus, the estimation of the essential matrix in orthogonal-selection is more reliable than that of random-selection.

Fig. 12 shows the comparison of the mean residual errors between orthogonal- and random- selections under three different noise levels in feature measurement. It indicates that the larger noise levels, the larger was the mean residual error.

Table 1 and Fig. 13 show the comparison of computation time between orthogonal- and random- selections for the estimation of an essential matrix. The orthogonal selection costs an additional 9 ms in computation time. However, combining the residual error of Fig. 12 and the computation time of Fig. 13, for the same residual error, prolongs the random-selection for the computation of an essential matrix compared with the orthogonal-selection. For instance, for the noise level of 0.5 pixels, the random-selection takes 54 ms, compared with 27 ms for the orthogonal-selection.

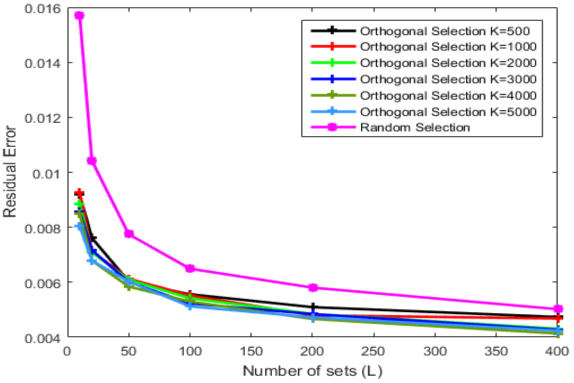


FIGURE 14. Residual error with different *K* and *L*.

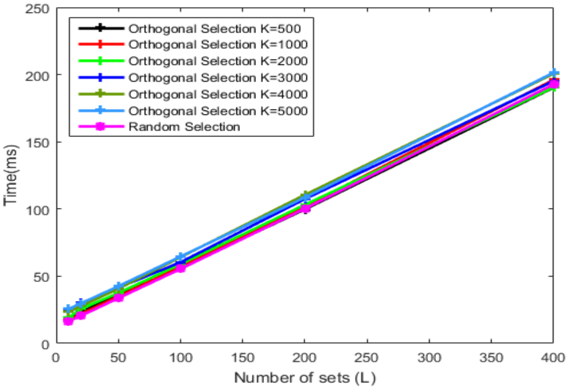


FIGURE 15. Computation time under different *K* and *L*.

3) ACCURACY AND EFFICIENCY TRADE-OFF

In ***Algorithm I***, K is the number of five-point sets that are generated randomly while L represents the number of selected sets providing the highest orthogonality indices. The increase in K or L in the orthogonal-selection improves accuracy at the cost of computational complexity.

To determine the designed K and L for trade-off between accuracy and efficiency, we tested orthogonal-selection with different Ks and Ls. We measured the residual error and computational time as shown in Figs. 14 and 15 with K ranging from 500 to 5000 and L from 10 to 400. In general, the higher K yields the lower residual error at higher computation cost. The additional time needed for feature selection increases linearly compared with the value of K and a lower improvement in accuracy. The residual errors at K over 1000 are quite similar and reduced with reference to the number of L selected sets. However, the reduction varies quickly with L from 10 to 50 and slowly from 50 to 400. The time increase along L is larger than along$K$. The residual error of orthogonal-selection at L was 50, which was similar to that of random-selection at L equal to 200. Computation time at L value of 50 was three time faster than at an L of 200. Therefore, L values ranging from 50 to 100 and K values around 1000 span a range of optimal values in terms of accuracy and efficiency.

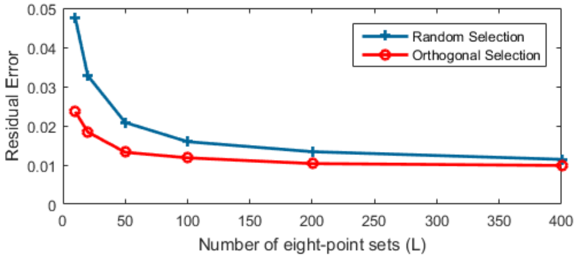


FIGURE 16. The residual error of orthogonal and random-selection with different set numbers.

B. FUNDAMENTAL AND HOMOGHRAPY MATRIX-BASED VISUAL ODOMETRY

1) FUNDAMENTAL MATRIX

To evaluate the fundamental matrix, the residual error described in (23) was slightly modified as follows:

$$min\left(\left\|\frac{F}{\|F\|}-\frac{F_{GT}}{\|F_{GT}\|}\right\|,\left\|\frac{F}{\|F\|}+\frac{F_{GT}}{\|F_{GT}\|}\right\|\right) \qquad (24)$$

The residual errors of orthogonal and random selections were measured under varying numbers of eight-point sets, with L values such as 10, 20, 50, 100, 200, and 400. The blue curve represents the residual error of the random selection while the red curve represents the orthogonal selection errors. At each number in the eight-point set, the residual error of the orthogonal selection was lower than in random selection, suggesting that the orthogonality index-based feature selection improved the VO accuracy.

2) HOMOGRAPHY MATRIX

To evaluate the homography matrix, the residual error associated with the evaluation of the fundamental matrix (24) used was as follows:

$$min\left(\left\|\frac{H}{\|H\|}-\frac{H_{GT}}{\|H_{GT}\|}\right\|,\left\|\frac{H}{\|H\|}+\frac{H_{GT}}{\|H_{GT}\|}\right\|\right) \qquad (25)$$

To guaranty the four selected features in a co-planar set, the 3D points were generated on a plane. We compared the orthogonal- and random- selections using different numbers of the selected set L : 10, 20, 50, 100, 200, and 400. Fig. 17 shows the residual errors of two approaches under different sets selected. Notably, the residual error of the orthogonal-selection was lower than in random-selection with all set numbers.

C. STRUCTURE FROM MOTION-BASED VISUAL ODOMETRY

Compared with other approaches, BA-based VO optimizes iteratively the camera poses by minimizing the re-projection error for a large number of points. Based on the results derived from ***Algorithm I*** for selection of feature sets to estimate the fundamental matrix, a sub-set including N pairs of features for SfM was gathered by gradually adding sets of eight points providing the highest orthogonality index as follows:

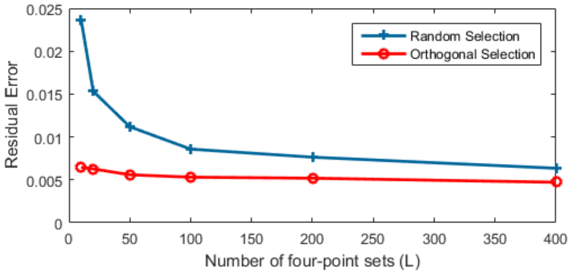


FIGURE 17. Residual error of homography matrix for orthogonal- and random-selection with different number of sets.

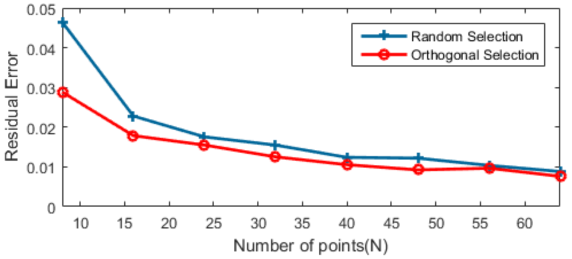


FIGURE 18. The residual error in the orthogonal–and random–selection with respect to the varying number of sets using SfM.

- ***Step 0***: The constraint vectors for all correspondences were generated by equation (14) and the non-selected status was marked for all the constraint vectors generated.
- ***Step 1***: The K sets of eight random constraint vectors were generated from non-selected-status vectors.
- ***Step 2***: Orthogonality indices associated with K sets were calculated, and the sets were sorted based on their orthogonal scores from maximum to minimum values.
- ***Step 3***: The set of eight vectors associated with the highest orthogonality index was added to the selected list and the selected status was marked.
- ***Step 4***: The steps 1 to 3 were repeated until the number of correspondences in the selected list was equal N.

Since the rotations and translations are refined by SfM, we calculated the essential matrix and used the residual error in (23) to compare the orthogonal selection with the random selection. We measured the residual errors of two approaches with N ranging from 8 to 64. As shown in Fig. 18, the residual error in either random or orthogonal selection decreased with the increase in the number of points. However, the residual error of orthogonal-selection was less than in random-selection with the same number of points. The discrepancy between two approaches is reduced along with the number of points.

D. SECTION SUMMARY

Based on the above statistical analysis, we draw the following conclusions:

1. The accuracy of essential matrix estimation depends upon the noise level in feature measurements as well as the number and orthogonality indices of the selected five-point feature sets.
2. Under a fixed noise level, the mean of the residual error and its variance of an estimated essential matrix decrease monotonically with the increase in the mean orthogonality index, as shown in Figs. 6 and 7. As shown in Fig. 11, the distribution of the residual error varies in 5000 tests involving different numbers of five-point feature sets.
3. The mean of the residual error decreases with the increase in the number of five-point feature sets for random-selection, as shown in Fig. 8, reducing the gap with orthogonal-selection. This decrease is attributed to the increase in the number of five-point feature sets in random-selection, and the increased probability of selection of five-point feature sets with higher orthogonality indices.
4. As indicated in Fig. 8 a smaller number of five-point feature sets selected using orthogonal-selection may be associated with a mean residual error equivalent to a larger number of five-point feature sets selected by random-selection. For instance, 20 sets derived using orthogonal-selection were commensurate with 100 sets derived from random-selection.
5. The proposed optimal feature selection based on the orthogonality index may be applicable to various VO methods, such as those based on essential, fundamental, and homography matrices as well as the SfM.

V. EXPERIMENTAL RESULTS

A. STEREO-BASED VISUAL ODOMETRY

The stereo-based VO proposed here involves two steps. In the first step, the five-point algorithm [7] was applied to a pair of two left images, L_1 *and* L_2, and another pair of first right and second left images, R_1 *and* L_2. This step resulted in the estimation of rotation and translation up to the scale. In the second step, the absolute scale was quickly estimated from the loop closure formed by the two translation vectors and the camera baseline vector.

1) FEATURE EXTRACTION

The input into our VO algorithm corresponds to features correlating the four images in the previous and current stereo camera frames. We adopted the feature detector and matcher employed by Geiger in VISO-2 [10] due to its 36 ms speed and feature repeatability. This feature detector enhanced the performance of VO with KITTI dataset in [8] [10]. In particular, the corner features in images were extracted by utilizing 5 × 5 blob and corner masks. Additionally, the matching was carried out using the sum of absolute differences (SAD) error metric to compare 11x11 block windows of horizontal and vertical Sobel filter responses to the two feature points detected. To speed-up matching, Sobel responses is quantized to 8 bits and the differences is summed over a sparse set of 16 locations instead of being summed over the whole block window. The extracted features are assigned to four classes (blob max, blob min, corner max, and corner min) and the matching process is done on the same class to reduce the computational time. At this stage, some of the outliers were rejected by circular matching [10], suggesting that each feature needs to be matched between left and right images of two consecutive frames, requiring four matches per feature. The remaining outliers were removed by RANSAC [17]. Finally, the bucketing technique was used to divide the corresponding features into 50 × 50 grids and selecting only a limited number of features in each bucket. This step guaranties uniform distribution of the selected features.

2) ROTATION ESTIMATION

The parameter E represents a 3 x 3 matrix including eight unknowns and an unobservable scale, which satisfies the five epipolar constraints of five correspondences in equation (9). Equations (7), (8), and (9) are extended to 10 cubic constraints, and then to a ten-degree polynomial. As a result, a maximum of 10 essential matrix solutions was obtained for any five-point set. The solution yielding the highest number of inliers was selected as a set representative. In order to guarantee success, multiple five-point feature sets were used to generate hypotheses. The set with the best pre-emptive score and the largest number of inliers was selected as the optimal solution. The essential matrix obtained was decomposed into four rotation-translation pair solutions. The correct pair reconstructed the maximum number of 3D points in the front of both cameras. In order to improve the precision of rotation estimation, a fusion step was added. Assuming the rotation between the current frame in time t and the frames in times *t-1*, *t-2* were denoted as $t^{t-2}q$ and $t^{t-1}q$, which were estimated from the five-point algorithm. The rotation between the frames in *t-1* and *t-2*, denoted as $t-1^{t-2}q$, was calculated in the previous step. The final rotation $t^{t-1}q$ was refined via spherical linear interpolation (SLERP) [23] using two measurements: directly calculated by $t^{t-1}q^1$ and indirectly calculated via $t^{t-1}q^2 = {}^{t-1}_{t-2}q * t^{t-2}q$.

3) TRANSLATION ESTIMATION

The relative orientation between two frames was carried out as above. Here, we introduce a novel method to estimate the translation from two translational scales by solving the linear equations. Assume that the five-point algorithm is applied to two pairs of images, $L_1 - L_2$ and $R_1 - L_2$, in order to obtain two normalized translations $L_1 t_{L_2}$ and $R_1 t_{L_2}$ where $L_1 t_{L_2}, {}^{R_1}t_{L_2}$ represent translation directions from L_2 to L_1 and from L_2 to R_1, respectively. Here, we present a linear approach to compute the absolute scale for these vectors. Three cameras L_1, L_2, and R_1 satisfy a closed-loop constraint, as shown in Fig. 19.

In other words,

$$\alpha t_1 - \beta t_2 = B \tag{26}$$

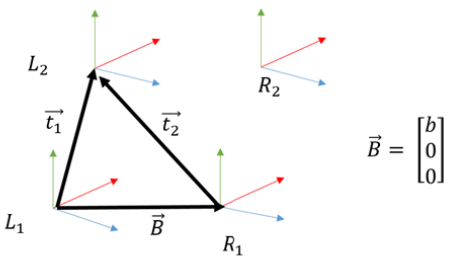


FIGURE 19. Translation scale is estimated by a loop constraint between three camera L_1, R_1, L_2.

where B is the baseline vector,$t_1 = {}^{L_1}t_{L_2}, t_2 = {}^{R_1}t_{L_2}$. Equation (26) is rewritten in detail as

$$\alpha \begin{bmatrix} t_{1x} \\ t_{1y} \\ t_{1z} \end{bmatrix} - \beta \begin{bmatrix} t_{2x} \\ t_{2y} \\ t_{2z} \end{bmatrix} = \begin{bmatrix} b \\ 0 \\ 0 \end{bmatrix} \tag{27}$$

Equation (27) is expressed as a function of α and β

$$\begin{bmatrix} t_{1x} & -t_{2x} \\ t_{1y} & -t_{2y} \\ t_{1z} & -t_{2z} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} b \\ 0 \\ 0 \end{bmatrix} \tag{28}$$

Equation (28) is a linear equation with two unknown variables $X = [\alpha\beta]^T$ and

$$AX = b \tag{29}$$

Thus, X is easily solved using the pseudo-inverse relationship as follows:

$$X = \left(A^T A\right)^{-1} A^T b \tag{30}$$

Equation (30) can be written in a closed form described as follows:

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = b \begin{bmatrix} t_{1x}\left(t_{1x}^2 + t_{1y}^2 + t_{1z}^2\right) - t_{2x}(t_{1x}t_{2x} + t_{1y}t_{2y} + t_{1z}t_{2z}) \\ t_{1x}\left(t_{1x}t_{2x} + t_{1y}t_{2y} + t_{1z}t_{2z}\right) - t_{2x}(t_{2x}^2 + t_{2y}^2 + t_{2z}^2) \end{bmatrix} \tag{31}$$

In order to further improve the accuracy of estimated translation, it was optimized by minimizing the re-projection error with known rotation. The features in the previous frame were reconstructed in 3D, and projected to the current frame using estimated rotation and translation above. The re-projection error in pixel between the projected point and the corresponding observation was measured. The feature points with re-projection errors under a defined threshold were selected. They represent inputs for the refinement step, which was carried out by minimizing the summation of re-projection error with objective function:

$$\sum\nolimits_{k=1}^{n} \left(x_k^l - P_L(X_k, R, t)\right)^2 + \left(x_k^l - P_R(X_k, R, t)\right)^2, \tag{32}$$

where P_L, P_R represent projection matrices obtained by the estimated rotation and translation. The 3D point X_k is the reconstructed point in the previous frame. x_k^l, x_k^l are the observations in the frame.

B. REAL DATASET EVALUATION

In order to consolidate the comparison by simulations, we applied orthogonal and random selections to two publicly available datasets: 1) the autonomous car driving KITTI dataset [24], [25] and 2) the 10 km Devon Island dataset [26] for space mission.

1) KITTI DATASET EVALUATION

The KITTI dataset was collected by driving under different traffic scenarios, which are widely used for evaluating autonomous driving algorithms. The 22 total sections were equally divided into training and testing datasets. The training dataset including 11 sections provides rotational/translational ground-truth. The testing dataset containing the remaining 11 sections maintains the ground-truth private for fair and public comparison. Both datasets accommodate challenging aspects such as different lighting, shadow conditions, and dynamic moving objects. We evaluated the proposed orthogonal selection compared with random selection in the training dataset with different numbers of five-point feature sets associate with average rotational/translational errors. Subsequently, we showed in detail the errors of the 100 testing datasets in terms of the path length provided by the KITTI leaderboard.

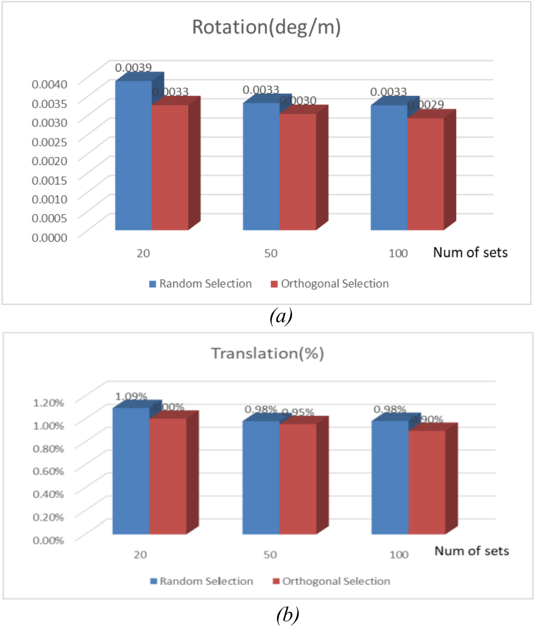
In order to evaluate the performance of the VO approaches, the KITTI benchmark provides a tool for measuring rotational/translational error metrics [24]. RMSEs of ground-truth are computed from all possible subsequences of length (100,200…800 meters) as defined in the paper [25].

2) EVALUATION OF KITTI TRAINING DATASET

We analyzed random and orthogonal selection using the training dataset with varying number of five-point sets such as 20, 50, and 100. The RMSEs of all cases were computed and displayed in Fig.20.

In general, the transformation errors of the two approaches share a steady but significant decline from 20 to 100 sets. The rotational/translational errors of orthogonal selection are less than in random selection under all three cases. Specifically, the proposed method enhances the accuracy of translation and rotation by 9% and 16%, respectively, when 20 sets are used. In addition, the errors of orthogonal selection using 20 sets (1.0% /0.0033 deg/m) are equivalent to the errors of random selection using 100 sets.

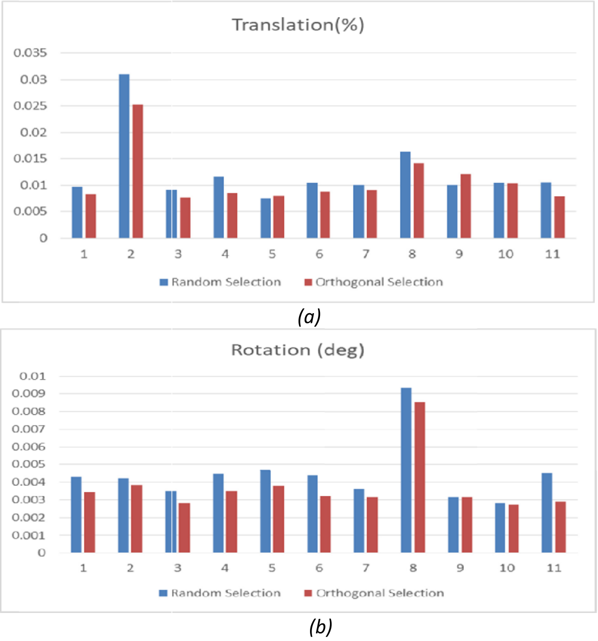
The differences in RMSEs between orthogonal and random selection are highlighted with 20 sets in Fig. 21, which shows the specific errors of the 11 individual training sections. The RMSEs of orthogonal selection, shown in red, are lower than in random selection, shown in blue, generally, for both rotation and translation. The translational errors of the two approaches at an individual section are around 1% except for two sections: 1 and 8. Further, the rotational error of section 8 was also higher than that of the other. The errors in the two sections are large due to the presence of challenging aspects, such as the object movement and the extremely high speed up to 100 km/h.



(a)

(b)

FIGURE 20. Average translational and rotational RMSEs of the KITTI training dataset for the two approaches. (a) Average rotational error. (b) Average translational error.



(a)

(b)

FIGURE 21. Average translational and rotational errors of 11 training sections for two approaches. The orthogonal selection errors are less than in random selection at most sections. (a) Translation error. (b) Rotation error.

Based on the statistical data involving the training dataset, it is obvious that the translational/rotational errors of the feature selection based on orthogonality index were fewer than in the conventional approach. It enhanced the accuracy of translation and rotation by around 9% and 16%, respectively.

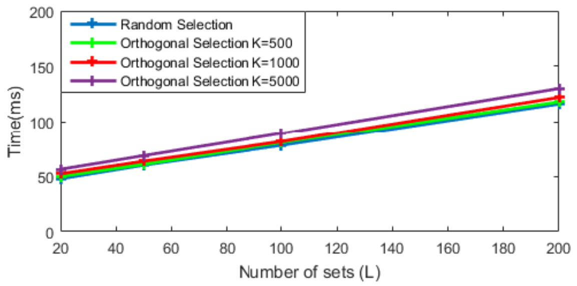


FIGURE 22. Computational time in orthogonal and random selection under different *K* and *L* values.

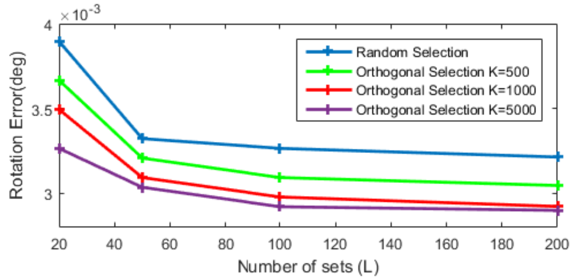


FIGURE 23. Average rotational errors of orthogonal- and random selection under different *K* and *L* values.

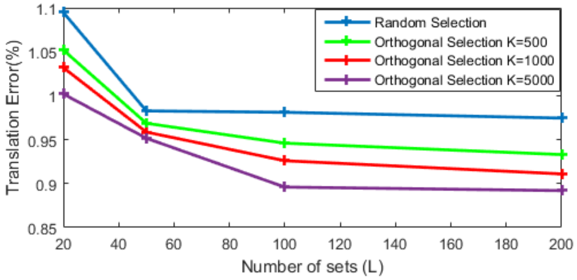


FIGURE 24. Average translational errors of orthogonal and random selection with respect to different *K* and *L* values.

3) ACCURACY AND EFFICIENCY TRADE-OFF

To balance the accuracy and cost, we performed the stereo-based VO with different numbers of generated sets Ks and selected sets Ls:*K*ranged from 500 to 5000 and *L* is from 20 to 200. The computational time, rotational error and translational error are presented in Figs 22, 23 and 24, respectively. The selection of the highest orthogonality indices ranged from 1 ms to 9 ms almost linearly with the increase in *K* from 500 to 5000. Since *K* is larger than 1000, the accuracy of rotation and translation did not improve significantly. The computational time increased gradually along with the number of selected sets *L*. However, the rotational/translational error was reduced exponentially with the increased *L*. The reduction was rapid when *L* ranged from 20 to 50 and was not significant when *L* exceeded 100. A *K* value around 1000 and *L* ranging from 50 to 100 are trade-off points between the efficiency and accuracy.

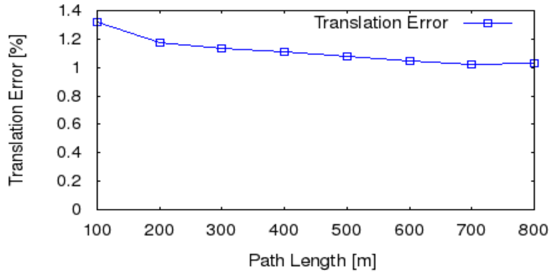


FIGURE 25. Average translation error with respect to the path length of the KITTI test dataset.

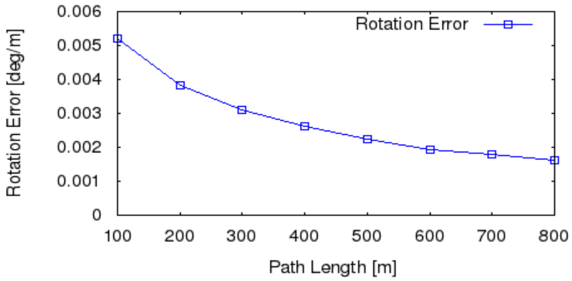


FIGURE 26. Average rotational error with respect to the path length of KITTI test dataset.

4) EVALUATION OF KITTI TESTING DATASETS

We applied VO supported by the proposed orthogonality index in the KITTI testing dataset, and uploaded the estimation of the complete trajectories onto the KITTI leaderboard for evaluation. As shown in Figs. 25 and 26, the average translational and rotational errors varied along with the path length from the KITTI webpage. The rotational and translational errors were reduced from 1.3%/0.005 deg/m at 100 meters to 1%/0.002 deg/m at 800 m. Our approach yielded an average translational/rotational RMSE around 1.13%/0.0030 deg/m for all possible lengths.

In Table 2, we summarize our approach compared to other popular approaches published in the KITTI leaderboard.

This table lists the translational and rotational errors of the proposed and other approaches. The KITTI leaderboard orders the methods based on translational error, resulting in a 1.13% translational error with our approach, which is less than in other popular VO approaches such as ORB_SLAM2 with 1.15%, MFI with 1.3%, and VISO2 with 2.44%. Additionally, it was also better than SSLAM that provides 1.57%, which used robust key points and key-frame selection. Our approach was slightly less accurate than SOFT at 1.03%, which carefully selects long life and stable features and tracking based on four categories. SOFT2 represents an extended version of SOFT based on additional closed-loop constraint to reduce the translational error from 1.03% to 0.65%.

Fig. 27 visualizes the trajectory of section 12 of our method compared with the ground-truth and other popular approaches. Our trajectory almost overlaps with that of SOFT and was closer to the ground-truth than the other approaches. This trajectory is a representative example of the average errors shown in Table 2.

TABLE 2. Errors compared with conventional approaches.

	Translational Error	Rotational Error
[ISRI_VO]Ours[27]	**1.13%**	**0.0030(deg/m)**
SOFT2[28]	*0.65%*	0.0014(deg/m)
SOFT[8]	1.03%	0.0029(deg/m)
ORB_SLAM2[14]	1.15%	0.0027(deg/m)
MFI[19]	1.3%	0.0030(deg/m)
SSLAM[11]	1.57%	0.0044(deg/m)
VISO2[10]	2.44%	0.0114(deg/m)

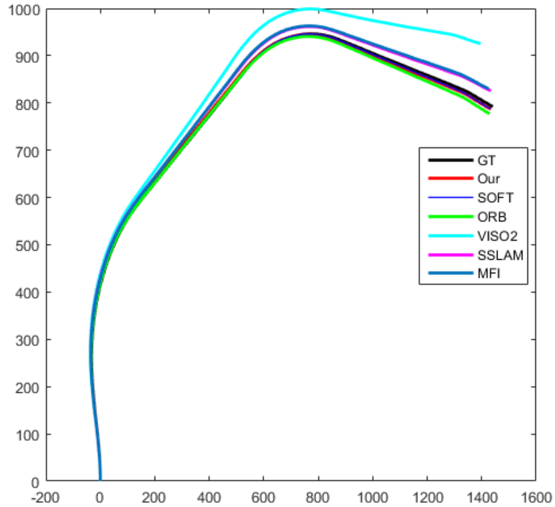


FIGURE 27. Trajectory of section 12 in our method and other approaches compared with ground truth. Our trajectory in red almost overlaps with SOFT in blue, which is the closest to GT in black. VISO2 trajectory is the furthest. SSLAM and MFI are close to each other.

5) DEVON ISLAND DATASET EVALUATION

The Devon Island dataset [26] spans a 10-km distance across a Mars analog site containing 23 individual sections, each approximately 500 m in length and located in the high Arctic regions of Canada. The resulting images are coupled with sun/inclinometer sensors and excellent ground-truth position. At the starting point of each individual section, the rover stops for several minutes to collect adequate sun/inclinometer data to generate ground-truth orientation ranging from the camera to the topo-centric frame. Thus, the rotations of complete individual sections are calculated, without any orientation information between two arbitrary camera frames. In order to evaluate the proposed method with Devon Island data, we conducted a comparison of the feature selections using orthogonality index and the conventional one, and an additional comparison of the proposed VO with the Lambert's approach [13].

6) PERFORMANCE OF FEATURE SELECTION BASED ON ORTHOGONALITY INDEX

A statistical comparison of rotational and translational errors between random and orthogonal selections was performed in order to clarify the benefits of orthogonality index, using the estimated frame-to-frame transformation with three different sets of 20, 50, and 100. Because every single frame does not have an orientation ground truth, excluding the starting frame of each section, it is impossible to compare the relative rotation between two arbitrary frames suggesting that the error metric used to evaluate the KITTI benchmark cannot be applied here. Therefore, we slightly modified the error metric to fit the missing orientation. First, two different translational metric errors were measured: the average translational error covering the full path and the average translational error with different lengths. Second, the average rotational error was calculated. The average translational error over the full path can be easily obtained using the following equation

$$e_{avg} = \frac{\sum_{i=1}^{23} e_i}{23}, \tag{33}$$

in which e_i is the location error of the last frame of section i to the ground-truth position.

The average translational error with different lengths based on individual sections, at stop points such as 100, 200...700 meters, was calculated using equation (34). N_{count} represents the sum of all possible stop points; e_{il} defines the difference in meters between the estimated location and ground-truth at the stop point.

$$e_{avg} = \frac{\sum_{i=1}^{23} \sum_{l=100}^{700} e_i}{N_{count}} \tag{34}$$

As mentioned previously, in the absence of orientation ground at every frame, the rotational error was evaluated based on the starting-frame ground truths, which specifically calculates the difference between the estimation and the ground truth. The rotation ground-truth of section i, ${}^{C_i}_{C_{i+1}}R_{GT}$, was calculated using two ground-truth rotations derived from camera frames to the topo-centric frame. The camera frame C_i represents the first frame of section i. The rotation ${}^{C_i}_{C_{i+1}}R_{VO}$ is estimated by accumulating single rotations. The rotational error of section i is defined as $R_e = {}^{C_i}_{C_{i+1}}R_{GT}\,{}^{C_i}_{C_{i+1}}R_{VO}^{-1}$. In the ideal case, matrix R_e is a 3 × 3 identity matrix. However, in real-world situations, the rotational error can be expressed as follows:

$$O^{rot} = arccos\left(\frac{trace\,(R_e) - 1}{2}\right) \tag{35}$$

This rotation error metric was used to compare 22 individual sections using two approaches, with no possible ground truth calculation involving the last section. The average rotational error was also computed as follows:

$$\mathbf{O}^{rot}_{avg} = \frac{\sum_{i=1}^{22} \mathbf{O}^{rot}_{i}}{22} \tag{36}$$

The three error metrics are discussed below.

As shown in Fig. 28, illustrating the average error covering the full-length path, the horizontal axis represents the variation in the number of sampled subsets, while the vertical axis carries the mean translational errors involving the two

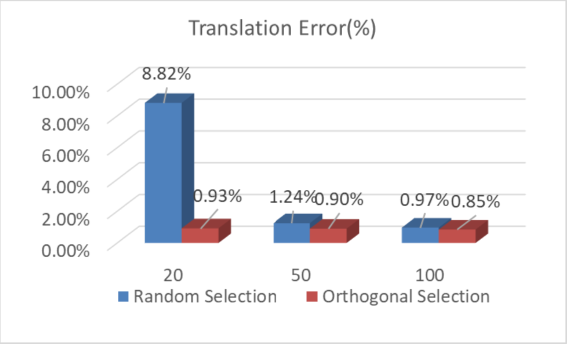


FIGURE 28. Average translational error with respect to the full-path length of the Devon Island dataset.

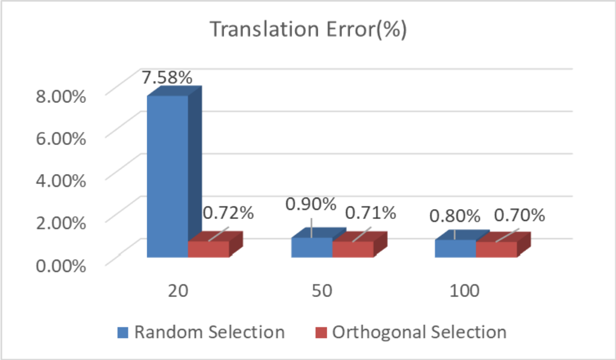


FIGURE 29. Average translational error with respect to the path lengths of Devon Island dataset.

approaches. The random selection method using 20 sets displays the highest percentage of translational errors at 8.8%. Using 50 and 100 sets, this error was reduced continuously to 1.2% and 0.97%, respectively. Orthogonal selection also showed a downtrend with the number of sampled sets. However, the largest percentage of translation error was only 0.93% considering 20 sampled sets, which decreased to 0.9% and 0.85% with 50 and 100 sets, respectively. This result suggests that the average translational error spanning the full length of the random method was larger than that of the proposed approach.

Fig. 29 illustrates the average translational errors with different lengths. The translational error of random-selection at 20 sets was 7.5%, which was reduced to 0.9% at 50 and 0.8% at 100 sets. These errors associated with orthogonal selection were 0.72%, 0.71%, and 0.70%, respectively. The trend in translational error with different lengths was similar to the full-length in terms of downtrend, shape of the graph, and other parameters

Fig. 30 shows the average rotational error. The largest error of rotation still belongs to random selection using 20 sets with 0.027 deg/m, which was reduced to 0.005 deg/m with 50 sets and to 0.044 deg/m with 100 sets, respectively. The rotational error of orthogonal selection also showed a downward trend; however, it was slower and more stable, starting with 0.0045 deg/m at 20 sets and ending at 100 sets with 0.0040 deg/m.

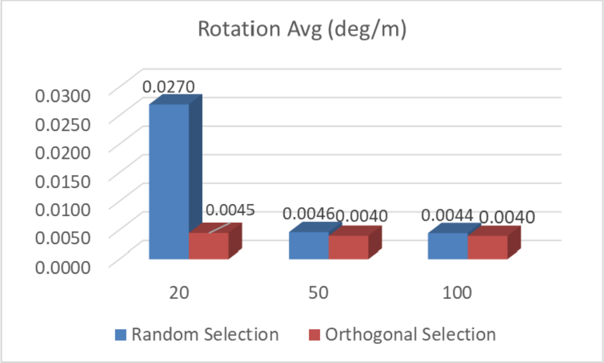
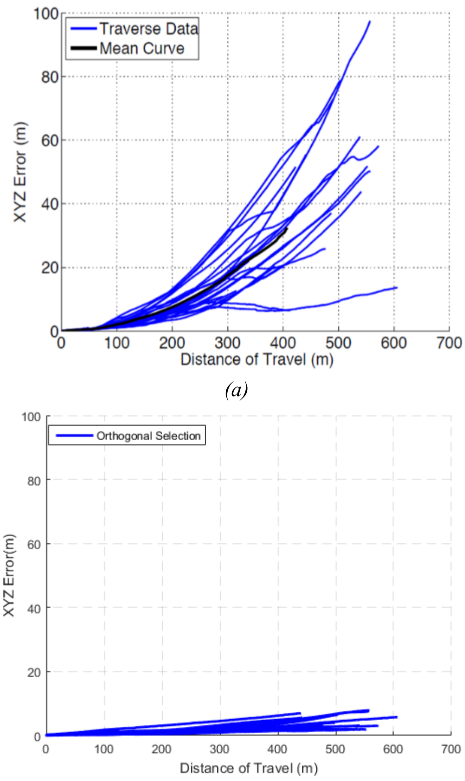


FIGURE 30. Average rotational error with respect to the full path length of Devon Island dataset.



(a)

(b)

FIGURE 31. Comparison of XYZ error with respect to travel distance under two approaches. (a) Conventional *VO's XYZ* error [13]. (b) Our *XYZ* error.

The above statistical results valid the theory discussed in simulation, suggesting that the feature selection with orthogonality index was better than in random selection in terms of accuracy, especially when using fewer sets. Random selection with 20 sampled sets provides enormous amounts of translational and rotational errors due to the breakdown of essential matrix estimation at specific frames in certain sections. This finding indicates that even with fewer sets, orthogonal selection guarantees the success of estimation. Therefore, when the errors over the full-path length are considered, the translational enhancement increased the accuracy of the 50 and 100 sets by 20% and 10%, respectively. The rotational enhancement of 50 sets was 13%. Moreover, the accuracy of orthogonal selection at 20 sets was similar to that of random selection at 100 sets. The translational error of the proposed method was 0.9% over the full-length path.

7) COMPARISON TO CONVENTIONAL APPROACH

A study reported from the top performance of the conventional method [13] evaluated 23 individual sections with VO.

In order to show the benefit of the proposed method, we followed the same strategy used by Andrew Lambert for comparison, by measuring the translational error in XYZ with regard to the distance traveled. Fig. 31(a) shows the trends in Lambert's VO translational error while Fig. 31(b) describes our own. The translational errors for all sections in the proposed method were always less than 10 m; otherwise, these errors involving Lambert's method were around 40 m at 500 m. Clearly, the average translational error of Lambert's VO with 8% was higher compared with our method, at around 0.9%.

VI. CONCLUSION

A new approach for the selection of optimal features in VO is presented by introducing the orthogonality index associated with a set of equations and constraints involved in VO. Compared with the conventional methods that resort to a large number of heuristically or randomly selected features, the proposed method relies on a mathematical formalism by minimizing the extent of uncertainty involved in estimating the essential, fundamental or homography matrix for VO. The effectiveness of the proposed method was verified via statistical simulations, which demonstrated a definitive negative correlation between the orthogonality index and the residual error estimated. The simulations also establish that the proposed optimal feature selection reduced residual errors significantly compared with conventional random selection using various approaches based on VO utilizing the essential, fundamental or homography matrix as well as the structure derived from motion. Experiments with KITTI dataset show that the proposed orthogonal selection outperforms the random selection by nearly 10% for both rotation and translation estimation, resulting in an average translational error of 1.13%. Experiments with Devon Island dataset for stereo-visual odometry also indicate that our method based on the proposed optimal feature selection resulted in an average translational error of 0.9%, exceeding the top performance of the conventional method [13]. In the future, we plan to widen the scope of applications utilizing the proposed orthogonality index-based optimal feature selection.

AUTHOR CONTRIBUTIONS

Sukhan Lee proposed the original concept of orthogonality index with theoretical analysis; Nguyen Huu Hung designed and performed the simulations and experiments with the implementation of a prototype.

REFERENCES

[1] D. Nistér, O. Naroditsky, and J. Bergen, "Visual odometry," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, Jun./Jul. 2004, p. I.

[2] A. Kelly, "Linearized error propagation in odometry," *Int. J. Robot. Res.*, vol. 23, no. 2, pp. 179–218, Feb. 2004.

[3] M. O. A. Aqel, M. H. Marhaban, M. I. Saripan, and N. B. Ismail, "Review of visual odometry: Types, approaches, challenges, and applications," *SpringerPlus*, vol. 5, no. 1, p. 1897, Dec. 2016.

[4] S. Poddar, R. Kottath, and V. Karar, "Motion estimation made easy: Evolution and trends in visual odometry," in *Recent Advances in Computer Vision*, Cham, Switzerland: Springer, 2019, pp. 305–331.

[5] D. Scaramuzza and F. Fraundorfer, "Visual odometry [tutorial]," *IEEE Robot. Automat. Mag.*, vol. 18, no. 4, pp. 80–92, Dec. 2011.

[6] F. Fraundorfer and D. Scaramuzza, "Visual odometry: Part II: Matching, robustness, optimization, and applications," *IEEE Robot. Automat. Mag.* vol. 19, no. 2, pp. 78–90, Jun. 2012.

[7] D. Nistér, "An efficient solution to the five-point relative pose problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 6, pp. 756–770, Jun. 2004.

[8] I. Cvišić and I. Petrović, "Stereo odometry based on careful feature selection and tracking," in *Proc. Eur. Conf. Mobile Robots (ECMR)*, Sep. 2015, pp. 1–6.

[9] B. Kitt, A. Geiger, and H. Lategahn, "Visual odometry based on stereo image sequences with RANSAC-based outlier rejection scheme," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2010, pp. 486–492.

[10] A. Geiger, J. Ziegler, and C. Stiller, "StereoScan: Dense 3D reconstruction in real-time," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2011, pp. 963–968.

[11] M. Fanfani, F. Bellavia, and C. Colombo, "Accurate keyframe selection and keypoint tracking for robust visual odometry," *Mach. Vis. Appl.*, vol. 27, no. 6, pp. 833–844, Aug. 2016.

[12] J.-P. Tardif, Y. Pavlidis, and K. Daniilidis, "Monocular visual odometry in urban environments using an omnidirectional camera," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2008, pp. 2531–2538.

[13] A. Lambert, P. Furgale, T. D. Barfoot, and J. Enright, "Field testing of visual odometry aided by a sun sensor and inclinometer," *J. Field Robot.*, vol. 29, no. 3, pp. 426–444, May/Jun. 2012.

[14] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An open-source SLAM system for monocular, stereo,and RGB-D cameras," *IEEE Trans. Robot.* vol. 3, no. 5, pp. 1255–1262, Oct. 2017.

[15] A. Schmidt, M. Kraft, and A. Kasiński, "An evaluation of image feature detectors and descriptors for robot navigation," in *Proc. Int. Conf. Comput. Vis. Graph.* Berlin, Germany: Springer, 2010, pp. 251–259.

[16] Y. Jiang, Y. Xu, and Y. Liu, "Performance evaluation of feature detection and matching in stereo visual odometry," *Neurocomputing*, vol. 120, pp. 380–390, Nov. 2013.

[17] M. A. Fischler and C. B. Robert, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM* vol. 24, no. 6, pp. 381–395, Jun. 1981.

[18] M. Buczko and W. Volker, "How to distinguish inliers from outliers in visual odometry for high-speed automotive applications," in *Proc. Intell. Vehicles Symp. (IV)*, Jun. 2016, pp. 1–6.

[19] H. Badino, A. Yamamoto, and T. Kanade, "Visual odometry by multi-frame feature integration," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Dec. 2013, pp. 222–229.

[20] A. Björck, "Numerics of gram-schmidt orthogonalization," *Linear Algebra Appl.* vols. 197–198, pp. 297–316, Jan./Feb. 1994.

[21] R. I. Hartley, "In defence of the 8-point algorithm," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jun. 1995, pp. 1064–1070.

[22] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[23] E. B. Dam, K. Martin, and M. Lillholm, *Quaternions, Interpolation and Animation*, Vol. 2. Copenhagen, Denmark: Datalogisk Institut, Københavns Universitet, 1998.

[24] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, 2013.

[25] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3354–3361.

[26] P. Furgale, P. Carle, J. Enright, and T. D. Barfoot, "The Devon Island rover navigation dataset," *Int. J. Robot. Res.*, vol. 31, no. 6, pp. 707–713, 2012.

[27] *Our Evaluation Results on KITTI Dataset*. [Online]. Available: http://www.cvlibs.net/datasets/KITTI/eval_odometry_detail.php?&result=efdc138ff1c6d279abcfdf9105b7f6ba7553071c

[28] I. Cvisic, J. Cesis, I. Markovic, and I. Petrovic, "SOFT-SLAM: Computationally efficient stereo visual simultaneous localization and mapping for autonomous unmanned aerial vehicles," *J. Field Robot.*, vol. 35, no. 4, pp. 578–595, Jun. 2017.

HUU HUNG NGUYEN was born in Thanh Hóa, Vietnam, in 1984. He received the B.S. degree in applied mathematics-informatics from the Hanoi University of Science and Hanoi National University, in 2006, and the M.S. degree in electrical engineering from Sungkyunkwan University, South Korea, in 2013.

From 2007 to 2011, he was a Research Assistant in technology and electronics with Le Quy Don Technical University, Hanoi, Vietnam. Since 2011, he has been a Research Assistant with the Intelligent Systems Research Institute, Sungkyunkwan University. His research interests include SLAM, visual odometry, 3D object recognition, pose estimation, calibration, and modeling.

SUKHAN LEE was received the B.S. and M.S. degrees in electrical engineering from Seoul National University, in 1972 and 1974, respectively, and the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, in 1982.

From 1983 to 1997, he was with the Department of Electrical Engineering and the Department of Computer Science, University of Southern California. From 1990 to 1997, he was with the Jet Propulsion Laboratory, California Institute of Technology, as a Senior Member of Technical Staff. From 1998 to 2003, he was the Executive Vice President and a Chief Research Officer at the Samsung Advanced Institute of Technology. Since 2003, he has been a Professor of information and communication engineering and a WCU Professor of interaction science with Sungkyunkwan University, where he has been the Director of the Intelligent Systems Research Institute. Since 2011, he has been the Dean of the Graduate School, Sungkyunkwan University. He is currently a Fellow of the Korean National Academy of Science and Technology. His research interests include cognitive robotics, intelligent systems, and micro/nano electro-mechanical systems.