


Uncertainty Hypervolume in Point Feature-Based Visual Odometry

InJun Mun¹ ^a and Sukhan Lee^{2,*} ^b

*Intelligent Systems Research Institute, Department of Artificial Intelligence,
Sungkyunkwan University, Suwon 16419, South Korea*

Keywords: Localization, Visual Odometry, Essential Matrix, Optimal Feature Selection, Uncertainty Hypervolume, Bucketing.

Abstract: Visual odometry based on point feature matching has been well-established. Notably, methods based on essential and fundamental as well as homography matrices have been widely used. It is known that the accuracy of visual odometry is affected by the choice of matched feature point pairs. However, no mathematically rigorous formula relating the choice of feature point pairs to the uncertainty involved in visual odometry is available. Instead, point selection heuristics based on feature point distribution combined with RANSAC-based refinement are mostly adopted to ensure accuracy. In this paper, we present “Uncertainty Hypervolume” as a rigorous mathematical formula that relates the selected feature point pairs to the uncertainty of visual odometry. The uncertainty hypervolume associated with selected feature point pairs provides a precise metric for evaluating the selected feature point pairs and the resulting visual odometry. This metric is useful in practice not only for selecting the best feature point pairs but also for selecting poor feature point pairs available for visual odometry. Furthermore, it accurately identifies the uncertainty in visual odometry, which helps better manage the performance of visual odometry applications.

1 INTRODUCTION


Visual odometry (Nistér et al., 2004) is a fundamental technique in computer vision and robotics that facilitates estimating camera movement by analyzing sequential images. The accuracy of pose estimation, a critical aspect of visual odometry, depends not only on the chosen pose estimation method but also significantly on the quality of the selected correspondence feature points. S. Poddar, R. Kottath, and V. Karar (Poddar et al., 2019) conducted a comprehensive review of feature selection strategies for visual odometry, outlining key steps including feature detection, description, inlier/outlier detection, feature distribution, and consideration of feature quality. This multi-step process emphasizes the intricate relationship between the accuracy of pose estimation and the characteristics of the selected feature points.


In feature selection, the initial removal of outliers is paramount, as mismatched feature pairs can lead to

erroneous pose estimation. To solve this issue, pioneering work by Fischler (Fischler et al., 1981) introduced the random sample consensus (RANSAC) algorithm, which uses geometric constraints to remove outliers from the feature set effectively.

However, the effort to improve accuracy in pose estimation extends beyond the methodological level. Researchers have recognized that the distribution and uniformity of corresponding feature points in space also play a critical role in determining visual odometry performance (Cvišić et al., 2015). As pointed out in Poddar's review, traditional feature selection methods often result in a non-uniform distribution of feature points across the image. As a result, clusters of closely spaced feature points can lead to suboptimal pose estimation results.

To overcome this limitation and achieve improved accuracy, an innovative approach has emerged. This approach, known as the bucketing technique (Zhang et al., 1995, Kitt et al., 2010) attempts to achieve uniformity in the distribution of correspondence feature points. By partitioning the

^a  <https://orcid.org/0009-0009-8524-0719>

^b  <https://orcid.org/0000-0002-1281-6889>

* Corresponding author

image into a grid of $M \times M$ buckets and selecting only a small number of features from each bucket, the bucketing technique ensures a well-distributed selection of features. In particular, this uniform distribution of features has the potential to improve both the accuracy and computational efficiency of pose estimation.

The previously mentioned approaches for feature selection aimed to select dependable features through the removal of outliers or consideration of feature point distribution. While these methods have proven effective in enhancing the accuracy of VO, they do not ensure optimality in terms of mathematical formalism, which guarantees the minimum uncertainty of VO.

Recently, the concept of the Orthogonality Index (Nguyen and Lee, 2019) has been introduced to analytically derive optimal feature selection. This approach demonstrates optimal feature selection through a well-defined mathematical format instead of random selection. The process increases the orthogonal exponent of individual equations and applies constraints to computation to reduce uncertainty when estimating Essential, Fundamental, or Homography matrices associated with visual odometry. However, while the Orthogonality Index provides a mathematical method for optimal feature selection, they do not account for uncertainty in feature points due to measurement or other noise. This issue must be addressed as it significantly impacts VO estimation. Therefore, a method that reflects these factors is necessary to ensure optimal feature selection.

To this end, our study capitalizes on insights gained from simulation experiments, which have shown that the measurement error variance and the spatial distribution of the extracted feature points significantly affect pose estimation. We propose a novel approach that incorporates both of these factors.

Our approach can be summarized as follows: If the matched feature point pairs are well-matched with minimal measurement error and are uniformly distributed throughout the image, the estimated essential matrix is expected to be close to the ground truth essential matrix. However, due to the uncertainty of the matched feature point pairs used to estimate the essential matrix and the error of the equations generated using them, the estimated essential matrix forms a stochastically distributed distribution centered on the ground truth (GT). We experimentally demonstrate that the degree of dispersion depends on the magnitude of the uncertainty in estimating the essential matrix. We

found that the spatial distribution they form should be taken into account when selecting matching feature point pairs, and present a novel "Uncertainty Hypervolume" approach that takes both into account. The estimated essential matrix is stochastically distributed around the reference ground truth essential matrix, and we quantify this with hypervolume. Through experiments, we show a significant correlation between hypervolume and the error of the pose derived from the essential matrix. Based on these results, we propose a mathematically well-structured Uncertainty Hypervolume based approach for feature point pair selection to obtain the optimal solution.

In the following sections, we detail our methodology, experimental setup, and results, culminating in a comprehensive analysis of the interplay between feature selection, spatial distribution, and pose estimation accuracy.

2 PROBLEM DEFINITION AND APPROACH

2.1 Preliminary

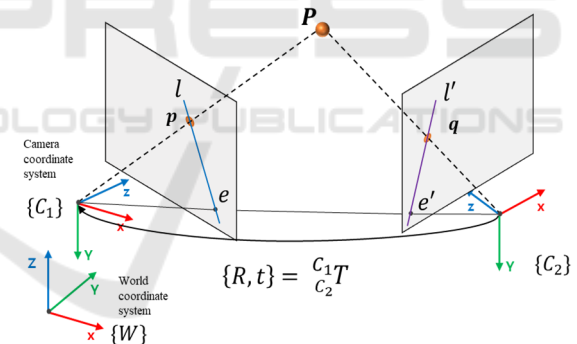


Figure 1: Epipolar Geometry, A 3D point P is projected onto the normalized image plane of each camera at p and q . The points e and e' where the line connecting the two camera origins and the image plane meet are called epipole, and the straight lines l and l' connecting the projection points and the epipole are called epiline (epipolar line).

In epipolar geometry (Deriche et al., 1994), given a point P in space, cameras C_1 and C_2 view the point P from two different perspectives. The point P is then projected onto the normalized image plane of each camera C_1 and C_2 as p and q (p and q are homogeneous normalized image coordinates). It is known that there is always a 3×3 essential matrix (Nistér, 2004), E between the projected points p and q that satisfies the epipolar constraint $p^T E q = 0$.

This essential matrix obeys the following constraints.

$$\det(E) = 0 \tag{1}$$

$$2EE^T E - \text{tr}(EE^T)E = 0 \tag{2}$$

The second expression is a matrix constraint that gives nine equations for the elements of E . However, only two of them are algebraically independent. Thus, with the two essential matrix constraints mentioned above, we can determine the essential matrix with only five corresponding point pairs (Deriche et al., 1994).

Once E is determined, the rotation matrix R and the translation vector t can be obtained by performing a Singular Value Decomposition (SVD).

While R has 3 degrees of freedom and t has 3 degrees of freedom, if we consider the essential matrix as a projection element, it has 5 degrees of freedom with the scale factor removed. Therefore, we can estimate E with five pairs of corresponding feature points and the Essential Matrix constraint.

The epipolar constraint $p^T E q = 0$ can be expressed simply as follows.

$$v \hat{E} = 0 \tag{3}$$

where,

$$v = [p_1 q_1, p_2 q_1, p_3 q_1, p_1 q_2, p_2 q_2, p_3 q_2, p_1 q_3, p_2 q_3, p_3 q_3] \tag{4}$$

$$\text{and } \hat{E} = [E_{11}, E_{12}, E_{13}, E_{21}, E_{22}, E_{23}, E_{31}, E_{32}, E_{33}]^T \tag{5}$$

E can be determined based on the five pairs of corresponding feature points, p and q that define the following 5x9 matrix equation:

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

$$\text{where, } A = [v_1 \ v_2 \ v_3 \ v_4 \ v_5]^T \tag{7}$$

Then, VO between the two camera frames can be derived from E obtained by (1), (2), and (6).

2.2 Problem Definition

The accuracy of the estimated essential matrix determines the accuracy of the transformation relationship between the two cameras. This is equivalent to the performance of Visual Odometry. In VO, we use pairs of corresponding feature points that match in both image planes to compute the essential matrix. In other words, it is obvious that the accuracy of the estimated essential matrix will increase if the pairs of corresponding feature points with good quality and evenly distributed in the image plane are selected and computed. In this paper, we investigate how the accuracy of VO is affected by selecting well-distributed and high-quality corresponding feature

points when estimating the essential matrix. We propose a new approach to feature point selection using a canonicalization metric called "**Uncertainty Hypervolume**".

2.3 Approach

Corresponding pairs of feature points (p_i, q_i) have uncertainties due to measurement error, matching error, noise error, etc. In the epipolar constraint of (3), the solution of E lies in the space perpendicular to v_i . The uncertainty of v_i leads to the uncertainty of E , and the estimated E is stochastically distributed around the GT due to the uncertainty of v_i . The solution subspace formed around the GT changes in size as a function of the error associated with v_i , i.e., the larger the uncertainty of the corresponding pair of feature points, the more stochastically spread the solution subspace becomes. This is equivalent to saying that the volume size of the solution subspace represents the uncertainty. From now on, we will refer to the size of the solution subspace in higher dimensions as the "Hypervolume". Our goal is to choose (p_i, q_i) for v_i such that the size of this hypervolume is minimized (i.e., we choose v_i such that the uncertainty of the solution is small). To achieve our goal, we will explore how the uncertainty of a corresponding pair of feature points affects the uncertainty of the solution subspace, and more specifically, we will define and explain the concept of a hypervolume.

Consider the simplest quadratic form of the problem (the solution of the Essential matrix we want to find is high-dimensional, with 9 dimensions. So, we extend the concepts from lower to higher dimensions).

$$\begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \tag{8}$$

Let $A = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix}$, $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $c = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$ and the solution we want to find is X . Knowing A and c , there is only one solution ($a_1 \neq a_2, b_1 \neq b_2, A \neq 0$). However, if we consider the case where there is uncertainty due to the error of A , it is equivalent to (9).

$$\begin{bmatrix} a_1 \pm \Delta a_1 & b_1 \pm \Delta b_1 \\ a_2 \pm \Delta a_2 & b_2 \pm \Delta b_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \tag{9}$$

$$A'_i = \begin{bmatrix} a_1 \pm \Delta a_1 & b_1 \pm \Delta b_1 \\ a_2 \pm \Delta a_2 & b_2 \pm \Delta b_2 \end{bmatrix} \tag{10}$$

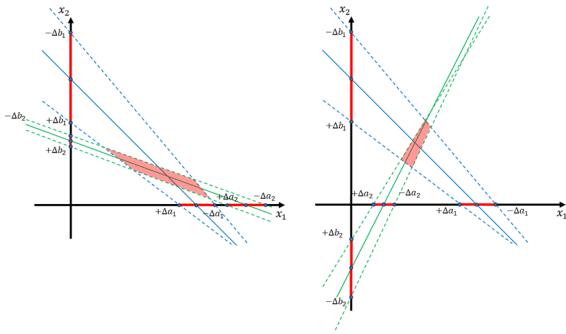


Figure 2: (Left) Graph plotted when the equation of two linear lines has a negative slope. The intercept is determined by $\pm\Delta a_1$, $\pm\Delta b_1$, $\pm\Delta a_2$, and $\pm\Delta b_2$, which is the variance due to uncertainty. (Right) When one linear equation has a positive slope and another has a negative slope. There are a total of four possible cases where A'_i , ($i = 1, \dots, 4$) can be determined, but the equations of this line have four vertices, and the trapezoid is the hypervolume.

The magnitude of the error is denoted by Δa_1 , Δb_1 , Δa_2 , and Δb_2 for the elements of A . The meaning of \pm is the variance of the error, which corresponds to the uncertainty by having a value in the range. Fig. 2 shows the graph of (9). In the graph, the solid line is the equation of the original straight line before adding the error, and the equation of the straight line represented by the dashed line is the equation of A'_i ($i = 1, \dots, 4$) when the error variance is maximum.

$$\begin{aligned} A'_1 &= \begin{bmatrix} +\Delta a_1 & +\Delta b_1 \\ +\Delta a_2 & +\Delta b_2 \end{bmatrix}, & A'_2 &= \begin{bmatrix} +\Delta a_1 & +\Delta b_1 \\ -\Delta a_2 & -\Delta b_2 \end{bmatrix} \\ A'_3 &= \begin{bmatrix} -\Delta a_1 & -\Delta b_1 \\ +\Delta a_2 & +\Delta b_2 \end{bmatrix}, & A'_4 &= \begin{bmatrix} -\Delta a_1 & -\Delta b_1 \\ -\Delta a_2 & -\Delta b_2 \end{bmatrix} \end{aligned} \quad (11)$$

The solution of the original system of equations without error is the intersection of the equations of the two straight lines represented by the solid lines and is determined to be one. However, the solution of the system of equations with the uncertainty given by the error will probabilistically lie within a trapezoid whose vertices are the intersections of the equations of the dotted lines shown in the graph. The larger the error, the greater the width of this trapezoid, i.e., the greater the uncertainty of the solution. The "Uncertainty hypervolume" defined in the previous section 2.3 corresponds to the area of the trapezoid in this problem. Thus, by calculating the size of the hypervolume, we can quantitatively represent the uncertainty. The hypervolume we are talking about forms an n -dimensional hypercube depending on the dimension of the problem to be solved, which is two-dimensional in the case of equation (9), so it becomes a two-dimensional trapezoid, that is, the area of the

shape. In this case, the number of vertices that form the boundary of the shape is 2^n , where n is the number of unknown variables. In the next higher dimension, three-dimensional, we can think of the hypervolume as a three-dimensional cube, which is a crumpled cube with eight vertices (one dimension: line, two dimensions: face, three dimensions: cube, ..., n -dimensional: n -hypercube).

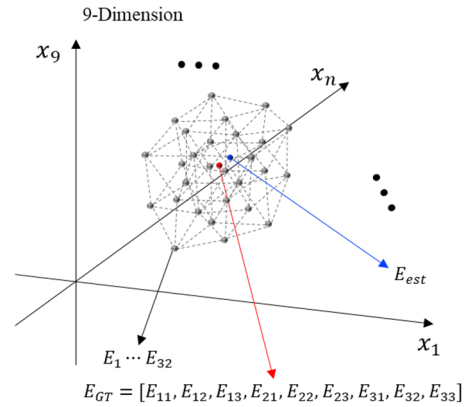


Figure 3: E_{est} is the essential matrix computed with five pairs of corresponding feature points. E_1, \dots, E_{32} , obtained by (13), form the vertices of a hypercube, which is a clustering of groups that have the maximum variance and are probabilistically likely to be the solution subspace. Its hypervolume reflects the uncertainty, and the goal is to find the corresponding feature point set that minimizes it.

Let's return to our original problem and extend the concept of hypervolume defined in a low-dimensional space to a higher-dimensional domain. The essential matrix we want to find is a 3×3 matrix with 9 elements, as shown in (5). The solution of E exists somewhere in the 9-dimensional space and must be singularized. However, due to the uncertainty caused by the errors of v_i , the corresponding feature point pairs, the estimated solution of E will be stochastically distributed around the ground truth E . In the case of the 5-point algorithm we use for VO estimation, the solution can be obtained using only 5 pairs of corresponding feature points due to the additional constraints (1) and (2). Therefore, using equation (6) to express the error variance to represent the uncertainty in the manner of (9), $2^5 = 32$ pairs of v'_i , ($i = 1, \dots, 32$) are generated (13). If the Essential matrix is estimated using this as input to the 5-point algorithm, E_i ($i = 1, \dots, 32$) is generated. This means that the manifold formed by the two constraints projects the solution subspace that exists in the 9-dimension to a lower dimension in the 5-dimension. Therefore, the E_i ($i = 1, \dots, 32$) calculated using this method are stochastically distributed around the

ground truth E and form the outermost vertices of the solution subspace. Fig. 3 shows the projection of E in 9-dimensional space into 3-dimensional space (not an exact projection, but an approximation for illustration purposes). The uncertainty due to multiple error factors is represented by a 5-dimensional hypercube with 32 vertices, and the solution probabilistically exists inside the hypercube. The volume of the hypercube is the hypervolume, and its size represents the degree of uncertainty. We try to minimize the uncertainty of VO estimation by selecting a set of feature points with the minimum hypervolume size.

2.4 Hypervolume Calculation Using Qhull Algorithm

In the previous section, we detailed the procedure for acquiring 32 vertices $E_i (i = 1, \dots, 32)$ that constitute the hypervolume and its significance. Additionally, it was proposed that the size of the hypervolume corresponds to the level of uncertainty, which can be quantified through its acquisition. We utilized the MATLAB function *convexhulln* to obtain the hypervolume with the 32 vertices E_i that comprise its outermost layer. The function is founded on the Qhull algorithm, which functions in the following manner.

- a. *Point Sorting*: First, use the technique of aligning data points appropriately. Because data points can be randomly distributed in higher-dimensional spaces, setting up the ordered order makes it more efficient in subsequent steps. Sorted data helps you calculate convex shells.
- b. *Centrum Location*: Calculates the center position of the data. This is used to clip points based on the center position and to quickly calculate the convex shell. The center position can be related to the mean or median of high-dimensional data.
- c. *Create Point Clipping*: Use the center position to clip data points and apply techniques to remove unnecessary points. This reduces unnecessary calculations and optimizes memory usage. In high-dimensional data, many points may not contribute to the formation of convex shells.
- d. *Convex Hull of Clipped Points*: Calculates the convex shell for clipped points. This forms most of the final convex shell. The process of calculating convex shells for clipped data is efficient.
- e. *Return Results*: Finally, the Qhull algorithm returns the calculated convex shell. This gives results as convex polygons or convex polygons

surrounding a given data point in a high-dimensional space.

Through these various geometric and computational techniques, the Qhull algorithm effectively computes the convex shells of high-dimensional data. And then, The optimal feature selection based on the proposed hypervolume method is described as follows:

Algorithm 1: Hypervolume-based optimal feature selection.

Data: L correspondence feature point sets

Result: five-point sets with the lowest hypervolume

Step 1. Random selection of five feature point sets, $\{ (p_i, q_i), i = 1, \dots, 5 \}$ from the L feature pairs detected from the images of two camera views subject to VO. It is worth noting that a bucketing approach can be incorporated into this step to improve the initial selection of five feature point sets.

Step 2. Generate the corresponding coefficient vectors, $\{ v_i, i = 1, \dots, 5 \}$. Then, compute the essential matrix E_{est} using the five points algorithm with $\{ v_i, i = 1, \dots, 5 \}$ as input.

Step 3. Generate 32 pairs of $v'_i = (p'_i, q'_i), \{ i = 1, \dots, 32 \}$ taking into account the error of the corresponding feature points to $\{ v_i, i = 1, \dots, 5 \}$. With this, the 32-vertex essential matrix is computed. And, compute the hypervolume of the hypercube composed of these 32 vertices using the Convex Hull algorithm.

Step 4. Select feature sets with hypervolume values less than a threshold K . Estimate VO using the best set of selected feature points.

$$p'_i = \begin{bmatrix} p_x^1 \pm e_x^1 & p_x^2 \pm e_x^2 & p_x^3 \pm e_x^3 & p_x^4 \pm e_x^4 & p_x^5 \pm e_x^5 \\ p_y^1 \pm e_y^1 & p_y^2 \pm e_y^2 & p_y^3 \pm e_y^3 & p_y^4 \pm e_y^4 & p_y^5 \pm e_y^5 \end{bmatrix}$$

$$q'_i = \begin{bmatrix} q_x^1 \pm e_x^6 & q_x^2 \pm e_x^7 & q_x^3 \pm e_x^8 & q_x^4 \pm e_x^9 & q_x^5 \pm e_x^{10} \\ q_y^1 \pm e_y^6 & q_y^2 \pm e_y^7 & q_y^3 \pm e_y^8 & q_y^4 \pm e_y^9 & q_y^5 \pm e_y^{10} \end{bmatrix}$$

$$v'_i = (p'_i, q'_i), \{ i = 1, \dots, 32 \} \quad (12)$$

In Step 1 of Algorithm 1, partitioning the image into grids using the bucketing technique and extracting feature points from each grid region resulted in improved performance regarding running time and estimation error. The use of the bucketing method proves to be more efficient in finding solutions over multiple iterations. This can be explained by the fact that, as shown in Section 2.3, we

empirically verified that the pose estimation error is minimized when a well-spread distribution of feature points is used as input. In the following, we showcase our proposed approach and evaluate its practical effectiveness.

$$\begin{aligned}
 p'_i &= \begin{bmatrix} p_x^1 + e_x^1 & p_x^2 + e_x^2 & p_x^3 + e_x^3 & p_x^4 + e_x^4 & p_x^5 + e_x^5 \\ p_y^1 + e_y^1 & p_y^2 + e_y^2 & p_y^3 + e_y^3 & p_y^4 + e_y^4 & p_y^5 + e_y^5 \end{bmatrix} \\
 &= \begin{bmatrix} p_x^1 - e_x^1 & p_x^2 + e_x^2 & p_x^3 + e_x^3 & p_x^4 + e_x^4 & p_x^5 + e_x^5 \\ p_y^1 - e_y^1 & p_y^2 + e_y^2 & p_y^3 + e_y^3 & p_y^4 + e_y^4 & p_y^5 + e_y^5 \end{bmatrix} \\
 &= \begin{bmatrix} p_x^1 - e_x^1 & p_x^2 - e_x^2 & p_x^3 - e_x^3 & p_x^4 - e_x^4 & p_x^5 - e_x^5 \\ p_y^1 - e_y^1 & p_y^2 - e_y^2 & p_y^3 - e_y^3 & p_y^4 - e_y^4 & p_y^5 - e_y^5 \end{bmatrix} \\
 q'_i &= \begin{bmatrix} q_x^1 + e_x^6 & q_x^2 + e_x^7 & q_x^3 + e_x^8 & q_x^4 + e_x^9 & q_x^5 + e_x^{10} \\ q_y^1 + e_y^6 & q_y^2 + e_y^7 & q_y^3 + e_y^8 & q_y^4 + e_y^9 & q_y^5 + e_y^{10} \end{bmatrix} \\
 &= \begin{bmatrix} q_x^1 - e_x^6 & q_x^2 + e_x^7 & q_x^3 + e_x^8 & q_x^4 + e_x^9 & q_x^5 + e_x^{10} \\ q_y^1 - e_y^6 & q_y^2 + e_y^7 & q_y^3 + e_y^8 & q_y^4 + e_y^9 & q_y^5 + e_y^{10} \end{bmatrix} \\
 &= \begin{bmatrix} q_x^1 - e_x^6 & q_x^2 - e_x^7 & q_x^3 - e_x^8 & q_x^4 - e_x^9 & q_x^5 - e_x^{10} \\ q_y^1 - e_y^6 & q_y^2 - e_y^7 & q_y^3 - e_y^8 & q_y^4 - e_y^9 & q_y^5 - e_y^{10} \end{bmatrix}
 \end{aligned} \tag{13}$$

3 EXPERIMENTS

In this section, we assess the effectiveness of our suggested methodology using real data. The subsequent text provides an overview of the experimental setting. We conducted all experiments utilizing parallel computation on a computer equipped with an Intel Core i5-9400F CPU operating at 2.9 GHz in MATLAB. The experimental data was evaluated using the public RGB-D TUM datasets *freiburg1_desk*. In this evaluation, we evaluate the relationship of the estimation accuracy of the Essential matrix with the hypervolume using real data. The input points used in this evaluation were extracted as ORB feature points (Rublee et al., 2011) using the *detectORBFeatures* function of the Computer Vision Toolbox, and the matching feature points used to estimate the essential matrix were obtained using the *matchFeatures* function.

3.1 Real Data

To evaluate our proposed approach to real-world images, we used the TUM-RGBD dataset. This data provides the RGB and depth image data and the ground truth trajectory data for evaluating the Visual Odometry and Visual SLAM systems. All data is at a full frame rate of 30Hz and the camera sensor, a Microsoft Kinect sensor, has a resolution of 640x480. Next section we demonstrate the effectiveness of our proposed optimal feature selection using the hypervolume in the following experiments.

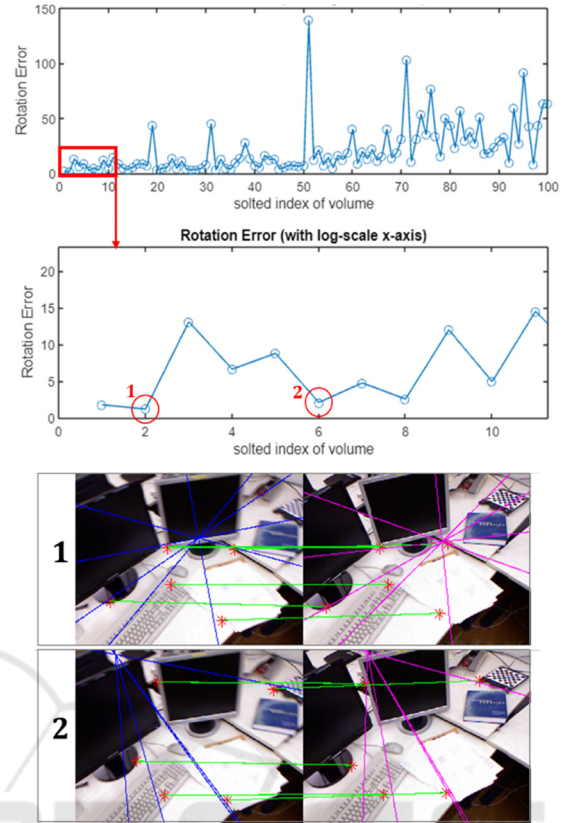


Figure 4: (Top) Graph showing rotation error and correlation when hypervolume is small. (Bottom) Distribution of selected 5 pairs of corresponding feature points with small rotation error among cases with small hypervolume.

3.2 The Effect of Hypervolume Based Optimal Feature Selection

After completing Step 4 of Algorithm 1, the rotation matrix calculated from the hypervolume and essential matrix was evaluated by analyzing the correlation using the error with the ground truth. The data was accumulated and analyzed through 100 iterations. Fig. 4(Top) is a graph showing the relationship between rotation error and hypervolume. For the rotation error, the essential matrix estimated by the five-point algorithm and the rotation matrix between the two images were obtained from the ground truth trajectory provided by the TUM RGB-D dataset. The x-axis of the graph is represented by an index based on the size of the hypervolume, and the y-axis is the rotation error. The general trend is that the larger the hypervolume, the larger the rotation error. First, let's look at the 10 data with the smallest hypervolume. Among them, No. 1 and No. 2 are the cases with the

smallest rotation error, and the selected matching feature point set is shown in Fig. 4(Bottom).

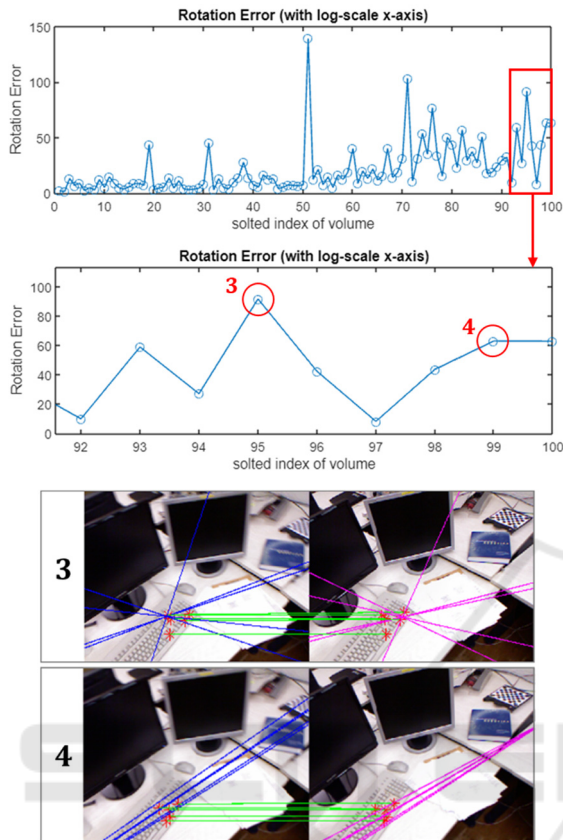


Figure 5: (Top) Graph showing rotation error and correlation when hypervolume is large. (Bottom) Distribution of selected 5 pairs of corresponding feature points in the case with large rotation error among the cases with large hypervolume.

We can see that the five pairs of corresponding feature points are evenly distributed among each other. This is the same result we found experimentally in Section 3. The second case is when the hypervolume is large and the rotation error is the largest. As shown in Fig. 5, the five pairs of corresponding feature points have a clustered distribution. This confirms the correlation of different rotation errors with hypervolume size and proves the validity of our proposed approach.

3.3 Threshold for Hypervolume Selection

When estimating the essential matrix using our proposed method, we used multiple iterations to select a set of 5 pairs of corresponding feature points when the hypervolume is less than a threshold K

value. The criterion for selecting K depends on the number of matching points, which we found through experimentation. In our case, if the number of matching points is 150 or less, we selected a set of 5 pairs of corresponding feature points when the value of logscale applied to the hypervolume is less than -32, and if the number of matching points is more than 150, we set $K > -36$.

3.4 Comparison with RANSAC Algorithm Using KITTI Odometry Benchmark Dataset

The following evaluates the performance of the proposed method by comparing the Absolute Pose Error (APE) with the commonly used RANSAC algorithm using the KITTI Odometry Benchmark dataset (Geiger et al., 2012). In the previous section, it is the same as the environment that evaluates the correlation between Uncertainty Hypervolume and pose estimation using the TUM dataset. The algorithm for calculating VO used *mono_vo*, and the feature point extraction method used the FAST algorithm. The feature point matching method was performed using a KLT tracker and a 5-point Algorithm for motion estimation. The Scale Factor was extracted and used from the Ground Truth provided by the KITTI dataset. The method of selecting the optimal feature point set with the proposed Uncertainty Hypervolume method is the same as Algorithm 2.

The difference from Algorithm 1 is that posture estimation cannot be performed because the optimal set of feature points cannot be selected depending on the threshold value during the VO process. Therefore, in the process of selecting the optimal set of feature points, as in Algorithm 2, the Uncertainty Hypervolume measurement index value calculated by several extracts is sorted, and then VO is used as input data for estimation by selecting the lower 10% set with a small value. In addition, in this dataset experiment, the Uncertainty Hypervolume measurement index using the Rotation Matrix obtained using it as Roll, Pitch, Yaw, and its size as an indicator performed better than the Uncertainty Hypervolume measurement index using Algorithm 1. The following are the results of evaluating the KITTI Odometry Benchmark 00, 02, 03, 04, and 05 sequences.

The results of the 00 sequence are shown in Fig. 6. Although both results have large errors due to drift and estimation errors by scale factor, it can be seen that APE's RMSE and Mean are relatively less

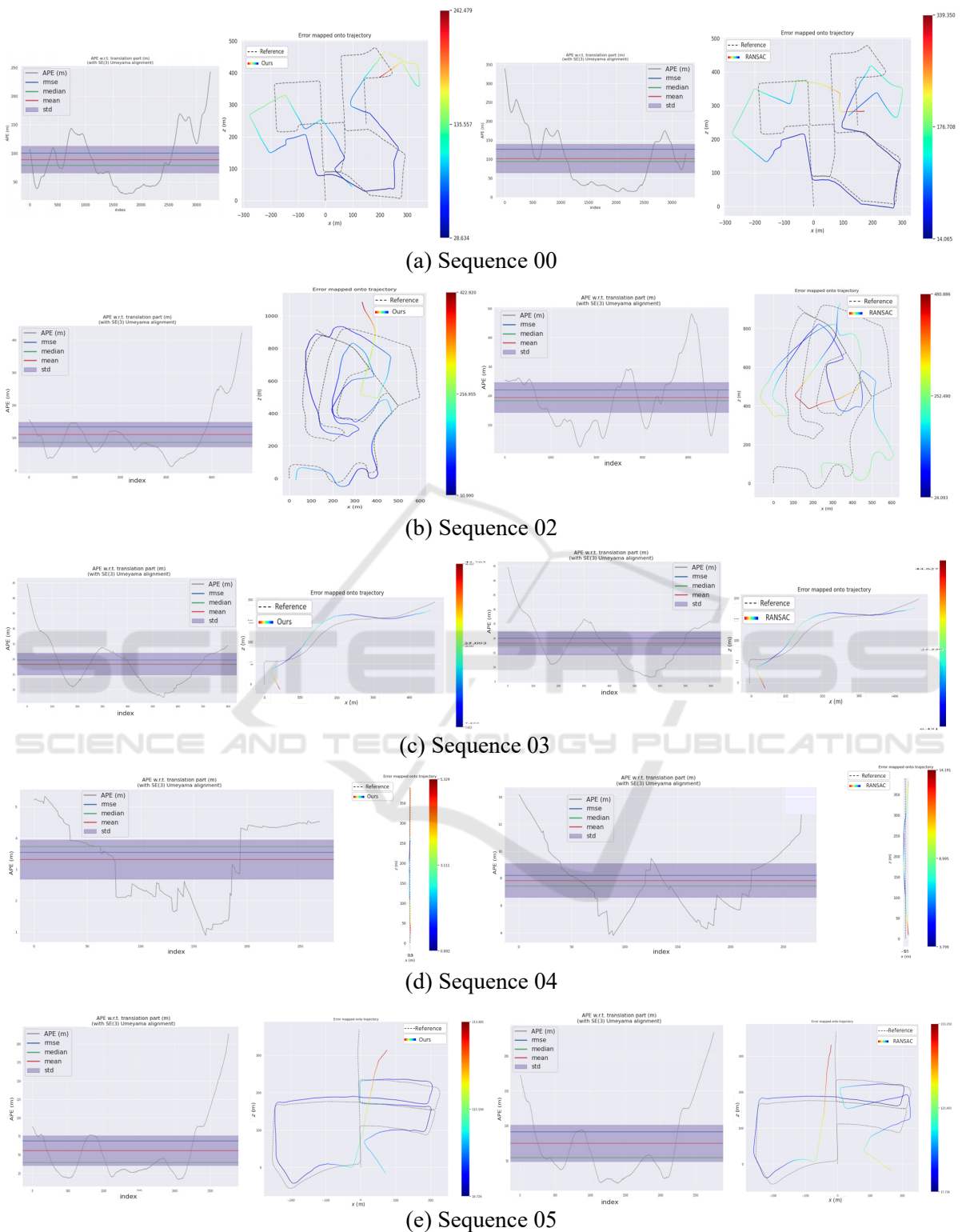


Figure 6: The KITTI Odometry Benchmark dataset (00, 02, 03, 04, 05) APE Results. Each sequence includes a graph of the translation part APE, RMSE, Median, Mean, and Std, and a graph of the Error Mapped ontology. The left is when the proposed Uncertified Hypervolume method is applied, and the right is the RANSAC method. It can be seen that the overall proposed method has a smaller error than the case of RANSAC. Performance comparisons are summarized in Table 1.

in the proposed method compared to RANSAC using all feature points. In VO, the sensitivity to rotation is high, so when calculating the Uncertainty Hypervolume, it is better to use it to convert it to Rotation Matrix, convert it to Roll, Pitch, and Yaw, and then use the difference from the reference Rotation Matrix as Uncertainty Hypervolume. In the simplified part of the Uncertainty Hypervolume calculation mentioned above, the measurement indicators obtained using the Rotation Matrix instead of the Essential Matrix are used. On the 02 sequence, the same method as the 00 sequence shows slightly better results. Next, for the 03 sequence, the proposed method for RMSE shows better results, but for Mean, RANSAC has a slight advantage. Next, it can be seen from the graph that the results of the proposed method for both the 04 and 05 sequences clearly show good performance.

Table 1: Comparison of the Uncertainty Hypervolume Method with the RANSAC Method.

Sequence	Uncertainty Hypervolume Method APE (m)		RANSAC Method APE (m)	
	RMSE	Mean	RMSE	Mean
00	100.66	89.07	126.85	102.07
02	133.74	109.84	220.29	194.16
03	19.73	18.35	19.98	18.27
04	3.54	3.31	8.19	7.81
05	68.93	55.66	91.53	74.92

4 CONCLUSIONS

This study proposes the selection of optimal feature points based on Uncertainty Hypervolume, a new approach for estimating the Essential Matrix for visual odometry. Through the pioneers' previous research and simulation experiments, it was found that uncertainty due to various errors in the selected corresponding feature point pair affects posture estimation, and that better performance can be obtained in VO if a set of feature point pairs well distributed in space is selected without clumping or forming lines. Based on this, the uncertainty in VO estimation is quantified by Uncertainty Hypervolume, a new measurement index that considers the error of the selected corresponding feature point pair and the distribution they form. Using actual data, it was confirmed that selecting a feature point set with a small measurement index had a smaller error with the Ground Truth value. The proposed method can work

effectively even though there are many features extracted with large errors due to low visibility or bad weather conditions. It can provide robustness in VO because only features with high quality are collected and used for VO estimation. The future work includes the optimization of algorithms to improve the computational efficiency while maintaining the performance advantage of the proposed approach over other conventional methods that rely on inlier feature points and RANSAC.

Algorithm 2: Uncertainty hypervolume-based optimal feature selection for visual odometry.

Data: L correspondence feature point sets

Result: A set of five-feature point pairs from the bottom 10% with low Uncertainty Hypervolume

Step 1. Randomly select K sets of five-feature point sets, $\{(p_i, q_i), i = 1, \dots, 5\}$ from the feature pairs detected from the images of two camera views subject to VO. Incorporating a bucketing approach into this step to improve the initial selection of the five feature sets provides a way to select well-distributed feature sets with few attempts.

Step 2. Generate the corresponding coefficient vectors, $\{v_i, i = 1, \dots, 5\}$. Then, compute the essential matrix using the five-point algorithm with $\{v_i, i = 1, \dots, 5\}$ as input.

Step 3. Generate 32 pairs of $v'_i = (p'_i, q'_i), \{i = 1, \dots, 32\}$ taking into account the error of the corresponding feature points to $\{v_i, i = 1, \dots, 5\}$. With this, the 32-vertices Essential matrix is computed. And, the rotation matrix is estimated using the essential matrix that constitutes the vertex. The matrix is then converted to Roll, Pitch, and Yaw. The difference between this and the reference rotation, obtained with the reference essential matrix, is taken as the uncertainty hypervolume.

Step 4. Select a set of five feature point pairs from the bottom 10% with low Uncertainty Hypervolume. The VO is estimated using the best set of five selected feature point pairs.

ACKNOWLEDGMENTS

This research was funded, in part, by the “Intelligent Manufacturing Solution under Edge-Brain Framework” (Grant 2022-0-00067 and IITP-2022-0-00187) project, in part, by the Artificial Intelligence Graduate School Program, Grant No. 2019-0-00421, and by ICT

Consilience Program, IITP-2020-0-01821, of the Institute of Information and Communication Technology Planning & Evaluation (IITP), sponsored by the Korean Ministry of Science and Information Technology (MSIT).

REFERENCES

- D. Nistér, N. Oleg and B. James, "Visual odometry." *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. Vol. 1. IEEE, pp I-I, 2004.
- S. Poddar, R. Kottath, and V.Karar, "Motion Estimation Made Easy: Evolution and Trends in Visual Odometry" . In *Recent Advances in Computer Vision* (pp. 305-331). Springer, Cham, 2019.
- M.A. Fischler and C.B. Robert, "Random sample consensus: an automated cartography." *Communications of the ACM* 24.6 (1981): 381-395.
- Igor Cvišić and Ivan Petrović, "Stereo odometry based on careful feature selection and tracking," in *Proc. Eur. Conf. Mobile Robots (ECMR)*, Sep. 2015.
- Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," *Artificial Intelligence*, vol. 78, no.1-2, pp. 87-119, 1995.
- B. Kitt, A. Geiger, H. Lategahn, "Visual odometry based on stereo image sequences with RANSAC-based outlier rejection scheme," in *Intelligent Vehicles Symposium, 2010*, pp. 486-492
- H. H. Nguyen, and S. Lee, Orthogonality Index Based Optimal Feature Selection for Visual Odometry. *IEEE Access*, 7, 6228 4-62299, 2019.
- R. Deriche, Z. Zhang, Q.-T. Luong and O. Faugeras, Robust recovery of the epipolar geometry for an uncalibrated stereo rig, in: *Proceedings Third European Conference on Computer Vision I, Stockholm, Sweden (1994)* 567-576.
- Nister, D. "An Efficient Solution to the Five-Point Relative Pose Problem." *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Volume 26, Issue 6, June 2004.
- Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). ORB: An efficient alternative to SIFT or SURF. *2011 International Conference on Computer Vision*, 2564-2571.
- Geiger, A., Lenz, P., & Urtasun, R. (2012). Are we ready for autonomous driving? The KITTI vision benchmark suite. *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3354-3361.