Received: 24 July 2021 | Revised: 16 April 2022 | Accepted: 19 June 2022

DOI: 10.33012/navi.542

ORIGINAL ARTICLE

Measurement Error Detection for Stereo Visual Odometry Integrity

Yuanwen Fu | **Shizhuang Wang** | **Yawei Zhai** | **Xingqun Zhan** | **Xin Zhang**

School of Aeronautics and Astronautics, Shanghai Jiao Tong University, Shanghai, China

Correspondence
Xin Zhang
School of Aeronautics and Astronautics, Shanghai Jiao Tong University, Shanghai, China
Email: xin.zhang@sjtu.edu.cn

Abstract

Integrity, a safety-of-life framework from civil aviation for satellite navigation, is greatly under-explored in visual navigation. A new two-factor approach to rejecting measurement outliers is proposed for navigation integrity in stereo visual odometry (VO). In contrast to other treatments using reprojection error as measurement residuals, our choice of landmark matching error inherently connects navigation solutions and integrity monitoring. We propose two methods to detect large measurement residuals that cannot otherwise be identified by existing outlier rejection methods in state-of-the-art VO pipelines. By rejecting these outliers, measurement residuals can be bounded by the distribution overbounding method that provides fundamental inputs for integrity computations. We evaluate our methods using an open-source data set. Overbounding performance is improved in terms of tightness, computational efficiency, and most important of all, scenario tolerance. This could be a good starting point for developing future integrity monitoring algorithms for visual navigation and in particular, stereo VO.

Keywords

integrity, landmark matching error, measurement error, outlier rejection, overbounding, visual odometry

1 | INTRODUCTION

Future autonomous systems are expected to bring significant convenience to people. To ensure operational safety, navigation systems must continuously provide positioning solutions with high integrity while also meeting accuracy requirements. In addition, because those systems are applied over various scenarios, the navigation systems also must provide high robustness against environmental changes. Here, the term *scenario* refers to the environment around the camera (such as various light conditions, weather, and moving objects), and it does not include the subject (such as a vehicle) mounting the camera. Therefore, providing high integrity and robustness for future autonomous systems across various scenarios is one of our key research challenges. Integrity is a quantifiable performance metric used to set certifiable requirements on an individual subsystem to ensure a level of safety for the overall system (Kelly & Davis, 1994). It is a key metric

that measures navigation safety for safety-critical applications (Blanch et al., 2007; Brown, 1992; Zhai et al., 2018, 2020). This concept was originally introduced in aviation to measure the trust of the navigation information. It also includes the ability of the system to provide timely alerts to users when results from the navigation subsystem cannot be trusted.

The particular interest of this paper is stereo visual odometry (VO), which has been identified as one of the main navigation sensors to support safety-critical autonomous systems. It aims to estimate the ego-motion of a camera by identifying the projected movement of landmarks in consecutive frames. Typical VO pipelines include feature-based approaches, direct approaches, and hybrids of feature-based and direct approaches (Jiang et al., 2013; Naroditsky et al., 2012; Scaramuzza & Siegwart, 2008). We focus on the Oriented FAST and Rotated BRIEF (ORB; Rublee et al., 2011) feature-based approach due to its prevalence. ORB is an efficient feature-extraction approach based on features from accelerated segment testing (FAST; Rosten & Drummond, 2006) corner point detection and the binary robust independent elementary features (BRIEF; Calonder et al., 2010) descriptor. The workflow of this feature-based approach is presented in Figure 1 (Cumani, 2011; Mur-Artal & Tardos, 2017) with six primary steps.

The first step is preprocessing. More specifically, it is to calibrate the camera's distortion parameters and internal parameters. The second step is feature extraction. The term *feature* here means a distinctive pixel. There are many feature-extraction methods, including Harris and Stephens' work (1988), ORB, Bay et al.'s (2006) speeded up robust features (SURF), and distinctive image features from Lowe's (2004) scale-invariant keypoints (SIFT). The third step is associating features extracted from consecutive frames, for which there is a mismatch limit check. The fourth step is recovering the depth information of features, which has been lost in the mapping from landmarks to features. Such methods include the sum of absolute difference (SAD; Szeliski, 2010), semi-global block matching (SGBM; Hirschmüller, 2008) and graph cut (GC; Li & Chen, 2004). The fifth step is outlier rejection, and its main method is by random sample consensus (RANSAC; Fischler & Bolles, 1981). The sixth step is to estimate the pose of camera, for which we utilize the methods of bundle adjustment and singular value decomposition (SVD). Among these steps, there are multiple tests to reject landmark pairs with large matching residuals, which are denoted as *checks* here.

Although there are many existing works on improving VO accuracy and robustness (Nistér et al., 2004; Scaramuzza & Fraundorfer, 2011), there are very few studies on VO integrity. Among the small amount of relevant research (Li & Waslander, 2019; Wang et al., 2020a; Zhu et al., 2020), all use the Gaussian distribution to describe visual measurement errors without giving reasonable proof. It is worth noting that Zhu et al. (2019) strictly derived the geometric error model from a

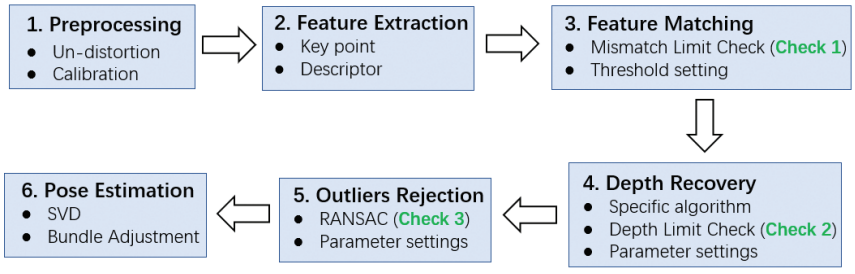

FIGURE 1 A block diagram of our feature-based approach

chessboard-like visual feature. The researchers define 2D features to act as visual measurement. The feature point is the projection of the landmark on the image plane, which loses depth information. Thus, the researchers fail to faithfully characterize *landmark matching errors* (later defined in Equation [2]), let alone across diverse scenarios. In contrast, we define a 3D landmark matching pair as the visual measurement and the difference between pair members as residual, for the simple reason that the rotation matrix and translation vector are computed by landmark matching pairs in stereo VO. Overbounding such visual measurement errors will be the first step in developing fault detection and exclusion (FDE) algorithms, just as those found in advanced receiver autonomous integrity monitoring (ARAIM) for civil aviation (Walter et al., 2019) are currently the starting point for future visual navigation integrity frameworks.

The main contributions are as follows. Firstly, the definition of the stereo VO measurement residual, in the form of landmark matching error, is proposed. Secondly, we propose two real-time detection methods against large errors that otherwise cannot be identified by the current VO's outlier rejection methods. In particular, we propose two methods to check feature distinctiveness and motion constraints. Thirdly, we evaluate our method in scenarios of various natures and show that, by using the proposed methods, the overbound becomes much more effective (tighter and requiring less computation) and scenario tolerant, which could be a good reference for developing future integrity monitoring algorithms for stereo VO.

In the rest of this paper, we first formulate the problem by discussing the conventional outlier rejection methods and current overbounding challenges in Section 2. We describe our proposed detection methods for large errors in Section 3. Extensive experiments are conducted in Section 4, and we conclude the paper in Section 5.

2 | PROBLEM FORMULATION

In this section, our choice of measurement model is, first, justified. Conventional approaches to outlier rejection are then investigated. This includes three checks (i.e., landmark mismatch thresholding, depth thresholding, and random sample consensus [RANSAC]). Then, we show that some large measurement errors cannot be removed by using these conventional checks using real examples. The discovery of the correspondence between these errors and their sources sets the stage for the two proposed checks in Section 3.

2.1 | Measurement Model

Several potential measurement models are readily available that allow us to leverage prior work in ARAIM when developing integrity concepts and methods. For the specific problem of VO, two kinds of relative measurement models exist: reprojection error and landmark matching error. These are, respectively, defined as:

$$p_k - \pi\left(T_{k,k-1} \cdot P_{k-1}\right) \tag{1}$$

and:

$$P_k - T_{k,k-1} \cdot P_{k-1} \tag{2}$$

Here, a 3D point P_{k-1} in world frame at epoch $k-1$ is projected onto the imaging plane and output as expected pixel coordinates $\pi\left(T_{k,k-1}\cdot P_{k-1}\right)$ at epoch k. This is subtracted from measurement p_k of the same 3D point at epoch k to obtain the reprojection error. $T_{k,k-1}$ is the relative pose between these two epochs. Landmark matching error is defined as the difference between the same 3D point at epoch k and its expected value $T_{k,k-1}\cdot P_{k-1}$.

Our choice of landmark matching error is justified by the fact that it is closer to the navigation end solution than the reprojection error in the stereo VO pipeline. Instead of using a reprojection error method which may include erroneous camera matrix or distortion models, we reformulated landmark matching error into the measurement residual used throughout this work. In contrast to the measurement models found in other state-of-the-art treatments on VO integrity, our choice inherently connects the navigation solution and integrity monitoring by recognizing that, unlike 2D features, 3D landmarks preserve scale information, which is the major advantage of stereo VO over monocular VO.

2.2 | Conventional Approaches to Outlier Rejection in VO

In this section, we will look into the conventional checks in state-of-the-art VO pipelines in detail. As shown in Figure 1, there are three checks in the VO workflow.

2.2.1 | Mismatch Thresholding

In the ORB feature method, the strategy of brute force matching is adopted in the feature matching step. Specifically, there is feature set $\mathcal{P}=\left\{p_1,p_2,\cdots,p_n\right\}$ and feature set $\mathcal{Q}=\left\{q_1,q_2,\cdots,q_m\right\}$, which represent the features extracted from the current frame and the previous frame, respectively. We take a random feature p_i from the feature set $\mathcal{P}$ and find a feature q_j in the feature set $\mathcal{Q}$, which has the smallest Hamming distance (Szeliski, 2010) from p_i compared to the rest of the features in the feature set $\mathcal{Q}$.

The Hamming distance between feature p_i and feature q_j is constrained in Check 1 as:

$$\left\{(X,Y,Z)\middle|distance\left(p_i,q_j\right)\le Threshold\,1\right\} \tag{3}$$

where:

$$Threshold\,1=max\left(30,\,2*mindist\right) \tag{4}$$

When the Hamming distance is larger than the *Threshold* 1 computed by Equation (4), the match $\left(p_i,q_j\right)$ is rejected. In Equation (4), *mindist* refers to the minimum value among all the distances of matched feature point pairs. The essence of Check 1 is to exclude mismatch events through the distinctiveness of feature. Moreover, please note that *Threshold* 1 is not a fixed value; rather, it is adaptive to input frames. As shown in Equation (4), *Threshold* 1 is equal to the

maximum value of 30 and two times the mindist (Gao & Zhang, 2021). The minimum value of Threshold 1 (i.e., 30 pixels), corresponds to a medium resolution (e.g., 720×480 pixels) image, decent camera motions, and mediocre lighting conditions. Tougher scenarios should expect a larger mindist, and therefore a larger Threshold 1.

2.2.2 | *Depth Thresholding*

The landmark coordinates, (X, Y, Z), in the camera frame can be calculated by Equation (5) and Equation (6). The feature (u, v) is the projection of the landmark (X, Y, Z). In Equation (5), u_l and u_r represent the abscissas of feature in the left and right camera pixel planes, respectively, and d represents the disparity. In Equation (6), f_x and f_y are the focal lengths; c_x and c_y are the coordinates of the optical center's projection onto the image; and b is the baseline between the left and right cameras. The parameters including f_x, f_y, c_x, c_y, and b are only related to the camera, itself, and thus termed *intrinsic* parameters.

$$d = u_l - u_r \tag{5}$$

$$\begin{cases} Z = f_x\, b/d \\ X = \left(u - c_x\right)/f_x \cdot Z \\ Y = \left(v - c_y\right)/f_y \cdot Z \end{cases} \tag{6}$$

In order to show the camera projection and disparity in an intuitive way, we designed Figure 2(a) and 2(b), respectively. Figure 2(a) shows the projection model with one camera and Figure 2(b) shows the disparity between two camera views.

The key to computing the landmark is to find the correspondence between the left and right pixels. This is a challenging task because there are many potential distractions including: optical distortion and noises (brightness, hue, and saturation misalignment), specular reflection on a smooth surface, projection reduction,

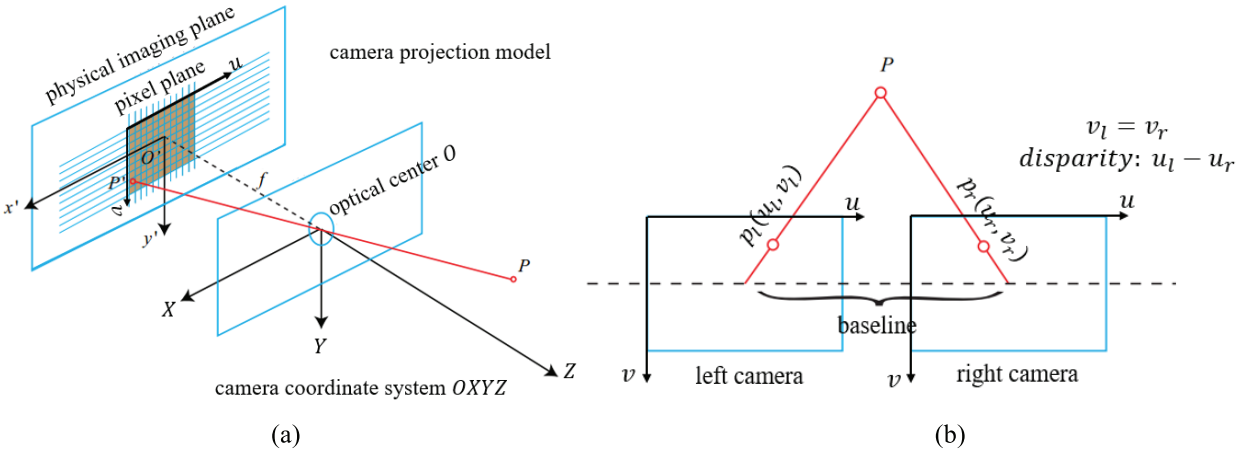

FIGURE 2 The pinhole camera model and parameters: (a) is the projection model for a single camera and the intrinsic parameters and (b) is a stereo camera with baseline and disparity.

perspective distortion, low or repeated textures, and transparent objects, as well as overlapping and discontinuity. The most common cases are object boundary and fine structure, in which it is difficult to find correspondence between the left and right pixels. Other practical problems arise from the difference or reflection in recording and lighting.

The SGBM algorithm strikes a good balance between efficiency and accuracy and is, therefore, examined in this paper. The Open-Source Computer Vision Library (OpenCV; Bradski, 2000) implemented the SGBM algorithm and encapsulated it into the cv::StereoSGBM class. However, we find that it has two shortcomings. Firstly, the calculated disparities often exceed the range that has been preset in the OpenCV function. Secondly, the same disparity disturbance at different depths leads to different depth errors. More specifically, larger depth is susceptible to larger error under the same disparity perturbation. Therefore, the disparity and depth are constrained by Check 2:

$$\left\{(X,Y,Z)\middle| d\in\left[d_{min},d_{max}\right],\ Z\le Threshold\ 2\right\} \tag{7}$$

where (X, Y, Z) is the landmark, Z refers to the depth between the landmark and the camera, d is the disparity that is depicted in Figure 2, and d_{min} and d_{max} are the lower and upper bounds of the disparity, respectively. Furthermore, d_{min} represents the starting point of the epipolar line search in the right image and d_{max} represents the maximum search boundary. These two parameters have little to do with the scenario, but are related to the size of the image. In this paper, d_{min} is set to 0 and d_{max} is set to 64. They remain the same for all 10 scenarios that are used to model visual measurement errors.

The landmark is abandoned when its computed disparity is not in the predefined range (that is, $d \notin [d_{min}, d_{max}]$). The landmark is also abandoned when its depth is larger than Threshold 2 (that is, $Z >$ Threshold 2). Threshold 2 is a parameter that should be given before the VO execution. Normally, it should be somewhere between 100 and 200 meters. Check 2 is a strong check when Threshold 2 is set close to 100, and it is a weak check when Threshold 2 is set close to 200. When objects in the scenario (such as Old Town) are far from the camera, we recommend using a weak check. With closer objects (such as the Hospital), using a strong check could be better. In general, Check 2 discards large depth error events with a low performance.

2.2.3 | *RANSAC*

The conventional Check 1 and Check 2 only help to exclude large mismatch events and large depth error events, but there are no sanity checks on moving object events. Therefore, a RANSAC step is inserted into VO, denoted as Check 3. A brief description of the RANSAC algorithm appears in Section 1 where it was first introduced (Fischler & Bolles, 1981). The basic idea of RANSAC is to use random sets of data points to fit models and use the rest of the data points to verify these models. Data points in the model with the highest consensus are selected to form the inlier set; the remaining data points are defined as outliers. In the context of stereo VO, the data point in question is the matched landmark pairs and the model is the transformation matrix T, which describes rotation R and translation t in three dimensions. The RANSAC algorithm is implemented in the following three steps.

First, we calculate the transformation matrix. The transformation matrix is computed with four landmark pairs following a direct linear transformation,

ignoring the internal constraints of the rotation matrix, which is formulated in Equation (8) as:

$$\begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 \\ X_2 & Y_2 & Z_2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & X_2 & Y_2 & Z_2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X_2 & Y_2 & Z_2 & 1 \\ X_3 & Y_3 & Z_3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & X_3 & Y_3 & Z_3 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X_3 & Y_3 & Z_3 & 1 \\ X_4 & Y_4 & Z_4 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & X_4 & Y_4 & Z_4 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X_4 & Y_4 & Z_4 & 1 \end{bmatrix} \cdot \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \\ r_6 \\ r_7 \\ r_8 \\ r_9 \\ t_1 \\ t_2 \\ t_3 \end{bmatrix} = \begin{bmatrix} X_1' \\ Y_1' \\ Z_1' \\ X_2' \\ Y_2' \\ Z_2' \\ X_3' \\ Y_3' \\ Z_3' \\ X_4' \\ Y_4' \\ Z_4' \end{bmatrix} \tag{8}$$

where points $P_i\left(X_i, Y_i, Z_i\right) i = 1, 2, 3, 4$ are the landmarks observed in the previous frame; points $Q_i\left(X_i^{'}, Y_i^{'}, Z_i^{'}\right) i = 1, 2, 3, 4$ are the landmarks in the current frame; $r_1 \sim r_9$ are the elements of the rotation matrix, R; and $t_1 \sim t_3$ are the elements of the translation vector t. Then, we project R solved by Equation (8) into the Euclidean orthogonal space using Equation (9):

$$R' = \left(RR^T\right)^{-\frac{1}{2}} R \tag{9}$$

Substitute the value of R' into the Equation (8), and we arrive at Equation (10):

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} = \begin{bmatrix} X_1' - r_1X_1 - r_2Y_1 - r_3Z_1 \\ Y_1' - r_4X_1 - r_5Y_1 - r_6Z_1 \\ Z_1' - r_7X_1 - r_8Y_1 - r_9Z_1 \\ X_2' - r_1X_2 - r_2Y_2 - r_3Z_2 \\ Y_2' - r_4X_2 - r_5Y_2 - r_6Z_2 \\ Z_2' - r_7X_2 - r_8Y_2 - r_9Z_2 \\ X_3' - r_1X_3 - r_2Y_3 - r_3Z_3 \\ Y_3' - r_4X_3 - r_5Y_3 - r_6Z_3 \\ Z_3' - r_7X_3 - r_8Y_3 - r_9Z_3 \\ X_4' - r_1X_4 - r_2Y_4 - r_3Z_4 \\ Y_4' - r_4X_4 - r_5Y_4 - r_6Z_4 \\ Z_4' - r_7X_4 - r_8Y_4 - r_9Z_4 \end{bmatrix} \tag{10}$$

The updated translation vector can be obtained by solving Equation (10) with any least-squares method.

Next, we use this model and a threshold (named Threshold 3) to determine whether a landmark matching pair is an inlier or outlier. To be more specific, we first calculate the landmark matching error, *err*, according to Equation (11). The resultant *err* value is then compared against Threshold 3. If the *err* value is smaller than the threshold, the landmark matching pair is claimed an inlier; otherwise, it is an outlier.

$$err = \left\| q_i - \left(R' p_i + t' \right) \right\| \tag{11}$$

The above two steps are repeated until an optimal inlier set is found, which refers to the inlier set with the largest number of inliers.

There are two important parameters in the RANSAC step, which are the number of iterations and Threshold 3. The more iterations we have, the more likely we are to identify inliers from outliers correctly. A general rule is that noisy scenarios (bad weather, low light conditions, etc.) necessitate more iterations. However, due to limited computing power, RANSAC iterations cannot be set infinitely large. In this paper, we set RANSAC iterations to be 500 for the 10 scenarios, since this value strikes a balance between accuracy and efficiency according to our previous work (Fu et al., 2020). Threshold 3 is used to determine whether a landmark matching pair is an inlier or outlier. Therefore, it is related to the statistical properties of inliers, and has nothing to do with the scenario. Different scenarios may have different outliers, but their inliers are similar in terms of a specific set of landmarks. Our previous work (Fu et al., 2020) shows that setting Threshold 3 to 1 meter admits great performance according to sensitivity analyses on integrity.

2.3 | Overbounding Measurement Error

Overbounding is a statistical concept that was developed in the early 2000s to deal with navigation algorithms that required modeling unknown error distributions. This technique arose because the common usage of Gaussian-distributed random variables was insufficient to adequately describe many error distributions that were vital to the utilization of new navigation technologies. Conceptually, overbounding provides a replacement statistical model of a random variable whose true probability distribution function is unknown. The probability of an error as computed by the overbound is always greater than or equal to the true probability. Let X be a random variable with cumulative distribution function (CDF), $F(x)$. Then, $B(x)$ is the overbound of X with Equation (12):

$$\begin{aligned} B(x) &\geq F(x), \ \forall F(x) < 0.5 \\ B(x) &\leq F(x), \ \forall F(x) > 0.5 \end{aligned} \tag{12}$$

We present the basic methodologies on developing error models of visual measurements using an open-source data set in this section. We explore the TartanAir data set (Wang et al., 2020b) because, compared to other data sets, it covers much more diverse scenarios (18 in total), each of which include two distinctive difficulty levels/modes (easy and hard). There are 48 sequences in total. The images are collected in urban, rural, natural, and indoor environments, so they represent various scenarios covering challenging viewpoints and diverse motion patterns subject to changes of light, weather, moving objects, etc.

The accuracy of VO depends directly on the landmark matching pairs. Therefore, the landmark matching errors are characterized in the following experiments. The landmark matching error is defined in Equation (13), which is the matrix form of Equation (2):

$$\begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} = \begin{bmatrix} X_2 \\ Y_2 \\ Z_2 \end{bmatrix} - \left(R_{21} \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \end{bmatrix} + t_{21} \right) \tag{13}$$

where $\left[X_2, Y_2, Z_2\right]^T$ and $\left[X_1, Y_1, Z_1\right]^T$ are the corresponding landmarks in the current frame and previous frame, respectively. R_{21} and t_{21} are the ground truth values of the rotation matrix and translation vector from the previous frame to the current frame, respectively. Finally, $\left[\Delta X, \Delta Y, \Delta Z\right]^T$ is the landmark matching error.

To characterize the landmark errors, we designed the following experiment with the TartanAir data set. The data is divided into two levels (easy and hard) in terms of motion patterns. We chose the hard mode because the images captured in this mode represent corner cases that may push current algorithms to their limits. These cases include, but are not limited to, moving objects—intensive and violent actions mixed with significant rolling and yaw motions. For each image, we acquire landmark matching pairs and conduct the three aforementioned conventional checks in an attempt to reject the faulty landmark matching pairs, followed by overbounding.

We employed two example scenarios to show the inability of conventional checks in terms of overbounding. Test shots of the two scenarios are shown in Figure 3, where the left and right pictures refer to data sequences *Carwelding* and *Neighborhood*, respectively. For the Carwelding case, there are lots of moving objects (i.e., robot arms and frame structures for cars on the production line). Moreover, because of the strong light caused by the electric arc during welding, there are large illumination changes in the consecutive frames. These events significantly impact the landmark error. For the Neighborhood case, because of repetitive textures such as roof, road, and leaves, the VO front-end was expected to receive many mismatching events, which could also lead to dramatically large landmark matching errors. These two data sequences cover all the negative factors that would be encountered in the real world and, therefore, we use their landmark matching results to find the shortcomings of conventional checks.

Figure 4 presents the error analysis results for two example scenarios. The three figures in the left column correspond to the errors in x, y, and z directions for the Carwelding Scenario, whereas the right three are for the Neighborhood Scenario. The error profiles are presented in terms of folded CDF plots. This is a very mature method in GPS error overbounding, and readers are referred to previous works along this line (Decleene, 2000; Larson, 2018; Rife et al., 2004, 2006; Wang et al., 2021) for a more detailed explanation of this principle. In each figure, the blue dotted curve corresponds to the folded CDF of the true data. The green curve represents a normal distribution with a mean and standard deviation of the distribution.

FIGURE 3 Two example scenarios for preliminary analysis: Carwelding (left) and Neighborhood (right)

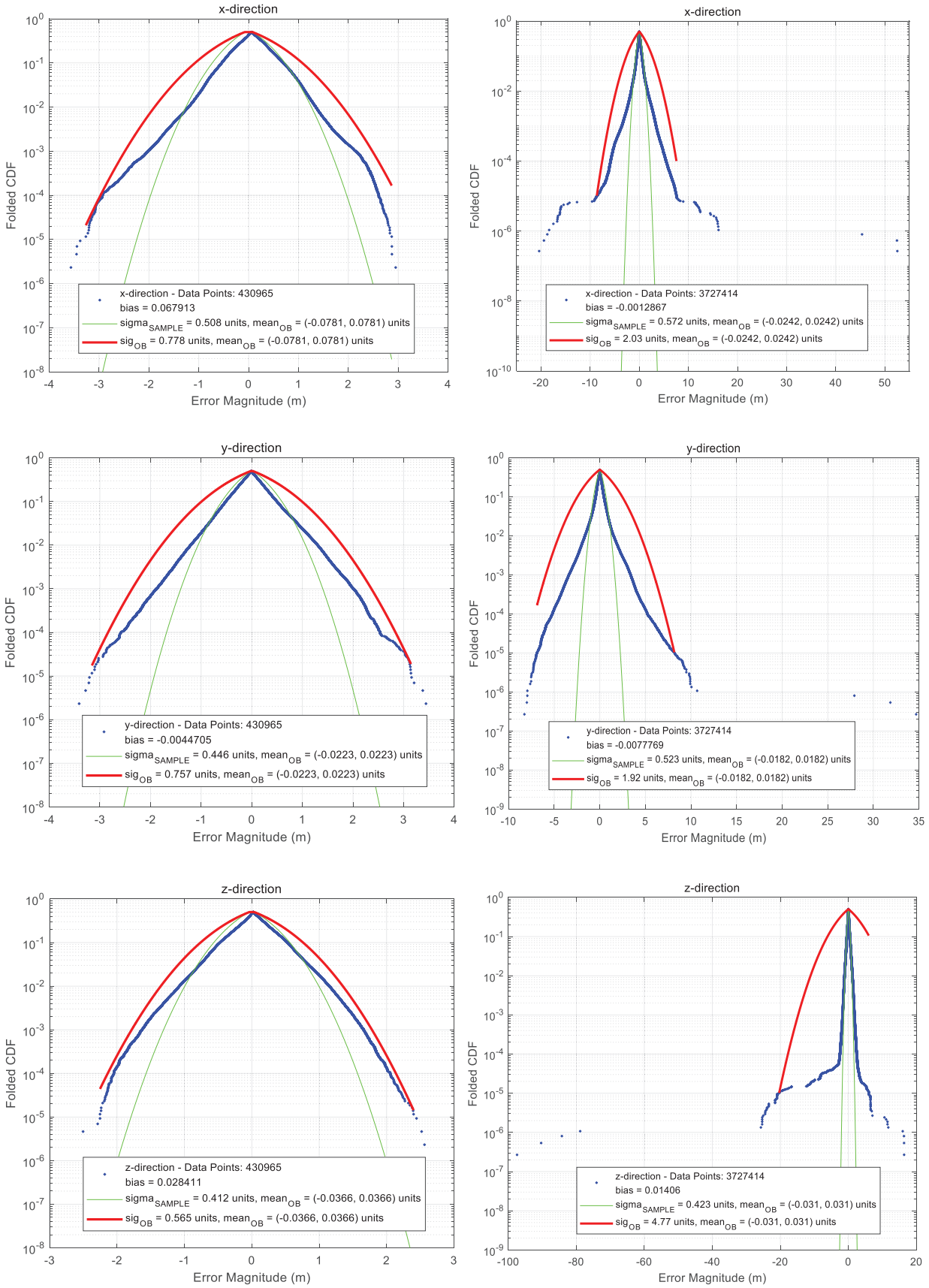

FIGURE 4 Folded cumulative density function (CDF) results of the landmark matching errors in *x*, *y*, and *z* directions (top to bottom) for the two example scenarios: Carwelding (left) and Neighborhood (right)

TABLE 1
Three-Axis Overbounding Standard Deviations of the Landmark Matching Errors

Scenarios	ΔX	ΔY	ΔZ
Carwelding	0.778 m	0.757 m	0.565 m
Neighborhood	2.03 m	1.92 m	4.77 m

The red envelope is the minimal overbounding folded CDF, which is further addressed in the following section. The results are summarized in Table 1.

It can be seen from Figure 4 and Table 1 that the overbounding sigmas of both scenarios in each axial direction are much larger than the expected values, which is especially remarkable for the Neighborhood case. Although there are three conventional checks to reject outliers, there are still a considerable number of faulty landmark matching pairs in the final measurement set due to the fact that landmarks are susceptible to exceedingly large errors and low efficiency of the three checks. In addition, the Neighborhood Scenario is subject to a significantly larger overbounding sigma than the Carwelding Scenario, which is caused due to many mismatch events and large depth error events.

2.4 | Preliminary Analysis

In order to further figure out the different behaviors of the overboundings of landmark matching errors in two different scenarios, we further explore the distributions of the landmark matching errors of these two scenarios. We ordered the scenarios in descending order according to magnitude $\sqrt{(\Delta X)^2+(\Delta Y)^2+(\Delta Z)^2}$ and drew the curves in Figure 5 accordingly. Within each subplot, the horizontal axis indexes the ordered measurements while the vertical axis corresponds to the log_{10} (magnitude) of the landmark matching errors in meters.

Since the two scenarios contained a different number of images, their abscissa ranges were different. It can be noted that the error range in the Neighborhood Scenario spans much wider than that of the Carwelding Scenario. Meanwhile, there were a few unusually large residuals in the Neighborhood Scenario.

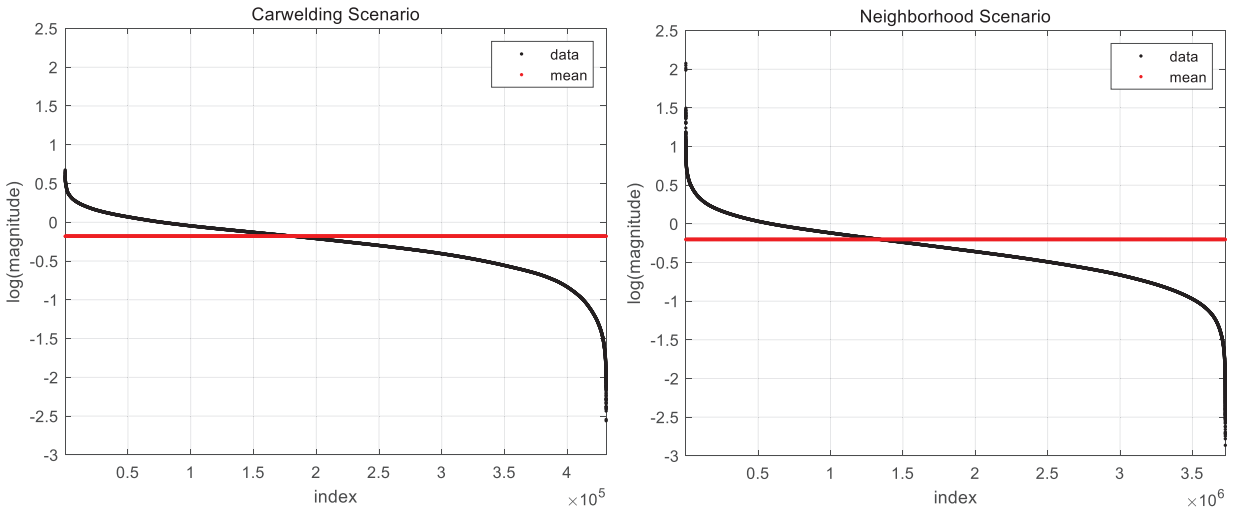

FIGURE 5 The magnitudes of landmark matching errors of the Carwelding Scenario (left) and the Neighborhood Scenario (right)

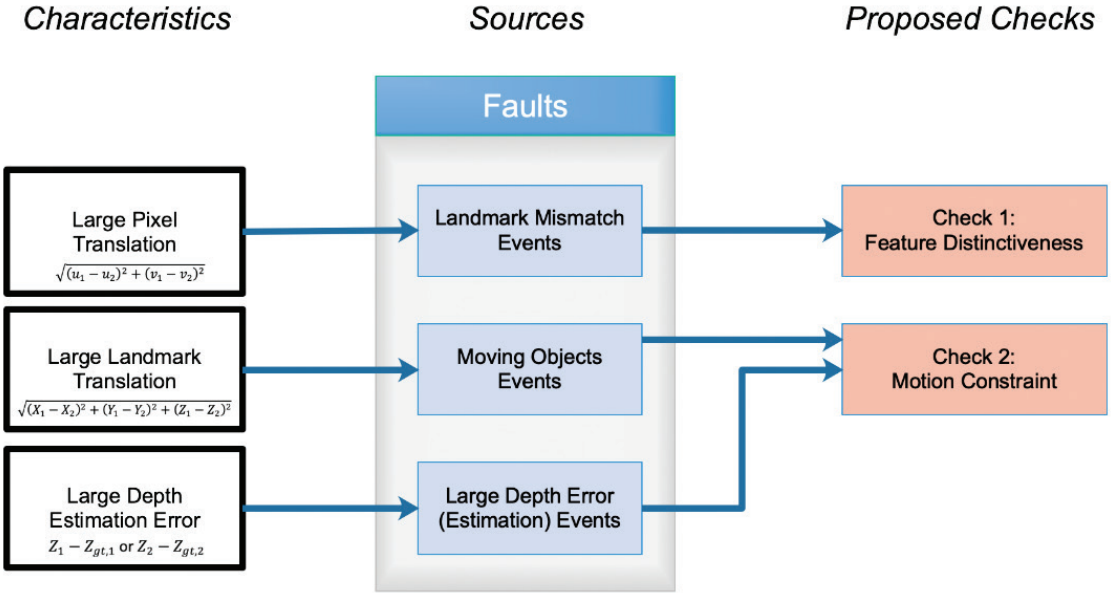

FIGURE 6 Error characteristics, outlier sources, and their proposed checks

Suppose we obtained the attributes associated with landmark matching error, including:

$$u_1, v_1, X_1, Y_1, Z_1 u_2, v_2, X_2, Y_2, Z_2, Z_{gt,1}, Z_{gt,2}, \Delta X, \Delta Y, \Delta Z$$

where u_1, v_1 represents the feature point in the previous frame; X_1, Y_1, Z_1 represents the landmark in the previous frame; u_2, v_2 represents the feature point in the current frame; X_2, Y_2, Z_2 represents the landmark in the current frame; $Z_{gt,1}$ represents the ground truth depth of landmarks in the previous frame; $Z_{gt,2}$ represents the ground-truth depth of landmark in the current frame; and $\Delta X, \Delta Y, \Delta Z$ represents the landmark matching error computed by Equation (13). We find that data with large matching errors share certain common characteristics, namely:

- large depth estimation errors (i.e., large $Z_1 - Z_{gt,1}$ and $Z_2 - Z_{gt,2}$);
- large displacement of features/pixels, $\sqrt{(u_1-u_2)^2+(v_1-v_2)^2}$; and
- large displacement or translation of landmarks, $\sqrt{(X_1-X_2)^2+(Y_1-Y_2)^2+(Z_1-Z_2)^2}$.

These properties are mainly attributed to feature mismatch events, landmark movement events, and large depth error events, which are defined as faults in Figure 6. They are rarely occurring unknown deterministic errors that cannot be modeled by Gaussian white noise (Hafez et al., 2020). Conventional checks fail to identify remnant large measurement errors due to these faults across scenarios of different natures, leading to a limitation of existing CDF overbounding methods.

3 | THE PROPOSED METHOD

We, thus, propose two checks to reject the faults according to their causes shown in Figure 6 and insert them into the existing VO workflow in Figure 7.

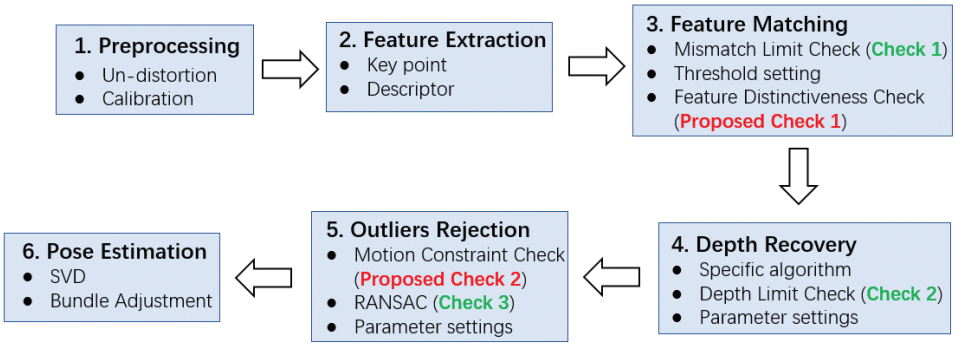

FIGURE 7 Proposed checks in the VO workflow updated from Figure 1

Specifically, a feature distinctiveness check was inserted into the Feature Matching step to address the large displacement of features and a motion constraint check was inserted into the Outlier Rejection step to address large depth estimation errors and large displacement of landmarks.

3.1 | Proposed Check 1: Feature Distinctiveness Check

Proposed Check 1 was added to reject mismatch events in the Feature Matching step. Feature Matching is based on descriptors, which are one of the attributes of each feature and are usually considered to be a unique identity of the feature. The BRIEF descriptor is computed in Equation (14; Calonder et al., 2010):

$$descriptor[i] = \begin{cases} 1, \text{ if } I(x_{i1}) > I(x_{i2}) \\ 0, \text{ if } I(x_{i1}) \le I(x_{i2}) \end{cases} \quad i \text{ from } 0 \text{ to } 127 \tag{14}$$

where *descriptor* [*i*] is the *i*-th bit of the 128-dimensional descriptor; x_{i1} and x_{i2} are the pixels around the keypoints x_1 and x_2, respectively; $I(\cdot)$ denotes the intensity function; and pixel sets $\{x_{11}, x_{21}, x_{31}, \cdots, x_{n1}\}$ and $\{x_{12}, x_{22}, x_{32}, \cdots, x_{n2}\}$ follow a particular pattern (Calonder et al., 2010).

The features that have similar descriptors are more likely to correspond to the same landmark in the real world. Given identical recording procedures and illumination in the generation process of images, the same landmark in the real world should form two features with identical descriptors. However, this is not the case in practical applications.

Due to complex distortions, such as specular surfaces, different viewpoints, illumination changes, and so on, the descriptors are not exactly the same. Therefore, Conventional Check 1 (i.e., the Mismatch Limit Check) is usually conducted in the Feature Matching process as a mitigating countermeasure. If the difference between two descriptors is below Threshold 1, the corresponding feature points would be considered as a correct match. However, conventional Check 1 tends to fail if two or more features from the same image have numerically comparable descriptors. This occasion occurs in the scenarios where there are resemblant objects (e.g., cars, grasses, windows) or repetitive textures (e.g., walls, roads). Based on the discussion above, there is a high risk of mismatch in these cases.

Therefore, we designed Proposed Check 1 to cope with these similar features. The main idea is that the tested feature pair will be discarded if it is less distinctive.

In a quantitative way, the ratio of optimal distance to suboptimal distance would exceed Threshold 4, which means that there is an unselected feature in the current frame that is similar to a feature in the previous frame (see Equation (15)).

Threshold 4 is a ratio value from 0 to 1. The closer it gets to zero, the stricter Proposed Check 1 is. *Optimal distance*, thus, refers to the shortest Hamming distance and *suboptimal distance* refers to the second shortest Hamming distance. Only the feature pair whose ratio is less than or equal to Threshold 4 would be preserved. In Equation (15), *distance* 1 is the optimal distance and *distance* 2 is the suboptimal distance:

$$\frac{distance\,1}{distance\,2} \leq Threshold\,4 \tag{15}$$

From the above introduction, we can deduce that Proposed Check 1 works by excluding weak feature point pairs. However, when sufficient feature points cannot be extracted from the image, setting Threshold 4 too small will lead to fewer matched feature pairs, which would affect the accuracy and continuity of VO. Therefore, the setting of Threshold 4 is a trade-off between VO integrity, accuracy, and continuity.

Three experiments were designed to quantitatively analyze the effects of Threshold 4 on VO continuity, accuracy, and integrity, individually. Over the course of the three experiments, Threshold 4 was set to 0.9, 0.7, 0.5, 0.3, and 0.1, respectively, for the five landmark matching pairs required for the RANSAC process in the VO algorithm. VO cannot work when the number of landmark matching pairs is less than five, which renders an unsolvable frame.

The data input of the three experiments was the P010 Sequence in the Forest Winter Scenario, which contained 1,576 frames in total. We chose the P010 Sequence in the Forest Winter Scenario because it contained the largest number of images among all sequences of the scenarios.

In the first experiment, we explored the effect of Threshold 4 on VO continuity. Specifically, the Forest Winter P010 Sequence was input into the VO algorithm, and we computed the number of landmark matching pairs in each frame under different Threshold 4 settings. The statistical result of unsolvable frames is shown in Table 2, and the change of the number of landmark matching pairs over frame epoch is shown in Figure 8. Note that a value of zero in Figure 8 means that the number of landmark matching pairs is no more than four.

It can be seen from Table 2 that the smaller Threshold 4 is, the larger the number of unsolvable frames will be. When Threshold 4 was set to 0.1, all frames were unsolvable. As reflected in Table 2, there was a wide gap between values 0.5 and 0.3. A similar conclusion can be drawn from Figure 8, where the number of landmark matching pairs varied greatly among the different frames. Note that the number of landmark matching pairs of all frames was zero when Threshold 4 was 0.1. Combining Table 2 and Figure 8, we can find that Threshold 4 has a great influence on VO continuity and that the number of unsolvable frames is sensitive to Threshold 4. For the sake of continuity, it is advisable to set Threshold 4 between 0.5 and 0.9.

TABLE 2
Number of Unsolvable Frames with Different Threshold 4 Values

Threshold 4	0.9	0.7	0.5	0.3	0.1
count	43	116	316	1080	1576

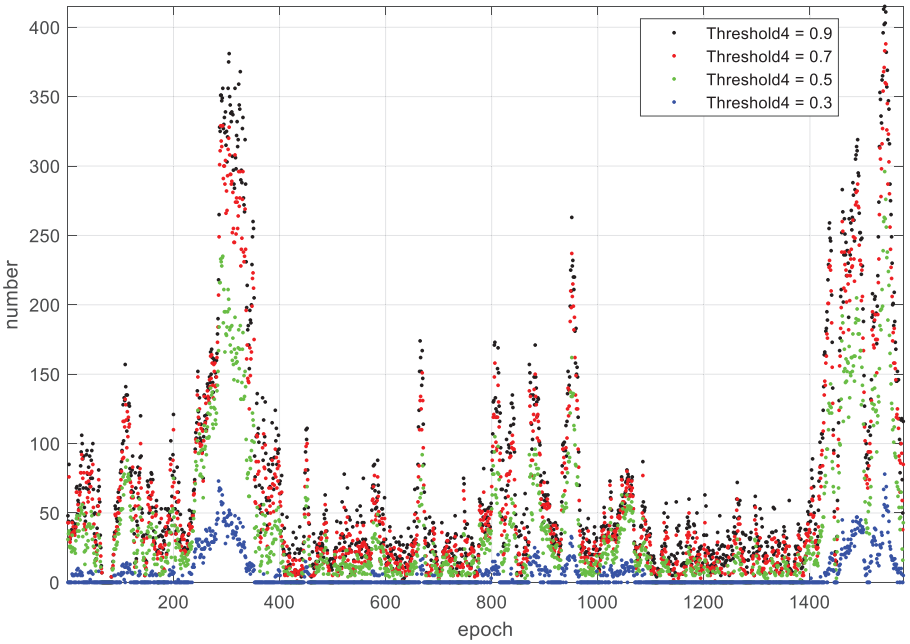

FIGURE 8 Number of landmark matching pairs per epoch for different settings of *Threshold* 4

In the second experiment, we explored the effect of *Threshold* 4 on VO accuracy. To be specific, data was input into the VO algorithm and pose ground truth was used to calculate the magnitude of the Lie algebraic error of each frame under different Threshold 4 settings according to Equation (16). Then, Absolute Pose Error (APE) was calculated according to Equation (17):

$$e_i = \left\| \log\left(T_{gt,i}^{-1} T_{esti,i} \right)^{\vee} \right\|_2 \tag{16}$$

$$APE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} e_i^2} \tag{17}$$

where N represents the total number of poses; e_i represents the pose error in the Lie algebraic meaning; $T_{gt,\,i}$ represents the ground truth of the transformation matrix; $T_{esti,\,i}$ represents the estimated transformation matrix; and $\log(\cdot)^{\vee}$ represents the logarithmic transformation that converts the Lie group into Lie algebra. APE is actually the root-mean-square error of the pose Lie algebra value, which can describe the estimated errors of rotation and translation.

Table 3 lists the results of APE. When Threshold 4 was 0.1, VO did not work on all frames. So, APE could not be calculated in this case. Figure 9 shows the curve of the pose error changing with the frame. The data point with a value of -0.1 indicated that VO was not working on that frame and that pose error could not be calculated.

As can be seen from Table 3, with the decrease of Threshold 4, the pose error first decreases and then increases. The reason for this is that some landmark matching pairs with large residuals were excluded at the beginning, which reduced the error of the estimated pose. Then landmark matching pairs with small residuals were also excluded, and the reduction in the number of effective measurements led to a larger error in the estimated pose.

TABLE 3
APE with Different Threshold 4 Values

Threshold 4	0.9	0.7	0.5	0.3	0.1
APE	1.901	1.399	1.381	2.264	N/A

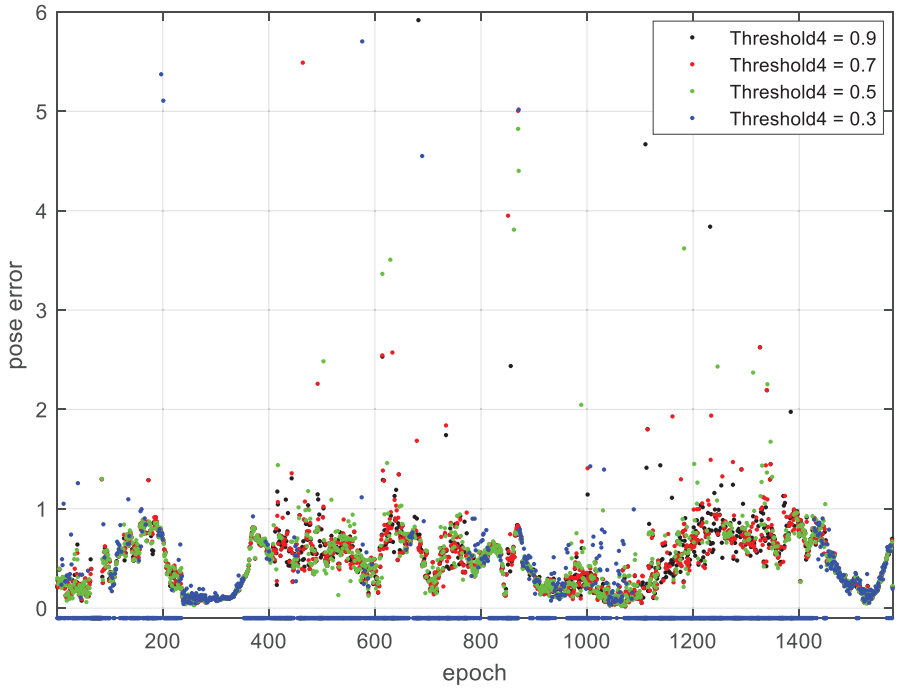

FIGURE 9 Pose error per epoch for different settings of Threshold 4

By comparing Figure 8 and Figure 9, it can be seen that the pose error of the frame with a large number of landmark matching pairs was relatively small. The same trend is shown with different values of Threshold 4. According to the analysis in Table 3 and Figure 9, VO accuracy is rather insensitive to Threshold 4. For the sake of accuracy, it is appropriate to set the value of Threshold 4 between 0.5 and 0.7.

In the third experiment, we explored the effect of Threshold 4 on VO integrity. Specifically, data was input into the VO algorithm and the landmark matching error was calculated with pose ground truth. The overbounding sigma with the same expected fault probability was calculated.

Figure 10 shows the results of the experiment, noting that the expected fault probability was set to 10^{-3} given the small amount of data points. When Threshold 4 was set to 0.1, the number of landmark matching pairs in all frames was no more than four, which means that VO could not work. Consequently, the result of the 0.1 setting is not graphically depicted. As can be seen from the figure, the smaller Threshold 4 is, the smaller the overbounding sigma will be. In general, the overbounding sigma is insensitive to Threshold 4.

Considering these three experiments, setting Threshold 4 between 0.5 and 0.7 could achieve a good balance between continuity, accuracy, and integrity of VO. When Threshold 4 is close to 0.5, Proposed Check 1 is a strong check; however, when it takes a value close to 0.7, Proposed Check 1 is a weak check. Scenarios have some impact on the setting of Threshold 4. To be specific, when there are rich features in the scenario (such as at Factory Day), we recommend using a strong check; otherwise in cases like the Office Scenario, using a weak check could be better.

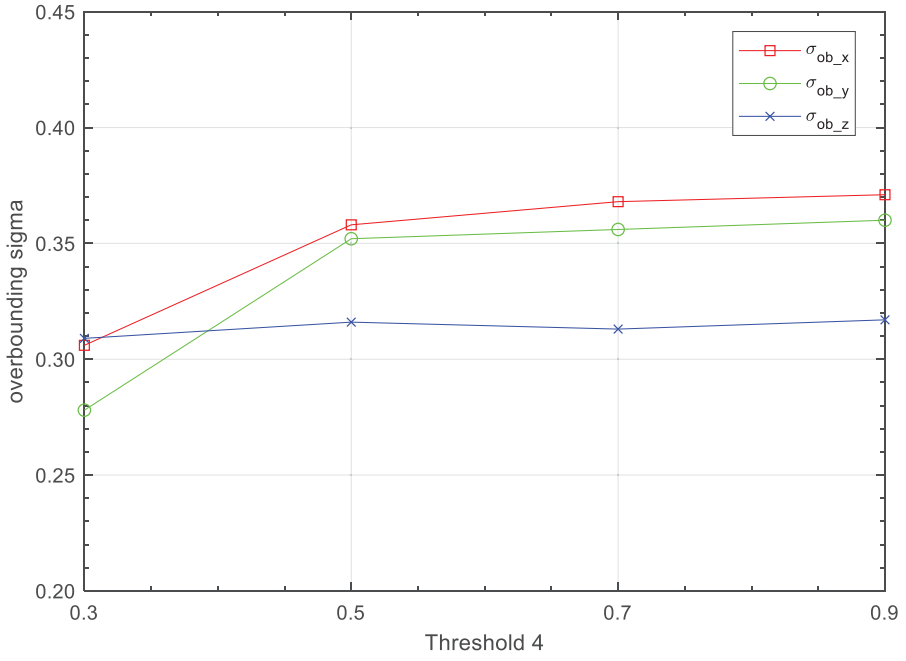

FIGURE 10 Overbounding sigma with different Threshold 4 values

3.2 | Proposed Check 2: Motion Constraint Check

In practice, there are still a considerable number of faulty landmark matching pairs in the landmark pair set after all the three conventional checks. As indicated by Section 2.4, most of these are caused by large depth error events. The reasons are briefly stated as follows.

In the design of Conventional Check 2, only relatively loose restrictions on disparity and depth are imposed, which could not exclude large depth error events efficiently. It is important to note that it is a common phenomenon that the depth obtained by the SGBM algorithm is wrong in the real world, especially outdoors. Conventional Check 3 eliminates the outliers in probabilistic senses, but there are still a considerable number of outliers in the final inlier set due to the large number of measurements.

In order to reject these faulty landmark matching pairs on an absolute scale, we added Proposed Check 2 just before Conventional Check 3. In view of the following two facts, the motion of successive frames is small. The first fact is that the working premise of VO requests overlapping consecutive frames. The second is that the motion of the agent (such as cars, drones, handheld devices, and robots) should be constrained by dynamics and kinematics in the real world. That is, in a very short time of 0.1 s for 10 Hz or 0.03 s for 30 Hz, the displacement of the agent should be small. The core idea of Proposed Check 2 is to constrain the movement between two successive frames, as shown in Equation (18):

$$\left\| P_k - P_{k-1} \right\| \leq Threshold\ 5 \tag{18}$$

where $P_{k-1}\left(X, Y, Z\right)$ and $P_k\left(X', Y', Z'\right)$ represent the coordinates of an identical landmark in the camera coordinate system at epoch $k-1$ and epoch k, respectively. Note that it was considered that the camera was static, while the landmark was kinematic. Therefore, the translation of the camera can be represented by the

coordinate differences of the corresponding landmarks. More specifically, when the camera experiences the general case of motion, $P_k - P_{k-1}$ can be expressed as:

$$P_k - P_{k-1} = RP_{k-1} + t - P_{k-1} \approx \left(I + \theta\times\right)P_{k-1} + t - P_{k-1} = \theta \times P_{k-1} + t \tag{19}$$

where R is the rotational matrix; t is the translation vector; $\theta = [\alpha, \beta, \gamma]^T$ is the rotation represented by Euler angles; and $\times$ represents the antisymmetric operator.

$$\theta\times = \begin{bmatrix} 0 & -\gamma & \beta \\ \gamma & 0 & -\alpha \\ -\beta & \alpha & 0 \end{bmatrix} \tag{20}$$

Therefore, the left side of Equation (18) is equal to $\left\|\theta \times P_{k-1} + t\right\|$, and Threshold 5 is its upper bound. At this point, Threshold 5 constrains the rotation θ and the translation t. It is acceptable that Proposed Check 2 may exclude a few inliers at times because there are, overall, a lot of measurements in the context of stereo VO.

Threshold 5 is heavily correlated with pose changes. If the motion of camera between two consecutive frames is large (such as a car), then Threshold 5 should be set high. If the motion of the camera between two consecutive frames is small (such as a pedestrian handheld device), then Threshold 5 should be set low.

In order to quantitatively explore the displacement between consecutive frames with different forms of agents, we computed the displacements between two consecutive frames using the KITTI data set (Geiger et al., 2012) and the TartanAir data set, in which the agent of the former was a car and the agent of the latter was an unmanned aerial vehicle. All sequences (11 in total) in the KITTI data set and the 10 most representative sequences in the TartanAir data set were selected to compute displacement.

First, we analyzed the data sets as follows. Both the KITTI data set and TartanAir data set provided ground truth of the camera's pose. Therefore, we could calculate the translation vector between two adjacent frames. Then, we counted the magnitude of the translation vector and the number of occurrences, which were then presented using histograms in Figure 11.

Second, the results of the histogram analyses were used as prior information on the range within which our target motion resided. From the two data sets that contained the richest variation of scenarios on this topic (i.e., KITTI and TartanAir),

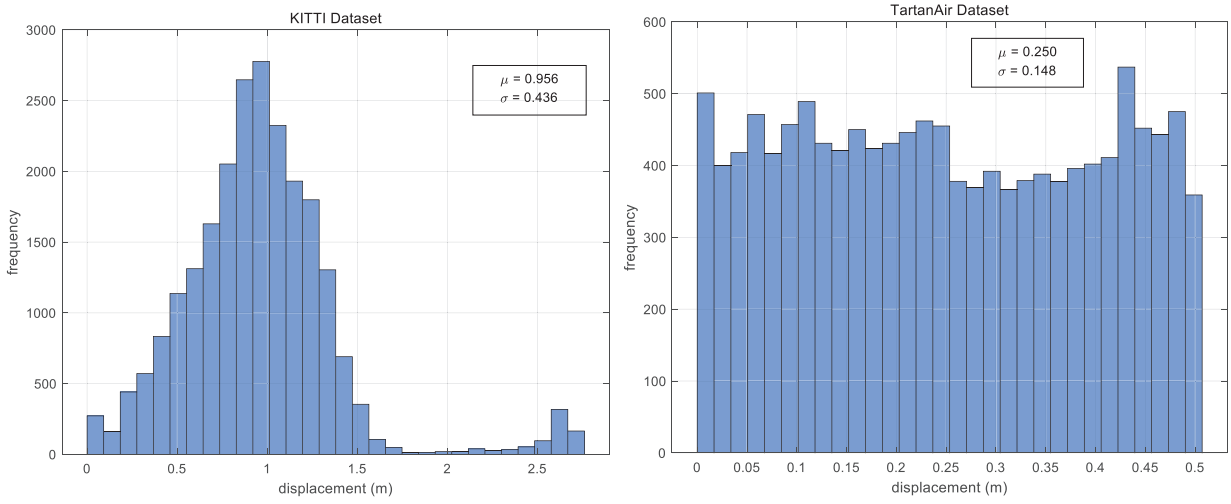

FIGURE 11 Displacement between adjacent frames with two different agents

we concluded that if the landmark were to move, it would move between 0 to 1 meter. Threshold 5, as an upper limit, should thus be chosen as a reasonably scaled-up value, which turns our attention to each instance of the five investigated values (i.e., 0.6, 0.8, 1.5, 3.0, and 5.0). It can be seen from Figure 11 that the displacement in the KITTI data set were mainly concentrated near 1 meter, while the displacement in TartanAir data set were approximately evenly distributed between 0 and 0.5. In general, there were some differences between the displacement of different agents.

In order to further analyze how to select an appropriate value for Threshold 5 that can provide VO with good integrity without causing serious damage to continuity and accuracy, we conducted three experiments to quantitatively analyze the sensitivity of accuracy, continuity, and integrity of VO to Threshold 5.

In the first experiment, the influence of Threshold 5 on VO continuity was examined. The Forest Winter P010 Sequence was used as data input for the VO algorithm and the number of landmark matching pairs in each frame under different Threshold 5 values was counted. The number of unsolvable frames is shown in Table 4, and the curve of the number of landmark matching pairs over frame is shown in Figure 12.

From Table 4, it is easy to see that the number of unsolvable frames decreases as the value of Threshold 5 increases. When Threshold 5 is greater than 3.0, the number of unsolvable frames remains unchanged at 174. A similar conclusion can be drawn from Figure 12. Moreover, the number of landmark matching pairs varies greatly from frame to frame and they all show the same trend. Combining the data from Table 4 with Figure 12, it can be deduced that Threshold 5 has a small impact on VO continuity, and the number of unsolvable frames has an overall low sensitivity to the Threshold 5 value. For the sake of VO continuity, it is advisable to set Threshold 5 close to 3.0.

TABLE 4
Number of Unsolvable Frames with Different Threshold 5 Values

Threshold 5	0.6	0.8	1.5	3.0	5.0
count	299	221	192	174	174

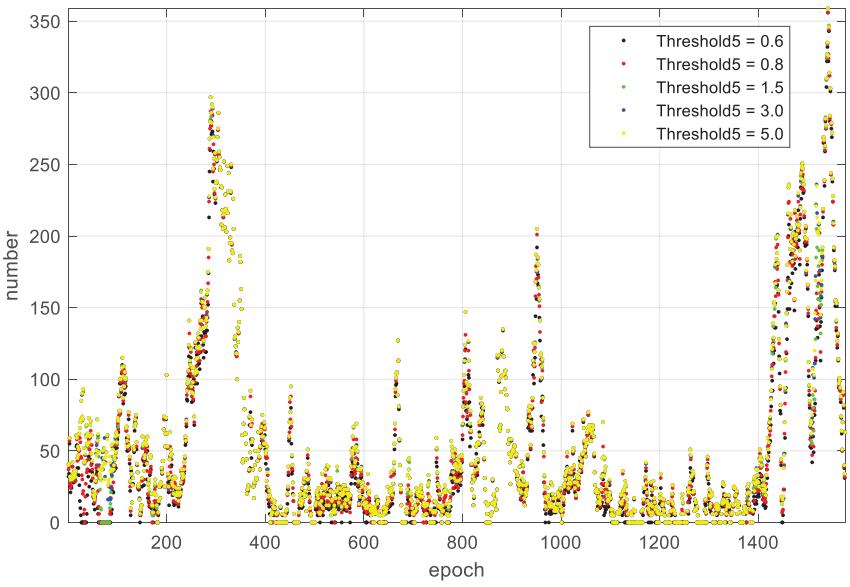

FIGURE 12 Number of landmark matching pairs per epoch for different settings of Threshold 5

TABLE 5
APE With Different Threshold 5 Values

Threshold 5	0.6	0.8	1.5	3.0	5.0
APE	1.145	0.842	1.164	1.645	1.628

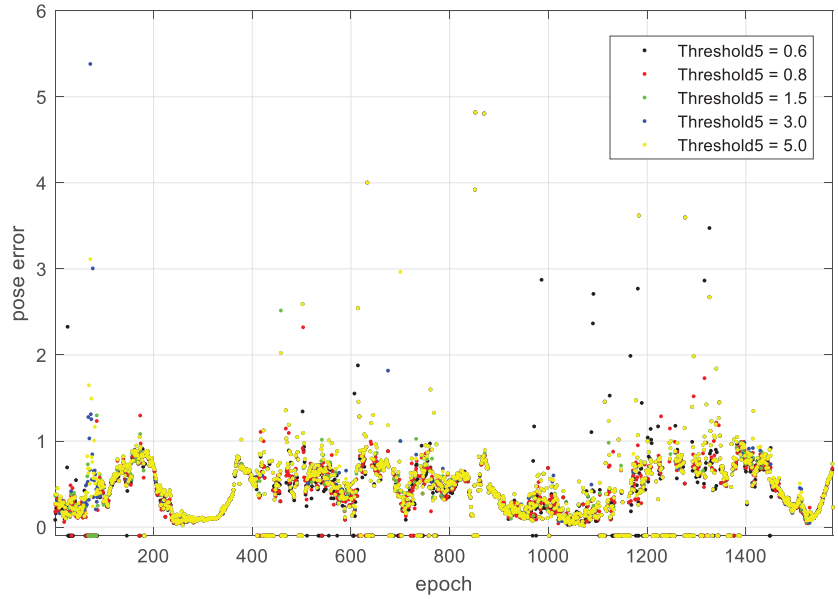

FIGURE 13 Pose error per epoch for different settings of Threshold 5

In the second experiment, we examined the influence of the Threshold 5 on VO accuracy. We calculated the landmark matching error for each frame in the form of Lie algebra magnitude and, then, calculated the APE. The results are shown in Table 5 and Figure 13. It can be seen from Table 5 that APE decreases first and, then, increases as Threshold 5 values increases. The reason for this trend is that the increase of useful measurements enhanced the accuracy of VO. Later, the accuracy is actually reduced when faults are mixed in.

By comparing Figure 12 and Figure 13, it can be seen that the pose error of the frame with a large number of landmark matching pairs is small. Moreover, the pose error under different values of Threshold 5 shows the same trend. Combining the data from Table 5 with Figure 13, we can deduce that VO accuracy is overall insensitive to Threshold 5. For the sake of accuracy, it is advisable to set the threshold between 0.6 and 1.5.

In the third experiment, we explored the influence of Threshold 5 on VO integrity by calculating overbounding sigmas with different values of Threshold 5. Figure 14 shows the results of the experiment. For the sake of reducing the amount of data points, we set the expected fault probability at 10^{-3}. As can be seen from the figure, the larger the Threshold 5 value is, the larger the overbounding sigma will be. The overbounding sigma is, thus, sensitive to Threshold 5. For the sake of VO integrity, it is advisable to set Threshold 5 near 1.5.

Taking into account all three experiments, setting Threshold 5 around 1.5 could strike a good balance between continuity, accuracy, and integrity of VO.

It deserves mentioning that Threshold 5 has little to do with the scenario. Rather, it is heavily correlated with pose changes of the camera. If the motion of the camera between two consecutive frames is large (such as car-mounted equipment), then Threshold 5 should be set high; otherwise, if the motion of the camera between two

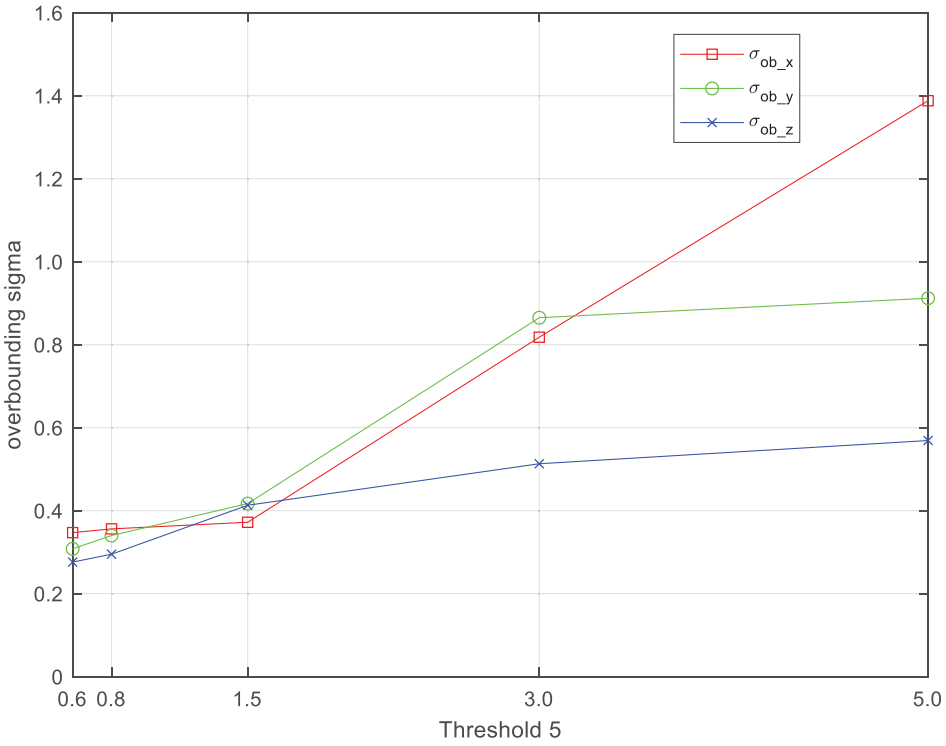

FIGURE 14 Overbounding sigma with different Threshold 5 values

consecutive frames is smaller (such as with a handheld device), Threshold 5 should take on a small value.

4 | EXPERIMENTAL VALIDATION

For the newly proposed checks, we first conducted ablation tests to show the effects of the individual check. We, then, proceeded to show the combined effects of the two proposed checks. Using examples, we also show that CDF overbounding greatly improves in terms of tightness, computational efficiency, and most important of all, scenario tolerance.

4.1 | Ablation Study

In order to analyze the contribution of Proposed Check 1 and Proposed Check 2 to overbounding landmark matching error, we performed an ablation study on them. Specifically, we used the entire set of data sequences included in the Carwelding Scenario and Neighborhood Scenario as inputs for the VO pipeline. There were four sequences in the Carwelding Scenario and 18 sequences in the Neighborhood Scenario. These two scenarios were initially used in Section 2.3 to compare the overbounding measurement error of the conventional checks.

We inserted the newly Proposed Check 1 and Proposed Check 2 into the VO algorithm, calculated the landmark matching errors, rendered the folded CDFs, and obtained the overbounding sigmas. The results are shown in Figure 15 and Figure 16, with the expected fault probability set to 10^{-5}. Their corresponding overbounding sigmas are summarized in Table 6 and Table 7.

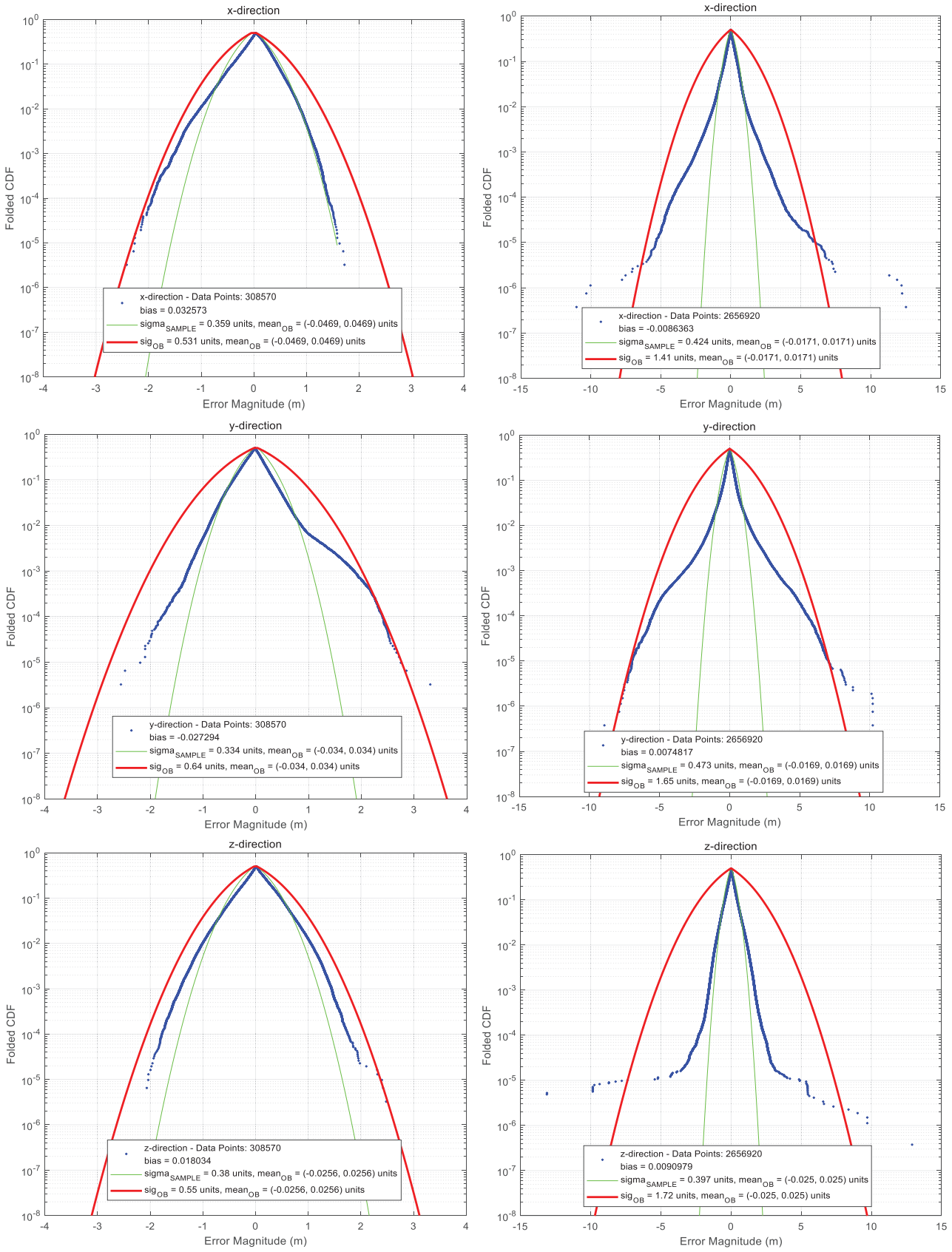

FIGURE 15 Folded CDFs of landmark matching errors with Proposed Check 1 in x, y, and z directions (top to bottom) for the Carwelding Scenario (left) and the Neighborhood Scenario (right)

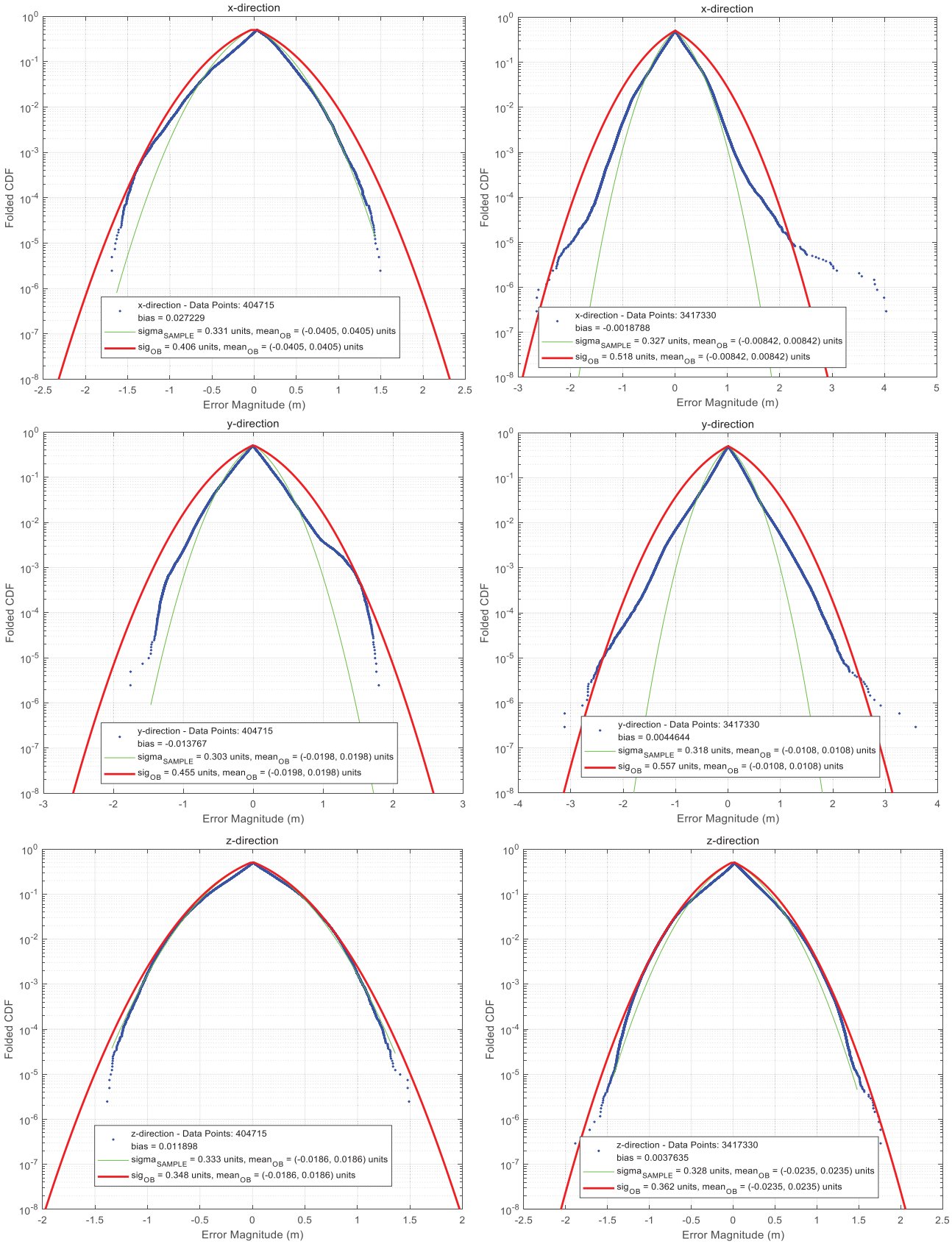

FIGURE 16 Folded CDFs of landmark matching errors with Proposed Check 2 in *x*, *y*, and *z* directions (top to bottom) for the Carwelding Scenario (left) and the Neighborhood Scenario (right)

TABLE 6
Three-Axis Overbounding Standard Deviations of Landmark Matching Errors With Proposed Check 1

Scenarios	ΔX	ΔY	ΔZ
Carwelding	0.531 m	0.640 m	0.550 m
Neighborhood	1.41 m	1.65 m	1.72 m

Figure 15 is the result of adding Proposed Check 1 only and Figure 16 is the result of adding Proposed Check 2 only. As can be seen from Figure 15 and Table 6, there is still a large gap between the landmark matching errors in the two scenarios when only Proposed Check 1 was added.

TABLE 7
Three-Axis Overbounding Standard Deviations of Landmark Matching Errors With Proposed Check 2

Scenarios	ΔX	ΔY	ΔZ
Carwelding	0.406 m	0.455 m	0.348 m
Neighborhood	0.518 m	0.557 m	0.362 m

As shown in Figure 16 and Table 7, the overbounding distributions of landmark matching errors in the two scenarios were similar when Proposed Check 2 was added.

We also took these two scenarios as data input, executed VO with both of these two proposed checks to get the landmark matching errors, and overbounded these errors to the same probability of 10^{-5}. The results are shown in Figure 17 and Table 8. Consequently, the overbounding sigmas of the two scenarios were similar; the difference between overbounding sigmas in these two scenarios had entirely disappeared.

TABLE 8
Three-Axis Overbounding Standard Deviations of Landmark Matching Errors With Proposed Checks 1 and 2

Scenarios	ΔX	ΔY	ΔZ
Carwelding	0.469 m	0.499 m	0.367 m
Neighborhood	0.537 m	0.590 m	0.362 m

By comparing Figure 15, Figure 16, and Figure 17, we found that Proposed Check 1 contributes to the general model of landmark matching error, but Proposed Check 2 is the key to stabilize the landmark matching errors. In order to show this point more directly, we compared the overbounding sigmas of VO without the proposed checks, VO with Proposed Check 1, VO with Proposed Check 2, and VO with both proposed checks in Figure 18.

The results of VO with the proposed checks turned off and on are given in Figure 4 (off) and Figure 17 (on), respectively. By comparing the results of traditional VO and VO with Proposed Check 1, it can be seen that Proposed Check 1 reduces the difference of overbounding sigmas between the Carwelding Scenario and Neighborhood Scenario, but in a less prominent way than Proposed Check 2 does. Proposed Check 1 was designed to exist at the Feature Matching step in the front-end while Proposed Check 2 is a downstream module that exists in the

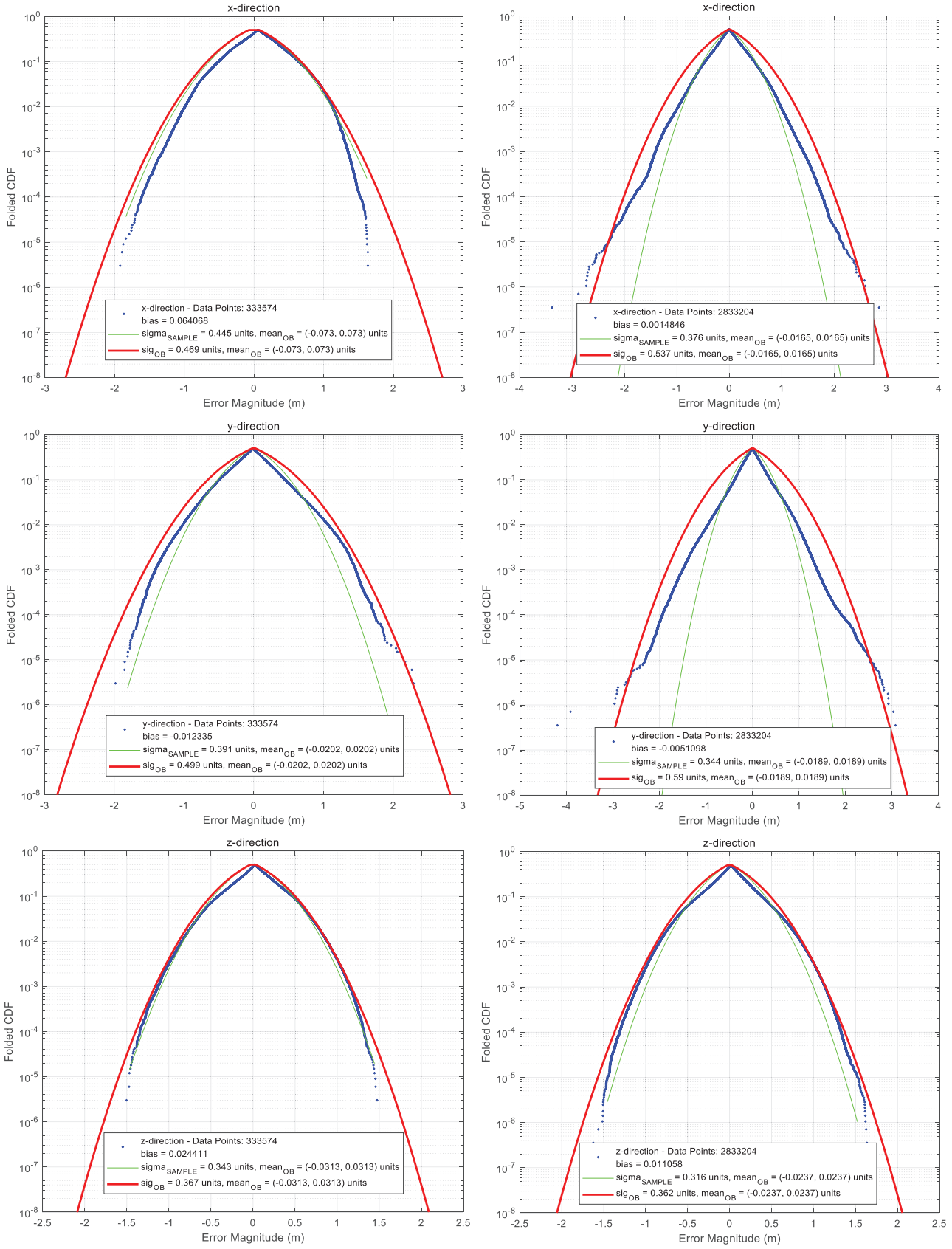

FIGURE 17 Folded CDFs of landmark matching errors with Proposed Check 1 and Proposed Check 2 in x, y, and z directions (top to bottom) for the Carwelding Scenario (left) and the Neighborhood Scenario (right)

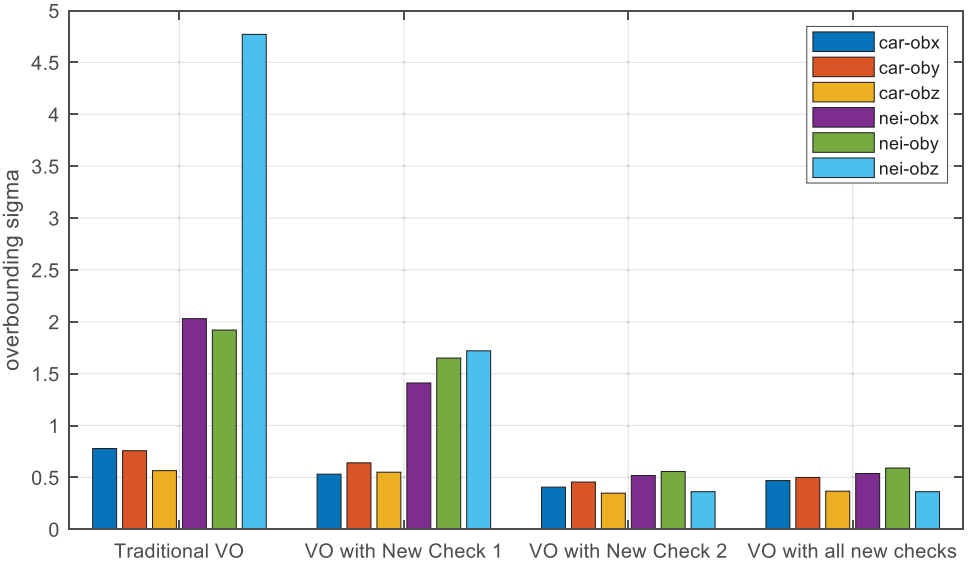

FIGURE 18 Overbounding sigmas of VO with different configurations

back-end. They are, thus, designed to deal with undermining factors of different natures: Proposed Check 1 deals with measurement outliers due to feature mismatches by resemblant objects or repetitive textures; Proposed Check 2 deals with measurement outliers caused by landmark movement or wrong depth estimation, which is unknown to the estimator. In this regard, the two proposed checks are complementary; neither of them can replace the other.

4.2 | Scenario Tolerance Test

To capture the VO sensitivity to operational scenario changes, 10 scenarios were investigated. These scenarios are illustrated in Figure 19, including the Carwelding, Factory Day, Factory Night, Forest Autumn, Forest Winter, Hospital, Neighborhood, Office, Old Town, and Rainy Day scenarios. The selection of these scenarios was based on potential impacting factors.

For example, the comparison between the Factory Day Scenario and Factory Night Scenario tested how lighting condition impacts landmark matching errors. The comparison between the Forest Autumn Scenario and Forest Winter Scenario shows how seasonal variation can influence landmark matching errors. The comparison between the Rainy Day Scenario and the Factory Day Scenario illustrates how weather affects landmark matching errors.

All the sequences included in these scenarios were processed. Only the hard level was considered to capture the worst-case scenario. With the two proposed checks inserted into the VO pipeline, the landmark matching error model could be better established. In this section, we directly show the final results of the model.

In the case of ARAIM by which future VO integrity concepts will be inspired, a fault probability was introduced along with an overbounding standard deviation. For example, the GPS constellation service provider (CSP) commits a 10^{-5} fault probability (Department of Defense, 2020), for which the fault was defined as a large error whose magnitude exceeded $4.42\sigma_{URA}$ (Walter et al., 2019). In the context of stereo VO, we considered landmark matching pairs with large depth error events, feature mismatch events, and landmark movement events as faults because they typically cause large landmark matching residuals. Back in Figure 17, all the red curves represented the minimal overbounding CDF plots.

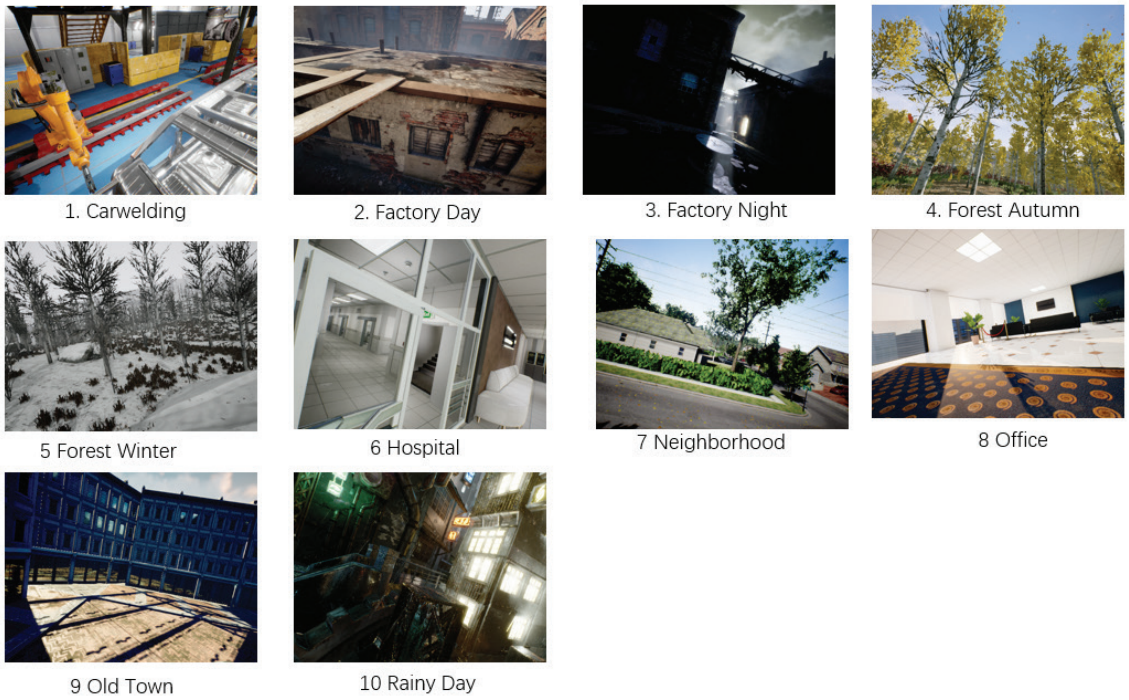

FIGURE 19 Investigated operational scenarios for landmark matching error sensitivity analysis

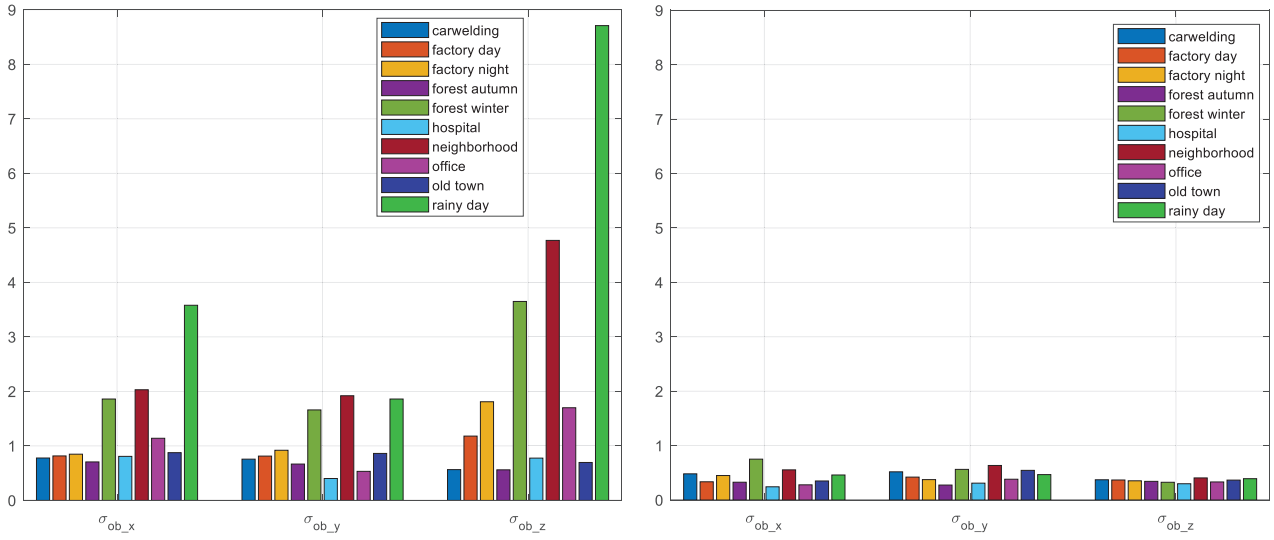

FIGURE 20 Overbounding standard deviations of the landmark matching errors for all scenarios without proposed checks (a) and with proposed checks (b)

Because a $P_{VO} = 10^{-5}$ is employed in those figures, these curves were all bounded to a 10^{-5} probability. Then, a minimal bounding standard deviation was established for nominal errors within the probability. The same approach can be applied to other P_{VO} values, such as 10^{-3} or 10^{-4}. Therefore, the landmark errors are bound using a paired P_{VO} and σ_{VO}.

With and without the proposed checks, the overbounding standard deviations of the landmark matching errors were obtained and are summarized in Figure 20, bounded to the 10^{-5} probability. It can be seen that these two proposed checks make

the overbounding sigma much more scenario tolerant, so it is safe to lump data from different scenarios together. Figure 21 shows the folded CDF plots of landmark matching errors with a 10^{-5} fault probability and Table 9 generalizes these results by providing more fault probability options. Most importantly, we provided recommended overbounding sigma values based on our data processing results. Due to the endless challenge scenarios, the recommended values were obtained by scaling up the calculated results for more conservative considerations. It should

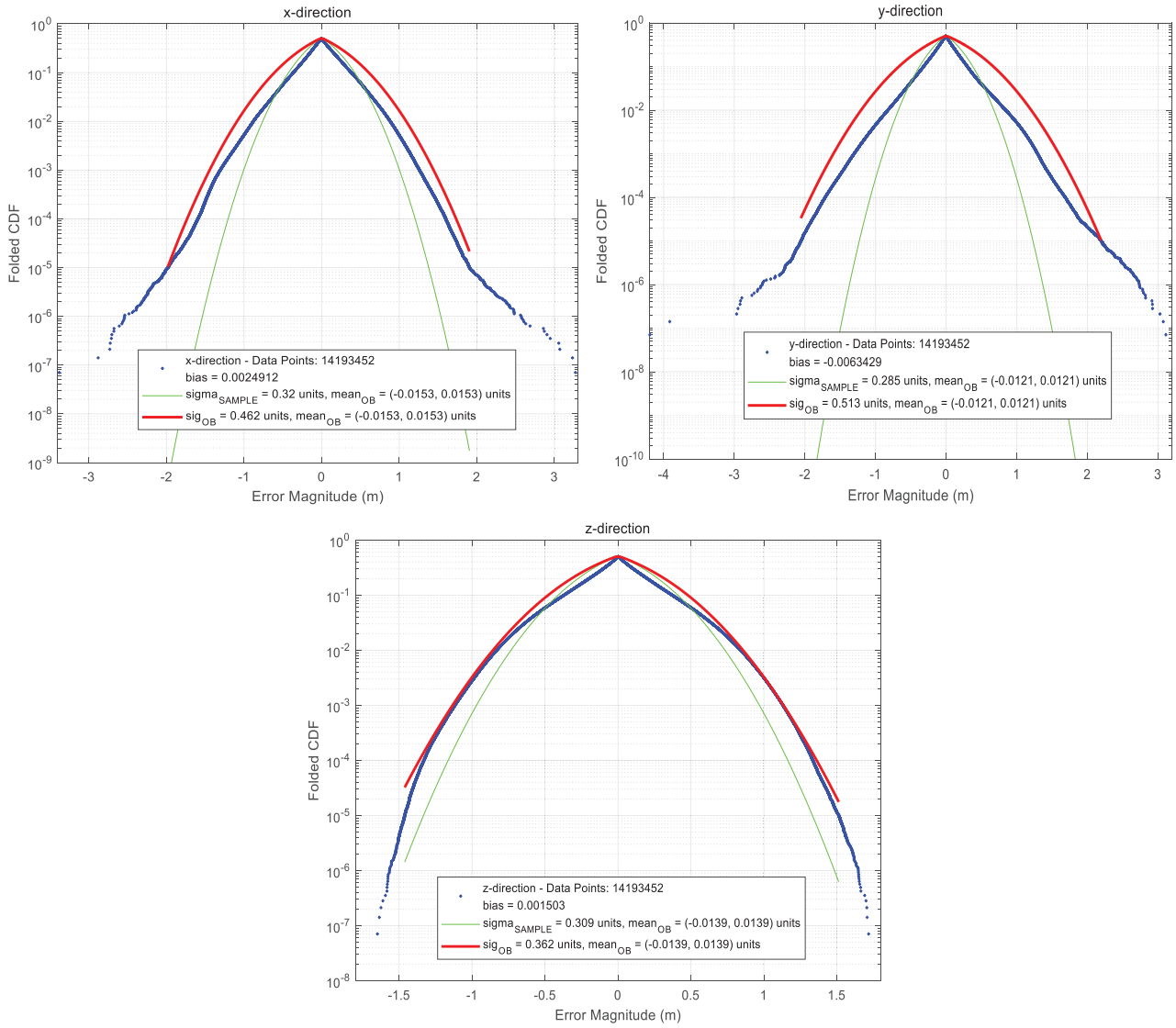

FIGURE 21 Folded CDF results of the landmark matching errors with the proposed checks

TABLE 9
Overbounding Pairs for Landmark Matching Errors With Proposed Checks

Fault Probability	Actual / Recommended σ_x (m)	Actual / Recommended σ_y (m)	Actual / Recommended σ_z (m)
10^{-5}	0.462 / 0.6	0.513 / 0.6	0.362 / 0.5
10^{-4}	0.427 / 0.5	0.458 / 0.6	0.362 / 0.5
10^{-3}	0.419 / 0.5	0.419 / 0.5	0.362 / 0.5

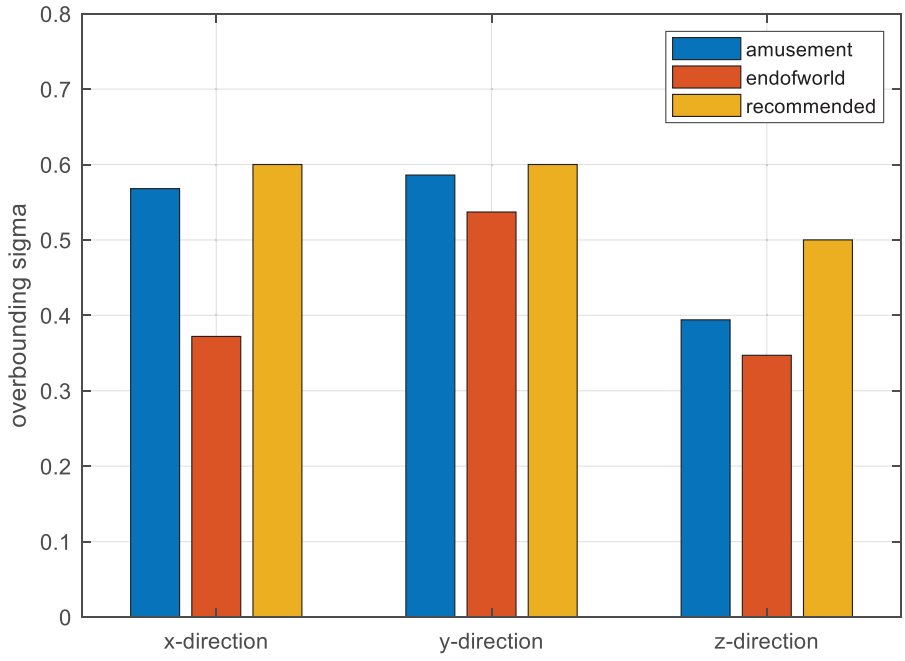

FIGURE 22 Comparison between the overbounding sigmas of test scenarios and their recommended values

be noted that all the error models were derived based on data. We have shown that the error overbound is independent of operational scenarios. This is owed to the proposed checks, since they help to Gaussianize nominal errors by taming the fat-tailed distribution to a normal distribution.

Landmark matching error overbounding (see the second row of Table 9) was also validated using two new scenarios that were not involved in the error model development in an attempt to test the generalization capabilities of the proposed overbounding sigma. These are the Amusement Scenario and End of World Scenario in the TartanAir data set. Therefore, the overbound can be proven to be promising if the overbounding sigmas of these two scenarios are not greater than the recommended values. These two scenarios were not included in the error model derivation, so they are suitable for cross-validation. In addition, there are many similar features in the environment that may potentially lead to mismatch events. If the error models are validated using these two scenarios, the conservatism of the overbound are ensured. Figure 22 presents the comparison between the overbounding sigmas of the test scenarios and recommended values. The blue bar corresponds to the Amusement Scenario and the orange bar corresponds to the End of World Scenario. It can be seen that the blue and orange bars are perfectly bounded by the model, validating the model in the process.

5 | CONCLUSION

Integrity is a greatly under-investigated concept in visual navigation. We proposed a new approach to measurement error detection for the integrity of visual navigation and, in particular, stereo visual odometry (VO). As a first step, VO measurement residuals were defined as landmark matching error. This definition facilitated easy development of future visual navigation integrity concepts based on the current advanced receiver autonomous integrity monitoring (ARAIM) framework.

We, then, proceeded to propose two methods to detect large measurement residuals that otherwise could not be identified with the existing outlier rejection methods in state-of-the-art VO pipelines. By removing these large errors, measurement residuals were able to be better bounded by the common CDF overbounding methods used in ARAIM.

We evaluated our methods using the open-source data set TartanAir and showed that a tighter and more computationally effective overbound could be achieved. Most important of all, the proposed methods make the traditional overbounding method much more scenario tolerant.

These methods and findings are a good starting point for developing future integrity monitoring algorithms for visual navigation and, in particular, stereo VO. Of note, because Threshold 5 is heavily correlated with camera pose changes, it could be better obtained from an alternative source such as an IMU with proper extrinsic parameters (i.e., IMU-camera calibration). In the future, we plan to use this fusion framework to develop a fault detection and exclusion (FDE) algorithm and protection level calculation algorithms.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Grant Number: 62173227, 62103274 and 61403253).

REFERENCES

Bay, H., Tuytelaars, T., & Gool, L. Van. (2006). SURF: Speeded up robust features. In A. Leonardis, H. Bischof, & A. Pinz (Eds.), *European conference on computer vision* (Vol. 3951, pp. 404–417). Springer. https://doi.org/10.1007/11744023_32

Blanch, J., Ene, A., Walter, T., & Enge, P. (2007). An optimized multiple hypothesis RAIM algorithm for vertical guidance. *Proc. of the 20th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS 2007)*, Fort Worth, TX, 2924–2933. https://www.ion.org/publications/abstract.cfm?articleID=7644

Bradski, G. (2000). The OpenCV Library. https://opencv.org/

Brown, R. G. (1992). A baseline GPS RAIM scheme and a note on the equivalence of three RAIM methods. *NAVIGATION*, *39*(3), 301–316. https://doi.org/10.1002/j.2161-4296.1992.tb02278.x

Calonder, M., Lepetit, V., Strecha, C., & Fua, P. (2010). BRIEF: Binary robust independent elementary features. In K. Daniilidis, P. Maragos, & N. Paragios (Eds.), *European conference on computer vision* (Vol. 6314, 778–792). Springer. https://doi.org/10.1007/978-3-642-15561-1_56

Cumani, A. (2011). Feature localization refinement for improved visual odometry accuracy. *International Journal of Circuits, Systems and Signal Processing*, *5*(2), 151–158. https://www.naun.org/main/NAUN/circuitssystemssignal/19-679.pdf

DeCleene, B. (2000). Defining pseudorange integrity––Overbounding. *Proc. of the 13th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS 2000),* Salt Lake City, UT, 1916–1924. https://www.ion.org/publications/abstract.cfm?articleID=1603

Department of Defense. (2020). *Global Positioning System standard positioning service performance standard* (5th Ed.). https://www.gps.gov/technical/ps/

Fischler, M. A., & Bolles, R. C. (1987). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. In M. A. Fischler & O. Firschein (Eds.), *Readings in computer vision* (726–740). https://doi.org/10.1016/B978-0-08-051581-6.50070-2

Fu, Y., Wang, S., Zhai, Y., & Zhan, X. (2020). Visual odometry errors and fault distinction for integrity monitoring. *Aerospace Systems*, *3*(4), 265–274. https://doi.org/10.1007/s42401-020-00062-x

Gao, X., & Zhang, T. (2021). *Introduction to visual SLAM: from theory to practice*. Springer.

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*, Providence, RI, 3354–3361. https://doi.org/10.1109/CVPR.2012.6248074

Hafez, O. A., Arana, G. D., Joerger, M., & Spenko, M. (2020). Quantifying robot localization safety: A new integrity monitoring method for fixed-lag smoothing. *IEEE Robotics and Automation Letters*, *5*(2), 3182–3189. https://doi.org/10.1109/LRA.2020.2975769

Harris, C., & Stephens, M. (1988). A combined corner and edge detector. *Proc. of the Alvey Vision Conference*, *15*(50), 147–151. https://doi.org/10.5244/C.2.23

Hirschmüller, H. (2008). Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(2), 328–341. https://doi.org/10.1109/TPAMI.2007.1166

Jiang, Y., Xu, Y., & Liu, Y. (2013). Performance evaluation of feature detection and matching in stereo visual odometry. *Neurocomputing*, *120*, 380–390. https://doi.org/10.1016/j.neucom.2012.06.055

Kelly, R. J., & Davis, J. M. (1994). Required navigation performance (RNP) for precision approach and landing with GNSS application. *NAVIGATION*, *41*(1), 1–30. https://doi.org/10.1002/j.2161-4296.1994.tb02320.x

Larson, J. D. (2018). *Gaussian-Pareto overbounding: A method for managing risk in safety-critical navigation systems* [Doctoral dissertation, University of Minnesota]. University of Minnesota Digital Conservancy. https://conservancy.umn.edu/handle/11299/200312

Li, C., & Waslander, S. L. (2019). Visual measurement integrity monitoring for UAV localization. *2019 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR),* Würzburg, Germany. https://doi.org/10.1109/SSRR.2019.8848975

Hong, L., & Chen, G. (2004). Segment-based stereo matching using graph cuts. *Proc. of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, DC. https://doi.org/10.1109/CVPR.2004.1315016

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

Mur-Artal, R., & Tardós, J. D. (2017). ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras. *IEEE Transactions on Robotics*, *33*(5), 1255–1262. https://doi.org/10.1109/TRO.2017.2705103

Naroditsky, O., Zhou, X. S., Gallier, J., Roumeliotis, S. I., & Daniilidis, K. (2012). Two efficient solutions for visual odometry using directional correspondence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(4), 818–824. https://doi.org/10.1109/TPAMI.2011.226

Nistér, D., Naroditsky, O., & Bergen, J. (2004). Visual odometry. *Proc. of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, DC. https://doi.org/10.1109/cvpr.2004.1315094

Rife, J., Pullen, S., Enge, P., & Pervan, B. (2006). Paired overbounding for nonideal LAAS and WAAS error distributions. *IEEE Transactions on Aerospace and Electronic Systems*, *42*(4), 1386–1395. https://doi.org/10.1109/TAES.2006.314579

Rife, J., Walter, T., & Blanch, J. (2004). Overbounding SBAS and GBAS error distributions with excess-mass functions. *International Symposium on GNSS/GPS*. https://www.researchgate.net/publication/228724631_Overbounding_SBAS_and_GBAS_Error_Distributions_with_Excess-Mass_Functions

Rosten, E., & Drummond, T. (2006). Machine learning for high-speed corner detection. *Proc. of the 9th European Conference on Computer Vision*, Berlin, Germany, 430–443. https://doi.org/10.1007/11744023_34

Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). ORB: An efficient alternative to SIFT or SURF. *2011 International Conference on Computer Vision*, Barcelona, Spain. https://doi.org/10.1109/ICCV.2011.6126544

Scaramuzza, D., & Fraundorfer, F. (2011). Visual odometry [Tutorial]. *IEEE Robotics and Automation Magazine*, *18*(4), 80–92. https://doi.org/10.1109/MRA.2011.943233

Scaramuzza, D., & Siegwart, R. (2008). Appearance-guided monocular omnidirectional visual odometry for outdoor ground vehicles. *IEEE Transactions on Robotics*, *24*(5), 1015–1026. https://doi.org/10.1109/TRO.2008.2004490

Szeliski, R. (2010). *Computer vision algorithms and applications*. Springer-Verlag London Limited.

Walter, T., Blanch, J., Gunning, K., Joerger, M., & Pervan, B. (2019). Determination of fault probabilities for ARAIM. *IEEE Transactions on Aerospace and Electronic Systems*, *55*(6), 3505–3516. https://doi.org/10.1109/TAES.2019.2909727

Wang, S., Zhai, Y., & Zhan, X. (2021). Characterizing BDS signal-in-space performance from integrity perspective. *NAVIGATION*, *68*(1), 157–183. https://doi.org/10.1002/navi.409

Wang, S., Zhan, X., Fu, Y., & Zhai, Y. (2020a). Feature-based visual navigation integrity monitoring for urban autonomous platforms. *Aerospace Systems*, *3*(3), 167–179. https://doi.org/10.1007/s42401-020-00057-8

Wang, W., Zhu, D., Wang, X., Hu, Y., Qiu, Y., Wang, C., Hu, Y., Kapoor, A., & Scherer, S. (2020b). TartanAir: A dataset to push the limits of visual SLAM. *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Las Vegas, NV. https://doi.org/10.1109/IROS45743.2020.9341801

Zhai, Y., Joerger, M., & Pervan, B. (2018). Fault exclusion in multi-constellation global navigation satellite systems. *Journal of Navigation*, *71*(6), 1281–1298. https://doi.org/10.1017/S0373463318000383

Zhai, Y., Patel, J., Zhan, X., Joerger, M., & Pervan, B. (2020). An advanced receiver autonomous integrity monitoring (ARAIM) ground monitor design to estimate satellite orbits and clocks. *Journal of Navigation*, *73*(5), 1087–1105. https://doi.org/10.1017/S0373463320000181

Zhu, C., Joerger, M., & Meurer, M. (2020). Quantifying feature association error in camera-based positioning. *2020 IEEE/ION Position, Location and Navigation Symposium (PLANS),* Portland, OR. https://doi.org/10.1109/PLANS46316.2020.9109919

Zhu, C., Steinmetz, C., Belabbas, B., & Meurer, M. (2019). Feature error model for integrity of pattern-based visual positioning. *Proc. of the 32nd International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS+ 2019)*, Miami, FL, 2254–2268. https://doi.org/10.33012/2019.16956

How to cite this article: Fu, Y., Wang, S., Zhai, Y., Zhan, X., & Zhang, X. (2022). Measurement error detection for stereo visual odometry integrity. *NAVIGATION, 69*(4). https://doi.org/10.33012/navi.542