

# Integrity of Visual Navigation—Developments, Challenges, and Prospects

Chen Zhu<sup>1</sup> | Michael Meurer<sup>1,2</sup> | Christoph Günther<sup>1,3</sup>

<sup>1</sup> Institute of Communications and Navigation, German Aerospace Center (DLR), Oberpfaffenhofen, Germany

<sup>2</sup> Chair of Navigation, RWTH Aachen University, Aachen, Germany

<sup>3</sup> Chair of Communications and Navigation, Technical University of Munich (TUM), Munich, Germany

## Correspondence

Chen Zhu, Institute of Communications and Navigation, German Aerospace Center (DLR), Oberpfaffenhofen, Germany  
Email: [Chen.Zhu@dlr.de](mailto:Chen.Zhu@dlr.de)

## Summary

Camera-based visual navigation has great potential for various applications, especially in satellite-signal-degenerated environments. However, the lack of integrity protection has constrained its utilization in safety-critical applications. Integrity characterizes the quality of the information that a navigation system delivers. Integrity frameworks have been developed over decades for satellite navigation, and continue to play an essential role in safety-critical applications like civil aviation. Nevertheless, there are several challenges to quantify the risks associated with visual navigation. Over the last few years, several approaches to tackle these challenges have been investigated. These developments are the first steps toward a reliable visual positioning framework with integrity monitoring capabilities. In this paper, we review the current status, particular challenges, and development trends in visual positioning integrity monitoring. In addition, we propose a preliminary framework so that the future developments on visual navigation integrity can benefit from a systematic approach.

## Keywords

integrity, robust, safety, vision

## 1 | INTRODUCTION

Global navigation satellite systems (GNSSs) are widely used in applications nowadays. From maritime transportation to civil aviation, from route guidance to location-based services on smartphones, GNSS is a core technique for positioning. However, significant challenges with new applications in specific environments persist. Autonomous driving, unmanned aerial vehicles (UAV), and urban air mobility (UAM) require reliable navigation in deep urban areas. In these environments, the performance of GNSS is degenerated due to shadowing by tall buildings and strong multipath signals. It is difficult to solve these problems using GNSS technologies only. Multiple sensor fusion is necessary to deliver sufficient navigation performance required by such applications.

Visual navigation using cameras as part of a multi-sensor solution offers a number of advantages in such applications. Firstly, the number of visual cues for camera-based localization algorithms is large in dense urban environments where GNSS faces its greatest challenges. The associated detailed visual information leads to precise estimates of the camera's pose (position and orientation). Secondly, the

development of machine-learning techniques has made enormous progress in recent years and transforms the complex patterns seen in images into a perception of the environment (Liu et al., 2017). In addition, this is closest to human perception and is, thus, the easiest to interpret by humans for both interaction and verification purposes. Visual *simultaneous localization and mapping* (SLAM) is a key element in the estimation process. It estimates the relative pose and corresponding map simultaneously, and has been widely used in robotics over the past decades (Cadena et al., 2016). Last but not least, digital cameras do not cost much and are easy to integrate into a system. In many cases, they are already present due to requirements from other tasks such as monitoring and detection, which provides a solid foundation for the visual navigation technique's wide application.

Broadly speaking, visual navigation covers various topics and areas including perceptions, path planning, positioning, and control, etc. In this survey, we focus on the positioning and localization (i.e., pose estimation using camera measurements). Cameras are actively used for positioning in applications such as virtual reality, augmented reality, and relative pose estimation in factories already. However, challenges remain for safety-critical applications such as autonomous driving and urban air mobility (air-taxi); the lack of a mature integrity monitoring framework being the most significant obstacle.

In a safety-of-life context, the performance of a given navigation system is characterized by four aspects: accuracy, availability, continuity, and integrity (Langley, 1999). According to the definition from the European Space Agency (ESA), "Integrity is the measure of the trust that can be placed in the correctness of the information supplied by a navigation system. Integrity includes the ability of the system to provide timely warnings to users when the system should not be used for navigation," (European Space Agency, 2011). In other words, an *integrity risk* is present whenever the system reports a position with sufficient confidence but the actual error exceeds an acceptable bound.

In order to ensure safety, the navigation system needs to monitor the integrity of the parameters, usually by setting up some specific test statistics, and to conservatively calculate a *protection level* (PL) which guarantees that inside the bound of the PL, the occurrence probability of an undetected error does not exceed the maximum allowable integrity risk. By comparing the PL value with the *alert limit* (AL) set according to application requirements, the system should issue an alert and should be marked *unavailable* if the PL is larger than the AL value. Intuitively, if the PL calculation is overconservative (in which false alerts occur frequently), the continuity of the navigation solution would be interrupted. In practice, the PL is calculated according to the maximum allowable integrity risk as well as the minimum availability and continuity requirements. Since integrity failures may cause fatal accidents in safety-critical applications, the tolerated integrity risk probability is extremely low in such scenarios. Table 1 lists a few examples of the integrity requirements' order for safety-critical navigation applications, according to its standards and reports (EUSPA, 2021a, 2021b, 2021c; RTCA, 2004).

In order to satisfy both accuracy and integrity requirements in challenging environments such as urban canyons, the integrity monitoring for a multi-sensor navigation system is essential. This further requires the monitoring and fault detections for each individual sensor and their corresponding processing procedures. For GNSS, the core integrity monitoring methods are already well investigated and have been practiced for years (Blanch et al., 2015; Joerger et al., 2014). However, for camera-based navigation techniques, a common agreement on framework is still missing. From our point of view, developing a framework for visual navigation integrity monitoring is an important step toward reliable solutions and the future

**TABLE 1**  
Examples of Positioning Integrity Requirements for Safety-Critical Applications

Application	Integrity risk	Alert limit
CAT I landing <sup>1</sup>	$\leq 10^{-7}$	$\leq 10\text{m}$
CAT II/III landing	$\leq 10^{-9}$	$\leq 10\text{m}$
Autonomous driving	$\leq 10^{-6}$	$\leq 2.5\text{m}$
Train track change	$\leq 10^{-8}$	$\leq 2.5\text{m}$

standardization of techniques for safety-critical applications. In the following sections of this paper, we briefly introduce the basics of visual navigation techniques. Then, a brief review of integrity concepts and discussion on particular challenges for visual positioning integrity are presented, followed by a review of the current development states of vision integrity monitoring approaches. Furthermore, we propose a preliminary integrity description framework for visual positioning that stresses monitoring error sources in various domains. The framework focuses on feature-based methods given that there are additional unsolved challenges in integrity description of machine-learning-based approaches, which will be discussed in more detail in the following sections. Last but not least, we discuss the prospects for the research community to further develop integrity-oriented methods in the future.

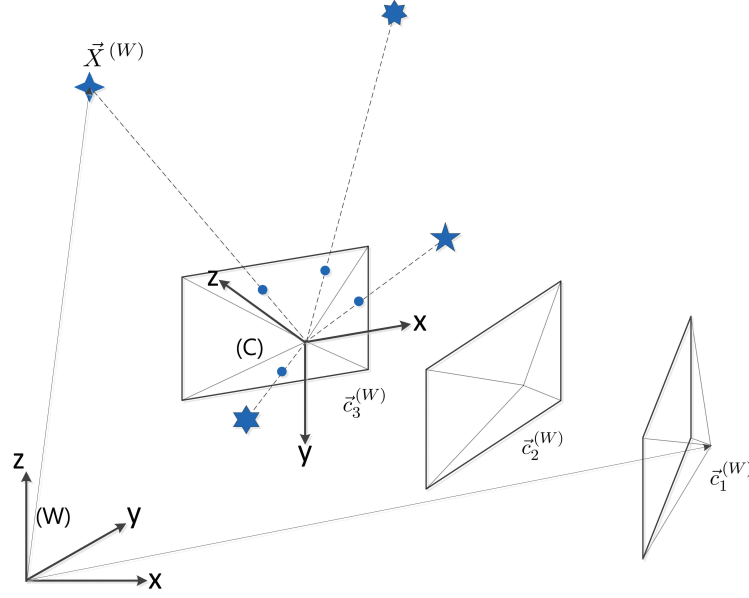
## 2 | THE BASICS OF VISUAL NAVIGATION

In this section, the basic concepts of visual navigation and common error sources are briefly introduced. For more detailed background content, one can refer to the classic book by Hartley and Zisserman, *Multiple View Geometry in Computer Vision* (2003), and more detailed review sections in the study by Zhu (2020). In the remainder of this paper, a superscript with parentheses <sup>( $\cdot$ )</sup> is used to denote the reference frame in which vectors are represented. Vectors and matrices are denoted with bold symbols, while scalars are not. Vectors such as  $\vec{\mathbf{X}}_j^{(W)} \in \mathbb{R}^3$  with geometric meanings are written with an arrow.

The raw sensor measurement of digital cameras is a discrete image  $\mathbf{I}$ , which represents an amplitude measure of the illuminance during the exposure time on the image plane  $\Omega \subset \mathbb{R}^2$ . For visual navigation applications, it is essential to extract geometric information from the luminance of the images, and the processing should be in real time. The gain afforded by using color information is usually insignificant compared to monochrome luminance, because processing colored images requires at least three times as much computational power as processing grayscale images with the same resolution. As a result, for navigation purposes, grayscale images are typically used (i.e.,  $\mathbf{I} \in \mathbb{Z}^{N_h \times N_w}$  denoting the amount of intensity values at each pixel with  $N_h$  and  $N_w$  as the number of the pixels in the dimensions of the height and the width, respectively). The intensity values in the image  $\mathbf{I}$  are noisy due to various error sources. We model measurement noise as additive noise on top of the error-free image  $\mathbf{I}_0$  as:

$$\mathbf{I} = \mathbf{I}_0 + \mathbf{n}_I \quad (1)$$

<sup>1</sup>Category I/II/III (CAT I/II/III) are defined by the ICAO (International Civil Aviation Organization) and FAA (Federal Aviation Administration) to categorize operations of precision instrument approach and landing. Different categories have distinct parameters including decision heights and visibility. For example, a CAT I landing has a decision height no lower than 200 feet (60 meters) and with a visibility of no less than 800 meters or a runway visible range of no less than 550 meters. A higher category operation has more stringent demands on the navigation system.



**FIGURE 1** Reference frames in visual positioning; the blue stars with different shapes illustrate four feature points in the 3D space, and the round dots are their 2D projection in the image plane (the projection rays are denoted by dashed lines).

The noise  $n_i$  is normally referred to as *photometric noise* in the literature, since it represents the raw error in the image intensity values.

The location of pixels in the image plane and the corresponding light sources follows projective geometry. A simple example with four points in the space projected to the image plane is shown in Figure 1. The origin of the camera reference frame (C) is set at the optical center. The projection model is dependent on the applied lens. A review of different camera models and selection methods can be found in the works by Polic et al. (2020) and Sturm et al. (2011). Most systems can be well approximated using the classic pinhole model (fisheye lenses are the only common optical system that may significantly depart from that). For the pinhole camera model, the two-dimensional (2D) position of an image point  $\mathbf{u}_i = [u_{i,x}, u_{i,y}]^T \in \Omega \subset \mathbb{R}^2$  can be related to the corresponding 3D point  $[X_x^{(C)}, X_y^{(C)}, X_z^{(C)}]^T$  as:

$$\mathbf{u}_i = \left[ f \frac{X_x^{(C)}}{X_z^{(C)}} + \frac{N_w}{2}, f \frac{X_y^{(C)}}{X_z^{(C)}} + \frac{N_h}{2} \right]^T \quad (2)$$

with  $f$  denoting the focal length. Potential lens distortions would be corrected using a model as described in studies by Fryer and Brown (1986) and Zhao et al. (2020).

The transformation between the camera frame (C) and the world reference frames (W), e.g., the local east-north-up (ENU) frame, is dependent on the position and attitude of the camera as shown in Figure 1. For a camera at position  $\bar{\mathbf{c}}^{(W)} \in \mathbb{R}^3$  and with attitude represented by  $\mathbf{R}_{(C \rightarrow W)} \in \mathbf{SO}(3)$  in the world frame, the projective geometry of a 3D point in the world frame is described by:

$$\tilde{\mathbf{u}}_i = d_i \begin{bmatrix} \mathbf{u}_i \\ 1 \end{bmatrix} = \mathbf{K} \left[ \mathbf{R}_{(W \rightarrow C)} \mid -\mathbf{R}_{(W \rightarrow C)} \bar{\mathbf{c}}^{(W)} \right] \begin{bmatrix} \tilde{\mathbf{X}}_i^{(W)} \\ 1 \end{bmatrix} = \mathbf{P} \tilde{\mathbf{X}}_i^{(W)} \quad (3)$$

where  $\tilde{\mathbf{X}}_i^{(W)} \in \mathbb{P}^3$  is expressed in homogeneous coordinates in the world frame, and  $\mathbf{K}$  is the intrinsic matrix that can be estimated by camera calibration methods

(Heikkila & Silven, 1997; Zhang, 2000). However, calibration is not perfect in practice. As a result, calibration bias remains in each measurement. The magnitude of the calibration bias is dependent on the applied optical system. In most cases, such calibration bias cannot be ignored and must be properly processed in integrity monitoring. By reorganizing Equation (3) with variables in Euclidean space, we denote the 2D coordinates of a point in the image by a function of the camera pose and the 3D location of the point as:

$$\mathbf{u}_i = \pi(\bar{\mathbf{X}}_i^{(W)}, \mathbf{x}) \quad (4)$$

where  $\mathbf{x} \in \mathbb{R}^6$  is the camera pose parameterized with a six degrees-of-freedom (DoF) vector.

The camera pose  $\mathbf{x}$  can be estimated given a set of associated 3D points  $\{\bar{\mathbf{X}}_i^{(W)}\}$  and their corresponding 2D projections in the image  $\{\mathbf{u}_i\}$ . The coordinates of the 3D points can be obtained by using a georeferenced landmark or a map database, or by estimating the depth of the consistently tracked points in visual odometry or SLAM approaches. If the 3D coordinates are obtained from a georeferenced map database, the pose of the camera can be estimated using measurements at each time snapshot. For SLAM-type approaches, the camera tracks the points when moving and estimates the 3D point depth by using techniques such as triangulation. In such cases, the pose estimation is relative over time, and the absolute positioning of the camera is available only if the camera position is known in the world frame at some time instant (e.g., at the starting point).

For visual odometry or SLAM methods, the error propagation in the depth estimation must be additionally considered, and the global scale must be estimated if a monocular camera is applied, normally with the aid of other sensors like inertial sensors (Lynen et al., 2020; Qin et al., 2018) or ranging measurements (Zhu et al., 2017, 2018). In both cases, a Bayesian filter like an extended Kalman filter (EKF) and smoothing method like incremental smoothing and mapping (iSAM; Kaess et al., 2011) can be applied to retain the smoothness and continuity of the estimation process.

There are generally three categories of relative pose estimation approaches: direct methods, indirect methods, and end-to-end machine-learning-based methods. The direct and indirect approaches are based on the aforementioned projective geometry model, and the machine-learning approaches implicitly extract geometric information from the images using data-based training. Engel et al. (2018) reviewed different categories of physics-model-based approaches. Direct methods utilize pixel intensity values directly to estimate the camera motion (e.g., see the state-of-the-art work [Engel et al., 2014, 2018; Gao et al., 2018]). It is assumed that the points corresponding to the visible pixels form continuous surfaces in the 3D space and their luminance is invariant over short time.

Indirect methods such as ORB-SLAM (Mur-Artal & Tardos, 2017) and SOFT-SLAM (Cvišić et al., 2017) first apply feature detectors to locate features of interest in the image (e.g., corners, edges, light blobs, or high-level features like objects detected by pattern recognition methods) so that a set of geometric measurements  $\{\boldsymbol{\mu}_i = S(\mathbf{I}) \subset \Omega | i = 1, \dots, N_p\}$  can be extracted from the image intensities and are matched with the 3D coordinates. It should be mentioned that though data-driven deep learning techniques are also applied to detect features (e.g., the SuperPoint feature proposed by DeTone et al. [2018]), these approaches are still feature-based rather than end-to-end, considering their difference in the main processing procedure.

End-to-end machine-learning-based methods apply deep learning techniques to estimate relative pose. The methods normally do not execute any explicit feature extraction or association. Instead, deep neural networks are trained using a huge set of training data with known ground truth camera poses to learn the relation between the measurement images and the camera poses. Important end-to-end approaches include VLocNet++ (Radwan et al., 2018), EssNet (Zhou et al., 2020), PoseNet (Kendall et al., 2015), and its variants (Kendall & Cipolla, 2016). A survey of this category of methods can be found in the review by Chen et al. (2020).

Since, in practice, it is very difficult to validate the lighting-invariance assumption of direct methods and the output integrity of the deep neural networks (the problem is described in the following sections), we focus on the feature-based visual navigation methods (i.e., indirect methods) for safety-critical applications. In the feature extraction process, the photometric noise  $\mathbf{n}_i$  in  $\mathbf{I}$  is propagated to the geometric error  $\mathbf{n}_i$  for the  $i$ -th point. The noisy measurement equation can be expressed as:

$$\boldsymbol{\mu}_i = \mathbf{u}_i + \mathbf{n}_i + \mathbf{b}_i = \pi_i(\mathbf{x}) + \mathbf{n}_i + \mathbf{b}_i \quad (5)$$

where  $\mathbf{b}_i$  is the nominal bias in the point location caused by calibration error. The  $N_p$  2D measurements need to be associated with the set of 3D points, i.e., for each  $\boldsymbol{\mu}_i$ , it would be associated with a 3D candidate  $\bar{\mathbf{X}}_{a_i}^{(W)}$ . Erroneous matching may occur in this process. If the association is correct,  $a_i = i$  for this feature.

By stacking all the  $N_p$  successfully associated feature points as illustrated in Figure 1 into a vector, the position and orientation of the camera can be estimated by solving the following nonlinear optimization problem iteratively:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\boldsymbol{\mu} - \pi(\mathbf{x})\|_{\Sigma_n}^2 \quad (6)$$

with:

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \vdots \\ \boldsymbol{\mu}_{N_p} \end{bmatrix}, \pi(\mathbf{x}) = \begin{bmatrix} \pi_1(\mathbf{x}) \\ \vdots \\ \pi_{N_p}(\mathbf{x}) \end{bmatrix}, \mathbf{n} = \begin{bmatrix} \mathbf{n}_1 \\ \vdots \\ \mathbf{n}_{N_p} \end{bmatrix} \quad (7)$$

$\Sigma_n$  is the covariance matrix of noise vector  $\mathbf{n}$ . This equation is solved by linearization.

Equation (6) is the fundamental optimization problem for visual pose estimation. It can be observed from its definition in Equations (3) and (4) that the measurement function  $\pi(\mathbf{x})$  is highly nonlinear. Therefore, a good initial estimate of  $\mathbf{x}$  is essential to ensure that the linearization in the iterative process does not lead to local optima. In noise-free cases,  $\mathbf{x}$  can be solved directly by using  $N_p > 3$  measurements  $\{\mathbf{u}_i | i = 1, \dots, N_p\}$ , which is called the Perspective-n-Point (PnP) problem. A review of PnP problem solutions is found in the study by Lu (2018). The state-of-the-art methods are EPnP (Lepetit et al., 2009) and the recently proposed EOPnP (Zhou & Kaess, 2019).

Figure 2 summarizes how the geometric information flows in the corresponding error space propagation in the fundamental processing chain of feature-based visual navigation methods. At the beginning of processing, the geometric information is hidden in the raw measurement image which contains a huge amount of information provided by the intensity values of all pixels and their spatial relations.

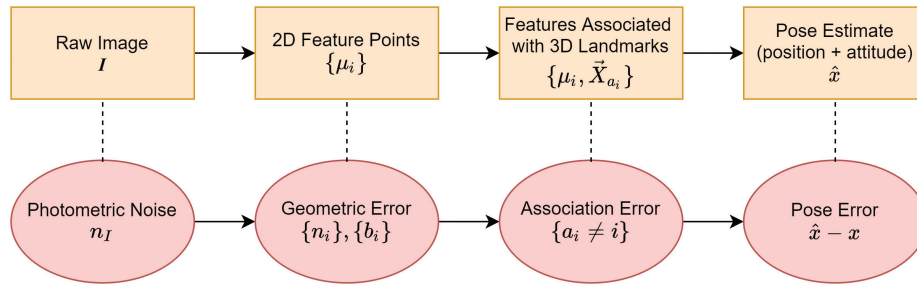


FIGURE 2 Fundamental procedure and error space propagation of visual positioning

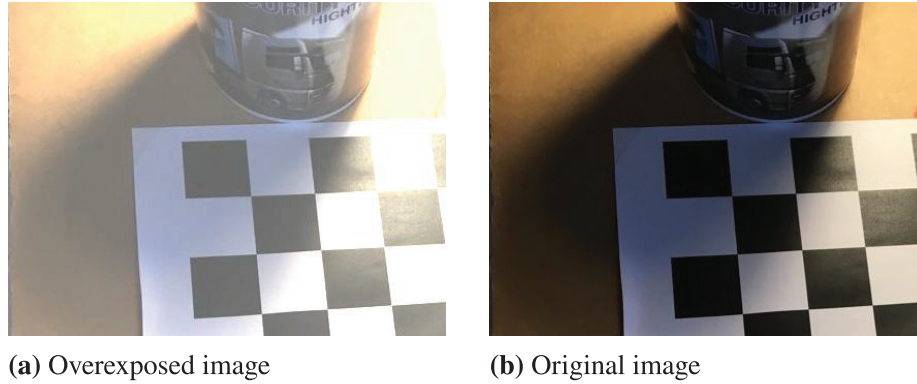


FIGURE 3 An example of overexposure effect

The corresponding error is photometric noise in the intensity domain. Then, feature points are detected from the raw image in order to extract the most important geometric information for positioning. The error can be expressed as a 2D geometric error in the feature locations. On top of the 2D error, binary association faults may occur when matching the 2D feature locations to the 3D point coordinates. With matched pairs of 2D features and 3D points, the 6-DoF camera pose can be estimated. The error in the pose estimate is propagated from the earlier stages and additional error sources can be introduced during the estimation process.

For a specific visual navigation task or algorithm, the exact processing chain varies and there can be additional important modules (sequential estimation, point triangulation for SLAM, image retrieval for map-based positioning, etc.). Nevertheless, the aforementioned procedure is fundamental for most feature-based visual navigation methods, and properly addressing the integrity for such a fundamental procedure is essential, but already a challenging task. Consequently, this short summary has focused on the common procedure without addressing certain aspects. The integrity concept discussion in the following portion of this work also has the same focus.

There are error sources in every domain in the fundamental process of visual positioning. Here we provide a few illustrative examples of some common fault sources. In the raw image (i.e., pixel intensity domain), photometric noise  $n_I$  can be modeled as a random variable caused by effects like sensor thermal noise and lens blur. On top of that, there are other erroneous effects in the intensity domain that affect the visual positioning performance, such as the overexposure shown in Figure 3 and motion blur shown in Figure 4.

Figure 5 illustrates an example of incorrect association in feature matching in visual navigation. The thin blue lines are successful matchings, and the thick red



(a) Blurry image



(b) Sharp image

**FIGURE 4** An example of motion blur effect; images from public data set (Seibold et al., 2017)



**FIGURE 5** An example of wrong association and outlier features; images from public data set (Geiger et al., 2012)



line is a significantly wrong association. If a pair of features with large biases in a 2D location or with an undetected wrong association are exploited in the pose estimation, the positioning result may also contain significant biases.

It should be mentioned that the faults caused by error sources may occur independently in different domains besides the error propagation in the processing phases. For example, the position estimation using measurements without any large bias or wrong association can still have significant error due to convergence failure in the nonlinear optimization. Therefore, error sources in the different phases of the algorithm must be taken into consideration in order to ensure the integrity of the pose estimation result. This topic will be discussed in more detail in following sections.

### 3 | INTRODUCTION OF INTEGRITY

#### 3.1 | Definitions of Integrity

In this section, we briefly review the basic definitions of *integrity* in the context of visual navigation. The integrity concept was first introduced by the civil aviation industry in order to quantify positioning and localization safety requirements. Integrity is a measure of trust that can be placed in the navigation system outputs, which is reflected by the probability defined as *integrity risk* (IR). Generally speaking, integrity risk is the probability that the position (or pose) estimation error is larger than the tolerable limit, while the navigation system, however, is not aware of the hazardous situation:

$$\text{IR} = p(\|\hat{\mathbf{x}} - \mathbf{x}\| > \text{AL}, m(\hat{\mathbf{x}}) \leq \text{AL}) \quad (8)$$

where  $\|\cdot\|$  denotes a metric of the positioning error, and AL denotes the alert limit (AL), which is defined as the maximum allowable error in the metric space. The AL is usually determined according to the operation requirements of the specific application. For instance, suppose an error metric is set as the absolute horizontal and vertical position error for civil aviation landing, and the AL is in the order of ten meters. For automobile applications such as autonomous driving, the AL should be significantly more stringent.  $m(\hat{\mathbf{x}})$  is a monitoring function which reflects the navigation system's belief in the estimation error. A functional navigation system for safety-critical applications must guarantee that  $\text{IR} \leq \text{IR}_{\max}$ , where  $\text{IR}_{\max}$  represents the maximum integrity risk requirements (as shown in Table 1) for various applications. The risky integrity events occur in practice, since normally the system and error models are not perfectly known, and the navigation algorithms are usually optimized for particular types of error, such as those based on the Gaussian noise assumption. A delicate design of the monitoring function is important for a navigation system to mitigate integrity risk.

In order to avoid integrity failures, the GNSS research community and the aviation industry developed integrity monitoring methods over the past decades. The integrity monitoring technique calculates, in real time, a conservative monitoring function as protection level (i.e.,  $m(\hat{\mathbf{x}}) = \text{PL}$ ). The PL is dependent on the measurement equation at the estimate point and on the continuity as well as availability requirements. It describes the worst-case error propagated to the position domain, given that the continuity and availability requirements are fulfilled. Ideally, PL should be an upperbound of the monitored error  $\|\hat{\mathbf{x}} - \mathbf{x}\|$  by modeling the error from various significant fault modes.

When  $PL \leq AL$ , the system operates nominally, provided the conservativeness in calculating PL. If  $PL > AL$ , the system triggers an alert within a required time (defined as time-to-alert), so that the vehicle can take operation (e.g., marking the navigation system as *unavailable* or changing the operational mode) before a hazardous situation occurs. Due to the long tail of real-life error distributions, there is a higher probability that the PL is smaller than the actual error. This case is defined as misleading information, i.e.,  $p_{MI} = p(PL \leq \|\hat{\mathbf{x}} - \mathbf{x}\|)$ . In the worst case, if the misleading information is provided by the monitor, and the actual error has exceeded the alert limit, it leads to a so-called *hazardous misleading information* (HMI) event. When designing a safety-critical navigation method, from the integrity aspect, a safe solution must fulfill that:

$$p_{HMI} = p(\|\hat{\mathbf{x}} - \mathbf{x}\| > AL, PL \leq AL) \leq IR_{max} \quad (9)$$

As shown in Figure 6, a tool named the Stanford-ESA Integrity Diagram has been designed to demonstrate the relations among positioning error, PL, and AL, and to illustrate the integrity of test data with convenience (Tossaint et al., 2007).

Besides the error monitoring, *fault detection and elimination* (FDE) methods are also of great importance for safe navigation solutions. The integrity risk after using  $K$  different FDE methods will become:

$$IR_{FDE} = p(\|\hat{\mathbf{x}} - \mathbf{x}\| > AL, m(\hat{\mathbf{x}}) \leq AL | FDE_1, \dots, FDE_K) \quad (10)$$

It should be mentioned that the integrity risk does not always decrease by adding more FDE methods. An improper FDE approach may introduce additional error sources and a corresponding integrity risk. As an example, if a set of data contains 60 percent unwanted outlier measurements with a constant large bias,

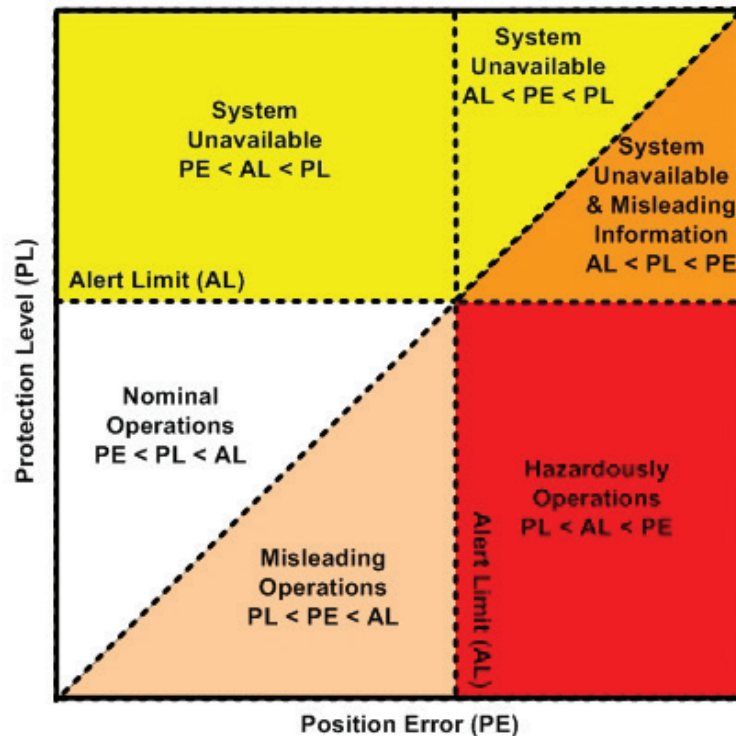


FIGURE 6 Concept of the Stanford-ESA Integrity Diagram (European Space Agency, 2011)

a simple FDE method that rejects the minority cluster of measurements would lead to worse results. Consequently, it is essential to quantify the error propagation in FDE processing and overbound the residual integrity risk with the aid of appropriate error models, so that it can be ensured that the integrity risk after FDE is still within the risk budget. This is a fundamental difference between integrity and other similar concepts.

The basics of visual navigation techniques have been under development for decades in computer vision and robotics research communities. In recent years, one of the most important aspects in the development of visual navigation and SLAM techniques has become *robustness*. Cadena et al. (2016) provide a good review of the development status of SLAM techniques and the future trend in robustness improvement. As a concept that is essential in robotics and other fields, robustness represents the capability of a system to recover from specific error and stay functional. The robustness concept has both similarities and differences with the integrity concept.

The function of detecting and recovering from failures (i.e., fault detection and elimination) is of great importance for both robustness and integrity. From a methodological point of view, both robust approaches and integrity monitoring approaches apply tools like statistical testing and outlier rejection in order to detect faults and improve system performance. Robustness targets the availability and continuity aspects of the system (i.e., the system should be able to recover when faults occur and deliver [ideally correct] solutions continuously, even under difficult conditions). Sometimes, methods to improve robustness may increase the integrity risk. Meanwhile, integrity focuses on the correctness of the solution with quantified criteria. In safety-critical applications, it is crucial to recognize a situation in which the estimates are unreliable instead of providing a best-effort output at high risk.

ORB-SLAM (Mur-Artal & Tardos, 2017) is, for example, widely used due to its robust performance and open-source availability. It includes different methods to optimize the robustness of the system. For instance, the algorithm initializes with two different geometrical models and exploits a statistical test to reject one of the hypotheses according to the residuals of the measurements. Similar techniques are also applied in integrity monitoring. However, the ORB-SLAM algorithm simply computes a score for both models and selects the higher score, without considering the probability that such a selection may be wrong (especially in the context of noisy measurements). The quantification of risk is the crucial part of the integrity assessment. A recently released version of the algorithm, ORB-SLAM3 (Campos et al., 2021), provides a new feature to re-localize the camera when motion tracking is interrupted with respect to the earlier established map. This feature is a typical example to demonstrate the efforts to maintain continuity. The risk of wrong re-localization is not quantified in the algorithm, and may lead to significant error.

### 3.2 | Particular Challenges for Vision Positioning Integrity

Adapting state-of-the-art integrity monitoring methods to visual navigation faces several challenges and requires new approaches to address them. GNSS receivers know the exact structure of each transmitted signal, while cameras are passive optical sensors that have little knowledge on what is captured during the exposure time. The detection of structures (features, objects, designed patterns, etc.) is challenging. A complex detection and association step is usually required.

In this process, outliers can occur due to various sources of error (e.g., repetitive patterns, occlusions, and moving objects) which all result in the impact that the applied feature is not the expected one or is not at the expected stationary position.

Abundant advanced detection methods have been and are still being developed for different problems. Quantifying errors remain an open issue, in particular in the case of deep-neural-network-based machine learning methods, which currently show the best performance in many detection problems. A *neural network* is capable of providing a distribution as its confidence on detection results, but the integrity of such output and the consequent impacts on the positioning error are difficult to monitor for data-driven models.

The research community has noticed such disadvantages and analyses have been undertaken accordingly to explain why the neural-network-based approaches usually have significant integrity issues, such as in the work from Sattler et al. (2019). In addition, many potential solutions are being investigated to solve the general output interpretability problem for deep learning. Current progress is reviewed in the survey papers by Bulusu et al. (2020) and Kumar et al. (2020). However, these approaches are still focusing on advanced training methods to improve the robustness of the model, and integrity concepts are not yet taken into consideration. A work worth noticing is from Sinha et al. (2018) that tries to quantify the robustness of the model, which can be a significant step toward more mature integrity analysis for deep learning in the future. In addition, the uncertainty modeling method proposed in the work by Kendall and Cipolla (2016) is also an inspiring approach to potentially provide a data-driven confidence in the output of a deep neural network, though its theoretical foundation is based on an assumed model in which the weights in the neural network are Gaussian random variables. Consequently, to the best of our knowledge, it still lacks tools to guarantee the integrity of the localization result for visual navigation methods based on data-driven machine learning. This is also the reason that the integrity description framework for end-to-end deep-learning-based approaches is not in the scope of detailed discussion in this work.

In addition, systematic biases are ubiquitous in feature extraction. For feature-based visual navigation methods, the exploited geometric measurements (i.e., feature locations) are obtained by feature detection from the measurement images. The computer vision algorithms need to distinguish the points of interest from the huge amount of other information in the image. The source of error in such detection processes is more like a kind of interference, rather than noise. As a result, errors are transformed from relatively easy-to-model photometric noise to complex geometric error, which may contain deterministic biases in addition to stochastic noise.

The introduction of biases from feature extraction occurs quite frequently. Large portions (sometimes even more than half) of the measurements are often biased. This depends on the applied algorithm. Figure 7 shows an illustrative example of such effects. In the image, the 20 strongest features are extracted using the Shi-Tomasi corner detector (Shi & Tomasi, 1994). The features are marked red with different levels of darkness to represent the strength measure. A feature marked with a lighter color is less significant for the feature detector. If detecting the high-contrast checkerboard corners is the goal, it can be seen that a few features from other parts of the image are more significant than some undetected checkerboard corners. The feature location error can be observed with ease for about half of the checkerboard corners. The systematic bias is more dominant than stochastic noise in the error, since the change of error is insignificant if the detection is

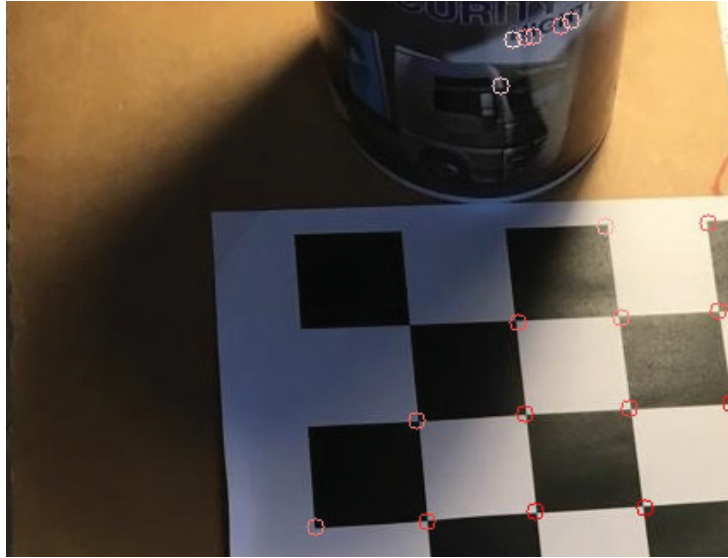


FIGURE 7 Different lighting conditions for local corners

repeated multiple times using the same detector. Consequently, innovative methods are required for bias detection and elimination.

Additionally, feature extraction algorithms are highly diverse. The choice of detection algorithm normally depends on specific tasks. The quantitative error model for a particular type of feature detector may not apply for other detectors. As a result, the error models become algorithm-specific and one cannot have a unified solution for all missions. Utilizing a proper error model for the specific task is one of the greatest challenges for visual navigation integrity monitoring.

Another significant challenge for vision integrity is that the performance strongly depends on lighting conditions. Figure 7 also provides an intuitive example to illustrate that. The error in feature detection is dependent on the local intensity distributions, which is influenced by lighting conditions. The geometric error distribution is different for the chessboard corner in the dark shadow and for the corners in the light. At the same time, the corner closest to the light source has a locally overexposed neighbor area, resulting in a different error distribution of the extraction result from the other corners with better contrast. Consequently, each measurement may have a unique error distribution dependent on local illuminance conditions. As a consequence, it is inappropriate to use a unified distribution to characterize the error everywhere in the image. This also makes a nominal error model (as the basis for statistical tests to detect faults) challenging. Some recent research such as the error model in the work by Zhu et al. (2019a) tries to solve this problem by extracting local intensity parameters and modeling the feature location error as a function of the parameters, achieving good performance for chessboard-like corners.

Additionally, the change of lighting conditions has a further impact on visual navigation performance. Since the view point or the light source may change, the intensity values in the images might vary over time. This could cause severe integrity issues for direct methods, which are based on the basic assumption that the illuminance of the scene stays constant over a short period of time. Consequently, if the lighting condition in the environment significantly changes with the view point or changes over short time periods, the direct methods are not a good choice.

The lighting changes also affect the performance of landmark association in map-based approaches.

The nonlinearity in the observation equation is another severe problem for visual positioning compared to GNSS. Due to the strong nonlinearity of the equations, visual positioning requires a good initial guess on the state estimates to prevent the optimization converging to a local minimum. Being trapped in a local minimum is often not easy to detect, since the residuals can be nearly as small as that of the global minimum. Other than the aforementioned points, the visual navigation processing complexity is high due to the millions of pixels in the measurement images, which also brings up challenges for real-time integrity monitoring.

Due to the above particularities of visual positioning techniques, one cannot simply transplant the GNSS integrity monitoring methods for camera-based positioning. Rather, we need a new framework to tackle specific problems.

## 4 | RECENT DEVELOPMENTS IN VISUAL POSITIONING INTEGRITY

As a new field that has not yet drawn much attention, the existing work on the integrity aspects of visual navigation is rather limited in number. We review some of the important publications on this topic in the following sections.

Mario and Rife (2010) proposed a simple integrity monitoring for camera-based lane detection. The method estimates the lane location with two independent vision algorithms and cross-validates the results to monitor the integrity of the lane detection output. An alert is triggered if the monitored parameter exceeds a pre-set threshold. Al Hage et al. (2019) proposed an approach in a similar context. The authors used GNSS measurements and visual measurements to track lane markers, while monitoring the residuals of the estimated position to remove faulty measurements. A protection level (PL) was calculated by exploiting the posterior covariance and a student's *t*-distribution overbound model.

Fu et al. (2015) adapted visual measurements to aid GNSS integrity monitoring for aircraft landing. The method first computed relative position between the camera and a landmark on the ground, and then transformed the result into additional synthetic range measurements, so that they could be combined with the GNSS measurements to exploit the state-of-the-art GNSS-based integrity monitoring approaches. Shytermeja et al. (2014) proposed an integrity monitoring architecture using GNSS, inertial measurement units (IMUs), and a fisheye camera for urban navigation. In the work, the camera is only used to check the GNSS signal line-of-sight in urban canyons.

The aforementioned approaches exploit cameras to solve particular navigation problems with integrity aspects. However, these methods have some limitations. The method in the work by Mario and Rife (2010) depended on ad-hoc parameters. For the other three reviewed publications, either the outputs from the camera were taken as nominal results for granted (Shytermeja et al., 2014), or the vision error was only monitored in a specific transformed domain (distance to lane in the approach by Al Hage et al. [2019] and virtual ranging in the method used by Fu et al. [2015]). As a result, misleading information may be provided to the navigation system in particular cases.

For example, it can be observed that, at some time instants, the sensor fusion error is even larger than the GNSS-only solution (e.g., in test results in the work by Al Hage et al. [2019]). This is probably due to the error model of the visual

positioning mismatching with reality and the integrity of the camera-only outputs not being properly monitored. In order to provide a navigation solution with sufficient integrity, the integrity of the visual navigation, itself, must be taken into consideration appropriately. As the importance of the topic has drawn more attention from the research community, a few pioneered and innovative methods addressing various aspects of the visual positioning integrity monitoring have been developed more recently.

Calhoun and Raquet (2016) and Calhoun et al. (2015) proposed a relative positioning method by matching rendered images with measurement images. An important innovative point of the work is that the image correspondence error was quantified and a protection level could be calculated for the relative position. Zhu and Taylor (2017) studied the integrity problem caused by correlated measurements in visual navigation. The authors propose exploiting inflated covariance estimation and the covariance intersection technique to obtain appropriate estimates.

Yang et al. (2018) investigated in detail the common causes of the ubiquitous feature-matching outliers in visual navigation and proposed a statistical error model for them. The authors provided an in-depth analysis of the scenarios that might result in feature association error in practice, and the performance limitation of some common outlier rejection methods was discussed. Based on the analysis, they applied a probabilistic data association filter exploiting the proposed outlier model to improve the integrity of vision-aided inertial navigation. Nevertheless, the work utilized some ad-hoc parameters (e.g. outlier percentage) for the specific test data sets, which required further improvement or explanation to generalize the model for universal scenarios.

Zhu et al. (2019b) proposed decoupling the geometric impact and the visual measurement error by using a six degrees-of-freedom dilution of precision, and proposed a feature error model for integrity monitoring purposes in their work (Zhu et al., 2019a). The approach modeled feature location error according to the local intensity distributions of the feature points for chessboard-like features (*X-junctions*). It provided a conservative scheme to reduce integrity risks when predicting measurement noise. Zhu et al. (2020) proposed a method to quantify the feature association error for integrity monitoring of visual navigation. This work quantified the probability of incorrect association when matching the measurements with known landmarks and exploited the probability to calculate the integrity risk caused by the fault mode.

Bhamidipati and Gao (2019, 2020) proposed a tightly coupled sensor fusion method with integrity-monitoring capability to integrate GNSS measurements and camera measurements. The direct visual SLAM technique was applied and sensor fusion was carried out with graph optimization. Outlier rejections were executed for the visual measurements and a PL for the position solution was calculated by considering error in measurements from both sensors. Wang et al. (2020) exploited similar outlier rejection methods for visual navigation error. In addition, the authors propose applying the widely used multiple hypothesis solution separation method (Blanch et al., 2007) from GNSS integrity monitoring to the feature point measurements for detecting multiple faults simultaneously.

A table (Table 2) is provided to summarize and compare the reviewed approaches with focus on visual positioning integrity aspects. It should be mentioned that current approaches are not yet an integrated mature solution to the visual navigation integrity monitoring problem, since most of them just consider a particular type of error source (e.g., large measurement biases).

**TABLE 2**  
Summary of Important Approaches on Visual Navigation Integrity

Reference	Tackled Problem	Sensor Setup	Integrity of Vision	Error Quantification	Monitored Vision Error Source
Mario and Rife (2010)	lane localization	camera	considered	w/ ad-hoc parameter	intensity domain detection error
Shytermeja et al. (2014)	urban navigation	camera+GNSS+IMU	not considered	only for GNSS	vision not monitored
Fu et al. (2015)	landing aircraft positioning	camera+GNSS	partially considered	addressed	high-level feature domain biases
Calhoun et al. (2015)	relative positioning	camera	considered	addressed	position domain nominal error
Zhu and Taylor (2017)	correlation of error	camera+IMU	considered	addressed	position domain covariance error
Yang et al. (2018)	error model	camera+IMU	considered	w/ ad-hoc parameter	feature association error
Al Hage et al. (2019)	lane localization	camera+GNSS	partially considered	only for GNSS	high-level feature domain biases
Zhu et al. (2019a 2019b)	error model & propagation	camera	considered	addressed	feature domain noise
Zhu et al. (2020)	landmark association	camera	considered	addressed	feature association error
Bhamidipati and Gao (2020)	urban navigation	camera+GNSS	considered	addressed	intensity domain biases
Wang et al. (2020)	urban navigation	camera	considered	addressed	feature domain biases



## 5 | AN INTEGRITY DESCRIPTION FRAMEWORK FOR VISUAL POSITIONING

As reviewed in the last section, many current research works (e.g., Al Hage [2019], Bhamidipati and Gao [2019, 2020], Fu et al. [2015], and Wang et al. [2020]), adapt state-of-the-art tools such as residual-based bias detection and multiple hypothesis solution separation developed for advanced receiver autonomous integrity monitoring (ARAIM; Blanch et al., 2015) in GNSS to visual positioning and camera-involved multi-sensor navigation. Given sufficient degrees of redundancy in the measurements, these methods can effectively detect the fault modes caused by large measurement biases by applying well-designed test statistics. In the state-of-the-art framework, for  $N$  measurement equations (defined in Equation [5] for visual positioning), the faults can be modeled as an  $N$ -dimensional bias vector added to the measurements:

$$\boldsymbol{\mu} = \boldsymbol{\pi}(\mathbf{x}) + \mathbf{b} + \mathbf{n} \quad (11)$$

and the probability that HMI can be decomposed using the  $2^N$  hypotheses (for each measurement, biased or not) as:

$$p(\text{HMI}) = \sum_{i=1}^{2^N-1} p(\text{HMI} | H_i) p(H_i) \leq \text{IR}_{\max} \quad (12)$$

where  $H_i$  denotes the hypothesis corresponding to the  $i$ -th fault mode and  $p(H_i)$  is its a priori probability. The model is sufficient for describing the main faults in GNSS-based positioning (in which context it is developed). However, there are more aspects to be taken into consideration in visual positioning, due to its passive sensing nature and high-nonlinearity in the measurement equation.

When associating features with 3D landmarks, incorrect associations can also result in integrity failures. Such faults are crucial when there are repetitive patterns in the scene, which are difficult to detect according to visual appearance alone. The positioning may have a large error if incorrect association occurs, even if all the feature locations were measured with good quality. For  $N_p$  points, there are maximum factorial of  $N_p$  association possibilities, which forms a permutation group. The association faults can be modeled as a permutation matrix multiplied to the measurement equations instead of a bias vector, as shown in the research by Zhu et al. (2020). In the work, the authors proposed a method to quantify landmark association error and calculate the corresponding integrity risk, which is a feasible integrity monitor for feature association error.

Furthermore, due to the strong nonlinearity in the visual positioning measurement equation (Equations [3] and [4]), pose estimation is sensitive to the initialization of the nonlinear optimization. With exactly the same feature location measurements and associations, the estimation results may differ for distinct initial guesses of the camera pose. Figure 8 illustrates a simple example of the impact of initialization on the pose estimation result. In the simulation, five feature points were visible from a camera located at the origin of the reference frame. The 2D measurements of the features are noise-free in the simulation. If the initial guess of the nonlinear optimization is close enough to the true position, the algorithm can generate the correct estimation results. However, if the initialization is not sufficiently accurate (e.g., at the position and orientation marked *initial* in the plot), the estimated camera pose is biased from the true value (as marked by *estimate* in the plot), even if all the measurements are noise-free and all the associations are

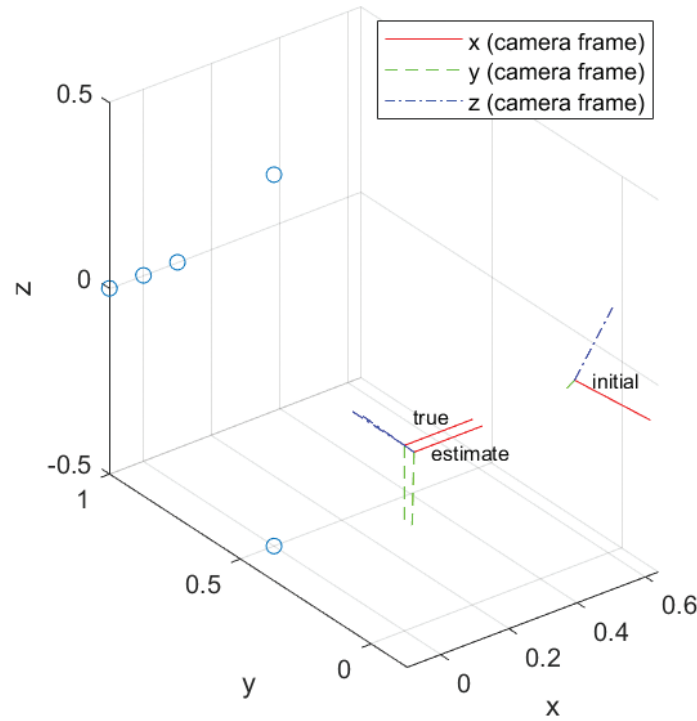


FIGURE 8 An example of initial guess impact on pose estimation; the blue circles illustrate the position of the five feature points used in the simulation.

correct. The error is caused by the convergence issue of nonlinear optimization. If the optimization solver is trapped in a local optimum, it may result in integrity failure, since, in some cases, it is challenging to distinguish local optima from the global optimum using the residuals. The impact of this important practical issue on positioning integrity has not been sufficiently discussed in literature.

In addition, it is important to identify the critical fault modes in visual positioning, since monitoring all possible fault modes may become computationally infeasible when the number of biased measurements is large. By monitoring the intensity domain, some crucial fault modes in visual navigation can be detected. For instance, Figure 9 illustrates an example of how the overexposure effect may influence feature extraction. The original image has a slightly overexposed area, and the top of the tree can be extracted as corner feature points (marked with red circles) for visual positioning. The lower image has artificially added brightness to simulate a stronger overexposure effect. It can be seen that there are also corners (marked with blue dots) at the top of the tree. However, the location of the corners has changed from the original image (red circles), since the edges of the corner are determined by the boundaries of the overexposure area. If the corners are consecutively used in tracking with exposure variations, the biases due to overexposure will affect the pose estimation of the camera. The corners look locally similar in both images and the biases are correlated for both corners, which makes the bias more challenging to detect in the feature domain. Nevertheless, the change in exposure can be easily detected by setting up a test to monitor the global intensity properties.

Another example is shown in Figure 10 to show the benefits of monitoring faults in the intensity domain. Due to motion blur, there are two sets of corners that are potentially extracted due to local intensity gradients. Assume that the blue set (marked with circles) are the correct locations at the time, and the red set (marked



(a) Original image

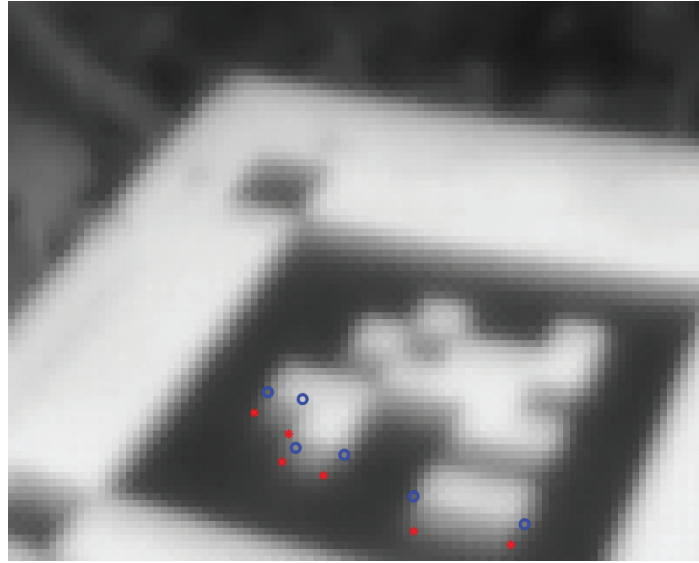


(b) Image with strong overexposure

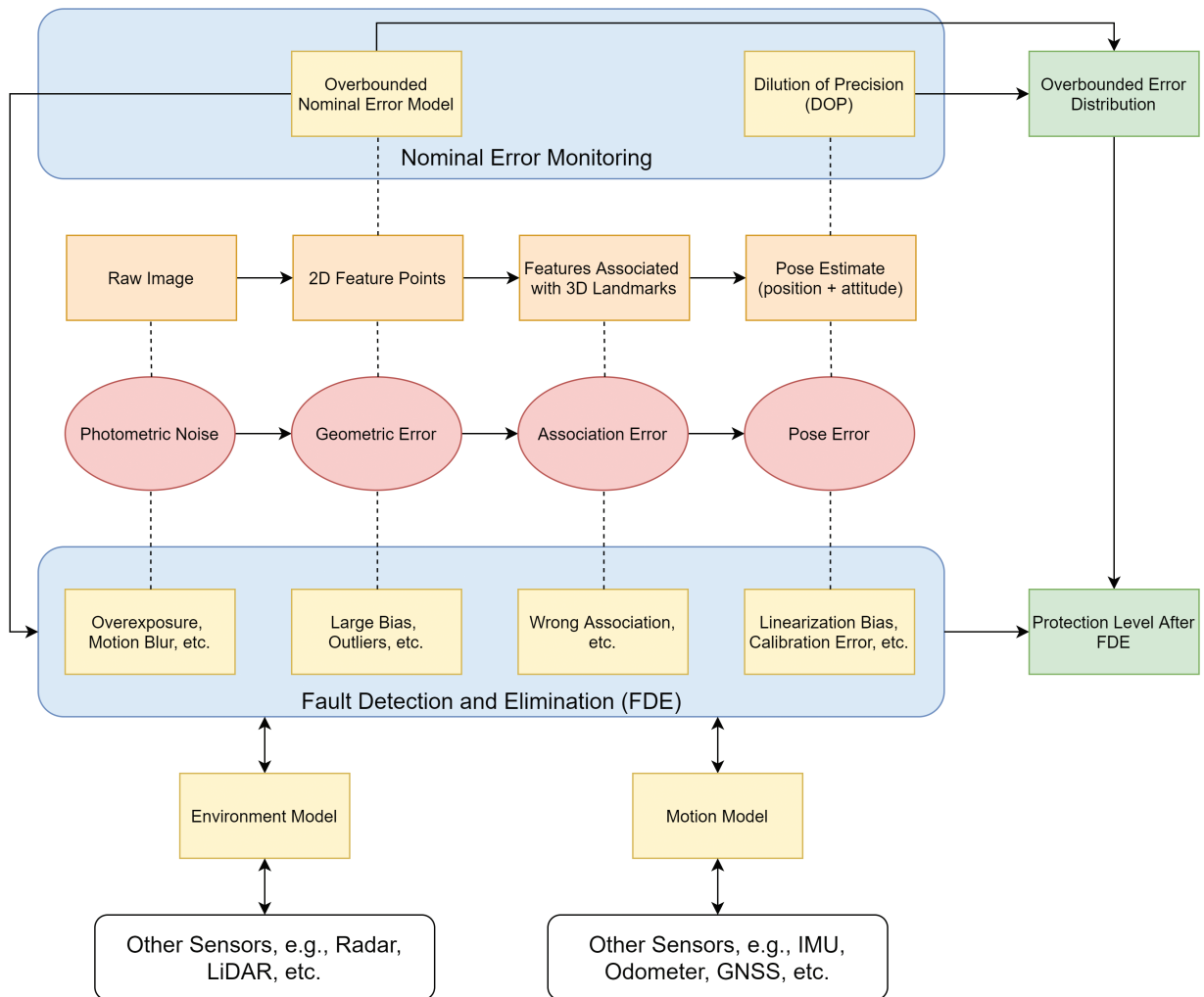
**FIGURE 9** Potential overexposure impacts; the red circles and blue dots denote the feature point location in the image with lower and higher overexposure respectively. Image from public data set (Delmerico et al., 2019).

with stars) is extracted due to the artificial intensity change due to motion blur. Since each point in the red set is biased by a similar translation from a corresponding blue set point, the association of features becomes ambiguous. As a result, the estimated position using the wrong set would be biased, and the error might not be reflected by the measurement residuals due to the strong correlation in the error. In the research by Zhu and Taylor (2017), the authors proposed a method to mitigate the risk caused by correlated error in measurements. A motion blur detector in the intensity domain is also a possible solution to avoid such integrity risks.

Considering the aforementioned specific effects for visual positioning, simply adapting the existing methods developed for GNSS is not the best solution for visual positioning integrity. Instead, it may be necessary to design the integrity framework smartly to include all the essential fault modes across different domains (not only the  $2^N$  additive bias hypotheses in Equation [12]). Therefore, we propose an integrity description framework illustrated in Figure 11 that monitors faults in multiple domains in the processing procedure, in order to inspire future development to cope with the specific challenges. Due to direct methods' strong reliance on uncontrollable lighting conditions, we focus on the feature-based methods in



**FIGURE 10** Potential motion blur impacts; the blue circles denote the original corner location, while the red stars denote the corners extracted due to motion blur. Image from public data set (Delmerico et al., 2019).



**FIGURE 11** Integrity framework for visual positioning

the context of safety-critical applications. The framework is based on the general core procedure of visual positioning shown in Figure 2.

In the diagram, the two rows in the middle (orange and red) demonstrate, respectively, the states' propagation and the error space propagation in the core procedure (more details are available in Figure 2). The potential faults are categorized according to the space in which the error lies at the corresponding phase of the processing. The integrity monitoring of the performance is separated into two parts: nominal error monitoring and fault detection and elimination. It should be mentioned that the separation is only for clarity. The two parts are tightly combined and function together in practice.

A major characteristic of the proposed framework is simultaneously monitoring the error in different domains during the procedure of the pose estimation. The nominal stochastic noise propagates from the intensity domain error  $\{n_j\}$  to feature location domain error  $\{i = 1, \dots, N_p \mid n_j\}$  and then to the position domain error  $\hat{\mathbf{x}} - \mathbf{x}$ . The nominal error propagation should be monitored in a conservative way (e.g., by using overbounded error model) so that the obtained error distribution of the estimates is not overoptimistic. The studies by Calhoun and Raquet (2016), Calhoun et al. (2015), and Zhu et al. (2019a, 2019b) are the result of development for nominal error integrity, which are capable of calculating PL in nominal operation for particular problems. Nevertheless, one challenge remaining in nominal error monitoring is that conservative measurement noise models are needed for different types of features. An error model is proposed in the work by Zhu et al. (2019a) for chessboard-like features (X-junctions), but there is still a lack of proper models for other types of features. For extracted features, a conservative geometric error model is required for monitoring nominal performance. An overbounded error distribution can be calculated to monitor the performance in nominal operation cases. By combining the model with the pose estimation results, a statistic test can be set up to validate the positioning result given the specific integrity and availability requirements, as demonstrated in the nominal error monitoring block in Figure 11. Meanwhile, the error model is also a basis for large bias detection in the geometric measurement domain.

The improvement of integrity in non-nominal cases relies on fault detection and elimination, as illustrated in the FDE block in Figure 11. For visual positioning, the faults can occur independently in various domains and may have different impacts on the positioning result. In the image intensity domain, there can be faults like overexposure or motion blur that can affect the following procedure. Such faults may cause failures in the following feature extraction step, which can be availability issues (e.g., not enough features are detected due to low contrast). At the same time, overexposure can also cause outliers or biases in the geometric measurements as in the example in Figure 9. Additionally, incorrect associations result in outliers and positioning error, as analyzed in detail by Yang et al. (2018) and Zhu et al. (2020). In pose estimation, errors such as linearization biases and imperfect calibration should also be monitored. The error introduced due to these model mismatches can result in overoptimistic error prediction. The FDE methods should monitor the errors in different domains and take the correlation and propagation of the error sources into consideration.

In order to detect faults, redundant measurements and quantified models (including the error model, environment model, and motion model) are necessary. It should be mentioned that although the deep-neural-network-based machine-learning methods lack integrity protection when used directly for positioning, they may play a significant role in the environment modeling that supports FDE. In a multi-sensor context, visual navigation can benefit from the information

from other sensors to obtain environmental and motion models. For example, inertial sensors can provide the vision system with a kinematic model, and radars and lidars can provide information of the surrounding environment. These models can be applied to monitor the integrity of visual positioning processing, as long as the integrity information of the model is taken into consideration. Meanwhile, the measurements and processing results from the vision system, as well as the corresponding integrity information, also contribute to the model, so that the integrity of the multi-sensor solution can be ensured. For this reason, the arrows are bidirectional in the diagram in Figure 11. It should be mentioned that when multiple sensors are used, imperfect time synchronization between the sensors would result in an additional error source in pose estimation. The bias caused by synchronization imperfection must also be considered by the protection level calculation.

According to the survey of existing development in vision integrity, it can be seen that currently implementations are available for a few elements in the framework. However, a mature solution is still missing for many blocks in the vision integrity framework. Most of the past developments focus on nominal error monitoring as well as bias and outlier detection from the feature location measurements. By proposing the multi-domain visual positioning integrity description framework, we want to raise attention to the development of innovative cross-domain integrity monitoring methods and consolidate the framework on a larger scale. More complete and advanced solutions still require an effort from the whole research community, so that the different elements in the framework can be combined and an integrated solution can be drawn.

## 6 | FUTURE TRENDS AND OPEN ISSUES

As mobility requirements grow, new applications such as autonomous driving and urban air mobility (air-taxis) demand high safety navigation solutions in challenging environments. As a consequence, the trend in navigation solutions is the usage of multiple sensor fusion with both high accuracy and high integrity, which is still an open problem. Various research communities have already started to pay great attention to developing reliable and safe algorithms in order to solve the problem. We believe that the integrity concept plays an important role in tackling the challenge, since it provides a quantified measure of risk for the navigation solution and the effectiveness has been verified in the years of usage in GNSS-based positioning for civil aviation.

Nevertheless, though cameras have significant advantages in applications, integrity monitoring is a challenging task due to the complexity of the relation between pixels in an image and pose estimation. Quantifying the error properly and detecting the incorrect results in all the visual navigation processing phases remains an open challenge. We hope that the decomposition proposed in this paper will contribute to the advancement of the solution. In this section, aspects that we believe would be important in further development are discussed.

### 6.1 | Investigation of Significant Fault Modes for Visual Navigation Methods

Current developments in visual navigation integrity are still focusing on the conventional fault modes of large measurement biases, for which the existing methods proposed for GNSS integrity monitoring can be adapted with ease. In order to

solve the crucial problems raised by the specific characters of the camera sensor, attention needs to be paid to the other vision-specific fault modes (e.g., the nonlinearity and convergence and error correlations). In addition, for measurement bias detection, there are also new challenges from visual positioning such as cases with a large number of biased measurements. Current approaches usually apply a random sample consensus (RANSAC; Fischler & Bolles, 1981) to cope with the issue, but the integrity risks in the random sampling process are not well investigated. New approaches may be required to solve the problem.

As mentioned in an earlier section, the computational complexity is another challenge that is more crucial for visual navigation integrity monitoring than for conventional GNSS-based solutions. For this direction, innovative methods that optimize the computational costs for real-time processing while ensuring the performance and safety are desired.

## 6.2 | The Importance of Error Quantification

One of the most powerful aspects of the integrity concept is its capability to ensure a quantified maximum risk probability. In order to achieve this goal, the error in all steps of processing must be quantified and their propagation needs to be understood. Nevertheless, properly quantifying the errors in 2D feature locations is definitely essential, but arguably the greatest challenge for integrity monitoring of feature-based visual navigation methods. Currently, most of the detection algorithms are based on a best-effort concept without providing any information on the error distribution of the results.

Furthermore, as shortly mentioned in Section 3.2, various feature detectors (SIFT [Lowe, 2004], SURF [Bay et al., 2006], ORB [Rublee et al., 2011], etc.) have been designed in the past years in order to fulfill the requirements for specific tasks in computer vision. However, detection repeatability is the main design focus of most detectors, and the processing procedures are usually complex and highly diverse in most approaches. The intricacy in processing makes a rigorous error characterization challenging, while the diversity in detection algorithms makes it infeasible to find a unified error model for all popular detectors. If an innovative feature detector can be designed, which provides an analytically quantifiable mapping between photometric information and the detected feature location while preserving sufficient detection repeatability, it would be a significant step toward mature integrity monitoring methods for visual navigation.

The feature error models that quantify error propagation in other processing steps are also necessary. For example, outlier rejection methods such as RANSAC and M-estimators (Meer et al., 1991) have been widely used in visual navigation over decades. Such tools require prior parameters dependent on the metadata of the measurements' data set, which are determined in an ad-hoc way today. According to given constraints on integrity risk, the appropriate determination of the parameters in the algorithms is definitely a non-trivial task.

## 6.3 | Re-Evaluation of Existing Algorithms from an Integrity Point of View

As computer vision techniques have been rapidly developing over the last few decades, there have been many powerful algorithms to cope with various problems and tasks. Nevertheless, the focus of algorithm development is normally on

improving availability and accuracy. An algorithm outperformed in accuracy is usually not valued as much as its more advanced competitors. However, higher accuracy and availability does not necessarily lead to better integrity. The complexity of a system usually makes fault detection and integrity monitoring more difficult. Consequently, if a vision algorithm provides a clear structure for error quantification and viable integrity checking options, small sacrifices in other performance factors such as accuracy would be acceptable, as long as the basic requirements on that aspect can be fulfilled. It would be of great value to revisit the existing visual navigation related algorithms with an integrity-oriented point of view and identify preferable solutions for safety-critical applications, which is not necessarily the state-of-the-art algorithms.

## 7 | CONCLUSION

The integrity aspect of visual navigation is an important topic for exploiting the technique in safety-critical applications. In this work, we provided a brief introduction to the basics of integrity concepts as well as the visual positioning core procedures, and reviewed the noticeable research works in the field. The development of vision integrity is still in its early stages. There are many particular challenges while also many opportunities for innovative research works. The further evolution of this topic with great potential calls for the attention and joint efforts of researchers in related fields.

## ACKNOWLEDGMENTS

The authors appreciate Markus Rippl and Daniel Gerbeth for their suggestions and supports.

## AUTHOR CONTRIBUTIONS

Chen Zhu proposed the main idea of the work and reviewed the literatures; Michael Meurer proposed part of the ideas; Christoph Günther proposed part of the ideas and the framework for sensor fusion.

## CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

## REFERENCES

- Al Hage, J., Xu, P., & Bonnifait, P. (2019). High integrity localization with multi-lane camera measurements. *2019 IEEE Intelligent Vehicles Symposium (IV)*, Paris, France. <https://doi.org/10.1109/ivs.2019.8813988>
- Bay, H., Tuytelaars, T., & Van Gool, L. (2006). SURF: Speeded up robust features. *Computer Vision – ECCV 2006*, Graz, Austria, 404–417. [https://doi.org/10.1007/11744023\\_32](https://doi.org/10.1007/11744023_32)
- Bhamidipati, S., & Gao, G. X. (2019). SLAM-based integrity monitoring using GPS and fish-eye camera. *Proc. of the 32nd International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS+ 2019)*. Miami, FL, 4116–4129. <https://doi.org/10.33012/2019.17117>
- Bhamidipati, S., & Gao, G. X. (2020). Integrity monitoring of Graph-SLAM using GPS and fish-eye camera. *NAVIGATION*, 67(3), 583–600. <https://doi.org/10.1002/navi.381>
- 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>
- Blanch, J., Walker, T., Enge, P., Lee, Y., Pervan, B., Rippl, M., Spletter, A., & Kropp, V. (2015). Baseline advanced RAIM user algorithm and possible improvements. *IEEE Transactions on Aerospace and Electronic Systems*, 51(1), 713–732. <https://doi.org/10.1109/taes.2014.130739>



- Bulusu, S., Kailkhura, B., Li, B., Varshney, P. K., & Song, D. (2020). Anomalous instance detection in deep learning: A survey. *42nd IEEE Symposium on Security and Privacy*. Princeton, NJ. <https://www.osti.gov/servlets/purl/1631092>
- Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I., & Leonard, J. J. (2016). Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on Robotics*, *32*(6), 1309–1332. <https://doi.org/10.1109/tro.2016.2624754>
- Calhoun, S. M., & Raquet, J. (2016). Integrity determination for a vision based precision relative navigation system. *2016 IEEE/ION Position, Location, and Navigation Symposium (PLANS)*, Savannah, GA. <https://doi.org/10.1109/plans.2016.7479713>
- Calhoun, S. M., Raquet, J., & Peterson, G. (2015). Vision-aided integrity monitor for precision relative navigation systems. *Proc. of the 2015 International Technical Meeting of the Institute of Navigation*, Dana Point, CA, 756–767. <https://www.ion.org/publications/abstract.cfm?articleID=12668>
- Campos, C., Elvira, R., Rodríguez, J. J. G., Montiel, J. M. M., & Tardós, J. D. (2021). ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and multimap SLAM. *IEEE Transactions on Robotics*, *37*(6), 1874–1890. <https://doi.org/10.1109/tro.2021.3075644>
- Chen, C., Wang, B., Lu, C. X., Trigoni, N., & Markham, A. (2020). *A survey on deep learning for localization and mapping: Towards the age of spatial machine intelligence*. <https://www.arxiv-vanity.com/papers/2006.12567>
- Cvišić, I., Česić, J., Marković, I., & Petrović, I. (2017). SOFT-SLAM: Computationally efficient stereo visual simultaneous localization and mapping for autonomous unmanned aerial vehicles. *Journal of Field Robotics*, *35*(4), 578–595. <https://doi.org/10.1002/rob.21762>
- Delmerico, J., Cieslewski, T., Rebecq, H., Faessler, M., & Scaramuzza, D. (2019). Are we ready for autonomous drone racing? the UZH-FPV drone racing dataset. *2019 International Conference on Robotics and Automation (ICRA)*, Montreal, Canada. <https://doi.org/10.1109/icra.2019.8793887>
- DeTone, D., Malisiewicz, T., & Rabinovich, A. (2018). Superpoint: Self-supervised interest point detection and description. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Salt Lake City, UT, 224–236. <https://doi.org/10.1109/cvprw.2018.00060>
- Engel, J., Koltun, V., & Cremers, D. (2018). Direct sparse odometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(3), 611–625. <https://doi.org/10.1109/tpami.2017.2658577>
- Engel, J., Schöps, T., & Cremers, D. (2014). LSD-SLAM: Large-Scale Direct Monocular SLAM. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer Vision – ECCV 2014*. Lecture Notes in Computer Science (Vol 8690, pp. 834–849). Springer. [https://doi.org/10.1007/978-3-319-10605-2\\_54](https://doi.org/10.1007/978-3-319-10605-2_54)
- European Space Agency. (2011). *Integrity*. Navipedia. <https://gssc.esa.int/navipedia/index.php/Integrity>
- EUSPA. (2021a). *Report on aviation user needs and requirements* (Report No. EUSPA-MKD-AV-UREQ-250287). European Union Agency for the Space Programme. [https://www.gsc-europa.eu/sites/default/files/sites/all/files/Report\\_on\\_User\\_Needs\\_and\\_Requirements\\_Aviation.pdf](https://www.gsc-europa.eu/sites/default/files/sites/all/files/Report_on_User_Needs_and_Requirements_Aviation.pdf)
- EUSPA. (2021b). *Report on rail user needs and requirements* (Report No. GSA-MKD-RL-UREQ-250286). European Union Agency for the Space Programme. [https://www.gsc-europa.eu/sites/default/files/sites/all/files/Report\\_on\\_User\\_Needs\\_and\\_Requirements\\_Rail.pdf](https://www.gsc-europa.eu/sites/default/files/sites/all/files/Report_on_User_Needs_and_Requirements_Rail.pdf)
- EUSPA. (2021c). *Report on road user needs and requirements* (Report No. GSA-MKD-RD-UREQ-250283). European Union Agency for the Space Programme. [https://www.gsc-europa.eu/sites/default/files/sites/all/files/Report\\_on\\_User\\_Needs\\_and\\_Requirements\\_Road.pdf](https://www.gsc-europa.eu/sites/default/files/sites/all/files/Report_on_User_Needs_and_Requirements_Road.pdf)
- Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, *24*(6), 381–395. <https://doi.org/10.1145/358669.358692>
- Fryer, J. G., & Brown, D. C. (1986). Lens distortion for close-range photogrammetry. *Photogrammetric Engineering and Remote Sensing*, *52*(1), 51–58. [https://www.asprs.org/wp-content/uploads/pers/1986journal/jan/1986\\_jan\\_51-58.pdf](https://www.asprs.org/wp-content/uploads/pers/1986journal/jan/1986_jan_51-58.pdf)
- Fu, L., Zhang, J., Li, R., Cao, X., & Wang, J. (2015). Vision-aided RAIM: A new method for GPS integrity monitoring in approach and landing phase. *Sensors*, *15*(9), 22854–22873. <https://doi.org/10.3390/s150922854>
- Gao, X., Wang, R., Demmel, N., & Cremers, D. (2018). LDSO: Direct sparse odometry with loop closure. *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Madrid, Spain, 2198–2204. <https://doi.org/10.1109/IROS.2018.8593376>
- 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. <https://doi.org/10.1109/cvpr.2012.6248074>
- Hartley, R., & Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511811685>

- Heikkila, J., & Silven, O. (1997). A four-step camera calibration procedure with implicit image correction. *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rick, 1106–1112. <https://doi.org/10.1109/cvpr.1997.609468>
- Joerger, M., Chan, F.-C., & Pervan, B. (2014). Solution separation versus residual-based RAIM. *NAVIGATION*, 61(4), 273–291. <https://doi.org/10.1002/navi.71>
- Kaess, M., Johannsson, H., Roberts, R., Ila, V., Leonard, J. J., & Dellaert, F. (2011). iSAM2: Incremental smoothing and mapping using the Bayes tree. *The International Journal of Robotics Research*, 31(2), 216–235. <https://doi.org/10.1177/02783649111430419>
- Kendall, A., & Cipolla, R. (2016). Modelling uncertainty in deep learning for camera relocalization. *2016 IEEE International Conference on Robotics and Automation (ICRA)*, Stockholm, Sweden, 4762–4769. <https://doi.org/10.1109/icra.2016.7487679>
- Kendall, A., Grimes, M., & Cipolla, R. (2015). PoseNet: A convolutional network for real-time 6-DOF camera relocalization. *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, 2938–2946. <https://doi.org/10.1109/iccv.2015.336>
- Kumar, A., Braud, T., Tarkoma, S., & Hui, P. (2020). Trustworthy AI in the age of pervasive computing and big data. *2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, Austin, TX. <https://doi.org/10.1109/percomworkshops48775.2020.9156127>
- Langley, R. B. (1999). The integrity of GPS. *GPS World*, 10(3), 60–63. <http://www2.unb.ca/gge/Resources/gpsworld.march99.pdf>
- Lepetit, V., Moreno-Noguer, F., & Fua, P. (2009). EPnP: An accurate o(n) solution to the PnP problem. *International Journal of Computer Vision*, 81(2). <https://doi.org/10.1007/s11263-008-0152-6>
- Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing*, 234, 11–26. <https://doi.org/10.1016/j.neucom.2016.12.038>
- 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>
- Lu, X. X. (2018). A review of solutions for perspective-n-point problem in camera pose estimation. *Journal of Physics: Conference Series*, 1087(5). <https://doi.org/10.1088/1742-6596/1087/5/052009>
- Lynen, S., Zeisl, B., Aiger, D., Bosse, M., Hesch, J., Pollefeys, M., Siegwart, R., & Sattler, T. (2020). Large-scale, real-time visual-inertial localization revisited. *The International Journal of Robotics Research*, 39(9), 1061–1084. <https://doi.org/10.1177/0278364920931151>
- Mario, C., & Rife, J. (2010). Integrity monitoring of vision-based automotive lane detection methods. *Proc. of the 23rd International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS 2010)*, Portland, OR, 245–255. <https://www.ion.org/publications/abstract.cfm?articleID=9152>
- Meer, P., Mintz, D., Rosenfeld, A., & Kim, D. Y. (1991). Robust regression methods for computer vision: A review. *International Journal of Computer Vision*, 6(1), 59–70. <https://doi.org/10.1007/bf00127126>
- 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>
- Polic, M., Steidl, S., Albl, C., Kukulova, Z., & Pajdla, T. (2020). Uncertainty based camera model selection. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA. <https://doi.org/10.1109/cvpr42600.2020.00603>
- Qin, T., Li, P., & Shen, S. (2018). VINS-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics*, 34(4), 1004–1020. <https://doi.org/10.1109/tro.2018.2853729>
- Radwan, N., Valada, A., & Burgard, W. (2018). VLocNet++: Deep multitask learning for semantic visual localization and odometry. *IEEE Robotics and Automation Letters*, 3(4), 4407–4414. <https://doi.org/10.1109/lra.2018.2869640>
- RTCA. (2004). *Minimum aviation system performance standards for local area augmentation system* (Report No. RTCA DO-245).
- 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>
- Sattler, T., Zhou, Q., Pollefeys, M., & Leal-Taixé, L. (2019). Understanding the limitations of CNN-based absolute camera pose regression. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA. <https://doi.org/10.1109/cvpr.2019.00342>
- Seibold, C., Hilsmann, A., & Eisert, P. (2017). Model-based motion blur estimation for the improvement of motion tracking. *Computer Vision and Image Understanding*, 160, 45–56. <https://doi.org/10.1016/j.cviu.2017.03.005>
- Shi, J., & Tomasi, C. (1994). Good features to track. *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, WA. <https://doi.org/10.1109/cvpr.1994.323794>

- Shytermeja, E., Garcia-Pena, A., & Julien, O. (2014). Proposed architecture for integrity monitoring of a GNSS/MEMS system with a fisheye camera in urban environment. *International Conference on Localization and GNSS*, Helsinki, Finland. <https://doi.org/10.1109/ICL-GNSS.2014.6934179>
- Sinha, A., Namkoong, H., Volpi, R., & Duchi, J. (2018). Certifying some distributional robustness with principled adversarial training. *6th International Conference on Learning Representations*, Vancouver, Canada.
- Sturm, P., Ramalingam, S., Tardif, J.-P., Gasparini, S., & Barreto, J. a. (2011). *Camera models and fundamental concepts used in geometric computer vision*, 6(1–2). <https://doi.org/10.1561/06000000023>
- Tossaint, M., Samson, J., Toran, F., Ventura-Traveset, J., Hernandez-Pajares, M., Juan, J. M., Sanz, J., & Ramos-Bosch, P. (2007). The Stanford – ESA integrity diagram: A new tool for the user domain SBAS integrity assessment. *NAVIGATION*, 54(2), 153–162. <https://doi.org/10.1002/j.2161-4296.2007.tb00401.x>
- Wang, S., Zhan, X., Fu, Y., & Zhai, Y. (2020). 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>
- Yang, C., Vadlamani, A., Soloviev, A., Veth, M., & Taylor, C. (2018). Feature matching error analysis and modeling for consistent estimation in vision-aided navigation. *NAVIGATION*, 65(4), 609–628. <https://doi.org/10.1002/navi.265>
- Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), 1330–1334. <https://doi.org/10.1109/34.888718>
- Zhao, H., Shi, Y., Tong, X., Ying, X., & Zha, H. (2020). A simple yet effective pipeline for radial distortion correction. *2020 IEEE International Conference on Image Processing (ICIP)*, Abu Dhabi, United Arab Emirates. <https://doi.org/10.1109/icip40778.2020.9191107>
- Zhou, L., & Kaess, M. (2019). An efficient and accurate algorithm for the perspective-n-point problem. *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Macau, China. <https://doi.org/10.1109/iros40897.2019.8968482>
- Zhou, Q., Sattler, T., Pollefeys, M., & Leal-Taixé, L. (2020). To learn or not to learn: Visual localization from essential matrices. *2020 IEEE International Conference on Robotics and Automation (ICRA)*, Paris, France. <https://doi.org/10.1109/icra40945.2020.9196607>
- Zhu, C. (2020). *Cooperative vision for swarm navigation* [Doctoral dissertation, Technische Universität München]. <https://mediatum.ub.tum.de/doc/1486567/1486567.pdf>
- Zhu, C., Giorgi, G., & Günther, C. (2017). Planar pose estimation using a camera and single-station ranging measurements. *Proc. of the 30th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS+ 2017)*, Portland, OR. <https://doi.org/10.33012/2017.15219>
- Zhu, C., Giorgi, G., & Günther, C. (2018). 2D relative pose and scale estimation with monocular cameras and ranging. *NAVIGATION*, 65(1), 25–33. <https://doi.org/10.1002/navi.223>
- 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. (2019a). 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>
- Zhu, C., Steinmetz, C., Belabbas, B., & Meurer, M. (2019b). Six degrees-of-freedom dilution of precision for integrity of camera-based localization. *Proc. of the 32nd International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS+ 2019)*, Miami, FL. <https://doi.org/10.33012/2019.17020>
- Zhu, Z., & Taylor, C. (2017). Conservative uncertainty estimation in map-based vision-aided navigation. *IEEE Transactions on Aerospace and Electronic Systems*, 53(2), 941–949. <https://doi.org/10.1109/taes.2017.2667278>

**How to cite this article:** Zhu, C., Meurer, M., & Günther, C. (2022) Integrity of visual navigation—Developments, challenges, and prospects. *NAVIGATION*, 69(2). <https://doi.org/10.33012/navi.518>