

REVIEW

Integrity of Visual Navigation - Developments, Challenges, and Prospects

Chen Zhu*¹ | Michael Meurer^{1,2} | Christoph Günther^{1,3}

¹Institute of Communications and Navigation, German Aerospace Center (DLR), Oberpfaffenhofen, Germany

²Chair of Navigation, RWTH Aachen University, Aachen, Germany

³Chair of Communications and Navigation, Technical University of Munich (TUM), Munich, Germany

Correspondence

Chen Zhu. Email: Chen.Zhu@dlr.de

Summary

Camera based visual navigation has great potentials for various applications, especially in satellite-signal-degenerated environments. However, the lack of integrity protections 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 since decades for satellite navigation, and are playing an essential role for 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 some challenges have been investigated. These developments are first steps towards a reliable visual positioning framework with integrity monitoring capabilities. In this paper, we review the current status, the particular challenges, and the 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, vision, safety, robust

1 | INTRODUCTION

Global Navigation Satellite Systems (GNSS) 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. Some 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 by using GNSS technologies only. A multiple sensor fusion is necessary to deliver sufficient navigation performance as required by such applications.

Visual navigation using cameras as part of a multi-sensor solution offer a number of advantages in such applications. Firstly, the number of visual cues for camera-based localization algorithms is large in those 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 allows to transform 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 easiest to interpret by humans both for interaction and for verification purposes. Visual Simultaneous Localization And Mapping (SLAM) is the key element in the estimation process. It estimates the relative pose and a map simultaneously, and is widely used in robotics in the past decades (Cadena et al., 2016). Last but not least, digital cameras do not cost much and are easy to integrate in 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 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, etc. already. However, the challenges remain for safety critical applications such as autonomous driving or urban air mobility (air-taxi), with the lack of a mature integrity monitoring framework being the most significant obstacle.

In a safety of life context, the performance of a navigation systems is characterized by four aspects: accuracy, availability, continuity, and integrity (Langley, 1999). According to the definition from 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 the safety, the navigation system needs to monitor the integrity of the parameters, usually by setting up some specific test statistics, and conservatively calculates a protection level (PL) which guarantees that inside the bound of PL, the occurrence probability of undetected error does not exceed the maximum allowed integrity risk. By comparing the PL value with an alert limit (AL) set according to the 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 so that false alerts occur frequently, the continuity of the navigation solution will be interrupted. In practice, the PL is calculated according to the maximum allowed 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 the standards and reports (EUSPA, 2021a 2021b 2021c; RTCA, 2004).

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

TABLE 1 Examples of positioning integrity requirements for safety-critical applications

In order to satisfy both the accuracy requirements and the 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 sensors and their corresponding processing procedures. For GNSS, the core integrity monitoring methods are already well investigated and applied in practice for years (Blanch et al., 2015; Joerger et al., 2014). However, for camera-based navigation techniques, a common agreement on the framework is still missing. To our point of view, developing a framework for visual navigation integrity monitoring is an important step towards reliable solutions and future standardization of the techniques for safety-critical applications. In the following sections of this paper, we will first briefly introduce the basics of visual navigation techniques. Then, a brief review of the integrity concepts and discussions 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 on 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 with more details in the following sections. Last but not least, we discuss the prospects for the research community to further develop the integrity-oriented methods in the future.

2 | BASICS OF VISUAL NAVIGATION

In this section, the basics of visual navigation and common error sources are very briefly introduced. For more detailed background contents, one can refer to the classic book (Hartley & Zisserman, 2003) and more detailed review sections in the study (C. Zhu, 2020). In the remainder of this paper, a superscript with parentheses ^(\cdot) is used to denote the reference frame in which the vector is represented. Vectors and matrices are

¹Category I/II/III (CAT I/II/III) are defined by 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, CAT I landing has a decision height not lower than 200 feet (60 meters) and with either a visibility of not less than 800 meters or a runway visible range of not less than 550

meters. A higher category operation has more stringent demands on the navigation system.

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 real-time. The gain using the color information is usually insignificant as compared to monochrome luminance, while processing colored images requires at least three times as much computational power as grayscale image 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 the measurement noise as additive noise on top of the error-free image \mathbf{I}_0 as:

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

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 Fig. 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 work (Polic et al., 2020; 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 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 (Fryer & Brown, 1986; 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 Fig. 1. For a camera at position $\vec{\mathbf{c}}^{(W)} \in \mathbb{R}^3$ 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

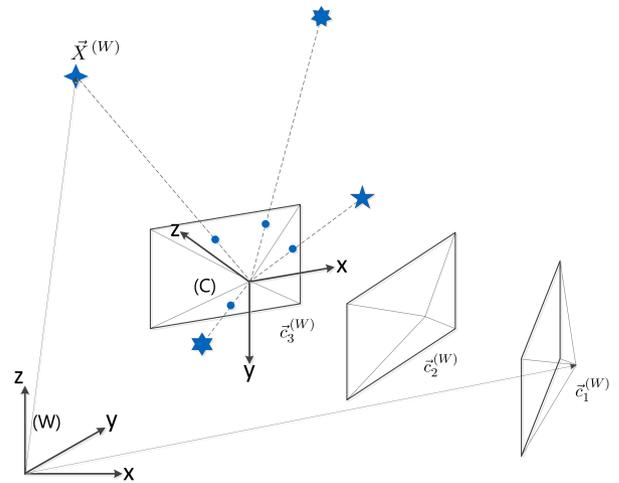


FIGURE 1 Reference frames in visual positioning. The blue stars with different shapes illustrate 4 feature points in the 3D space, and the round dots are their 2D projection in the image plane. The projection rays are denoted by dashed lines.

is described by

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

where $\vec{\mathbf{X}}_i^{(W)} \in \mathbb{P}^3$ 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). The calibration will not be perfect in practice. As a result, a calibration bias will remain 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 needs to be properly processed in integrity monitoring. By reorganizing Eqn. (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

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

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

The camera pose \mathbf{x} can be estimated given a set of associated 3D points $\{\vec{\mathbf{X}}_i^{(W)}\}$ 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 points 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 (C. Zhu et al., 2017 2018). In both cases, a Bayesian filter like extended Kalman filter (EKF) and smoothing methods like iSAM (incremental smoothing and mapping) (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 the geometric information from the images using data-based training. Engel et al. (2018) reviewed different categories of physic model based approaches. The direct methods utilize the pixel intensity values directly to estimate the camera motion, e.g., see the state-of-the-art work (Engel et al., 2018 2014; Gao et al., 2018). It is assumed that the points corresponding to the visible pixels form continuous surfaces in the 3D space and the luminance of them is invariant over short time. The 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 interests in the image, e.g., corners, edges, light blobs, or high-level features like objects detected by pattern recognition methods, etc., so that a set of geometric measurements $\{\boldsymbol{\mu}_i = \mathcal{S}(\mathbf{I}) \subset \Omega | i = 1, \dots, N_p\}$ are 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 instead of end-to-end learning methods considering their difference in the main processing procedure. The end-to-end machine learning based methods apply deep learning techniques to estimate the 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 (Q. 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 (Chen et al., 2020). Since in practice it is very difficult to validate the lighting-invariance assumption

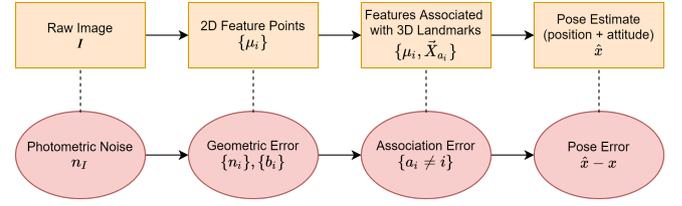


FIGURE 2 Fundamental procedure and error space propagation of visual positioning.

of the direct methods and the output integrity of the deep neural networks (the problem will be 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, etc. The N_p two dimensional 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 $\vec{\mathbf{X}}_{a_i}^{(W)}$. 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 Fig. 1 into a vector, the position and orientation of the camera can be estimated by solving the following nonlinear optimization problem iteratively

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

with

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

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

Eqn. (6) is the fundamental optimization problem for visual pose estimation. It can be observed from its definition in Eqn. (3) and (4) that the measurement function $\boldsymbol{\pi}(\mathbf{x})$ 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 Perspective-n-Point (PnP) problem. A review of PnP problem solutions is found in the study (Lu, 2018). The state-of-the-art methods are EPnP (Lepetit et al., 2009) and the recently proposed EOPnP (L. Zhou & Kaess, 2019).

Fig. 2 summarizes how the geometric information flows and the corresponding error space propagation in the fundamental processing chain of feature-based visual navigation

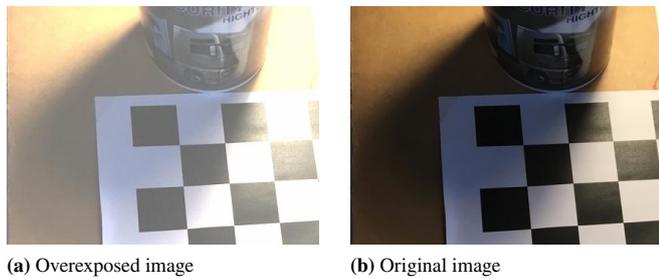


FIGURE 3 An example of overexposure effect.

methods. At the beginning of processing, the geometric information is hidden in the raw measurement image which contains huge amount of information provided by the intensity values of all pixels and their spatial relations. The corresponding error is photometric noise in intensity domain. Then, features points are detected from the raw image in order to extract most important geometric information for positioning. The error can be expressed as 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 points 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, e.g., sequential estimation, point triangulation for SLAM and 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 fundamental procedure is essential but already a challenging task. Consequently, this short summary has focused on the common procedure without addressing the particular aspects. The integrity concept discussions in the following part of this work also have 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 modelled as a random variable caused by effects like sensor thermal noise and lens blur, etc. On top of that, there are other erroneous effects in the intensity domain that affect the visual positioning performance, such as overexposure as shown in Fig. 3 and motion blur as shown in Fig. 4, etc.

Fig. 5 illustrates an example of wrong association in feature matching in visual navigation. The thin blue lines are successful matchings, and the thick red line is a significant wrong association. If a pair of feature with large biases in 2D location or with undetected wrong association are exploited in the pose



(a) Blurry image



(b) Sharp image

FIGURE 4 An example of motion blur effect. Images from public dataset (Seibold et al., 2017).



FIGURE 5 An example of wrong association and outlier features. Images from public dataset (Geiger et al., 2012).

estimation, the positioning result may also contain significant biases.

It should be mentioned that the faults caused by the 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, the error sources in all 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 is first introduced by civil aviation industry, in order to quantify the 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, i.e.,

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

where $\|\cdot\|$ denotes a metric of the positioning error, and AL denotes alert limit (AL), which is defined as the maximum allowed error in the metric space. The AL is usually determined according to the operation requirements of the specific application. For instance, the 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 $IR \leq IR_{max}$, where IR_{max} is the maximum integrity risk requirements, e.g., 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, e.g., based on 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 in the past decades. Integrity monitoring technique calculates in real-time a conservative monitoring function as protection level, i.e., $m(\hat{\mathbf{x}}) = PL$. 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 modelling 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

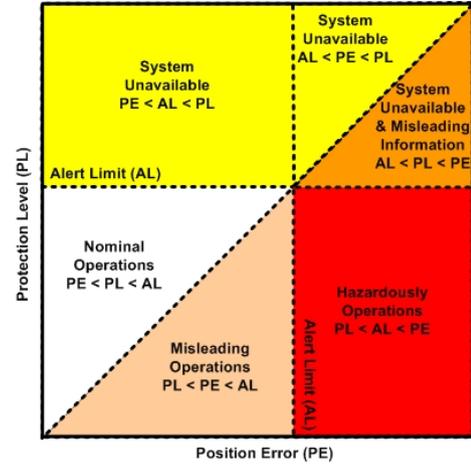


FIGURE 6 Concept of Stanford-ESA integrity diagram (European Space Agency, 2011).

will trigger 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 hazardous situation occurs. Due to the long tail of the real-life error distributions, there is a 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 the 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 Fig. 6, a tool named as Stanford-ESA 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 the corresponding integrity risk. As a toy example, if a set of data contains 60 percent unwanted outlier measurements with a constant large bias, a simple FDE method that rejects the minority cluster of measurements will lead to worse results. Consequently, it is essential to quantify the error propagation in the FDE processing and to 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 developments and improvements since 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 is the 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 to stay functional. The robustness concept has both common points and differences with the integrity concept.

As a common part, 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 apply tools like statistical testing and outlier rejection, etc., in order to detect faults and improve the system performance. Robustness targets at 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, the methods to improve robustness may increase the integrity risk. Meanwhile, integrity focuses on the correctness of the solution with quantified criterion. In safety critical applications, it is crucial to recognize 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 hypothesis 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 select 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 be able to re-localize the camera when the 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. GNSS receivers know the exact structure of the transmitted signal, while cameras are passive optical sensors that have little knowledge on what is captured during the exposure time. The detection of structures (e.g., 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, moving objects, etc., which all result in the impact that the applied feature is not the expected one or does not locate at the expected stationary position. Abundant advanced detection methods have been and are still being developed for different problems. Quantifying errors remains 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 the detection results, but the integrity of such output and consequent impacts on the positioning error is difficult to monitor for the data driven models. The research community has noticed such disadvantage and analyses are undertaken to explain the reason behind, such as in the work from Sattler et al. (2019). In addition, many potential solutions are being investigated to solve the general problem for deep learning. The current progress is reviewed in the survey papers (Bulusu et al., 2020; Kumar et al., 2020)). However, the 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 towards more mature integrity analysis for deep learning in the future. In addition, the uncertainty modelling method proposed in the work (Kendall & 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 that 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 discussions in this work.

In addition, systematic biases are ubiquitous in feature extraction. For feature based visual navigation methods, the exploited geometric measurements, i.e., the feature locations, are obtained by feature detection from the measurement images. The computer vision algorithms need to distinguish

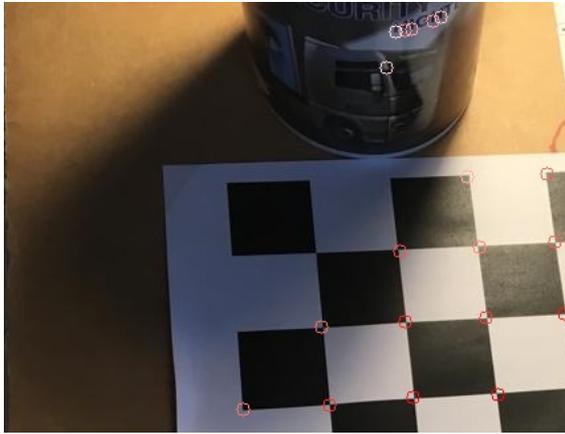


FIGURE 7 Different lighting conditions for local corners.

the points of interest from the huge amount of other information in the image. The source of error in such detection processes is more a kind of interference rather than noise. As a result, errors are transformed from relatively easy-to-model photometric noise to the 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. Fig. 7 shows an illustrative example of such effects. In the image, the 20 strongest features are extracted using Shi-Tomasi corner detector (Shi & Tomasi, 1994). The features are marked with red color with different darkness to represent the strength measure. A feature marked with 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 part 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 repeated multiple times using the same detector. Consequently, innovative methods are required for bias detection and elimination.

Additionally, the feature extraction algorithms are highly diverse. The choice of detection algorithm normally depends on the 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 the lighting conditions. Fig. 7 also provides an intuitive example to illustrate that. The

error in the feature detection is dependent on the local intensity distributions, which is influenced by the lighting conditions. The geometric error distribution is different for the checkerboard corner in the dark shadow and for the corners in the light. At the same time, the corner close 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 the 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 (C. Zhu et al., 2019a) tries to solve this problem by extracting the local intensity parameters and modelling the feature location error as a function of the parameters and has achieved good performance for checkerboard-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 may vary over time. This may cause severe integrity issues for direct methods, which are based on the basic assumption that the illuminance of the scene stays constant within a short period of time. Consequently, if the lighting condition in the environment may significantly change with the view point or may change 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. Besides 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. We rather need a new framework tackling the specific problems.

4 | RECENT DEVELOPMENTS OF 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.

Mario & 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-validate 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 similar contexts. The authors use GNSS measurements and visual measurements to track lane markers, while monitoring the residuals of the estimated position to remove faulty measurements. A protection level is calculated by exploiting the posterior covariance and a Student's t-distribution overbound model.

Fu et al. (2015) adapt the visual measurements to aid GNSS integrity monitoring for aircraft landing. The method first computes relative position between the camera and a landmark on the ground, and then transform the result into additional synthetic range measurements, so that they can 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 unit (IMU) 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 aid the solving of a particular navigation problem with integrity aspects. However, these methods have some limitations. The method in the work (Mario & Rife, 2010) depends on some ad-hoc parameters. For the other three reviewed publications, either the outputs from the camera are taken as nominal results for granted (Shytermeja et al., 2014), or the vision error is only monitored in a specific transformed domain (distance to lane in the approach (Al Hage et al., 2019) and virtual ranging in the method (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 (Al Hage et al., 2019). This is probably due to the error model of the visual positioning is mismatching with the reality and for the integrity of the camera-only outputs is not 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 recently.

Calhoun & Raquet (2016); 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 is quantified and a protection level can be calculated for the relative position. Z. Zhu & Taylor (2017) study the integrity problem caused by correlated measurements in visual navigation. The authors propose to exploit 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 the outliers. The authors provided an in-depth analysis of the scenarios that may result in feature association error in practice, and the performance limitation of some common outlier rejection methods is discussed. Based on the analysis, they apply a Probabilistic Data Association Filter (PDAF) 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 datasets, which requires further improvement or explanation to generalize the model for universal scenarios.

C. Zhu et al. (2019b) proposed to decouple 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 the work (C. Zhu et al., 2019a). The approach models the feature location error according to the local intensity distributions of the feature points for chessboard-like features ('X'-junctions). It provides a conservative scheme to reduce integrity risks when predicting measurement noise. C. Zhu et al. (2020) proposed a method to quantify the feature association error for integrity monitoring of visual navigation. This work quantifies the probability of wrong association when matching the measurements with known landmarks, and exploits the probability to calculate the integrity risk caused by the fault mode.

Bhamidipati & 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 is applied and the sensor fusion is carried out with graph optimization. Outlier rejections are executed for the visual measurements and a protection level for the position solution is calculated by considering error in measurements from both sensors. Wang et al. (2020) exploit similar outlier rejection methods for visual navigation error. In addition, the authors propose to apply widely used Multiple Hypothesis Solution Separation (MHSS) 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.

5 | AN INTEGRITY DESCRIPTION FRAMEWORK FOR VISUAL POSITIONING

As reviewed in the last section, many current research works, e.g., Al Hage et al. (2019); Bhamidipati & Gao (2019 2020); Fu et al. (2015); Wang et al. (2020), adapt state-of-the-art tools such as residual based bias detection and Multiple Hypothesis Solution Separation (MHSS) developed for ARAIM (Advanced Receiver Autonomous Integrity Monitor) (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 Eqn. (5) for visual positioning), the faults can be modeled as an N -dimensional bias vector added to the measurements, i.e.,

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

and the probability of 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 the features with 3D landmarks, wrong 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 the visual appearance. The positioning may have large error if wrong association occurs, even if all the feature locations are measured with good quality. For N_p points, there are maximum $N_p!$ association possibilities, which forms a permutation group. The association faults

Reference	Tackled Problem	Sensor Setup	Integrity of Vision	Error Quantification	Monitored Vision Error Source
Mario & 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
Z. Zhu & 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
C. Zhu et al. (2019a 2019b)	error model & propagation	camera	considered	addressed	feature domain noise
C. Zhu et al. (2020)	landmark association	camera	considered	addressed	feature association error
Bhamidipati & Gao (2020)	urban navigation	camera+GNSS	considered	addressed	intensity domain biases
Wang et al. (2020)	urban navigation	camera	considered	addressed	feature domain biases

TABLE 2 Summary of important approaches on visual navigation integrity

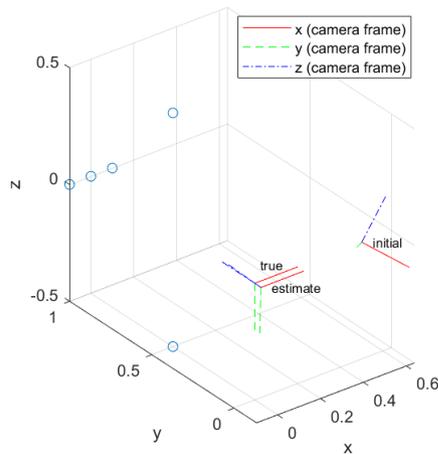


FIGURE 8 An example of initial guess impact on pose estimation. The blue circles illustrate the position of the 5 feature points used in the simulation.

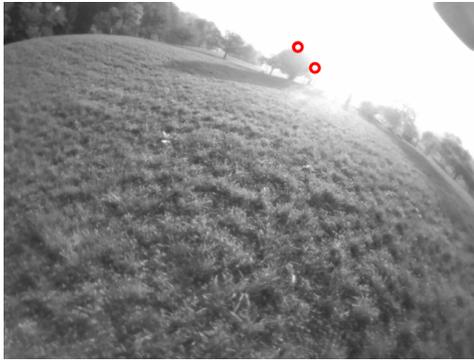
can be modeled as a permutation matrix multiplied to the measurement equations instead of a bias vector, as shown in the research (C. Zhu et al. (2020)). In the work, the authors propose a method to quantify the landmark association error and calculate the corresponding integrity risk, which is a feasible integrity monitor for feature association error.

Furthermore, due to the strong nonlinearity in the visual positioning measurement equation (Eqn. (3)(4)), the pose estimation is sensitive to the initialization of the nonlinear optimization. With exactly the same feature location measurements and associations, the estimation result may differ for distinct initial guesses of the camera pose. Fig. 8 illustrates a simple example of the impact of the initialization on the pose estimation result. In the simulation, five feature points are visible from a camera located at 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 correct estimation results. However, if the initialization is not sufficiently accurate, e.g., at the position and orientation marked as '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 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 literatures.

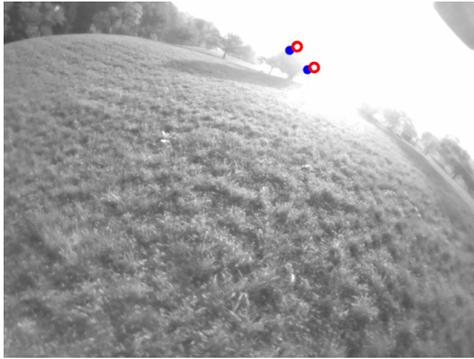
In addition, it is important to identify the critical fault modes in visual positioning, since monitoring all the possible fault modes may become computationally infeasible when the number of biased measurements is large. By monitoring the intensity domain, some crucial fault modes in visual navigation can be detected. For instance, Fig. 9 illustrates an example how the overexposure effect may influence the 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 is artificially added more 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 of exposure can be easily detected by setting up a test to monitor the global intensity properties.

Another example is shown in Fig. 10 to show the benefits of monitoring the 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 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 the features becomes ambiguous. As a result, the estimated position using the wrong set would be biased, and the error may not be reflected by the measurement residuals due to the strong correlation in the error. In the research (Z. Zhu & Taylor, 2017), the authors proposed a method to mitigate the risk caused by correlated error in measurements. A motion blur detector in 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 Eqn. (12)). Therefore, we propose an integrity description framework illustrated in Fig. 11 that monitors faults in multiple domains in the processing procedure, in order to inspire the future developments to cope with the specific challenges. Due to direct methods' strong reliance on uncontrollable lighting conditions, we focus on the feature based



(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 dataset (Delmerico et al., 2019).

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

In the diagram, the two rows in the middle (orange and red color) demonstrate respectively the states propagation and the error space propagation in the core procedure (more details are available in Fig. 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: the nominal error monitoring and the 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 $\{n_I\}$ to feature location domain error $\{i = 1, \dots, N_p | n_i\}$ 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

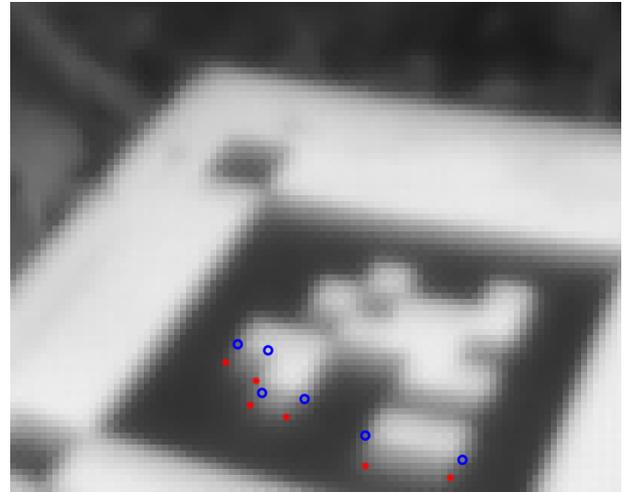


FIGURE 10 Potential motion blur impacts. The blue circles denote the original corner location, while the red stars denote the corners extracted due to motion blur. Image from public dataset (Delmerico et al., 2019).

error model, so that the obtained error distribution of the estimates is not overoptimistic. The studies (Calhoun & Raquet, 2016; Calhoun et al., 2015; C. Zhu et al., 2019a 2019b) are the developments for nominal error integrity, which are capable of calculating PL in nominal operation for some particular problems. Nevertheless, one challenge remaining in the nominal error monitoring is conservative measurement noise models for different types of features. An error model is proposed in the work (C. 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 the 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 result, 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 Fig. 11 . Meanwhile, the error model is also a basis for large bias detection in the geometric measurement domain.

The improvement of the integrity in non-nominal cases relies on fault detection and elimination as illustrated in the FDE block in Fig. 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 the

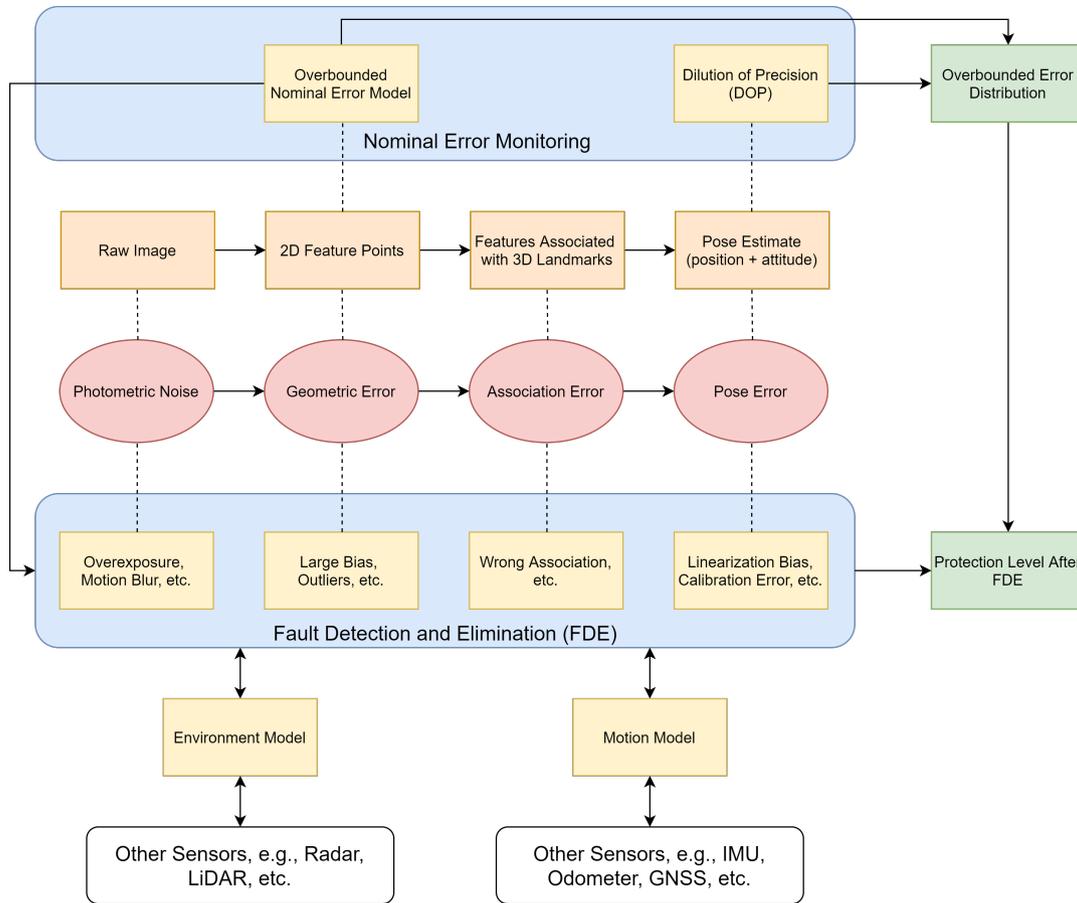


FIGURE 11 Integrity framework for visual positioning

example in Fig. 9. Additionally, wrong associations result in outliers and positioning error, as analyzed in detail by Yang et al. (2018) and C. Zhu et al. (2020). In the 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 error in different domains and take the correlation and propagation of the error sources into consideration.

In order to detect the faults, redundant measurements and quantified models (including error model, environment model and motion model, etc.) 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 modelling that supports FDE. In a multi-sensor context, the visual navigation can benefit from the information from other sensors to obtain environmental and motion models. For example, the inertial sensors can provide the vision system a kinematic model, and radars and LiDARs can provide information of the surrounding environment. These models can be

applied to monitor the integrity of the 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 Fig. 11. It should be mentioned that when multiple sensors are used, imperfect time synchronization between the sensors will result in an additional error source in pose estimation. The bias caused by synchronization imperfection must also be considered into the protection level calculation.

According to the survey of the existing developments on 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 the nominal error monitoring as well as the bias and outlier detection from the feature location measurements. By proposing the multi-domain visual positioning integrity description framework, we want to raise the attentions on 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 the mobility requirements grows, new applications such as autonomous driving and urban air mobility (air-taxis) demand high safety navigation solution in challenging environments. As a consequence, the trend of navigation solution 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 on 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, the 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 | Investigate the significant fault modes for visual navigation methods

Current developments in visual navigation integrity are still mainly 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 on the other vision-specific fault modes, e.g., the nonlinearity and convergence, error correlations, etc. In addition, for measurement bias detection, there are also new challenges from visual positioning, e.g., cases with a large number of biased measurements, etc. Current approaches usually apply RANSAC (RANDOM SAMPLE CONSENSUS) (Fischler & Bolles, 1981) to cope with the issue, but the integrity risks in the random process are not well investigated. New approaches may be required to solve the problem. As mentioned in the earlier section, the

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

6.2 | Value the importance of the error quantification

One of the most powerful character of the integrity concept is its capability of ensuring a quantified maximum risk probability. In order to achieve this goal, the error in all the steps of the 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 different feature detectors, e.g., 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 the most approaches. The intricacy in the processing makes a rigorous error characterization challenging, while the diversity in detection algorithms makes it unfeasible 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 the photometric information and the detected feature location while preserves sufficient detection repeatability, it would be a significant step towards mature integrity monitoring methods for visual navigation.

Besides the feature error models, quantifying the error propagation in other processing steps is also necessary. For example, outlier rejection methods such as RANSAC and M-estimators (Meer et al., 1991) are widely used in visual navigation since decades. Such tools require some prior parameters dependent on the metadata of the measurements dataset, 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-examine and re-evaluate existing algorithms from the integrity point of view

As the computer vision techniques have been rapidly developing since a few decades, there are many powerful algorithms

to cope with various problems and tasks. Nevertheless, the focus of the algorithm development is normally on improving availability and accuracy. An algorithm outperformed in accuracy is usually not valued as much as its advanced competitors. However, higher accuracy or 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 to 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 the integrity concepts as well as the visual positioning core procedures, and reviewed the noticeable research works in the field by now. The development of the vision integrity is still in an early state. There are many particular challenges while also many opportunities for innovative research works. The further evolution of this topic with great potentials calls for the attentions and joint efforts from researchers in the 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. In *2019 IEEE Intelligent Vehicles Symposium (IV)* (pp. 1232–1238). Paris, France. <https://doi.org/10.1109/ivs.2019.8813988>
- Bay, H., Tuytelaars, T., & Van Gool, L. (2006). SURF: Speeded Up Robust Features. In *Computer Vision – ECCV 2006* (pp. 404–417). Graz, Austria. https://doi.org/10.1007/11744023_32
- Bhamidipati, S., & Gao, G. X. (2019). SLAM-based integrity monitoring using GPS and fish-eye camera. In *Proceedings of the 32nd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2019)*. Miami, FL, USA. <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: Journal of the Institute of 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. In *Proceedings of the 20th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2007)* (pp. 2924–2933). Fort Worth, TX, USA.
- Blanch, J., Walker, T., Enge, P., Lee, Y., Pervan, B., Rippl, M., ... 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. In *42nd IEEE Symposium on Security and Privacy*. Princeton, NJ, United States.
- Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., ... 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. In *2016 IEEE/ION Position, Location and Navigation Symposium (PLANS)* (pp. 294–304). Savannah, GA, USA. <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. In *Proceedings of the 2015 International Technical Meeting of The Institute of Navigation* (pp. 756–767). Dana Point, CA, USA.

- 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 multi-map 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*.
- 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. In *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. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 224–236). Salt Lake City, UT, USA. <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 *Computer Vision - ECCV 2014* (Vol. 8690, pp. 834–849). https://doi.org/10.1007/978-3-319-10605-2_54
- European Space Agency. (2011). *Integrity*. Retrieved Accessed 20 Sept. 2020, from <https://gssc.esa.int/navipedia/index.php/Integrity>
- EUSPA. (2021a). *Report on aviation user needs and requirements* (3rd ed.; techreport No. EUSPA-MKD-AV-UREQ-250287).
- EUSPA. (2021b). *Report on rail user needs and requirements* (3rd ed.; Tech. Rep. No. GSA-MKD-RL-UREQ-250286).
- EUSPA. (2021c). *Report on road user needs and requirements* (2nd ed.; Tech. Rep. No. GSA-MKD-RD-UREQ-250283).
- 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.
- 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. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 2198–2204). Madrid, Spain. <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. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*. Providence, Rhode Island, USA. <https://doi.org/10.1109/cvpr.2012.6248074>
- Hartley, R., & Zisserman, A. (2003). *Multiple view geometry in computer vision*. <https://doi.org/10.1017/cbo9780511811685>
- Heikkilä, J., & Silven, O. (1997). A four-step camera calibration procedure with implicit image correction. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)* (p. 1106–1112). San Juan, Puerto Rico. <https://doi.org/10.1109/cvpr.1997.609468>
- Joerger, M., Chan, F.-C., & Pervan, B. (2014). Solution separation versus residual-based RAIM. *NAVIGATION: Journal of the Institute of 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/10.1177/0278364911430419>
- Kendall, A., & Cipolla, R. (2016). Modelling uncertainty in deep learning for camera relocalization. In *2016 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 4762–4769). Stockholm, Sweden. <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. In *2015 IEEE International Conference on Computer Vision (ICCV)* (pp. 2938–2946). Santiago, Chile. <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. In *2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)* (pp. 1–6). Austin, TX, USA. Retrieved from <https://doi.org/10.1109/percomworkshops48775.2020.9156127>
- Langley, R. B. (1999). The integrity of GPS. *GPS World*, 10(3), 60–63.

- 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), 155–166. <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, 01). 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, 052009. <https://doi.org/10.1088/1742-6596/1087/5/052009>
- Lynen, S., Zeisl, B., Aiger, D., Bosse, M., Hesch, J., Pollefeys, M., ... 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. In *Proceedings of the 23rd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2010)* (pp. 245–255). Portland, OR, USA.
- 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., & Tardos, 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. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Virtual. <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 (LAAS)* (2nd ed.; Tech. Rep. No. RTCA DO-245).
- Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). ORB: An efficient alternative to SIFT or SURF. In *2011 International Conference on Computer Vision (ICCV)* (pp. 2564–2571). Barcelona, Spain. <https://doi.org/10.1109/iccv.2011.6126544>
- Sattler, T., Zhou, Q., Pollefeys, M., & Leal-Taixe, L. (2019). Understanding the limitations of CNN-based absolute camera pose regression. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA. <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. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (p. 593-600). Seattle, WA, USA. <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. In *International Conference on Localization and GNSS 2014 (ICL-GNSS 2014)* (p. 1-6). 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. In *6th International Conference on Learning Representations (ICLR 2018)*. Vancouver, Canada.
- Sturm, P., Ramalingam, S., Gasparini, S., & Barreto, J. a. (2011). *Camera models and fundamental concepts used in geometric computer vision* (Vol. 6) (No. 1-2). <https://doi.org/10.1561/06000000023>
- Tossaint, M., Samson, J., Toran, F., Ventura-Traveset, J., Hernandez-Pajares, M., Juan, J., ... Ramos-Bosch, P. (2007). The Stanford - ESA integrity diagram: A new tool for the user domain SBAS integrity assessment. *NAVIGATION: Journal of the Institute of 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: Journal of The Institute of 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. In *2020 IEEE International Conference on Image Processing (ICIP)* (p. 878–882). Abu Dhabi, UAE. <https://doi.org/10.1109/icip40778.2020.9191107>
- Zhou, L., & Kaess, M. (2019). An Efficient and Accurate Algorithm for the Perspective-n-Point Problem. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 6245–6252). Macau, China. <https://doi.org/10.1109/iros40897.2019.8968482>
- Zhou, Q., Sattler, T., Pollefeys, M., & Leal-Taixe, L. (2020). To learn or not to learn: Visual localization from essential matrices. In *2020 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 3319–3326). Virtual. <https://doi.org/10.1109/icra40945.2020.9196607>
- Zhu, C. (2020). *Cooperative vision for swarm navigation*. .
- Zhu, C., Giorgi, G., & Günther, C. (2017). Planar pose estimation using a camera and single-station ranging measurements. In *Proceedings of the 30th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2017)*. Portland, OR, USA. <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: Journal of The Institute of 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. In *2020 IEEE/ION Position, Location and Navigation Symposium (PLANS)* (pp. 967–972). Portland, OR, USA. <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. In *Proceedings of the 32nd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2019)* (pp. 2254–2268). Miami, FL, USA. <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. In *Proceedings of the 32nd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2019)*. Miami, FL, USA. <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>