2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

A Counterfactual Reasoning Framework for Fault Diagnosis in Robot Perception Systems

Haeyoon Han[1], Mahdi Taheri[1], Soon-Jo Chung[1], and Fred Y. Hadaegh[1]

***Abstract*— Perception systems provide a rich understanding of the environment for autonomous systems, shaping decisions in all downstream modules. Hence, accurate detection and isolation of faults in perception systems is important. Faults in perception systems pose particular challenges: faults are often tied to the perceptual context of the environment, and errors in their multi-stage pipelines can propagate across modules. To address this, we adopt a counterfactual reasoning approach to propose a framework for fault detection and isolation (FDI) in perception systems. As opposed to relying on physical redundancy (i.e., having extra sensors), our approach utilizes analytical redundancy with counterfactual reasoning to construct perception reliability tests as causal outcomes influenced by system states and fault scenarios. Counterfactual reasoning generates reliability test results under hypothesized faults to update the belief over fault hypotheses. We derive both passive and active FDI methods. While the passive FDI can be achieved by belief updates, the active FDI approach is defined as a causal bandit problem, where we utilize Monte Carlo Tree Search (MCTS) with upper confidence bound (UCB) to find control inputs that maximize a detection and isolation metric, designated as Effective Information (EI). The mentioned metric quantifies the informativeness of control inputs for FDI. We demonstrate the approach in a robot exploration scenario, where a space robot performing vision-based navigation actively adjusts its attitude to increase EI and correctly isolate faults caused by sensor damage, dynamic scenes, and perceptual degradation.**

I. Introduction

Autonomous systems such as self-driving cars, unmanned aerial vehicles (UAV), and autonomous robots rely on perception systems to convert heterogeneous sensor measurements into a coherent representation of their surrounding environment [1]. The role of the perception system is to provide accurate and timely information on objects, terrain, and the surrounding environment so that higher-level modules in an autonomous system (e.g., localization, motion planning, and control) can guarantee safety and achieve mission objectives [2]. The combination of utilizing heterogeneous sensors (e.g., LiDAR, radar, cameras) and deep learning-based algorithms has led to recent advances in perception-based control. However, this has also resulted in an increased level of complexity in perception systems, which makes detecting their faults and algorithmic errors challenging [3], [4]. Considering the importance of a perception system in the guidance and control of an autonomous system, perception faults can result in the complete loss of a mission. For instance, on 6 June 2025, the Japanese lunar lander Resilience (Hakuto-R Mission 2) had a hard landing during its final descent on the Moon when its laser range finder began outputting erroneous altitude values in the last few kilometers before touchdown [5]. This highlights the need for accurate monitoring systems that can address the problem of fault detection and isolation (FDI) in perception systems.

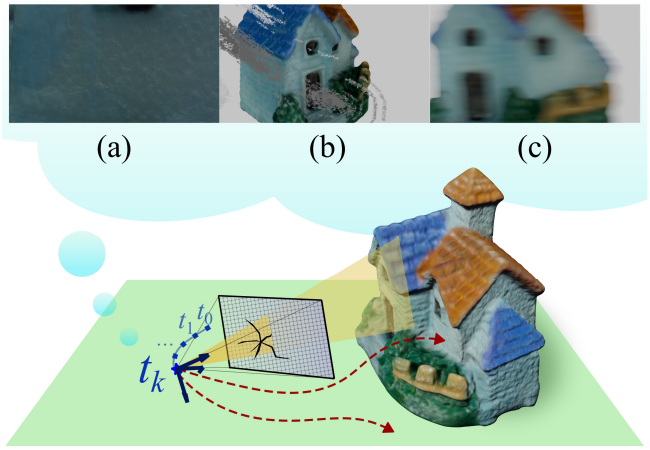

Fig. 1. Our FDI method predicts perception outputs under fault hypotheses and exploits control inputs to enhance fault detectability and isolability in robot perception systems. Examples of vision faults are (a) visual deprivation, (b) sensor damage, (c) motion blur. Adapted from OmniObject3D [6], licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/); changes: rendered with Blender.

The method presented in this paper can handle a broad range of fault and failure types, including both physical malfunctions and algorithmic errors in perception systems that cause deviations from their intended functionality. On the physical side, sensors can suffer calibration shifts, temporary occlusions, and environmental interference [7]. At the algorithmic level, deep neural networks (DNN) can misclassify objects due to distribution shifts (i.e., out-of-distribution inputs), and multi-sensor fusion can become erroneous due to calibration issues [8]–[10]. Moreover, faults that occur at an early stage of a perception system's pipeline propagate through it and do not remain isolated [11]. Hence, FDI methodologies that rely on physical redundancy may not be sufficient [12]. Thus, one needs to study and investigate FDI methodologies based on the available analytical redundancy in perception systems. Once a certain fault is detected and isolated, a fault recovery control can be implemented.

A. Related Work

The faults that occur in Simultaneous Localization and Mapping (SLAM) and Visual Inertial Odometry (VIO) systems are sensor faults [3], [7], tracking failures [15], [16], data association failures [17], [18], and filtering inconsistency problems [19]. Sensor faults are caused by hardware damage or software malfunction. Faults in front-end modules, such as tracking and data association failures, are often caused by visually deprived conditions (i.e., textureless surfaces and repetitive patterns), dynamic scenes (i.e., aggressive camera motion), and undesirable lighting conditions (i.e., high-contrast images). Lastly, the filtering

Haeyoon Han and Mahdi Taheri are co-first authors.

[1]Division of Engineering and Applied Science, California Institute of Technology (Caltech), Pasadena, CA 91125, {hhan3, mtaheri, sjchung, hadaegh}@caltech.edu.

This research is funded in part by the Technology Innovation Institute and the Defense Advanced Research Projects Agency (Learning Introspective Control).

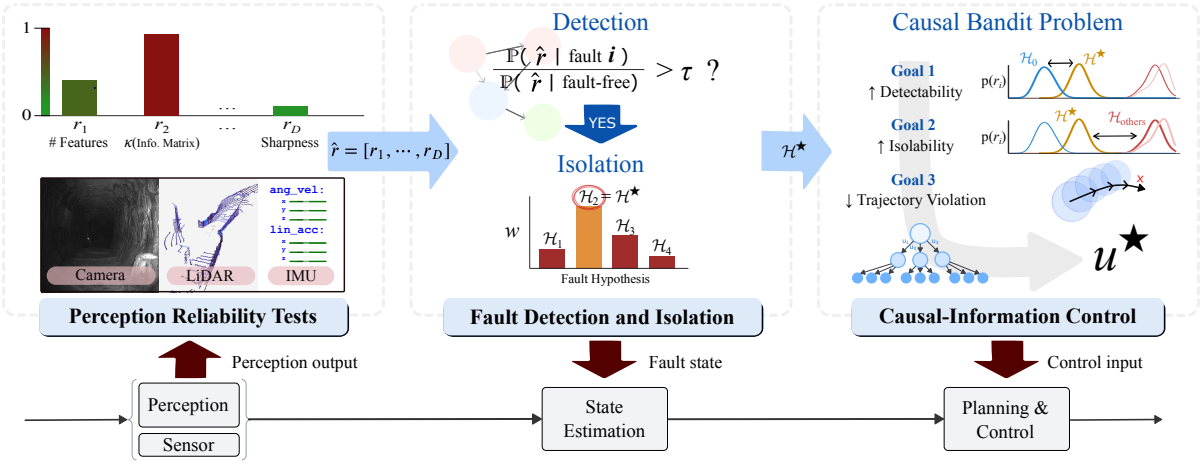

Fig. 2. Algorithm architecture overview. The FDI module compares the perception reliability test results with the fault-hypothesis distributions and generates control inputs that improve FDI performance. Camera and LiDAR data captured using Foxglove from CERBERUS DARPA Subterranean Challenge datasets [13], [14]

inconsistency problems, a type of fault in back-end modules, result from large inter-frame transformations that trigger the accumulation of linearization errors.

The work in [8] compares perception outputs with a predefined fault threshold for runtime monitoring. Additionally, [3] developed fault diagnostic graphs to associate errors with individual perception module outputs, as evaluated by diagnostic tests. Although these works enable FDI, they rely on having redundant sensors, which can be costly. To enhance the robustness of SLAM [20] developed image quality metrics to select confident features or scenes. Similarly, feature quality metrics that assess keypoint co-visibility between frames [3], [15], [21] and the dynamic scene metrics that leverage vehicle velocity [15], optical flow [22], and image sharpness [23], [24] have been proposed.

B. Contributions

We define perception reliability tests for various fault modes to capture differences between fault-free and fault-induced behaviors. We utilize the structural causal model (SCM) formalism of Pearl [25] and its operational rules for interventions and counterfactual queries, where we treat each hypothesized fault mode as an intervention on the perception pipeline. We then introduce and define an information-theoretic metric based on the Kullback–Leibler (KL) divergence between the reliability test results and those from a baseline fault-free case to measure the detectability and isolability of the hypothesized faults. This metric, designated as Effective Information (EI), captures how control inputs influence the reliability test results by affecting the autonomous system's state. To the best of our knowledge, this is the first work that studies the FDI as a counterfactual reasoning problem for a closed-loop autonomous system and also connects the informativeness of control inputs to the detection and isolation of the hypothesized faults. Finally, we show that finding the control input that helps maximizing the EI leads to having a causal bandit problem [26], where each action arm corresponds to an intervention on the control input that improves our FDI accuracy. A Monte-Carlo Tree Search (MCTS) approach with Upper Confidence Bound (UCB) [27], [28] that penalizes large deviations from primary mission objectives (e.g., tracking a trajectory) is employed to solve the mentioned causal bandit problem.

The main contributions of this paper are as follows.

1) We exploit analytical redundancy of the perception system and actively use control inputs for FDI by applying the do-operator from causal inference. This is achieved via a counterfactual reasoning approach, where it is analyzed how control inputs affect reliability test outcomes under various fault hypotheses. A quantitative detection and isolation metric measuring the informativeness of each control input for FDI is introduced.
2) We formulate the problem of selecting control inputs for FDI as a causal bandit problem. Using a MCTS strategy with UCB, we maximize a weighted reward function that prioritizes inputs informative about the most likely fault modes. In addition, our reward function penalizes large deviations from the desired trajectory of the system.
3) Our FDI method uses the distribution of reliability test results under various fault modes and accounts for the uncertainty inherent in the perception system's outputs. Thus, our method encodes more information than mean value and threshold-based FDI methods, which only reflect the central tendency of a distribution.

II. Preliminaries and Problem Statement

A. Structural Causal Models and do-operator

An SCM [25] describes a model M using a set of endogenous variables $V = \{X_1, \ldots, X_o\}$ generated by structural equations $X_i := m_i(\mathrm{Pa}_{X_i}, \hat{A}_i)$, where $\mathrm{Pa}_{X_i} \subseteq V \setminus \{X_i\}$ are the parents of X_i and $\hat{A}_i$ are exogenous variables, for $i = 1, \ldots, o$.

***Definition** 1 (do-operator):* An intervention that forcefully sets a subset of variables $X_S \subseteq V$ to values x_S is denoted by the do-operator, i.e., $\mathrm{do}(X_S = x_S)$.
Executing $\mathrm{do}(X_S = x_S)$ removes the structural equations for X_S and replaces them with $X_S = x_S$, which yields the model M_S. Consequently, the post-intervention distribution can be written as $p_M(v|\mathrm{do}(x_S)) = p_{M_S}(v)$. In other words, under model M, the distribution of outcome V after the intervention $\mathrm{do}(x_S)$ is the probability that model M_S assigns to V. It is worth noting that $p_M(v|\mathrm{do}(x_S))$ is different from having $p_M(v|x_S)$ in the sense that $\mathrm{do}(x_S)$ breaks incoming causal links into X_S rather than observing their values.

B. Effect Distribution (ED)

Let $\hat{V}$ be a variable of interest and $\hat{A} \in \hat{\mathcal{A}}$ an intervention variable (e.g., the control input, hypothesized faults) with the distribution $p_{\hat{A}}(\hat{a})$. The Effect Distribution (ED) [29] is the marginal distribution of $\hat{V}$ induced by randomizing $\hat{A}$ such that

$$\mathrm{ED}_{\hat{V}}(\hat{v}) = \mathbb{E}_{\hat{a}\sim p_{\hat{A}}}[p(\hat{v}|\mathrm{do}(\hat{a}))] = \int_{\hat{\mathcal{A}}} p(\hat{v}|\mathrm{do}(\hat{a}))p_{\hat{A}}(\hat{a})d\hat{a}. \tag{1}$$

The $\mathrm{ED}_{\hat{V}}$ captures the baseline behavior of $\hat{V}$ under all interventions we are able to perform on $\hat{A}$.

C. Effective Information (EI)

Suppose we actively sample $\hat{U} \sim p_{\hat{U}}$ and record the resulting $\hat{V}$. The Effective Information (EI) [29]–[31] from $\hat{A}$ to $\hat{V}$ can be defined as

$$\mathrm{EI}(\hat{A} \to \hat{V}) = \mathbb{E}_{\hat{a}\sim p_{\hat{A}}}[D_{\mathrm{KL}}\big(p(\hat{v}|\mathrm{do}(\hat{a}))\|\mathrm{ED}_{\hat{V}}(\hat{v})\big)], \tag{2}$$

where D_{KL} is the Kullback–Leibler (KL) divergence. Consequently, the EI of an individual $\hat{a}_i \in \hat{A}$ is $D_{\mathrm{KL}}\big(p(\hat{v}|\mathrm{do}(\hat{a}_i))\|\mathrm{ED}_{\hat{V}}(\hat{v})\big)$ which indicates the impact that $\hat{a}_i$ makes in the variable $\hat{v}$. The EI (2) is the average statistical distance (in the sense of the KL divergence) between the post-intervention distribution and the baseline $\mathrm{ED}_{\hat{V}}$.

D. System Model

We consider the following nonlinear control-affine system:

$$x(t+1) = f(x(t)) + g(x(t))u(t) + w(t), \tag{3}$$

where $x(t) \in \mathbb{R}^n$ is the state, $u(t) \in \mathbb{R}^m$ is the control input, and $w(t) \in \mathbb{R}^n$ denotes the process noise. Moreover, $f(x)$ and $g(x)$ are smooth functions.

The perception system integrates $S \in \mathbb{N}^+$ sensors (e.g., camera, LiDAR) in its pipeline, which consists of various layers (e.g., preprocessing, feature extraction), and $\mathbb{N}^+$ is the set of positive integers. The output of the perception system is given by

$$y(t) = (h \circ z)(x(t), e(t)) + \nu(t), \tag{4}$$

where $y(t) \in \mathbb{R}^p$ is the state to be estimated via the perception system, $z : \mathbb{R}^n \times \mathbb{R}^{n_e} \to \mathbb{R}^{n_z}$ is the function describing the output of perception sensors, $h : \mathbb{R}^{n_z} \to \mathbb{R}^p$ is a perception map from the sensor output to y, $e(t) \in \mathbb{R}^{n_e}$ represents unmeasured environmental states (e.g., lighting condition, weather changes, dynamic objects), $\nu(t) \in \mathbb{R}^p$ is the measurement noise. The definition of fault mode in perception systems considered in this paper is as follows.

***Definition** 2 (Fault Mode):* A fault mode is the function $t \to \delta_i(t)$ that represents the occurrence of a fault scenario $\mathcal{H}_i$, such that $\delta_i(t) = \mathbf{1}_{\mathcal{H}_i}(t), \quad i = 1, \ldots, M$, where $\mathbf{1}_{\mathcal{H}_i}(t) \in \{0, 1\}$ is the indicator function of the i-th fault hypothesis, i.e., $\mathcal{H}_i$, at time t, and $M \in \mathbb{N}^+$ is the number of considered fault scenarios.

Fault modes are inherently case-specific; representative examples include sensor damage, visually deprived conditions, and dynamic scenes. See Section VI-A for their relevance in vision-based asteroid exploration of a space robot. We consider the following assumption throughout this paper.

***Assumption** 1:* At any instance of time, there exists only one fault in the perception system (4), i.e., the cardinality of $\mathrm{supp}(\delta) = \{i \in \{1, \ldots, M\} \,|\, \delta_i \neq 0\}$ is equal to 1.

For brevity, we omit the explicit dependence of the variables on t in the remainder of the paper.

E. Problem Statement

Given the perception system (4), we develop and design perception reliability tests based on the system state $x(t)$ and internal variables of the perception system from $(h \circ z)(x(t), e(t))$. In order to avoid the direct comparison of high-dimensional visual information, the outputs of reliability tests are generated for various hypothesized fault scenarios and efficiently compared with observed test results. Consequently, we define a set of counterfactual hypotheses regarding faults in the perception system, i.e., δ_i, to study the detection and isolation of the hypothesized faults. To carry out the latter, we introduce a causal information-theoretic metric, designated as EI, that captures how distinguishable various hypothesized faults are for a certain control input. Finally, we find interventions, i.e., control input $u(t)$, that actively enhance the detection and isolation capabilities of our methodology for the hypothesized faults, and we formulate this problem within the causal bandit framework. Our proposed FDI methodology is shown in Fig. 2.

III. Perception Reliability Tests and Counterfactual Hypotheses

In this section, perception reliability tests are introduced as a method to derive essential diagnostics from complex perception outputs. As opposed to a direct comparison of the measured perception output with the nominal fault-free approximation of it, a set of perception reliability tests is developed as a measure of the perception system's well-being. Furthermore, the richness of visual information generated by perception sensors enables one to extract intermediate outputs that are more compact than raw sensor data but more informative than the final output alone. These intermediate outputs can be utilized in the set of reliability tests to carry out FDI for perception systems.

A. Perception Reliability Tests

The perception system first captures visual information from its sensors and subsequently processes this visual information through one or more algorithm modules. Let the output of the j-th module of the perception pipeline be $y^{(j)} = h^{(j)}\big(z(x, e)\big) = \mathcal{P}^{(j)}(x, e) \in \mathbb{R}^{n_{h^{(j)}}}$, where $h^{(j)} : \mathbb{R}^{n_z} \to \mathbb{R}^{n_{h^{(j)}}}$ is the function which maps the sensor output to the intermediate output $y^{(j)}$. Based on this intermediate output, a set of perception reliability tests is defined to evaluate the reliability of the perception output for each test.

***Definition** 3 (Perception Reliability Test):* A perception reliability test associated with module j in a perception system is a function

$$d : \underbrace{\mathbb{R}^{n_{h^{(j)}}} \times \mathbb{R}^{n_{h^{(j)}}} \times \cdots \times \mathbb{R}^{n_{h^{(j)}}}}_{W} \to \mathbb{R},$$

which maps output sequence $\{y^{(j)}(t)\}_{t=k+1-W}^{k}$ to a scalar value representing reliability measure.

The perception reliability test result $r(x_{k+1-W:k}, e) = d\big(\{\mathcal{P}^{(j)}(x_t, e)\}_{t=k+1-W}^{k}\big)$ quantifies the reliability of the output from a specific sensor or algorithmic module and is characterized by the length W-sized output derived from the state and the environmental states.

The choice of reliability tests is application-specific, but the way the perception reliability test is used for fault diagnosis is general enough to accommodate various types of reliability tests. The following example illustrates the perception reliability tests in practice.

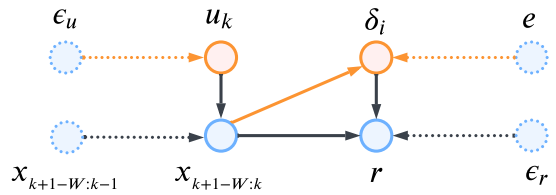

Fig. 3. Causal model for perception reliability test with correlation between fault mode, state, and environmental state. The dotted lines indicate the effect of exogenous variables, and the orange lines indicate the links removed after intervention $\mathrm{do}(\delta_j)$ and $\mathrm{do}(u_k)$.

Example 1: Visually deprived conditions can cause tracking or data association failures in vision-based SLAM and VIO systems. A perception reliability test that can verify such conditions involves counting the number of detected feature points in an image (e.g., low when the camera is directed toward a textureless surface). For instance, assuming prior knowledge of keypoint locations, the number of detected feature points $d(\mathcal{Q})$ can be computed from the state vector and the known keypoints as follows:

$$\mathcal{Q}(x,e) = \{i \in \mathbb{N} : C(e)\Pi(x)q_i \in \mathcal{F},\ q_i \in \mathcal{M} \subset \mathbb{R}^3\},$$
$$d(\mathcal{Q}) = |\mathcal{Q}(x,e)|, \tag{5}$$

where $\mathcal{Q}$ is the set of points' indices inside the field of view $\mathcal{F} \in \mathbb{R}^2$ defined with respect to the image plane, $C(e)$ is the feature point selection matrix that depends on the environmental conditions, $\Pi(x)$ is the projection matrix derived from the relative pose, $\mathcal{M}$ is the set of known keypoints, and $|\cdot|$ is the cardinality of the set.

As such, perception reliability tests are widely employed in practice, often combined with simple threshold-based scores or statistical evaluations, to verify the quality of visual information and validate the correctness of algorithmic results. Yet, the fundamental distinction is that our work utilizes reliability tests to identify the source of the fault, whereas other works use them to reject anomalous measurements or trigger fallback strategies to maintain safe behavior, regardless of the fault origin. For this purpose, we present a model that explicitly accounts for the fault mode dependency in modular and customizable reliability tests. Such a model bridges practical perception pipelines, including neural network–based approaches, with the proposed information-theoretic framework for fault diagnosis. A detailed example of reliability tests that are used in a camera-based perception pipeline is provided in Section VI.

B. Counterfactual Fault Hypotheses

Let us define $r = [r_1, \ldots, r_D]^\top \in \mathbb{R}^D$, where $r_i = d_i\big(\{\mathcal{P}^{(j)}(x_t, e)\}_{t=k+1}^{k+W}\big)$. As described in the previous subsection, each reliability test result r_i is sensitive to a set of fault modes given the state sequence, for $i = 1, \ldots, D$, where D is the number of tests. This implies that two distinct causes of the fault in the perception system may trigger the same reliability test, potentially leading to fault misclassification. In order to accurately isolate the faults, we employ the do-operator defined in Section II, which enables us to treat various hypothetical fault scenarios separately. We consider a healthy scenario and M fault hypotheses in the set $\mathcal{H} = \{\delta_0, \delta_1, \ldots, \delta_M\}$, where each δ_i corresponds to a fault scenario in our perception system, for $i = 1, \ldots, M$, and δ_0 is the healthy case, i.e., fault-free.

To evaluate the effect of fault modes and control inputs on the reliability test results, a structural causal model $M = \langle E, V, F\rangle$ is defined for the autonomous system at time $t = k$, where E and V denote the sets of exogenous variables and endogenous variables, respectively. Realizations of the endogenous variables are denoted by $u, x_{k+1-W:k}, \delta_j, r$, and those of the exogenous variables by $\epsilon_u, \epsilon_r, e$ with the observed states $x_{k+1-W:k-1}$. The set of functions that maps to V_i from $E \cup V \backslash V_i$ are $F = \{F_u, F_{x_{k+1-W:k}}, F_\delta, F_r\}$, with related variables shown in Fig. 3. As per the SCM in Fig. 3, there is $e \to \delta_j \to r$ and no direct $e \to r$ edge. Hence, under the intervention $\mathrm{do}(\delta_j)$, we remove all incoming arrows into δ_j (e.g., from x and e), and any association between δ_j and the environmental variable e is eliminated. The same method is used for the intervention $\mathrm{do}(u)$.

The effect of intervention due to the fault mode and the control input can be derived using the adjustment [25]. In our case, the conditional counterfactual due to the intervention $(\mathrm{do}(u), \mathrm{do}(\delta_j))$ on the reliability test result r is given by

$$p_{\mathrm{cf}}\big(r|x_{k+1-W:k-1}, \mathrm{do}(u), \mathrm{do}(\delta_j)\big) = \int_{\mathcal{X}} p(r|x_{k+1-W:k}, \delta_j) \cdot p(x_{k+1-W:k}|x_{k+1-W:k-1}, u) dx_{k+1-W:k}, \tag{6}$$

where $p(r|x_{k+1-W:k}, \delta_j)$ is the reliability test result's distribution given the system state and the fault mode, and $p(x_{k+1-W:k}|x_{k+1-W:k-1}, u_k)$ is the predicted belief at time k. The predicted belief can be obtained from the state estimation module, where our state estimator provides the abduction step in Pearl's abduction, action, and prediction counterfactual framework. On the other hand, $p(r|x_{k+1-W:k}, \delta_j)$, which encodes the probabilistic reliability test result under a given state sequence and fault mode, can be obtained from the experiment in controlled environments where the fault scenario has been set up, as we discuss further in Section VI. In what follows, all densities and information-theoretic metrics are counterfactual and implicitly conditioned on $x_{k+1-W:k-1}$. We also define $p(r|\mathrm{do}(u), \mathrm{do}(\delta_j)) := p_{\mathrm{cf}}(r|x_{k+1-W:k-1}, \mathrm{do}(u), \mathrm{do}(\delta_j))$.

IV. Information-Theoretic Diagnostic Metrics for Fault Detection and Isolation

Given the defined fault hypotheses in the previous section, we adopt the information-theoretic metric, Effective Information (EI), (see Section II) and modify it for our FDI problem. The modified EI measures the amount of information that is captured by our reliability test results due to an intervention $\mathrm{do}(u)$ in terms of the detection and isolation of each fault hypothesis. Hence, maximizing the EI would increase the probability of detecting and isolating the hypothesized faults.

A. Effect Distribution

One needs to first establish a baseline for normal behavior by redefining the Effect Distribution (ED) that was introduced in Section II. We define the ED as the expected distribution of the perception reliability test result vector r under fault-free conditions δ_0, marginalized over interventions in the space of admissible control inputs $\mathcal{U}$. One has

$$\mathrm{ED}(r|\mathrm{do}(\delta_0)) = \mathbb{E}_{u\sim\mathcal{U}}[p(r|\mathrm{do}(u), \mathrm{do}(\delta_0))]. \tag{7}$$

Remark 1: Since the exact computation of $\mathrm{ED}(r)$ is challenging, one can approximate it. Hence, under the assumption of having a Gaussian density for the reliability test distribution under each control input u, one has $p(r|\mathrm{do}(u), \mathrm{do}(\delta_0)) \approx \mathcal{N}(r; \mu, \Sigma)$, where μ and Σ are the mean value and the covariance. Consequently, the $\mathrm{ED}(r)$ in (7) can be approximated by drawing $\bar{K}$ samples and using a Gaussian Mixture Model (GMM) given by $\mathrm{ED}(r) \approx$

$\frac{1}{\bar{K}}\sum_{k=1}^{\bar{K}}\mathcal{N}(r;\,\mu_k,\Sigma_k)$ [32]. Alternatively, if the assumption of having Gaussian reliability test distributions is restrictive, one can fit a kernel density estimator directly to the $\bar{K}$ samples [33].

B. Fault Detectability Metric

The detectability of a fault δ_i is a measure of how different its signature in the reliability test r is from the fault-free baseline. In order to measure the difference between probability densities, we utilize the KL divergence. Consequently, we define the detectability of the i-th hypothesized fault under the control input u in the following form:

$$D^i_{\text{detect}}(u) = D_{\text{KL}}[p(r|\text{do}(u),\text{do}(\delta_i))||\,\text{ED}(r)]. \tag{8}$$

C. Fault Isolation Metric

In addition to the detection of faults, our fault diagnosis methodology can isolate various faults in the system. Correct isolation among faults results in accurately identifying which hypothesized fault in the set $\mathcal{H}$ is the true fault in the perception system. Thus, we define the isolation of the i-th fault under the control input u as the sum of the KL divergences between the reliability test distribution corresponding to δ_i and those of all other fault hypotheses, for $j \in \{1,...,M\}$, as given by

$$D^i_{\text{isolate}}(u) = \sum_{j\neq i} D_{\text{KL}}[p(r \mid \text{do}(u),\text{do}(\delta_i))\| \\ p(r \mid \text{do}(u),\text{do}(\delta_j))]. \tag{9}$$

***Remark** 2:* It is worth noting that due to the connection of the KL divergence with the notion of mutual information [34], $D^i_{\text{detect}}(u)$ quantifies how much a perception reliability test distribution $p(r \mid \text{do}(u),\text{do}(\delta_i))$ differs from the baseline ED with respect to the amount of information contained in r. Also, the KL divergence is inherently asymmetric, thus, it takes into account the causal direction of interventions. If these considerations are not critical for a particular application, the KL divergence in (8) and (9) can be replaced by an alternative distance measure, such as the Wasserstein metric.

D. Effective Information for Detection and Isolation

We combine $D^i_{\text{detect}}(u)$ in (8) and $D^i_{\text{isolate}}(u)$ given by (9) into a single metric that captures the total diagnostic value of a control input. The EI of a control input u to identify hypothesis δ_i is the sum of its detection and isolation metric, as expressed by

$$\text{EI}(u|\delta_i) = D^i_{\text{detect}}(u) + D^i_{\text{isolate}}(u). \tag{10}$$

The EI given in (10) quantifies how much an active intervention on u reveals diagnostic information in the sense of (7) and (10) about the fault δ_i.

V. Causal Bandit, Effective Information (EI), and Active Fault Detection and Isolation

Considering that EI measures both the detection and isolation of faults given an intervention $\text{do}(u)$, a causal bandit problem [35] is studied to maximize it. In order to investigate a solution for the mentioned causal bandit problem, we utilize a Monte Carlo Tree Search (MCTS) algorithm.

A. Maximizing EI in a Causal Bandit Problem

Our main objective in this section is to find a control input u^* that maximizes the EI and helps us to isolate the fault in the system. Hence, we define a weighted EI expressed by

$$\text{EI}_{\text{w}}(u) = \sum_{i=1}^{M} w_i \cdot \text{EI}(u|\delta_i), \tag{11}$$

where $w_i \in [0,1]$ such that $\sum_{i=1}^{M} w_i = 1$. After applying a control action u and observing a new reliability test $r' = r(t+1)$, these weights are updated via Bayes' rule in the following form:

$$w_i(t+1) = \frac{w_i(t)\cdot p(r'|\text{do}(u),\text{do}(\delta_i))}{\sum_{j=1}^{M} w_j(t)\cdot p(r'|\text{do}(u),\text{do}(\delta_j))}. \tag{12}$$

Maximizing the weighted EI in (11) by means of the control input u has two main outcomes. First, it drives the system to increase its information in the sense of $D^i_{\text{detect}}(u)$ and $D^i_{\text{isolate}}(u)$. Second, it accelerates the convergence of the hypothesis weights towards the true fault state. The latter is achieved by choosing a control action u^* that maximizes the information gain for the fault hypothesis with the highest weight. Moreover, the Bayesian weight update rule (12) ensures that the corresponding weight to the true fault hypothesis increases significantly relative to the others. Therefore, the fault can be identified by monitoring the hypothesis weights. Hence, $\hat{i}$ is the index of the identified fault as given by

$$\hat{i} = \arg\max_{i\in\{1,\dots,M\}} w_i. \tag{13}$$

The control action that maximizes (11) may not be aligned with our control objectives (e.g., tracking a desired trajectory). Therefore, we define the causal bandit problem in the following form:

$$u^* = \arg\max_{u\in\mathcal{U}} \Big(\sum_{i=1}^{M} w_i \cdot \text{EI}(u|\delta_i) - \lambda\cdot P(u)\Big), \tag{14}$$

where $\lambda > 0$ is a regularization parameter and $P(u) = \max(0, \|x - x_d\| - \epsilon)$ is a penalty for deviating from the reference trajectory $x_d \in \mathbb{R}^n$ beyond a tolerance $\epsilon > 0$.

B. MCTS for Active Fault Detection and Isolation

We employ an MCTS-UCB algorithm [27], [28] to solve the optimization problem (14). The MCTS is effective in handling and balancing the exploration and exploitation of large search spaces. The general MCTS procedure (tree expansion, UCB-based selection, search, simulate, and rollout) follows the standard formulation in [28]. We have implemented the MCTS given in Algorithm 1, where the reward used in simulation is aligned with our objective (14), i.e., $R(u) = \text{EI}_{\text{w}}(u) - \lambda P(u)$. Moreover, for the set of admissible control inputs $\mathcal{U}$, we employ progressive widening to manage tree growth.

C. Passive Fault Detection and Isolation

Considering that in certain applications actively modifying control inputs for FDI may be undesirable, we develop a passive methodology in this section. In the passive case, a fault is detected by evaluating the likelihood of an observed reliability test r' under a certain fault hypothesis. We adopt

Algorithm 1 Active Fault Detection and Isolation via MCTS

function SEARCH(x_0, w_0)
  initialize tree T with root history $h = []$
  $N(h) \leftarrow 0$, $Q(h, \cdot) \leftarrow 0$
  while planning budget remains **do**
    $R \leftarrow$ SIMULATE($x_0, w_0, h, 0$)
  end while
  return action u^* from the root child maximizing $Q(h, u)$
end function
function SIMULATE(x, w, h, d)
  if $d = k$ **then**
    return 0
  end if
  if $h \notin T$ **then**
    add node h to T, set $N(h) \leftarrow 0$, $Q(h, \cdot) \leftarrow 0$
    initialize empty child set $\text{Children}(h) = \{\emptyset\}$
    return ROLLOUT(x, w, h, d)
  end if
  if $|\text{Children}(h)| < \lceil N(h)^\alpha \rceil$ **then**
    sample new action $u_{\text{new}} \sim \mathcal{U}$
    add u_{new} to $\text{Children}(h)$, initialize $N(h, u_{\text{new}}) = 0$, $Q(h, u_{\text{new}}) = 0$
  end if
  choose $u' = \arg\max_{u \in \text{Children}(h)} Q(h, u) + c\sqrt{\ln N(h)/N(h, u)}$
  $x' \leftarrow$ predict system output for (x, u')
  $r \leftarrow EI_{w'}(u') - \lambda P(u')$
  $R \leftarrow r + \gamma$SIMULATE($x', w', h + [u'], d + 1$)
  $N(h, u') \leftarrow N(h, u') + 1$
  $Q(h, u') \leftarrow Q(h, u') + (R - Q(h, u'))/N(h, u')$
  return R
end function
function ROLLOUT(x, w, h, d)
  if $d = k$ **then**
    return 0
  end if
  sample $u' \sim \mathcal{U}$
  $x' \leftarrow$ predict system output for (x, u')
  $r \leftarrow EI_{w'}(u') - \lambda P(u')$
  return $r + \gamma$ ROLLOUT($x', w', h + [u'], d + 1$)
end function

a likelihood ratio test to compare the fault-free hypothesis $\mathcal{H}_0$ with the fault hypothesis $\mathcal{H}_j$, as follows:

$$j^* = \arg\max_j \frac{p\big(r' \mid \text{do}(\delta_j), \text{do}(u)\big)}{p\big(r' \mid \text{do}(\delta_0), \text{do}(u)\big)}.$$

A fault is then declared when the maximum likelihood ratio exceeds a threshold τ, which triggers the subsequent fault isolation algorithm based on weight updates as explained in Section V-A. This maximum likelihood ratio test is applied at every FDI time step to monitor the existence of faults.

VI. Simulation Experiments

A. Application to Vision-based Asteroid Exploration

This section illustrates the application of our fault diagnosis algorithm, where a spacecraft equipped with a monocular camera observes an asteroid and provides relative pose measurements. The spacecraft flies around a Sun-terminator orbit, and the target asteroid is modeled after 99942 Apophis using the publicly available shape model [36]. The spacecraft maintains a default nadir-pointing attitude, pointing its camera toward the asteroid's center. Fig. 4 (a) shows the simulation setup used in this study.

We assume the spacecraft employs vision-based perception for relative navigation during close-proximity operations near the asteroid. The vision-based perception framework used in this simulation consists of image formation, front-end, and back-end modules. The image formation module generates images from the relative states and known keypoints using a geometric vision model with a pinhole camera, the front-end extracts pixel locations of these keypoints, and the back-end estimates the relative pose of the spacecraft from 2D-3D point correspondence, using a Perspective-n-Point (PnP) solver with an Extended Kalman Filter (EKF).

TABLE I
RELIABILITY TESTS USED IN ASTEROID EXPLORATION SIMULATION

Name	Reliability Tests
No. Feat. Points	$s_1 = \lvert\mathcal{Q}(x_k, e)\rvert$
No. Station. Pix.	$s_2 = \lvert\mathcal{Q}(x_{k-1}, e) \cap Q(x_k, e)\rvert$
Ego Motion	$s_3 = \lVert v_k\rVert + \sqrt{2}H\lVert\omega_k\rVert$
Sharpness	$s_4 = \sum_{i \in \mathcal{Q}(x_k, e)} \frac{H_{\text{blur},i} v_{\text{app},i}}{s_1}$
Optical Flow	$s_5 = \sum_{i \in \mathcal{Q}(x_{k-1}, e) \cap \mathcal{Q}(x_k, e)} \frac{\lVert\big(\Pi(x_k) - \Pi(x_{k-1})\big) q_i\rVert}{s_1}$
Inter-frame TF.	$s_6 = \lVert x_k \ominus x_{k-1}\rVert$

We evaluate our passive and active FDI methods under multiple perception fault scenarios with this setup. The fault for this system is defined based on the sensor type (i.e., camera) and the characteristics of the environment (i.e., high-contrast lighting and comparatively slow target dynamics). Accordingly, we define three types of fault: dynamic scene ($\mathcal{H}_1$), sensor damage ($\mathcal{H}_2$), and visually deprived conditions, ($\mathcal{H}_3$), which have corresponding fault modes δ_1, δ_2, and δ_3, respectively. The test scenarios are as follows:

- Case 1: The camera is rotating along the camera principal axis due to disturbance torque.
- Case 2: 1.5 seconds after deployment, the camera lens is broken and locally generates random pixel values.
- Case 3: The spacecraft passes a region with limited visual information, while the outermost regions of the asteroid along the Sun-terminator orbit normal are unobservable due to extreme lighting conditions.

The fault scenarios to construct the probability distribution are generated from environmental conditions with (i) high relative angular velocity of the scene, (ii) a large number of occluded and wrongly placed pixels, and (iii) undetected pixels in high-illumination and low-illumination regions, for $\mathcal{H}_1$, $\mathcal{H}_2$, and $\mathcal{H}_3$, respectively. To build a probability density function for each hypothesis, we sample 30 fault scenarios for each hypothesis, given the state and the control input.

The perception reliability tests considered in this case study include: the number of detected features, the number of stationary features, weighted ego motion intensity, image sharpness, average of optical flow, and inter-frame transformation. The reliability tests are defined in the Table I, where $x_k = [p_k^\top, v_k^\top, \bar{q}_k^\top, \omega_k^\top]^\top$ is the state vector consisting of relative position, velocity, attitude, angular velocity, $\mathcal{Q}(x, e)$ is the set of inlier indices as defined in (5), H is the height of the image in terms of pixels, f is the focal length, t_{exp} is the exposure time, and $\ominus$ is the state difference operator that handles Euclidean components and quaternion. For calculating the sharpness, we use the following terms:

$$H_{\text{blur},i} = \frac{f t_{\text{exp}}}{z_{c,i}^2} \begin{bmatrix} z_c & 0 & -x_c \\ 0 & z_c & -y_c \end{bmatrix},$$
$$v_{\text{app},i} = T(\bar{q}_k)\big(- v_k - (q_i + T(\bar{q}_k)^\top p_k) \times \omega_k\big),$$

where (x_c, y_c, z_c) are the feature point location in camera frame and $T(\bar{q}_k)$ is the direction cosine matrix.

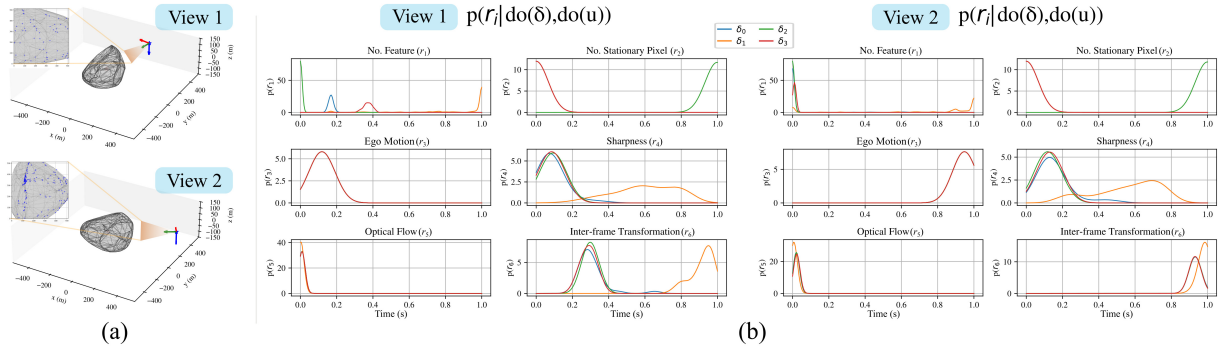

Fig. 4. Probability density function of normalized reliability test results for two different relative states. The distribution of $p(r|\mathrm{do}(\delta_j), \mathrm{do}(u))$ varies with scene context and ego motion, even under the same fault hypothesis. (a) Spacecraft's relative position and attitude with respect to the asteroid, (b) probability density of the normalized reliability test results for each case.

In addition, each reliability test is normalized to produce higher values under faults to ensure stable fault diagnosis. The first test—where higher s_1 indicates healthy—is normalized as $r_1 = 1 - \tanh(s_1/c_1)$, and the others—where lower values are healthy—are normalized as $r_j = \tanh(s_j/c_j), \quad j = 2, \cdots, 6$. Here, $c_1, \cdots c_6$ are scaling factors that emphasize the fault-relevant ranges of $s_1, \cdots, s_6$, hence, improving the sensitivity of FDI.

The calculation of effect of intervention follows directly from (6). For each FDI step, we compute $p(r|x_{k+1-W:k}, \delta_j)$ with $W = 2$ by sampling $x_k \sim p_{X_k}$ and evaluate r_i under each fault hypothesis across 30 randomized scenarios, as if obtained from offline experiments.

For active FDI, assuming the asteroid's rotation is negligible relative to the spacecraft's motion, we adopt quaternion-based attitude control, with rotational dynamics expressed as $\dot{\omega} = J_c^{-1}\big(u_{rot} - \omega \times (J_c\omega)\big)$, where J_c is the spacecraft's moment of inertia, ω is spacecraft's angular velocity, and u_{rot} is the attitude control input. Specifically, the control input along the camera's $x-$ and $z-$axes is perturbed, while the $y-$axis is constrained only for asteroid pointing. For additional simplicity, the spacecraft's actuation frame is assumed to be aligned with the camera axis.

B. Simulation Setup

The proposed FDI algorithm is implemented in Python, employing PyTorch3D for rendering, kernel density estimation to model reliability distributions, and MCTS for active FDI with rollouts based on prior control inputs. Control inputs are discretized within the reaction wheel torque limits.

It should be noted that the real-time feasibility of the proposed MCTS-based active FDI method is governed by the planning budget, tree depth, and rollout count, each of which can be adjusted to satisfy the timing constraints of a given platform. In the asteroid exploration scenario considered in this work, the relatively slow spacecraft dynamics provide a sufficiently large planning window, which allows the MCTS to converge to a high-quality control input within the required control cycle. The search can also be warm-started using prior control inputs as rollout seeds, which is a strategy already reflected in our implementation.

C. Reliability Test Distribution

The first simulation is designed to verify whether the distribution of the reliability test depends on the state, thereby demonstrating that having a fixed threshold for the reliability test result can lead to false alarms. Figure 4 describes that under fault interventions, the distribution of normalized reliability tests varies with scene context and robot motion. Across two exemplar views, the fault-free hypothesis concentrates most test values near zero, whereas each fault hypothesis typically activates multiple components of r with distinct sensitivities, enabling fault isolation: (i) dynamic scenes (δ_1) reduce the number of detected features (increasing r_1) and increase motion-related components—r_3, r_5, r_6; (ii) sensor damage (δ_2) induces increase in stationary pixels (r_2) and broadly degrades reliability; and (iii) visually deprived conditions (δ_3) decrease the number of detected features (increasing r_1) while leaving motion-related components relatively healthy. Nevertheless, even under the same fault hypothesis, the induced distributions differ across states, with Views 1 and 2 showing distinct patterns. This motivates active perception FDI: by adjusting the relative state via control, we can elicit more informative patterns in r to improve fault detectability and isolability.

D. FDI Performance Analysis

The second experiment is to compare our passive and active FDI methods with baseline FDI methods: threshold-based passive FDI and random control input-based FDI. The threshold-based passive FDI is defined as follows:

$$\delta_i = \delta_1 \mathbf{1}_{\{r_6 > 0.3\}} + \delta_2 \mathbf{1}_{\{r_6 \le 0.3, r_2 > 0.5\}} + \delta_3 \mathbf{1}_{\{r_6 \le 0.3, r_2 \le 0.5, r_1 > 0.02\}} + \delta_0 \mathbf{1}_{\{\text{otherwise}\}}.$$

Figure 5 shows the FDI results and EI values for each case, with the upper and lower rows corresponding to severe and mild faults, respectively. Severe faults in this simulation mean faults that induce larger 2-norm estimation errors in the perception system. In the severe fault cases, all FDI methods detect and isolate the fault within 0.4 seconds of onset, as indicated by the isolated fault mode plots. Passive and active FDI achieve comparable EI values, implying that diagnosability and isolability are already sufficient in these conditions. In the mild fault cases, only the active FDI method successfully diagnoses the fault mode while driving the system toward maximum EI. The threshold-based FDI is highly sensitive to scene context, as seen in performance variations across case 3, while our passive FDI fails to detect motion-related faults as in case 1. Although introducing ego-motion can improve performance in such cases, random control inputs cannot maximize EI and result in degraded performance compared to our active FDI method.

VII. Conclusion

This paper presented a perception-monitoring FDI framework based on counterfactual reasoning. The method utilizes intermediate perception outputs and reliability tests to extract diagnostic information without relying on sensor redundancy,

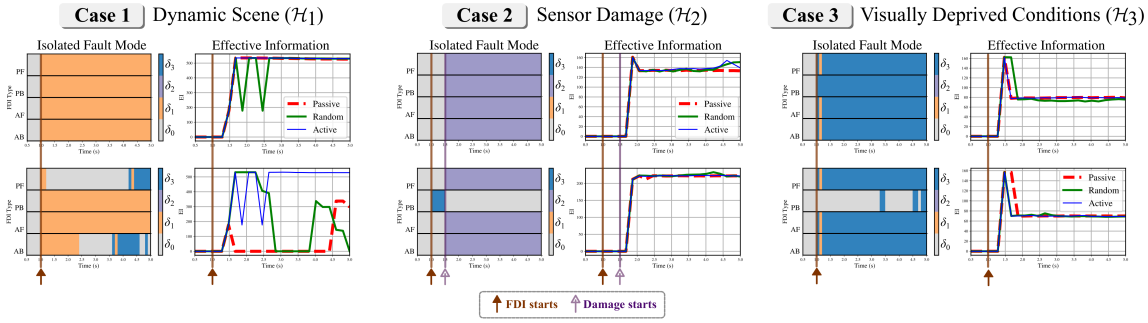

Fig. 5. FDI and EI comparison between different methods. The diagnosis process is initiated after 1 second. The upper row represents a severe fault, and the lower row represents a mild fault. PF is our Passive FDI, FB is the Passive Baseline, AF is our Active FDI, AB is the Active Baseline with random control input. The blue curve, corresponding to AF, demonstrates that MCTS-based control input yields higher EI, indicating improved fault diagnosability.

and fault hypotheses are evaluated through distributions derived from the fault and control interventions. To improve FDI performance, we introduced EI, which quantifies fault detectability and isolability via the KL divergence. Control inputs are optimized via solving a causal bandit problem using MCTS, to maximize EI while maintaining trajectory tracking. Results of simulation for an asteroid exploration scenario show that counterfactual reasoning combined with active control can render perception systems both more robust and self-explanatory. Beyond this, it establishes causal inference as a rigorous formalism to detect and isolate faults in complex perception pipelines.

REFERENCES

[1] P. Corke, W. Jachimczyk, and P. Remo, *Robotics, Vision and Control*, vol. 118 of *Springer Tracts in Advanced Robotics*. Cham: Springer Cham, 2023.

[2] R. Sinha, E. Schmerling, and M. Pavone, "Closing the loop on runtime monitors with fallback-safe mpc," in *Proc. IEEE Conf. Decis. Control (CDC)*, pp. 6533–6540, 2023.

[3] P. Antonante, H. G. Nilsen, and L. Carlone, "Monitoring of perception systems: Deterministic, probabilistic, and learning-based fault detection and identification," *Artificial Intelligence*, vol. 325, p. 103998, 2023.

[4] O. A. Hafez, M. Joerger, and M. Spenko, "Quantifying mobile robot localization safety for an ekf-based slam estimator: An integrity monitoring approach," *Int. J. Robot. Res.*, vol. 44, no. 6, pp. 972–988, 2025.

[5] "ispace releases technical cause analysis for HAKUTO-R Mission 2." Press release, June 2025. Accessed: 2025-07-03.

[6] T. Wu, J. Zhang, X. Fu, Y. Wang, L. P. Jiawei Ren, W. Wu, L. Yang, J. Wang, C. Qian, D. Lin, and Z. Liu, "Omniobject3d: Large-vocabulary 3d object dataset for realistic perception, reconstruction and generation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2023.

[7] T. Goelles, B. Schlager, and S. Muckenhuber, "Fault detection, isolation, identification and recovery (fdiir) methods for automotive perception sensors including a detailed literature survey for lidar," *Sensors*, vol. 20, no. 13, 2020.

[8] W. Hou, W. Li, and P. Li, "Fault diagnosis of the autonomous driving perception system based on information fusion," *Sensors*, vol. 23, no. 11, 2023.

[9] S. Mitra, C. Păsăreanu, P. Prabhakar, S. A. Seshia, R. Mangal, Y. Li, C. Watson, D. Gopinath, and H. Yu, "Formal verification techniques for vision-based autonomous systems–a survey," in *Principles of Verification: Cycling the Probabilistic Landscape: Essays Dedicated to Joost-Pieter Katoen on the Occasion of His 60th Birthday, Part III*, pp. 89–108, Springer, 2024.

[10] T. Ji, S. T. Vuppala, G. Chowdhary, and K. R. Driggs-Campbell, "Multi-modal anomaly detection for unstructured and uncertain environments," vol. abs/2012.08637, 2020.

[11] C. Hsieh, Y. Koh, Y. Li, and S. Mitra, "Assuring safety of vision-based swarm formation control," in *Proc. Amer. Control Conf. (ACC)*, pp. 3215–3222, 2024.

[12] M. O'Connell, J. Cho, M. Anderson, and S.-J. Chung, "Learning-based minimally-sensed fault-tolerant adaptive flight control," *IEEE Robot. Autom. Lett. (RA-L)*, vol. 9, no. 6, pp. 5198–5205, 2024.

[13] M. Tranzatto *et al.*, "Team cerberus wins the darpa subterranean challenge: Technical overview and lessons learned," 2022.

[14] M. Tranzatto *et al.*, "Cerberus in the darpa subterranean challenge," *Science Robotics*, vol. 7, no. 66, p. eabp9742, 2022.

[15] S. Rahman, R. DiPietro, D. Kedarisetti, and V. Kulathumani, "Large-scale indoor mapping with failure detection and recovery in slam," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, pp. 12294–12301, 2024.

[16] A. H. Qureshi, M. L. Anjum, W. Hussain, U. Muddassar, and S. Abbasi, "One step back, two steps forward: learning moves to recover from slam tracking failures," *Advanced Robotics*, vol. 38, no. 5, pp. 307–322, 2024.

[17] S. Pathak, A. Thomas, and V. Indelman, "A unified framework for data association aware robust belief space planning and perception," *Int. J. Robot. Res.*, vol. 37, no. 2-3, pp. 287–315, 2018.

[18] M. Shienman and V. Indelman, "D2a-bsp: Distilled data association belief space planning with performance guarantees under budget constraints," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, pp. 11058–11065, 2022.

[19] J. Zhang, Y. Tang, H. Wang, and K. Xu, "Asro-dio: Active subspace random optimization based depth inertial odometry," *IEEE Transactions on Robotics*, vol. 39, no. 2, pp. 1496–1508, 2023.

[20] X. Zhao, Z. Gao, H. Li, H. Ji, H. Yang, C. Li, H. Fang, and B. M. Chen, "How Challenging is a Challenge? CEMS: a Challenge Evaluation Module for SLAM Visual Perception," *J. Intell. Robot. Syst.*, vol. 110, p. 42, Mar. 2024.

[21] A. Samadzadeh and A. Nickabadi, "Srvio: Super robust visual inertial odometry for dynamic environments and challenging loop-closure conditions," *IEEE Trans. Robot.*, vol. 39, no. 4, pp. 2878–2891, 2023.

[22] S. Han and Z. Xi, "Dynamic scene semantics slam based on semantic segmentation," *IEEE Access*, vol. 8, pp. 43563–43570, 2020.

[23] J. Guo, R. Ni, and Y. Zhao, "Deblurslam: A novel visual slam system robust in blurring scene," in *Proc. IEEE Int. Conf. Virtual Reality (ICVR)*, pp. 62–68, 2021.

[24] F. Min, Z. Wu, D. Li, G. Wang, and N. Liu, "Coeb-slam: A robust vslam in dynamic environments combined object detection, epipolar geometry constraint, and blur filtering," *IEEE Sens. J.*, vol. 23, no. 21, pp. 26279–26291, 2023.

[25] J. Pearl, *Causality*. Cambridge University Press, 2 ed., 2009.

[26] E. Bareinboim, A. Forney, and J. Pearl, "Bandits with unobserved confounders: A causal approach," *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 28, 2015.

[27] L. Kocsis and C. Szepesvári, "Bandit based monte-carlo planning," in *Proc. Eur. Conf. Mach. Learn. (ECML)*, pp. 282–293, Springer, 2006.

[28] D. Silver and J. Veness, "Monte-carlo planning in large pomdps," *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 23, 2010.

[29] P. Chvykov and E. Hoel, "Causal geometry," *Entropy*, vol. 23, no. 1, p. 24, 2020.

[30] E. P. Hoel, "When the map is better than the territory," *Entropy*, vol. 19, no. 5, p. 188, 2017.

[31] G. Tononi and O. Sporns, "Measuring information integration," *BMC neuroscience*, vol. 4, pp. 1–20, 2003.

[32] D. A. Reynolds *et al.*, "Gaussian mixture models.," *Encyclopedia of biometrics*, vol. 741, no. 659-663, p. 3, 2009.

[33] E. Parzen, "On estimation of a probability density function and mode," *The annals of mathematical statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.

[34] T. M. Cover, *Elements of information theory*. John Wiley & Sons, 1999.

[35] F. Lattimore, T. Lattimore, and M. D. Reid, "Causal bandits: Learning good interventions via causal inference," *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 29, 2016.

[36] "Asteroid (99942) apophis." 3D Asteroid Catalogue. [Online]. Available: https://3d-asteroids.space/asteroids/99942-Apophis.