USENIX

THE ADVANCED COMPUTING
SYSTEMS ASSOCIATION

WIP: Evaluating the End-to-End Impact of False Localization Attacks on vSLAM-Based Autonomous Drones

Yuga Ebine, *Waseda University;* Kazuki Nomoto and Yuna Tanaka, *Waseda University, Deloitte Tohmatsu Cyber LLC;* Ryunosuke Kobayashi and Go Tsuruoka, *Waseda University;* Tatsuya Mori, *Waseda University, RIKEN AIP, NICT*

https://www.usenix.org/conference/vehiclesec25/presentation/ebine

This paper is included in the Proceedings of the 3rd USENIX Symposium on Vehicle Security and Privacy.

August 11–12, 2025 • Seattle, WA, USA

978-1-939133-49-6

Open access to the Proceedings of the 3rd USENIX Symposium on Vehicle Security and Privacy is sponsored by USENIX.

WIP: Evaluating the End-to-End Impact of False Localization Attacks on vSLAM-Based Autonomous Drones

Yuga Ebine
Waseda University

Kazuki Nomoto
Waseda University,
Deloitte Tohmatsu Cyber LLC

Yuna Tanaka
Waseda University,
Deloitte Tohmatsu Cyber LLC

Ryunosuke Kobayashi
Waseda University

Go Tsuruoka
Waseda University

Tatsuya Mori
Waseda University, RIKEN AIP, NICT

Abstract

Visual Simultaneous Localization and Mapping (vSLAM) is critical for autonomous navigation in self-driving vehicles, robotics, and drones, yet its security vulnerabilities remain largely unexplored. This study introduces *Phantom Path Attack*, an adversarial method that misguides drones using ORB SLAM3 by projecting deceptive video stimuli, leading to severe localization errors. Unlike previous attacks that rely on static adversarial inputs, *Phantom Path Attack* attack dynamically manipulates vSLAM's motion estimation, causing drones to deviate from their intended trajectory. We evaluate the impact of the attack through simulations and real camera experiments, demonstrating localization errors of up to 252 meters, while an end-to-end drone simulation reveals altitude deviations of 70 meters, ultimately leading to potential crashes. These findings reveal critical security risks in vSLAM-based systems and highlight the need for robust countermeasures, such as LiDAR/IMU sensor fusion and dynamic filtering of moving objects, to mitigate adversarial manipulation and improve resilience.

1 Introduction

Visual Simultaneous Localization and Mapping (vSLAM) plays a critical role in autonomous drone navigation, enabling precise localization and mapping from camera input. It supports applications in infrastructure inspection, search and rescue, and autonomous exploration [1]. However, the heavy reliance on visual data makes vSLAM vulnerable to adversarial perturbations, resulting in localization errors that pose serious safety and security risks [2, 3].

Recent research demonstrates vSLAM's vulnerability to environmental manipulation. For example, Chen et al. [2] introduced an adversarial patch attack that subtly alters visual features, significantly disrupting SLAM-based localization. However, this attack requires generating patches based on knowledge of machine learning, making it highly challenging. Moreover, while it can cause errors in localization, it cannot steer the model in an arbitrary direction. Similarly, Nassi et al. [3] demonstrated how projected phantom objects can deceive advanced driver assistance systems (ADAS) by inducing erroneous responses. These studies highlight the potential of perceptual attacks over traditional software-based cyberattacks. Wang et al. [4] propose an attack that induces errors in SLAM systems used in autonomous vehicles by utilizing an IR light-based approach. Nemcovsky et al. [5] showed that deploying a passive patch in the scene can substantially increase the error of a visual odometry model, posing a significant threat to autonomous drone navigation. Doe et al. [6] proposed a perceptual aliasing-based attack that exploits subtle ambiguities in visual features to disrupt loop closure detection. However, this attack is aimed at mapping in vSLAM, with a particular focus on loop closure, and thus differs in scope from the present study, which centers on localization. Fukunaga [7] introduced a random spoofing attack targeting LiDAR-based scan-matching SLAM systems. Although this study proposes an attack that adversely affects vehicle motion planning, its impact on Visual SLAM has not been evaluated. In addition, several studies have been conducted on drones [8–10]. . However, these studies have primarily focused on drone communication systems. While image-based attacks only introduce localization errors, our adversarial video attack can deliberately steer localization in any desired direction, significantly increasing the risk of drone crashes. Furthermore, this study focuses on the sensor security of autonomous flying drones.

Based on these observations, we present *Phantom Path Attack*, a novel adversarial attack that systematically misguides vSLAM-based drones (see Figure 1). Unlike previous methods that rely on static adversarial patches or projected phantom objects, our attack employs adversarially generated video projections to distort vSLAM's motion estimation. By projecting controlled visual stimuli into the drone's viewpoint, the attack constructs a fabricated trajectory, a "phantom path," leading to substantial localization errors and potential crashes.

To evaluate the impact of *Phantom Path Attack*, we conduct experiments under two distinct settings: (1) we eval-

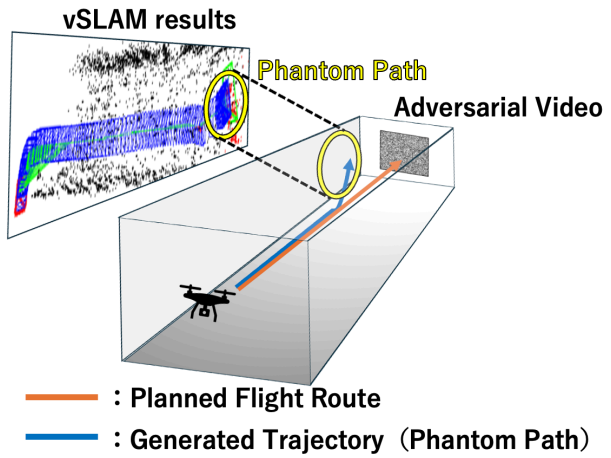


Figure 1: Illustration of the *Phantom Path Attack.*

uate the feasibility of the attack through the extensive experiments using a stand-alone vSLAM system, and (2) we perform an End-to-End (E2E) evaluation using a simulator to examine the impact on an autonomously navigating drone with the vSLAM system. For both experiments, we adopt ORB SLAM3 [11], a widely adopted open-source vSLAM framework, due to its extensive academic and commercial use, state-of-the-art performance, and reliance on feature-based visual cues, which makes it particularly vulnerable to adversarial manipulation. Our results show localization errors up to 252 meters in the Camera-only evaluation, and in the E2E setting, an altitude deviation of about 70 meters leads to a crash.

These findings reveal a critical security vulnerability in vSLAM-based drones. To mitigate this threat, we propose potential defensive measures, including sensor fusion techniques that integrate LiDAR or IMU data and dynamic filtering of uniformly moving objects. Our research highlights the threat of attacks targeting camera-based autonomous drones and proposes defensive techniques, thereby contributing to the enhancement of their safety and reliability. Our contributions can be summarized as follows:

- We introduce *Phantom Path Attack*, an adversarial method that exploits video projections to manipulate vSLAM-based localization, as illustrated in Figure 1.
- We show that the attack induces severe localization errors, with position estimation offsets up to 252 meters in ORB SLAM3.
- We perform the first end-to-end evaluation of adversarial attacks on vSLAM-powered autonomous drones, revealing controlled trajectory manipulation.
- We propose defense strategies using sensor fusion and anomaly detection to reinforce security in vSLAM-based autonomous systems.

2 Background

2.1 Autonomous Drones and vSLAM

Autonomous drones operate without external commands, using onboard sensors for localization and path planning. They are widely used in disaster rescue, agriculture, logistics, and infrastructure inspection [12]. However, localization is challenging in indoor environments or complex urban areas where GNSS signals are unavailable. To overcome this, vSLAM has emerged as a lightweight, camera-based localization solution [1]. By extracting and matching features, estimating poses, and constructing maps, vSLAM jointly provides environmental mapping and localization for large-scale areas. It consists of feature extraction, feature matching, pose estimation, and map construction. Representative algorithms include ORB SLAM3, LSD-SLAM, PTAM, and MonoSLAM, with ORB SLAM3 noted for its accuracy and efficiency [11, 13–15].

2.2 ORB SLAM3 Overview

ORB SLAM3 is a vSLAM algorithm that enables real-time localization and mapping using ORB features, making it suitable for autonomous drones operating in GPS-denied environments. It follows three key steps: tracking, where ORB features are extracted using FAST corner detection and BRIEF descriptors to estimate camera motion frame by frame; local mapping, which builds a 3D map by triangulating keyframes and refining the structure with bundle adjustment; and loop closure, which detects previously visited locations through feature matching and optimizes the global map to correct drift using PnP and bundle adjustment.

In ORB SLAM3, if the feature correspondences extracted from the input video fail to match, the system loses its ability to perform localization and build an environmental map. This phenomenon is called tracking loss. By leveraging monocular, stereo, or RGB-D cameras, ORB SLAM3 allows drones to navigate complex environments with high accuracy. Its ability to maintain stable positioning and correct drift makes it valuable for applications such as infrastructure inspection, search-and-rescue, and autonomous exploration.

3 Phantom Path Attack

3.1 Threat Model

The proposed attack induces localization errors in an autonomous drone using ORB SLAM3 under specific conditions, causing it to stray from its intended flight path or crash. For example, in an indoor security patrol scenario, an attacker can manipulate the drone's flight path or cause it to crash and avoid surveillance by projecting images onto nearby walls when the drone approaches them. In certain limited scenarios, the images contain distinctive patterns, but

because autonomous indoor drones are often deployed to reduce human labor and operate without constant human supervision [16], such attacks may go unnoticed in practice. It is assumed that the attacker knows the drone's planned route and its use of ORB SLAM3 for localization. They use a projector to generate and display adversarial videos on a suitable surface. This attack can be executed remotely, making it a stealthy and effective method to disrupt drone navigation. However, since the drone must approach the adversarial video quite closely, the scenarios in which the attack can be successfully carried out are limited.

3.2 Attack Overview and Procedure

Phantom Path Attack manipulates feature extraction of ORB SLAM3 via adversarial videos, exploiting its localization process. As shown in Figure 1, a high-contrast geometric video is projected onto a screen in the drone's environment. The vSLAM system extracts false feature points, causing localization errors and flight deviations.

The attack consists of three key steps:

Step 1: Creation of Adversarial Video: The adversarial video is designed to induce localization errors in the target drone while avoiding tracking loss in ORB SLAM3. To maximize feature extraction, the video employs strong brightness contrasts while avoiding repetitive patterns. The detailed generation process is described in Section 3.3.

Step 2: Placement of the Projection System: The attacker installs a projector at a strategic location along the drone's known flight path. If direct projection onto a surface is infeasible, a screen is used. The placement ensures that the drone's vSLAM system captures the adversarial video during operation.

Step 3: Projection of Adversarial Video: Once the setup is complete, the adversarial video is projected when the drone approaches the designated location. Continuous looping of the video eliminates the need for precise timing, ensuring sustained exposure to the adversarial patterns.

3.3 Adversarial Video

The adversarial video used in this study must maximize ORB SLAM3's feature extraction while minimizing tracking loss. We achieve this by creating fractal-noise-based black-and-white patterns moving consistently in one direction.

Feature Extraction and FAST Corner Detection: As detailed in Section 2.2, ORB SLAM3 uses ORB feature extraction for real-time localization and mapping. The FAST algorithm detects corners by evaluating each pixel against 16 surrounding pixels, identifying a corner if 9 consecutive pixels show sufficient contrast. Only local maxima proceed to BRIEF descriptor processing.

Generating the Adversarial Video: To disrupt localization while maintaining tracking, the adversarial video must both induce localization errors by disrupting feature extraction and prevent tracking loss. High-contrast black-and-white patterns increase detected corners, so repetitive patterns that hinder feature matching are avoided. This study uses fractal noise to create patterns with balanced randomness and contrast. Multi-resolution fractal noise is blended, Gaussian-blurred, and thresholded to produce Pattern 1, with Patterns 2 and 3 generated for comparison as shown in Figure 2.

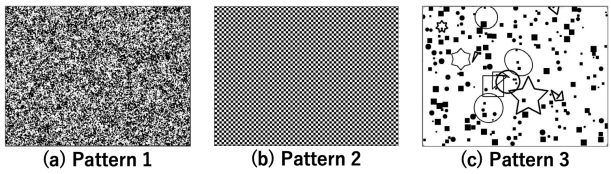


Figure 2: Three types of patterns composing adversarial videos: (a) Fractal noise, (b) Checkerboard Pattern, (c) Artistic Pattern.

Table 1: Number of Corner Detections per 900 Frames Using the FAST Algorithm for Each Video

Pattern 1	Pattern 2	Pattern 3
10,230,973	0	1,421,099

Comparison of Corner Detection Counts: To evaluate the effectiveness of each pattern, the FAST algorithm was applied to three types of adversarial videos, with corner detection performed five times per video. The average detection results, summarized in Table 1, confirm that Pattern 1 induces the highest number of detected corners, making it the most effective in causing localization errors.

4 Attack Evaluation on a Stand-Alone vSLAM System

4.1 Simulation-Based Evaluation

Setup and Procedure. As shown in Figure 3, we used Unity as the simulator and processed each topic on ROS2. Figure 4 illustrates the experimental environment, where we built a reproducible simulation setup. A virtual scene is created in which a static camera captures video input while an adversarial video is displayed on a screen. The adversarial videos, generated as described in Section 3.3, are projected onto the screen at predefined distances from the camera, specifically $d = 0.8, 0.9, \ldots, 1.3$ meters. These distances are chosen to analyze how variations in the screen-to-camera ratio affect the localization performance of ORB SLAM3.

The Unity camera model matches the Intel RealSense Depth Camera D435i's specifications, ensuring consistency in focal length and field of view. The camera remains fixed to prevent motion artifacts, while ORB SLAM3, running in a ROS2 framework, processes the video stream in real time

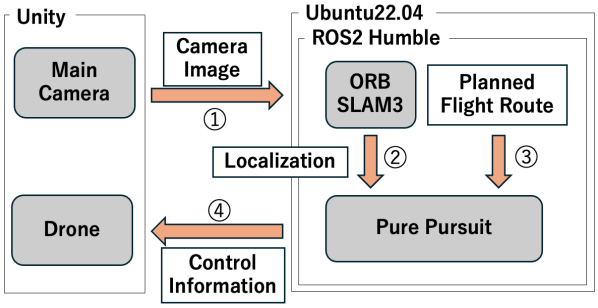
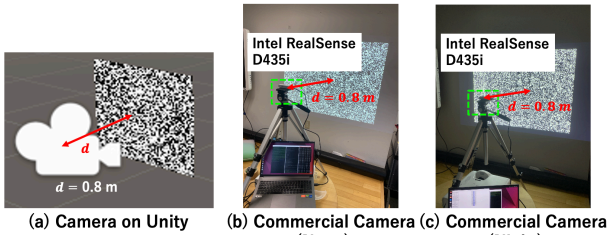


Figure 3: Experimental Setup.



(a) Camera on Unity (b) Commercial Camera (Noon) (c) Commercial Camera (Night)

Figure 4: Experimental Environment.

for localization and feature tracking. The adversarial video loops for 30 seconds while the camera records the scene, allowing ORB SLAM3 to build a trajectory and respond to perturbations. Localization performance is logged per frame, and displacement is measured by comparing the estimated camera positions at the start and end of the period, quantifying the adversarial video's impact on localization accuracy.

Results. Table 2 presents localization shifts under different attack conditions. Displacement varies with camera-screen distance, indicating that attack effectiveness depends on screen coverage in the camera's field of view. The largest shift, approximately 125 m, occurs at $d = 0.8$,m under Pattern 1, demonstrating the strongest adversarial effect. Except in cases of tracking loss, the attack consistently caused significant localization errors across all video patterns. Displacement remained similar, particularly at shorter distances where $d \leq 1.1$,m, as ORB SLAM3 continuously estimated incorrect trajectories. This suggests that adversarial perturbations effectively manipulate localization when the camera primarily relies on projected content.

As the camera moves further away from the screen, the effect of the adversarial perturbations diminishes, with displacement decreasing at $d = 1.2$ m and $d = 1.3$ m, and tracking loss occurring in some cases. This implies that when the adversarial video occupies a smaller portion of the camera's view, ORB SLAM3 can maintain localization accuracy by exploiting stable environmental features. Specifically, when $d = 1.1$ m, the screen occupies 54 % of the camera's field of view, while at $d = 1.2$ m, this proportion decreases to 46 %. This suggests that the attack's effectiveness may decline when the screen's coverage within the camera's view falls below approximately 50 %. Additionally, when using the Pattern 2 video, there were cases where localization and mapping mechanisms of ORB SLAM3 failed to initialize. This issue arises because ORB SLAM3 first detects corners using the FAST algorithm, which then extracts these corners as feature points. With Pattern 2, ORB SLAM3 sometimes failed to initialize because its checkerboard pattern lacked FAST-detectable corners, preventing feature extraction. These results confirm that adversarial videos reliably disrupt localization, provided the attack remains within an effective range.

Table 2: Localization Shift in Simulation-Based Evaluation.

Distance d (m)	The proportion of the screen occupying the camera's field of view	Pattern 1	Pattern 2	Pattern 3
0.8	103 %	124.7	-	∗ 38.7
0.9	81 %	114.9	-	115.2
1.0	66 %	96.8	∗ 0.5	104.8
1.1	54 %	81.0	-	88.0
1.2	46 %	∗ 12.7	∗ 0.0	92.6
1.3	39 %	∗ 18.3	-	∗ 12.5

∗ Tracking loss occurred midway.
- ORB SLAM3 did not start.

4.2 Real-World Evaluation

Setup and Procedure. Figure 4 shows the experimental setup for real-world evaluation, ensuring the attack's feasibility beyond simulation. An Intel RealSense D435i camera and an EPSON EB-L200SW projector display adversarial videos on a white screen at distances ranging from $d = 0.8$ m to 1.3 m from the camera, mirroring the simulation setup for direct comparison. To evaluate lighting effects on ORB SLAM3, experiments were conducted in bright (noon) and dark (night) environments. The camera recorded adversarial videos for 30 seconds per trial, and ORB SLAM3 in ROS2 processed the footage to extract localization trajectories and measure displacement, assessing the impact of adversarial perturbations in real-world conditions.

Results. Table 3 shows the localization shifts observed under real-world conditions. Similar to the simulation-based evaluation, the shift varies with screen-to-camera distance, with the largest shift occurring at $d = 0.8$,m for Pattern 3 in night conditions, reaching 252 m. This confirms that the adversarial videos effectively manipulate localization of ORB SLAM3 in practical settings. The attack consistently caused significant displacement across all adversarial video patterns at $d \leq 1.1$ m, with minor variations. At $d \geq 1.2$ m, displacement decreased, indicating reduced vulnerability when the adversary video occupies less of the camera's view. Higher attack effectiveness in real-world tests compared to simulation is due to the commercial camera capturing a larger portion of the screen, highlighting a performance gap between

Table 3: Localization Shift in Real-World Evaluation.

Distance d (m)	Pattern 1		Pattern 2		Pattern 3	
	Noon	Night	Noon	Night	Noon	Night
0.8	231.1	227.5	-	-	243.6	251.7
0.9	204.1	221.7	-	-	216.3	219.2
1.0	199.8	195.3	-	-	207.8	212.8
1.1	176.9	175.4	-	-	171.5	186.9
1.2	162.2	155.4	-	-	171.8	157.4
1.3	152.3	145.3	-	-	156.3	123.7

- ORB SLAM3 did not start.

the two camera systems. Displacement remains significant in both noon and night conditions but is slightly higher at night, especially for Pattern 3. This suggests lower lighting enhances the adversarial effect as perturbations stand out more against a darker background. The results confirm adversarial perturbations are highly effective in real-world, low-light scenarios, severely impacting ORB SLAM3's feature extraction and increasing localization errors.

5 End-to-End Evaluation Using a Simulator

This section evaluates the effectiveness of *Phantom Path Attack* in disrupting an autonomous drone's navigation in an E2E manner, extending the stand-alone evaluation presented in the previous section. While previous studies have examined attacks on vSLAM or its individual components, there has been limited evaluation of their effects on the entire autonomous flight system, from sensing to control. *Phantom Path Attack*, which has been validated in stand-alone evaluations, is now tested in a closed-loop system to assess its impact on drone navigation.

5.1 Experimental Setup

The experimental setup is illustrated in Figure 3. This experiment employs the pipeline spanning from ① to ④ in the figure. Among the adversarial video patterns, Pattern 1, which exhibited the highest number of detected corners using the FAST algorithm, is selected for the attack.

The experimental environment in both non-attack and attack scenarios is shown in Figure 5. The Unity simulator is used again, with a warehouse asset set up to simulate an indoor flight scenario. The drone is initially placed on the ground at 0 m, directly facing the screen, with a fixed distance of 60 m between them. The adversarial video is projected in a way that induces an upward shift in localization estimation, meaning the video content moves downward.

Upon simulation start, the drone takes off and is controlled via the Pure Pursuit algorithm with a PD controller. The Look Ahead Distance is set to 0.5 m. The algorithm continuously compares estimated localization of ORB SLAM3 with the predefined trajectory, adjusting movement to follow waypoints. The flight path consists of 175 waypoints, spaced 1 m apart at a height of 3 m. Waypoints are placed on warehouse assets in a straight line aligned with the drone's forward direction. A straight line connecting waypoints is published as the Path information topic. ORB SLAM3's self-position estimation is compared with the planned route using Pure Pursuit to adjust control information accordingly.

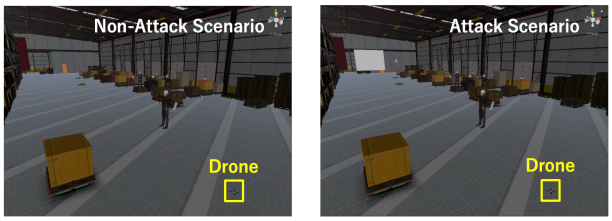


Figure 5: Simulation environment for the E2E evaluation. The setup includes a drone facing a screen projecting adversarial videos.

A simplified drone dynamics model applies independent acceleration in the x, y, and z directions. The Pure Pursuit algorithm adjusts only lateral and vertical movements, assuming default forward motion. A constant acceleration of 1.0 [$\mathrm{m\,s^{-2}}$] is applied forward, with a velocity limit of 5.0 [$\mathrm{m\,s^{-1}}$].

First, the drone's behavior without an attack is evaluated. Then, an adversarial video is projected on a screen ahead, and flight behavior is assessed for different screen sizes: (I) 9 m (height) × 12 m (width), (II) 6 m × 8 m, and (III) 3 m × 4 m. As the drone advances, the adversarial video enters its field of view, causing localization errors. The impact on flight behavior is analyzed by deviations from the planned trajectory.

5.2 Experimental Procedure

The experiment consists of two parts: a non-attack scenario, where no adversarial video is projected, and an attack scenario, where the adversarial video is displayed on the screen. In both scenarios, the drone starts from the initial position and follows the predefined waypoints based on the localization results from ORB SLAM3. The Pure Pursuit algorithm continuously corrects the drone's trajectory to align with the waypoints. The simulation runs for 60 seconds, during which the estimated localization results, the ground truth trajectory from the simulator, and the planned trajectory are recorded and compared.

5.3 Results

Figure 6 shows the ORB SLAM3 output and trajectory visualization in the non-attack scenario, serving as a baseline for evaluating attack effects. The planned trajectory (red)

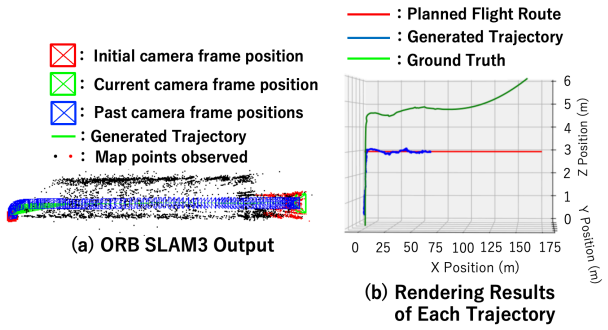


Figure 6: Rendering results of ORB SLAM3 and each trajectory in the non-attack scenario. The estimated localization aligns well with the planned trajectory.

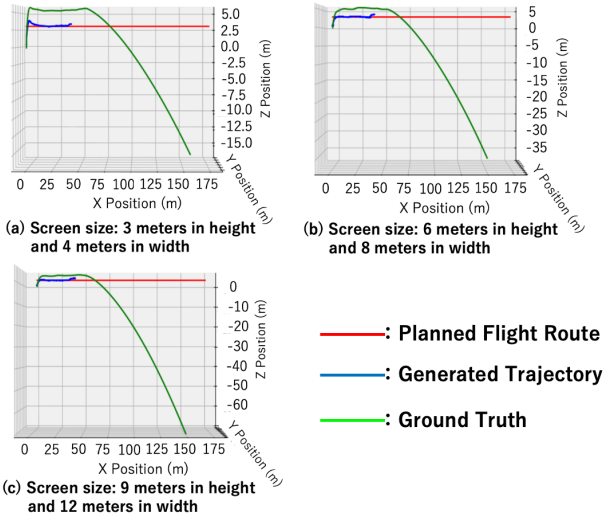


Figure 7: Generated trajectory deviations for different screen sizes during the attack (side view). Larger screens result in more significant downward deviations from the planned flight route.

represents the ideal path, the generated trajectory (blue) reflects localization of ORB SLAM3, and the ground truth trajectory (green) shows actual drone movement. The ground truth appears above the generated trajectory due to several-meter localization error of ORB SLAM3 in indoor environments [11]. Furtheremore, a slight time lag exists between takeoff and localization initiation. In the attack scenario, adversarial videos bias the drone's localization upward, causing downward trajectory corrections that may lead to a crash. With a 9 m $\times$ 12 m screen, the ground truth trajectory shifts downward by about 70 m from the initial position. Additionally, ORB SLAM3 loses tracking when the post-screen scene mismatches the pre-constructed map. Figure 7 shows trajectories under attack for different screen sizes. Larger screens cause greater downward deviations, increasing path divergence. This suggests *Phantom Path Attack* is amplified as adversarial content occupies more of the camera's view, leading to severe localization errors before tracking loss.

6 Discussion

6.1 Limitations

Variability Across vSLAM Systems This study evaluates *Phantom Path Attack* on ORB SLAM3, a widely used vSLAM system. Because different vSLAM systems use different feature extraction algorithms, the effectiveness of the attack may vary. Future work should evaluate its impact on other vSLAM systems. Since ORB SLAM3 is a feature-based method, it is necessary to test the effectiveness of *Phantom Path Attack* even in direct-method-based vSLAM.

Simulation Constraints and Real-World Applicability This study evaluated the impact of *Phantom Path Attack* on the control of autonomous drones in a simulated environment. However, the evaluation was based on a simplified simulation, and the effects of such attacks on the control systems of real-world commercial autonomous drones have not been assessed. The study uses a Pure Pursuit-based control algorithm, which is not widely adopted, so the attack's effects on drones with other control methods require further investigation. Moreover, the simulation employs an approximate motion model that may not fully reflect real-world behavior, necessitating validation in more realistic or actual environments. Although the attack was tested in simulation, its impact on commercial autonomous drones remains unclear, as these systems may use additional filtering or sensor fusion to mitigate such attacks. Therefore, assessing the effectiveness of *Phantom Path Attack* on commercial autonomous drones under real-world conditions remains a crucial direction for future research. To address this, we plan to evaluate the impact of the attack on self-localization and map generation using actual commercial drone platforms.

Attack Feasibility and Practical Scenarios The proposed attack method is transferable to indoor scenarios beyond security patrols, as long as the drone operates autonomously within an indoor environment. On the other hand, since the accuracy of ORB SLAM3 in outdoor environments is significantly lower, it is unlikely that autonomous drones used in outdoor scenarios rely on ORB SLAM3. Therefore, the feasibility of this attack in outdoor settings is considered to be low.

The current attack setup requires a relatively large projected video to induce localization errors. Furthermore, the current attack requires the drone to approach the adversarial projection to within approximately one meter, which constrains its applicability to specific scenarios. Nevertheless, since the effectiveness of the attack is influenced by the proportion of the adversarial image occupying the camera's field of view, it remains plausible that smaller projections could also yield successful attacks. Accordingly, further investi-

gation is warranted to evaluate this possibility. Identifying more practical scenarios, such as using smaller or adaptive projections, is an important research direction. The attack setup must also remain discreet to avoid detection. Advances in compact projectors improve stealth, especially in scenarios such as security drone patrols where third party observation is minimal.

6.2 Defense Measures

Hostile video causes significant localization errors, underscoring the need for countermeasures. Two defensive methods are being considered to address this issue. The first is adding sensors with sensor fusion, and the second is removing dynamic objects from the camera's field of view. For the first method, equipping the drone with IMU and LiDAR sensors allows for better movement estimation and reduces attack impact. However, LiDAR's weight makes it less suitable for drones, which are better equipped with only a camera if autonomous flight is possible. If the drone can autonomously fly with just a camera, it is preferable not to equip it with additional sensors in order to extend flight time. The second method involves removing dynamic objects from the camera feed used in ORB SLAM3, preventing adversarial image recognition and avoiding localization errors.

6.3 Ethical Considerations

This study does not exploit vulnerabilities in specific products, but rather demonstrates a fundamental weakness in the open source vSLAM system ORB SLAM3. Given the potential real-world impact, responsible disclosure procedures have been initiated to inform relevant stakeholders, including manufacturers of autonomous drones using ORB SLAM3. All real-world experiments were conducted in a controlled university environment with strict security precautions.

7 Conclusion

This paper introduces *Phantom Path Attack*, a novel adversarial attack that manipulates vSLAM-based localization by projecting adversarial video sequences into the drone's field of view. Through systematic evaluations, we show that the attack induces severe localization errors, causing estimated position shifts of up to 252 meters in a stand-alone ORB SLAM3 system and altitude deviations of approximately 60 meters in an end-to-end autonomous drone simulation, leading to unintended trajectory changes and potential crashes. These results reveal a critical safety risk in vSLAM-based autonomous navigation and emphasize the need for defensive measures such as sensor fusion with IMU and LiDAR to increase robustness. Future work should explore real-world applicability, refine attack scenarios for practicality, and evaluate mitigation strategies to ensure safe deployment of vSLAM-based autonomous systems.

Acknowledgments

This work was partially supported by JSPS KAKENHI 22H00519 and JST CREST JPMJCR23M4.

References

[1] Skydio. Skydio 2+. https://www.skydio.com/skydio-2-plus-enterprise.

[2] Baodong Chen, Wei Wang, Pascal Sikorski, and Ting Zhu. Adversary is on the Road: Attacks on Visual SLAM using Unnoticeable Adversarial Patch. In *33rd USENIX Security Symposium (USENIX Security 2024)*, pp. 905–922, 2024.

[3] Ben Nassi, Yisroel Mirsky, Dudi Nassi, Raz Ben-Netanel, Oleg Drokin, and Yuval Elovici. Phantom of the ADAS: Securing Advanced Driver-Assistance Systems from Split-Second Phantom Attacks. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security (CCS)*, pp. 293–308, 2020.

[4] Y. Wang, X. Yao, X. Liu, X. Li, P. Hao, and T. Zhu. I can see the light: Attacks on autonomous vehicles using invisible lights. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, Seoul, Republic of Korea, November 2021. ACM.

[5] Yaniv Nemcovsky, Matan Jacoby, Alex M. Bronstein, and Chaim Baskin. Physical passive patch adversarial attacks on visual odometry systems. In *Proceedings of the Asian Conference on Computer Vision (ACCV) 2022*, Seoul, Republic of Korea, October 2022. Springer.

[6] John Doe, Jane Smith, and Robert Brown. Perceptual aliasing++: Adversarial attack for visual slam front-end and back-end. *IEEE Robotics and Automation Letters*, Vol. 7, No. 2, pp. 1–1, April 2022.

[7] Masashi Fukunaga and Takeshi Sugawara. Random spoofing attack against lidar-based scan matching slam. In *Proceedings of the Symposium on Vehicle Security and Privacy (VehicleSec)*. Internet Society, 2024.

[8] Y. Son, H. Shin, D. Kim, Y. Park, J. Noh, K. Choi, J. Choi, and Y. Kim. Rocking drones with intentional sound noise on gyroscopic sensors. In *Proceedings of the 24th USENIX Security Symposium*, Washington, D.C., USA, August 2015. USENIX Association.

[9] Z. Wang, K. Wang, B. Yang, S. Li, and A. Pan. Sonic gun to smart devices: Your devices lose control under ultrasound/sound. In *Black Hat USA*, Las Vegas, NV, USA, July 2017. Black Hat.

[10] D. Davidson, H. Wu, R. Jellinek, T. Ristenpart, and V. Singh. Controlling uavs with sensor input spoofing attacks. In *WOOT*, Austin, TX, USA, August 2016. USENIX Association.

[11] Carlos Campos, Richard Elvira, Juan J. Gómez Rodríguez, José M. M. Montiel, and Juan D. Tardós. Orb-slam3: An accurate open-source library for visual, visual-inertial, and multimap slam. In *Proceedings of the IEEE Transactions on Robotics*, Vol. 37, pp. 1874–1890. IEEE, February 2021.

[12] U.S. Government Accountability Office. Drone operations. https://www.gao.gov/drone-operations.

[13] Jakob Engel, Thomas Schöps, and Daniel Cremers. Lsd-slam: Large-scale direct monocular slam. In *European Conference on Computer Vision (ECCV)*, pp. 834–849. Springer, 2014.

[14] Georg Klein and David Murray. Parallel tracking and mapping for small ar workspaces. In *ISMAR*, pp. 225–234. IEEE International Symposium on Mixed and Augmented Reality (IEEE), 2007.

[15] Andrew J. Davison, Ian D. Reid, Nicholas D. Molton, and Olivier Stasse. Monoslam: Real-time single camera slam. In *Computer Vision and Pattern Recognition (CVPR)*, pp. 1052–1067. IEEE, 2007.

[16] Security Industry Association. Surveillance and beyond: How autonomous indoor drones address security and operational challenges. https://www.securityindustry.org/2024/03/19/surveillance-and-beyond-how-autonomous-indoor-drones-address-security-and-operational-challenges/.