Rolling in the Deep: Exploiting Rolling Shutter Effect Against Stereo Depth Estimation in Drones

Dongfang Guo
Nanyang Technological University
Singapore
dongfang.guo@ntu.edu.sg

Rui Tan
Nanyang Technological University
Singapore
tanrui@ntu.edu.sg

Abstract

Stereo vision plays a critical role in enabling depth perception for drones, supporting navigation and obstacle avoidance in complex environments. However, the robustness and security of stereo vision systems remain largely underexplored. In this paper, we propose *Rolling in the Deep (RiD)*, a novel physical attack that exploits the rolling shutter effect (RSE) to inject imperceptible, structured perturbations into stereo image pairs. We analyze RSE formation in binocular camera setups and show how RSE-based perturbations can degrade deep learning-based stereo matching by exploiting model vulnerabilities and sensor misalignments, resulting in incorrect depth estimation. Preliminary results show the feasibility of *RiD* under realistic stereo configurations, revealing a new class of threats to drone perception systems.

CCS Concepts

• **Computer systems organization** → **Embedded and cyber-physical systems**; • **Security and privacy** → **Systems security**; **Side-channel analysis and countermeasures**.

Keywords

Stereo depth estimation, CMOS camera sensor, rolling shutter effect, adversarial attack

ACM Reference Format:
Dongfang Guo and Rui Tan. 2025. Rolling in the Deep: Exploiting Rolling Shutter Effect Against Stereo Depth Estimation in Drones. In *The 11th Workshop on Micro Aerial Vehicle Networks, Systems, and Applications (DroNet '25), June 23–27, 2025, Anaheim, CA, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3711875.3737660

1 Introduction

Micro aerial vehicles, commonly known as drones, have gained widespread adoption across diverse application domains, including aerial photography, express delivery, precision agriculture, infrastructure inspection, environmental monitoring, search-and-rescue operations, and military combat actions. Their increasing prominence is driven by advancements in autonomous navigation, lightweight sensor technology, and onboard computational capabilities. Accurate perception are fundamental to ensuring safe and effective operation, especially in dynamic, uncertain, and potentially hazardous environments.

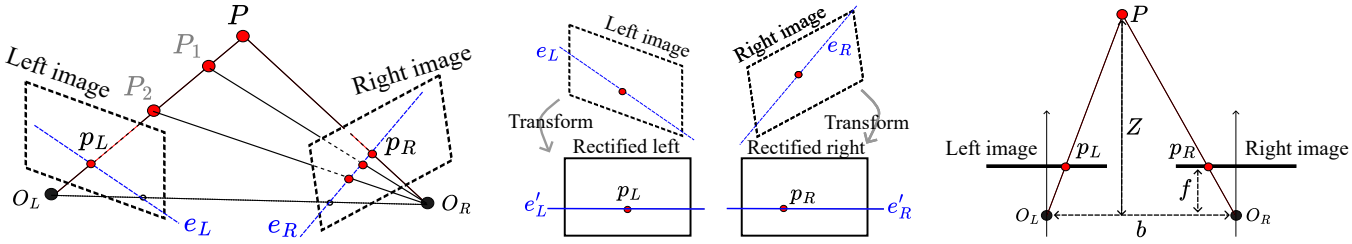
Among various perception tasks, depth estimation is one of the most foundamental and critical components for drones to peceive the surrounding targets for autonomous operation and obstacle avoidance. While ranging sensors like LiDAR and radar provide accurate depth measurements, their use in small drones is limited by cost, size, weight, and power constraints. In contrast, stereo vision estimates depth by computing pixel-wise disparity between images captured by two spatially separated cameras. Thanks to their favorable trade-offs in performance, cost, and weight, stereo cameras have become the *de facto* choice for depth sensing in many high-end commercial drone platforms, including the DJI Mavic and Phantom series [5, 6], Autel Evo II [1], Skydio 2/X2 [22, 23], and Parrot Anafi AI [19]. As such, the robustness of stereo camera-based depth estimation is critical for the safe and intelligent operation of drones. However, recent work [32] highlights a largely overlooked vulnerability in stereo vision: a physical attack that manipulates projected light beams and lens flare artifacts to induce false depth perceptions.

Building on this direction, we investigate a new class of optical attacks that exploit the inherent characteristics of image sensors. Most stereo camera systems in drones use complementary metal-oxide semiconductor (CMOS) sensors [14, 20, 25], which operate using a rolling shutter that exposes each scanline sequentially from top to bottom. Under high-frequency illumination changes, especially when flicker rates approach the shutter's scan frequency, this mechanism produces the *rolling shutter effect (RSE)*, leading to structured artifacts such as horizontal color stripes. Such RSE-induced distortions have been observed in commercial drones [33]. Recent studies have shown that adversarially modulated lighting can exploit RSE to create image perturbations that mislead deep learning models in single-camera tasks such as classification, object detection, and traffic light/sign recognitions [10, 11, 21, 30]. However, the implications of RSE-based attacks on stereo vision remain unexplored.

In this paper, we present *Rolling in the Deep (RiD)*, the first RSE-based physical attack targeting binocular stereo depth estimation. Specifically, *RiD* deploys an adversarial surface in the target drone's operational environment that reflects or emits flickering illumination. These controlled flickers generate adversarial color stripe patterns in the drone's stereo camera views, causing the estimated depth of the surface to be either increased or decreased. In the real world, the attack is stealthy, as the flickering frequencies exceed the perceptual limit of human vision and appear as benign illumination. Moreover, unlike prior work [32], *RiD* does not require

DroNet '25, Anaheim, CA, USA

ACM ISBN 979-8-4007-1453-5/25/06
https://doi.org/10.1145/3711875.3737660



(a) Camera stereo vision. (b) Stereo rectification. (c) Stereo depth estimation.

Figure 1: Image rectification and stereo depth estimation.

complex aiming maneuvers to direct attack lights into the camera lens. *RiD* could be used in military defense or personal property protection scenarios to disrupt unauthorized drone operations, either by denying access to protected areas or by misleading the drone into crashing or becoming trapped. The main contribution of this paper are summarized as follows:

- We propose *RiD*, a novel physical adversarial attack against stereo vision-based depth estimation in drones, leveraging the RSE.
- We analyze RSE formation in binocular camera setups and reveal how RSE-based perturbations can disrupt stereo matching by exploiting model vulnerability and/or temporal and spatial sensor misalignment.
- We conduct a proof-of-concept evaluation through simulation and demonstrate the feasibility of *RiD* under various stereo camera configurations.

Paper organization: Section 2 introduces the background and reviews related work. Section 3 analyzes RSE in stereo vision and presents the proposed *RiD* attack. Section 4 describes the evaluation setup and reports experimental results. Section 5 concludes the paper and outlines directions for future work.

2 Background and Related Work

2.1 Background

Stereo rectification & depth estimation. Fig. 1 illustrates the core principles of stereo vision systems. In Fig. 1a, a point P in the scene is projected onto the left and right image planes as p_L and p_R, respectively. The line connecting the two camera optical centers and the point P defines an epipolar plane, whose intersection with each image plane forms an epipolar line, denoted as e_L and e_R. According to epipolar geometry, the corresponding point of p_L in the right image must lie along the epipolar line e_R, and vice versa. This constraint reduces the correspondence search from 2D to 1D along the epipolar line, simplifying stereo matching. To further streamline correspondence, stereo vision systems apply *stereo rectification*, shown in Fig.1b, which transforms both images (typically via homographies) so that their epipolar lines become horizontal and aligned. Rectification serves as an equivalent (and more practical [18]) alternative to achieving perfect camera coplanarity, and is a standard preprocessing step in stereo vision systems. This is because even with high-precision hardware, maintaining perfect physical alignment is difficult in real-world setups due to factors such as imperfect calibration, mechanical tolerances, and

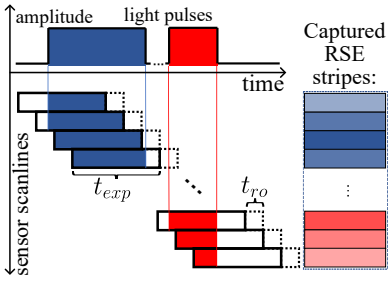


Figure 2: RSE.

misalignment between camera axes. After rectification, corresponding points lie on the same row in both images, allowing efficient disparity computation. As shown in Fig.1c, the depth Z of the 3D point P can then be calculated via triangulation from the disparity $d = p_L - p_R$, using the known focal length f and camera baseline b, i.e., $Z = (f \times b)/d$. In recent years, deep neural networks (DNNs) have become the state-of-the-art for stereo matching [3, 16, 29], leveraging learned features to predict disparity d, and most stereo datasets used for training stereo DNNs are well-calibrated and rectified [9, 17, 31]

Rolling shutter operation & effect. CMOS sensors with rolling shutters capture images row-by-row, with each scanline exposed for a duration t_{exp} and read out after a delay t_{ro}, as shown in Fig. 2. Because of this sequential exposure, different rows are captured at slightly different times. When the input light changes rapidly, some scanlines are exposed under different illumination conditions, resulting in visible artifacts known as the rolling shutter effect (RSE). These manifest as horizontal stripes in the image, where the stripe pattern depends on the timing and intensity of the light pulses during each scanline's exposure. The work [21] formulates the RSE from flickering LEDs as a mechanism for adversarial stripe optimization against object classification: Since each scanline is exposed at a different time, a time-varying attack light intensity $f(t)$ induces structured pixel variations across the image. The pixel value at pixel point (u, v) can be modeled as: $I(u, v) = I_{amb}(u, v) + \frac{I_{full}(u,v) - I_{amb}(u,v)}{t_{exp}} \int_{vt_{ro}}^{vt_{ro}+t_{exp}} f(t)\mathrm{d}t$ where $I_{amb}(u, v)$ is the pixel value under ambient illumination only, $I_{full}(u, v)$ is the image captured with both ambient and full LED illumination. By pre-capturing these reference images, the attacker isolates the

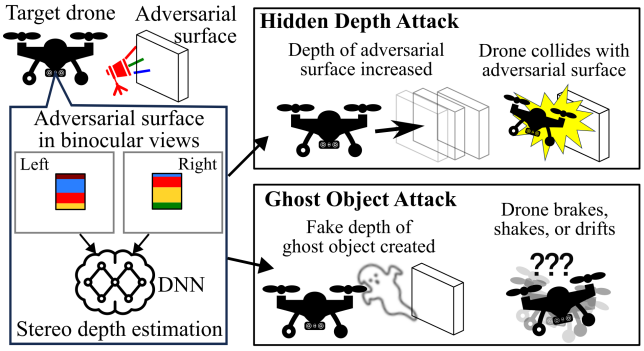


Figure 3: RiD attack.

LED's contribution. The attacker can then optimizes $f(t)$ by minimizing a loss function to spoof the classfier.

2.2 Related Work

Sensor attack against drones. Acoustic attacks inject resonant noise into gyroscopes [24, 27], but are limited by short effective ranges. Optical attacks manipulate visual inputs, such as projecting moving light patterns onto the ground [4], blinding vision sensors with lasers [8], or injecting deceptive light patterns into stereo cameras [32]. These methods often require precise, continuous aiming of the drone. The work uses physical adversarial patches to mislead onboard perception [12]. In contrast, *RiD* can be launched remotely without continuous aiming, and its adversarial patterns remain invisible to human eyes.

RSE exploitation for attacks. The work [21] manipulateds ambient illumination using flickering LEDs to mislead image classification. The works [11, 30] aim lasers into camera lens to create monochromatic stripes, interfering with object detection and traffic light color recognition tasks, respectively. A recent work [10] adjusts the timing of LED flickering to produce stable colored stripes on traffic signs, consistentlyu spoofing traffic sign recognition in autonomous vehicles. The above works only focus on single-camera tasks. In contrast, *RiD* investigates RSE-based attacks against binocular depth estimation in stereo vision.

Attacks against stereo depth estimation. Previous works like [2, 26, 28] have digitally injected pixel-level perturbations into stereo images to disrupt stereo depth estimation. However, physically deployable adversarial attacks on stereo vision, remain less explored. The work [32] directs deceptive light beams into stereo cameras to spoof the stereo depth estimation. A recent study [13] introduces printable adversarial patches against stereo matching. Differently, *RiD* exploits camera sensor's RSE to create invisible adversarial attack without continuous aiming.

3 *RiD* Attack

3.1 Threat Model

Attack form. Fig. 3 overviews *RiD* attack. The attacker can place an adversarial surface in the drone's operational environment to reflect or emit flickering illumination. For example, a reflective cover may be attached to surfaces to reflect light from an external LED source, or a flickering LED screen can be directly used to emit the adversarial light.

Attack objectives. *RiD* aims to disrupt drone operations by misleading binocular depth estimation, by either increasing or decreasing the perceived depth of the adversarial surface, which can lead to operational disruptions or even crashes. For example, (1) **Hidden Depth Attack (HDA)**: an increased depth estimation may cause a drone to unintentionally collide with the adversarial surface, potentially damaging or trapping it. (2) **Ghost Object Attack (GOA)**: a reduced depth estimate could create a perceived "ghost" object close to the drone, preventing access to certain critical areas or even causing instability, such as sudden braking, shaking or drifting during high-speed flights.

Attacker's knowledge. We consider two attacker knowledge scenarios: (1) Black-box attack: Without access to camera parameters or the stereo matching DNN model, the attacker can generate adversarial stripes using random flickering signals. (2) White-box attack: With knowledge of the camera and stereo matching DNN parameters, the attacker can apply gradient-based optimization to tailor the flickering signal for stronger attacks. Such white-box access may be obtained through open-source codebases, reverse engineering of commercial products, or social engineering of manufacturer personnel.

3.2 Analysis of RSE in Stereo Vision

In this section, we analyze how the RSE manifests in binocular stereo vision systems. We consider three representative cases:

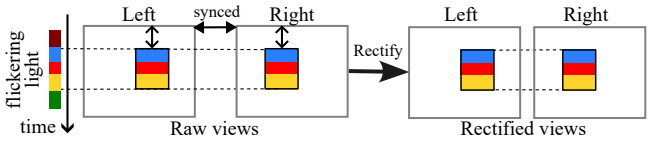


Figure 4: Case-①: Perfect sensor alignment and synchronization.

■ **Case-①: Perfect sensor alignment and synchronization.** In this case, the left and right cameras are identical in specifications and physically well-aligned (coplanar and horizontally aligned), and their exposures are tightly synchronized. This alignment ensures that the adversarial surface, appears at the same vertical position and with the same size in both views before rectification. Since both camera sensors expose corresponding rows at the same time, a flickering light will affect the same row index in both images simultaneously. As a result, the adversarial surface will exhibit identical colored stripe patterns in both views both before and after rectification, as illustrated in Fig. 4, preserving consistency in appearance across the stereo pair.

■ **Case-②: Perfect sensor synchronization with sensor misalignment.** This is the most realistic case. As discussed in §2.1, stereo cameras are often physically misaligned to some extent. For example, Fig. 5 shows simulated raw views obtained by reversing rectification using the calibration parameters of a KITTI Stereo 2015 dataset sample [17]. Objects (e.g., the wheel hub and wayfinding sign) in the two views appear at slightly different vertical positions.

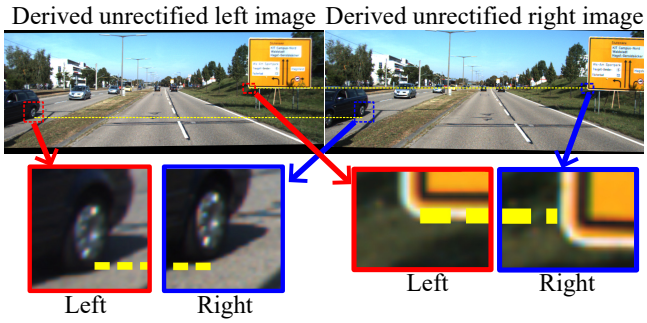


Figure 5: Example of derived unrectified views from KITTI dataset. Misalignments in raw images are common in stereo vision systems.

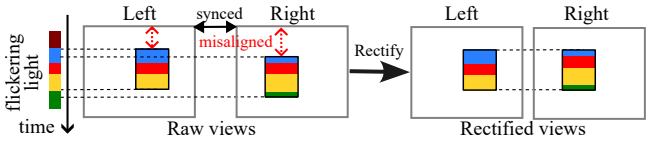


Figure 6: Case-②: Perfect sensor synchronization with sensor misalignment.

In this case, although the sensors are synchronized, physical misalignment causes them to capture different vertical portions of the scene at the same time. As a result, the adversarial surface may appear with variations in vertical position (e.g., due to vertical offset or tilt), apparent size (e.g., due to different target distances), or even shape (e.g., due to angular or perspective distortion). as exemplified in Fig. 6, flickering light then affects different scanlines of the adversarial surface in the two views, producing shifted RSE patterns. After rectification, the object becomes horizontally aligned, but the RSE patterns remain inconsistent across views.

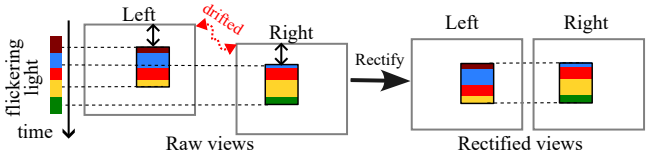


Figure 7: Case-③: Perfect sensor alignment with drifted synchronization.

■ **Case-③: Perfect sensor alignment with drifted synchronization.** While accurate stereo vision typically requires precise sensor synchronization [7], we consider a supplemental case where the cameras are physically aligned but loosely synchronized. In this scenario, the left and right sensors expose corresponding rows at slightly different times due to temporal drift. As a result, although the adversarial surface appears at the same vertical position in both raw views, flickering illumination affects the rows differently. This mismatch produces inconsistent RSE-induced color stripe patterns across the stereo pair after rectification, as illustrated in Fig. 7.

3.3 RSE-based Attack against Stereo Vision

In this section, we illustrate how *RiD* operates. Depending on the attacker's knowledge and capabilities, *RiD* can be performed in either a black-box or white-box setting.

3.3.1 Chances of Black-box attacks. In the black-box setting, *RiD* can exploit inherent vulnerabilities in stereo DNNs by introducing unnatural rolling shutter artifacts and leveraging spatial or temporal sensor misalignment.

HDA via unnatural stripy perturbations. Fig. 8 illustrates an example where random RSE-induced stripe patterns increase the perceived depth of the adversarial surface. This effect occurs even without optimization, likely because such patterns are absent from the model's training data, leading to mismatches in stereo correspondence. In our simulation, depth can be overestimated by up to tens of meters across a broad area, making parts of the surface appear as distant background. The drone may interpret these regions as passable gaps and attempt to fly through them, risking collision of the drone.

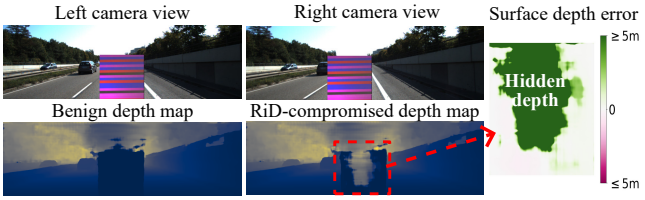
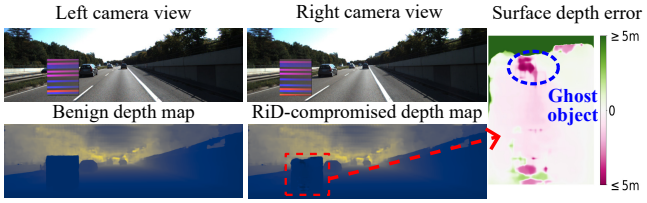


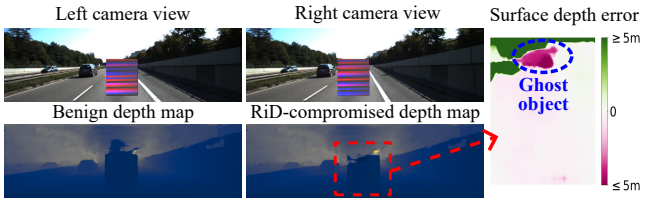
Figure 8: HDA from RSE artifacts (Case-①). Adversarial surface partially blended into the background.

GOA via misalignment. As stereo DNNs are typically trained on rectified data with aligned cues, misalignment between stereo views can cause GOA even without optimization. In Case-②, as shown in Fig.9a, slight spatial misalignment causes the same flicker signal to result in slightly mismatched stripe patterns on the adversarial surface. This disrupts visual consistency and causes the model to underestimate depth in parts of the adversarial surface. In Case-③, as shown in Fig.9b, GOA becomes even more easier to achieve. Synchronization drift leads to inherent misalignment in exposure timing, resulting in inter-view mismatches across the entire field of view, regardless of the adversarial surface location. In these two examples, the compromised depth can drop from 7 m to around 2 m which is the DNN's minimum detectable depth, which may trigger the false detection of near-field objects. Since drones typically react to the nearest perceived obstacle, such hallucinations can induce instability, such as drifting and shaking, and disrupting navigation and perception [32].

3.3.2 Enhanced attacks with white-box knowledge. We follow the RSE model described in §2.1 to define a differentiable flickering signal $f(t)$. With access to the stereo DNN and camera parameters (white-box setting), the attacker can further strengthen *RiD* by optimizing the flicker signal $f(t)$ through Projected Gradient Descent (PGD) [15], by solving $\text{argmin}_{f(t)}\, \ell(\mathcal{S}(I_L, I_R))$, where $\mathcal{S}$ is the stereo DNN that processes the stereo pair (I_L, I_R), and $\ell(.)$ is the loss function defined over the adversarial surface region $\mathcal{P}$.



(a) Case-②: Minor sensor misalignment induces GOA.



(b) Case-③: Drifted synchronization induces GOA.

Figure 9: GOA from inter-view mismatching.

Enhanced HDA. To enhance HDA, the objective is to reduce the predicted disparity values within $\mathcal{P}$, enlarging the overall depth of the surface. The loss is defined as:

$$\ell_{\text{HDA}} = \frac{1}{N} \sum_{(u,v)\in\mathcal{P}} \text{ReLU}\left(d(u,v) - (g_{\mathcal{P}} - m)\right)$$

where $d(u,v)$ is the predicted disparity at pixel (u,v), $g_{\mathcal{P}}$ is the ground-truth disparity for the surface (derived from the ground-truth depth of the patch), and m is a predefined margin. This encourages underestimation of disparity, which increases the perceived depth of the adversarial surface.

Enhanced GOA. For GOA, the goal is to increase disparity within the patch, reducing the perceived depth to simulate a phantom obstacle. The corresponding loss function is:

$$\ell_{\text{GOA}} = \text{ReLU}\left(D - \max_{(u,v)\in\mathcal{P}} d(u,v)\right)$$

where D is the upper limit of the DNN's output disparity. This encourages the DNN to predict high disparity values, making the object appear closer than it is.

4 Evaluation

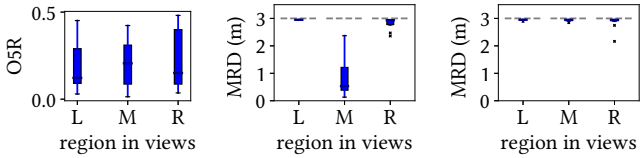
4.1 Evaluation Setup

Dataset and model. We evaluate the *RiD* attack using the KITTI Stereo 2015 dataset [17]. The stereo baseline is 0.54 m, and the camera height is approximately 1.65 m, which can simulate a camera setup of low-flying drones. Rectified image resolution is 1242×375, while raw images are captured at 1384×512. We adopt the widely used PSMNet [3] as the victim stereo matching model.

***RiD* implementation.** We model the attack surface as a rectangular object (size $1.6 \times 1.8\,\text{m}^2$) on the ground. Using the pinhole projection model and stereo disparity constraints, we project the object into both views to simulate its appearance. Following the RSE formulation in §2.1, we synthesize *RiD* attacks using real-world captured reference images I_{full} and I_{amb} of a white surface under full and ambient illumination, with flicker signals $f(t)$ to generate realistic RSE-induced perturbations. We simulate the three camera setup cases introduced in §3.2: (1) **Case-①**: Rectified images are scaled back to raw image sizes to simulate perfectly aligned raw views. The synchronized attack is synthesized in the raw images, which are then resized back to rectified dimensions for inference. (2) **Case-②**: Rectification is reversed using the dataset's calibration parameters to obtain naturally deviated raw views. The synchronized attack is synthesized in the raw domain, and the images are then re-rectified for inference. (3) **Case-③**: Based on Case-①, random temporal offsets are introduced between the stripe patterns in the left and right views to simulate synchronization drift. With these setups, the synthesized stereo pairs are passed to the victim model to generate disparity maps, which are then converted to depth. We evaluate attack effectiveness by analyzing depth estimation within the attack surface region.

Metrics. (1) **Over-5 Error Rate (O5R)**: for HDA, we measure the proportion of pixels on the adversarial surface where the estimated depth exceeds the ground truth by more than 5 m; (2) **Maximum Reduced Depth (MRD)**: for GOA, we define MRD as the largest depth underestimation observed on the adversarial surface.

4.2 Preliminary Results

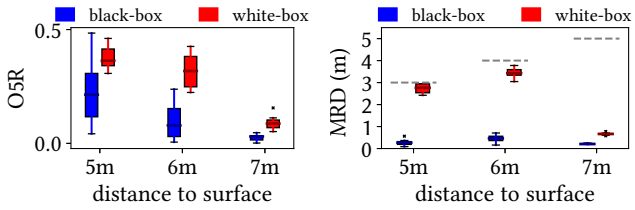


(a) Case-① HDA. **(b) Case-② GOA.** **(c) Case-③ GOA.**

Figure 10: Effectiveness of black-box attacks.

4.2.1 Effectiveness of black-box attacks. Fig.10 presents the performance of black-box *RiD*. The adversarial surface is placed at a distance of 5 m from the camera, positioned in the left (L), middle (M), or right (R) region of the image. In Case-①, black-box HDA achieves considerable O5R across all view regions, as shown in Fig. 10a. For Case-② and Case-③, we focus on GOA. As shown in Fig.10b, Case-② achieves high MRD in the left and right regions, approaching the upper bound (indicated by the grey line at the model's minimum perceivable depth of 2 m), while being less effective in the center. This is due to stronger sensor misalignment in the peripheral areas of the KITTI dataset, as discussed in §3.2. In Case-③, shown in Fig. 10c, MRD remains consistently high across all regions due to pervasive misalignment from temporal synchronization drift. These results provide preliminary evidence of the feasibility of black-box *RiD*.

4.2.2 White-box enhancement. Using the Case-① setting, we evaluate how white-box knowledge improves the effectiveness of *RiD*. Fig. 11 compares O5R and MRD under black-box and white-box settings across different surface distances. Fig. 11a shows that white-box optimization significantly enhances HDA effectiveness. Furthermore, as shown in Fig. 11b, while GOA is nearly infeasible under black-box conditions in Case-①, it becomes effective in the



(a) Case-① HDA. **(b) Case-① GOA.**

Figure 11: Enhancement by white-box optimization.

white-box setting, with MRD approaching the upper bound at 5 m and 6 m distances.

5 Conclusion and Future Work

We proposed *RiD*, a novel physical attack against stereo depth estimation on drones, exploiting the interaction between RSE and stereo matching process. *RiD* can increase or decrease the depth estimation of the adversarial surface, which may lead to operational disruptions or even crashes of the target drone. Our proof-of-concept simulations shows the feasibility of such attack under realistic stereo camera configurations.

Several directions remain open for further exploration. (1) We plan to improve the design of *RiD* to generate more robust and consistent attacks that can maintain effectiveness across diverse scenes and persist over time, under different environmental and motion conditions. (2) We plan to systematically evaluate how different factors, such as surface properties, camera configurations, environmental lighting, and stereo DNN architectures, affect the vulnerability and impact of *RiD*. (3) To better understand real-world implications, we plan to simulate the downstream effects of compromised depth perception on drone navigation using a full-featured drone simulator. Additionally, we intend to implement and test *RiD* on commercial off-the-shelf stereo cameras and drones, assessing its practicality, stability, and threat potential in physical deployment scenarios. (4) Finally, we plan to explore defense strategies to detect and counter RSE-based attacks in stereo vision systems, with the goal of improving the robustness and trustworthiness of depth perception in safety-critical applications.

Acknowledgments

This research/project is supported by the National Research Foundation Singapore and DSO National Laboratories under the AI Singapore Programme (AISG Award No: AISG2-GC-2023-006).

References

[1] Autel. 2025. *Autel Evo II series.* https://shop.autelrobotics.com/pages/evo-ii-collections
[2] Zachary Berger et al. 2022. Stereoscopic universal perturbations across different architectures and datasets. In *IEEE/CVF CVPR*. 15180–15190.
[3] Jia-Ren Chang et al. 2018. Pyramid stereo matching network. In *IEEE/CVF CVPR*. 5410–5418.
[4] Drew Davidson et al. 2016. Controlling UAVs with sensor input spoofing attacks. In *USENIX WOOT 16*.
[5] DJI. 2025. *DJI Mavic series.* https://www.dji.com/sg/search?q=mavic
[6] DJI. 2025. *DJI Phantom 4 Pro V2.0.* https://www.dji.com/sg/support/product/phantom-4-pro-v2
[7] FRAMOS. 2024. *Multi-Sensor Synchronization: SONY Rolling Shutter Sensors APPLICATION NOTE.* https://www.framos.com/framos-fsm-startup/downloads/AppNotes/FRAMOS-AppNote_MultiSensor-Synchronization.pdf
[8] Zhangjie Fu et al. 2021. Remote attacks on drones vision sensors: An empirical study. *IEEE TDSC* 19, 5 (2021), 3125–3135.
[9] Andreas Geiger et al. 2012. Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. In *CVPR*.
[10] Dongfang Guo et al. 2024. Invisible Optical Adversarial Stripes on Traffic Sign against Autonomous Vehicles. In *ACM MobiSys*. 534–546.
[11] Sebastian Köhler et al. 2021. They see me rollin': Inherent vulnerability of the rolling shutter in cmos image sensors. In *ACSAC*. 399–413.
[12] Taifeng Liu et al. 2023. RPAU: Fooling the eyes of UAVs via physical adversarial patches. *IEEE TITS* 25, 3 (2023), 2586–2598.
[13] Yang Liu et al. 2024. Physical Attack for Stereo Matching. In *ACM CVDL*. 1–5.
[14] Luxonis. 2025. *OpenCV AI Kit (OAK-D).* https://shop.luxonis.com/products/oak-d
[15] Aleksander Madry et al. 2018. Towards deep learning models resistant to adversarial attacks. In *ICLR*.
[16] Nikolaus Mayer et al. 2016. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In *IEEE CVPR*. 4040–4048.
[17] Moritz Menze et al. 2015. Joint 3D Estimation of Vehicles and Scene Flow. In *ISPRS ISA*.
[18] Daniel Oram. 2001. Rectification for any epipolar geometry.. In *BMVC*, Vol. 1. 653–662.
[19] Parrot. 2025. *Parrot Anfi AI.* https://www.parrot.com/en/drones/anafi-ai
[20] Intel RealSense. 2025. *Depth Camera D415.* https://www.intelrealsense.com/depth-camera-d415/
[21] Athena Sayles et al. 2021. Invisible perturbations: Physical adversarial examples exploiting the rolling shutter effect. In *IEEE/CVF CVPR*. 14666–14675.
[22] Skydio. 2025. *Skydio 2/2+.* https://support.skydio.com/hc/en-us/categories/360002358774-Skydio-2-2
[23] Skydio. 2025. *Skydio X2 Enterprise.* https://support.skydio.com/hc/en-us/categories/1260801636130-Skydio-X2-Enterprise
[24] Yunmok Son et al. 2015. Rocking drones with intentional sound noise on gyroscopic sensors. In *USENIX Security*. 881–896.
[25] StereoLabs. 2025. *ZED 2.* https://www.stereolabs.com/products/zed-2
[26] Pengfei Wang et al. 2024. Left-right Discrepancy for Adversarial Attack on Stereo Networks. *arXiv preprint arXiv:2401.07188* (2024).
[27] Zhengbo Wang et al. 2017. Sonic gun to smart devices: Your devices lose control under ultrasound/sound. *Black Hat USA* (2017), 1–50.
[28] Alex Wong et al. 2021. Stereopagnosia: Fooling stereo networks with adversarial perturbations. In *AAAI*, Vol. 35. 2879–2888.
[29] Haofei Xu et al. 2020. Aanet: Adaptive aggregation network for efficient stereo matching. In *IEEE/CVF CVPR*. 1959–1968.
[30] Chen Yan et al. 2022. Rolling colors: Adversarial laser exploits against traffic light recognition. In *USENIX Security*. 1957–1974.
[31] Guorun Yang et al. 2019. DrivingStereo: A Large-Scale Dataset for Stereo Matching in Autonomous Driving Scenarios. In *IEEE CVPR*.
[32] Ce Zhou et al. 2022. DoubleStar:Long-Range attack towards depth estimation based obstacle avoidance in autonomous systems. In *USENIX Security 22*. 1885–1902.
[33] Lukas Zmejevskis. 2024. *DJI Mavic 3 has no mechanical shutter. Sensor readout speed explained.* https://www.pix-pro.com/blog/dji-mavic-3-rolling-shutter