Demo: Controlling Depth Estimation in Stereo Cameras with Projected Patterns

Tetsu Ishizue[*2], *Nicholas U. Costagliola*[*1], *Sri Hrushikesh Varma Bhupathiraju*[1],
Ozora Sako[3], *Kentaro Yoshioka*, *Takeshi Sugawara*[2], *Sara Rampazzi*[1]

[1]*University of Florida*, [2]*The University of Electro-Communications*, [3]*Keio University*

Abstract

Stereo cameras are integrated into autonomous systems such as self-driving cars and drones to provide precise depth estimation more cost-effectively than LiDAR. In this demo, we reveal an intrinsic vulnerability in stereo cameras' image-pixel sampling that influences the outcomes of stereo matching algorithms. Attackers can exploit this vulnerability by using simple repeated patterns to achieve fine-grained control over the estimated depth of genuine obstacles. We determine the cause of the vulnerability and demonstrate precise manipulation of the estimated depth on the RealSense D435, a popular commercial stereo camera with a proprietary depth-matching algorithm. Our work highlights intrinsic vulnerabilities in stereo-based depth estimation algorithms and addresses critical limitations in autonomous system deployments.

1 Introduction

Stereo vision has emerged as a passive, lightweight, and cost-effective depth-sensing alternative to LiDARs in Autonomous Vehicles. These sensors mimic human depth perception using dual viewpoints and triangulation, delivering dense 3D data without sophisticated and expensive lasers or moving parts. At the heart of these systems lie Stereo Depth Estimation (SDE) algorithms, which fuse image data into depth maps of the surrounding environment.

Prior work has shown that deep-learning-based SDE algorithms are vulnerable to adversarial patches that induce uncontrolled depth-estimation errors [1, 2]. These works rely on complex optimization techniques that typically require scene- or image-specific optimization, limiting their applicability in practical scenarios. Our study instead uncovers a systematic vulnerability arising from the intrinsic image-pixel sampling process in these cameras without the need for adversarial examples or image-specific optimization techniques.

In this demo, we show how attackers can leverage this vulnerability to attack a RealSense D435, a popular commercial stereo camera equipped with a proprietary block-matching-based SDE. Specifically, the attack can systematically control the estimated depth, making legitimate obstacles appear far or close to the victim stereo camera. The full research work has been accepted at ACM CCS 2026.

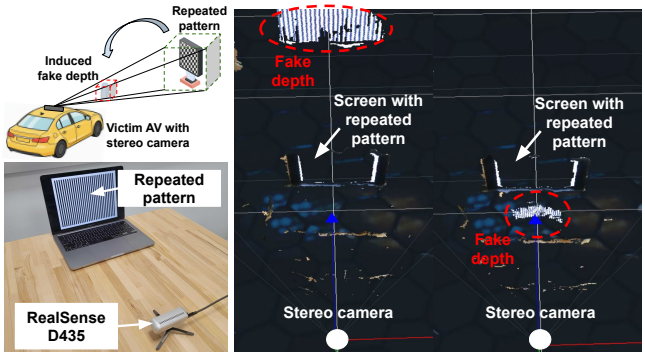


Figure 1: Overview of our attack and demo scenarios with a structured pattern displayed on a monitor (left). Induced controlled shifts in depth estimation, farther and closer to the victim stereo camera (right).

2 Vulnerability Overview

SDE algorithms estimate 3D depth by capturing two images from a stereo camera pair. They employ block matching, which computes disparity by comparing local pixel similarity within a fixed window. Regions with high similarity are matched, and depth is derived from the pixel offset between these matched regions in the two images. A larger pixel offset corresponds to a closer object, whereas a smaller offset indicates a distant one. The vulnerability arises from *sampling artifacts* that form during camera image processing. Sampling artifacts are pixel-level distortions introduced when continuous visual data is discretized into pixels on an image sensor. These artifacts inherently alter pixel intensities and the local features captured by the individual cameras in a stereo setup.

[*] denotes co-first authors.

In scenes containing repeated high-contrast boundaries, such as a black-and-white striped pattern, the resulting sampling artifacts vary in both pixel intensity and spatial position. This phenomenon is illustrated in Figure 2. These variations influence the pixel similarity scores between image regions and cause the SDE algorithm to perceive higher pixel similarity between spatially distinct regions in the pattern. This, in turn, results in a depth estimation error, depending on the pixel intensity of the sampling artifact. An attacker can adjust the pixel intensity of the sampling artifacts and the corresponding depth errors by systematically controlling the pattern.

3 Attack Goal and Threat Model

Attacker Goal. We consider stereo cameras deployed in autonomous systems. Unlike previous work, which focuses only on maximizing depth error for a given image pair and deep learning model, here the goal of the attacker is to mislead the victim stereo-based perception by consistently controlling the estimated depth of target object surfaces in the scene.

Underlying Principle and Scenario. Instead of generating physically optimized adversarial patches, the adversary can use a commercial projector or a monitor (e.g., roadside billboards) to display simple repeating shapes, such as stripes. These patterns can be projected onto target obstacle surfaces or roadside locations, including buildings, traffic signs, and the backs of vehicles. Displaying them for short durations (e.g., 0.5 sec) is sufficient to trigger the attacker's intended outcome. This manipulation causes obstacles to be perceived closer to or farther from their actual distance, causing downstream failures in autonomous decision-making. For example, self-driving cars may fail to brake and collide due to overestimating the distance to an obstacle, whereas drones can trigger abrupt maneuvers by underestimating the distance.

Assumptions. We assume that the adversary has knowledge of the stereo camera used by the victim autonomous system. This information can be inferred from publicly available camera manuals and data sheets. Alternatively, the attacker can perform a black-box study by acquiring a stereo camera similar to the one used in the victim vehicle. The attack is remote, not requiring any firmware or hardware access to the victim vehicle or stereo camera. We do assume an oracle access to the depth estimation algorithm employed by the victim stereo camera. Finally, we focus on simple and low-effort stripe pattern that an unskilled attacker can recreate without any adversarial machine learning technique.

4 Experiments

We verify the attack success on a commercial stereo camera widely used in autonomous systems (RealSense D435). Rather than inducing random depth errors, the attacker can systematically manipulate the depth of obstacles up to 2 m farther or 5 m closer to the victim's stereo camera. Our demo setup with the resulting manipulation in the estimated depth is illustrated in Figure 1. Evaluation with the SGM-DDC stereo-LiDAR fusion model reveals that the attack induces a fake depth error of up to 43 m, despite the use of a LiDAR sensor. In realistic driving scenarios at vehicle speeds up to 15 km/h, the attack persists for at least 0.5 sec, which is sufficient to trigger emergency braking in state-of-the-art autonomous driving frameworks, such as Autoware.

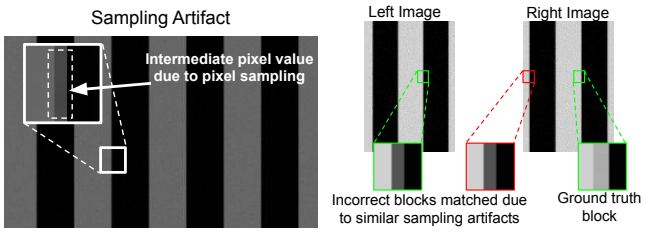


Figure 2: Sampling artifacts produce intermediate gray pixel intensity values at the stripe boundaries due to averaging across the pixel grid (left). Illustration of the incorrect block matching triggered by sampling artifacts (right).

5 Demonstration Plan

Demonstration of physical vulnerability. We will show the controlled sampling artifacts resulting from structured pattern display on a monitor.

Controlled manipulation of depth estimation. We will present live experimentation on the RealSense D435, demonstrating controlled manipulation of the estimated depth.

Demonstration of the impact in driving scenarios. We will show the videos of the attack performed in realistic driving scenarios with speeds up to 15 km/h and a projected pattern.

6 Acknowledgment

Generative AI tools were used for editorial purposes. All outputs were inspected to ensure accuracy and originality. We disclosed our findings to the vendor. This research was supported in part by the JST CREST JPMJCR23M4, JST Next-generation Edge AI Semiconductor JPM-JES2515, and JSPS KAKENHI 24K02940 and 24K14943.

References

[1] Hangcheng Liu, Xu Kuang, Xingshuo Han, Xingwan Wu, Haoran Ou, Shangwei Guo, Xingyi Huang, Tao Xiang, and Tianwei Zhang. Optimization-free patch attack on stereo depth estimation. *arXiv preprint arXiv:2506.17632*, 2025.

[2] Yang Liu, Jucai Zhai, Chihao Ma, Pengcheng Zeng, Xinan Wang, and Yong Zhao. Physical attack for stereo matching. In *Proceedings of the International Conference on Computer Vision and Deep Learning*, pages 1–5, 2024.