Poster: Target Switch Attacks on Gimbal-Stabilized Visual Tracking Systems via Acoustic Injection

Jiarui Li, Joseph Brewington, Qingzhao Zhang, Z. Morley Mao
University of Michigan
{*jiaruili, brewing, qzzhang, zmao*}*@umich.edu*

Abstract*—Gimbal-stabilized visual tracking is critical for modern autonomous systems such as Unmanned Aerial Vehicles (UAVs). While prior work shows acoustic signals can disturb gimbal internals, the impact of such attacks on real-world applications like UAV tracking and following remains underexplored. Also, existing demonstrations are largely laboratory-bound and overlook practical challenges for real-world attacks, such as object-motion uncertainty and runtime latency. To bridge this gap, we present `Banshee`, the first physically realizable attack that induces target switching in UAV visual tracking systems by exploiting acoustic vulnerabilities in gimbal-camera systems. `Banshee` generates carefully crafted acoustic waveforms that induce optimized adversarial gimbal oscillations, causing directionally biased camera-view drifts that break inter-frame target associations. This forces the onboard tracker to switch from the legitimate target to an attacker-chosen decoy, enabling tracking target switch or target loss. `Banshee` achieves a 98.3% success rate in simulation across two commercial gimbal systems and five trackers. Real-world benchtop and in-flight black-box attacks against a commercial drone across varied scenarios show an overall 95.5% attack success rate. Our results reveal a practical cross-domain vulnerability between acoustics and vision, highlighting the need for robust designs of gimbal systems and applications.

Abstract

Gimbal-stabilized visual tracking is a core capability of modern camera systems, enabling persistent, high-precision object following in dynamic scenes. As a prominent example, commercial Unmanned Aerial Vehicles (UAVs) widely deploy target following, which typically pairs a multi-axis gimbal with an onboard camera, combined with object tracking algorithms running in software, to enable active tracking and following on a selected mobile target [1], [2], [3], [4], [5], [6]. Gimbal-stabilized visual tracking enables applications such as autonomous filming, surveillance, and infrastructure inspection, but also creates a single point of failure: compromising this pipeline can lead to severe consequences, including flight hazards, loss of vehicle control, and tracking of false targets [7], [8], [9], [10], [11].

Gimbal systems commonly rely on real-time inertial measurement unit (IMU) feedback to mechanically stabilize onboard cameras during rapid motion [12], [13]. Prior research has shown that carefully crafted acoustic patterns, delivered via speakers, ultrasonic transducers, or even laser systems, can manipulate IMU readings and, in turn, disrupt gimbal stabilization [14], [15], [16], [17], [18]. While these studies establish the feasibility of influencing gimbal motion at the sensor level, they remain disconnected from real-world applications such as UAV target tracking and following, leaving their practical impact uncertain.

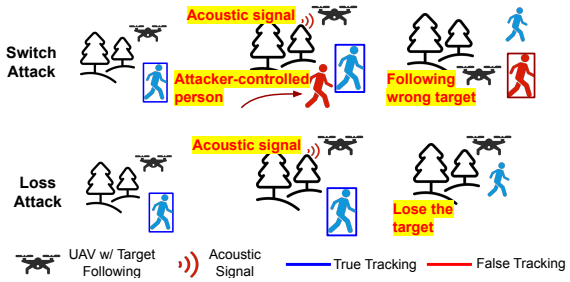


Figure 1: Illustration of `Banshee` on UAV target following. Crafted acoustic signals forces the UAV's visual tracking system into tracking a wrong target or lose track.

To address this gap, we propose `Banshee`, the first physical target switch attack against UAV visual tracking systems via acoustic injection, linking a hardware acoustic vulnerability with application level tracking weaknesses to enable end-to-end system compromise. Figure 1 illustrates two scenarios. (1) In the *Switch* attack, a UAV performing target following switches from its original target to an attacker-desired object, enabling the attacker to potentially steal the UAV from its owner or transfer tracking onto an unintended target. (2) In the *Loss* attack, the acoustic signal causes the UAV to lose its tracking target, e.g., allowing a suspect under surveillance to escape monitoring.

The attack has two stages. In *offline gimbal profiling*, the attacker uses an identical gimbal device to systematically characterize its acoustic response, deriving a gimbal acoustic response model that maps acoustic signals to gimbal motion. In the *online attack*, the attacker runs two loops simultanously. A *surrogate tracking* loop that runs a surrogate of the UAV onboard tracking mimicking the actual UAV tracking behavior using black-box knowledge. A planning-execution loop then optimizes a sequence of acoustic signals under physical and algorithmic constraints, leveraging

both the gimbal acoustic response model and the tracking surrogate. These crafted signals are then injected through a speaker or piezo transducers to redirect tracking away from its legitimate target.

Designing this application-aware acoustic attack raises several key challenges. First, the attack must achieve an empirically sufficient alignment between physical acoustic signals and their induced camera motion through the complex gimbal system to produce adversarial motion that disrupts tracking. Second, the attack must operate at runtime without prior knowledge of the UAV behavior, which motivates an adaptive online strategy that updates the surrogate tracker and signal injection plan as the scene evolves. Third, the attack must remain effective under real world uncertainties, including unknown future object motion, which we address with optimization algorithms that tolerate uncertainties.

Extensive experiments prove the practicality of the proposed attack. First, the offline profiling on built-in gimbals of two commercial UAV models shows that the precise run-time gimbal control is feasible. Second, we run large-scale simulation in Gazebo simulator, which deploys PX4-Autopilot flight stack, uses the profiled gimbal parameters, and tested the attack on five representative trackers and diverse scenarios. The results show that `Banshee` overall corrupts the tracking in 96.1% trials (including 80.6% Switch and 15.5% Loss), proving attack effectiveness and robustness. Finally, real-world experiments further validate successful black-box Switch on a commercial HighEnd-Drone, including a realistic exploit on the built-in object tracking of during flight.[1] We summarize our contributions below:

- We design `Banshee`, the first attack that uses acoustic injection to compromise UAV visual tracking. Our algorithm connects gimbal and tracking vulnerabilities, performs online optimization to achieve runtime attacks, and adapts to diverse real-world circumstances.
- We propose the first systematic method to achieve precise adversarial gimbal control via acoustic injection. Using an offline profiling technique and an online phase modulation routine, the adversary can control the gimbal rotation axis, direction, and angular motion in real time.
- We extensively evaluate `Banshee` in high-fidelity simulation (Gazebo + PX4-Autopilot), with realistic simulation of gimbal vulnerability. We also successfully perform real-world online attacks on a commercial HighEnd-Drone.

References

[1] Z. Han, R. Zhang, N. Pan, C. Xu, and F. Gao, "Fast-tracker: A robust aerial system for tracking agile target in cluttered environments," in *2021 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2021, pp. 328–334.

1. We anonymized the commercial product models. We have completed the disclosure responsibility to the vendor at the time of submission and will safely disclose the vulnerabilities upon official publication.

[2] H. Cheng, L. Lin, Z. Zheng, Y. Guan, and Z. Liu, "An autonomous vision-based target tracking system for rotorcraft unmanned aerial vehicles," in *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2017, pp. 1732–1738.

[3] A. Maalouf, N. Jadhav, K. M. Jatavallabhula, M. Chahine, D. M. Vogt, R. J. Wood, A. Torralba, and D. Rus, "Follow anything: Open-set detection, tracking, and following in real-time," *IEEE Robotics and Automation Letters*, vol. 9, no. 4, pp. 3283–3290, 2024.

[4] DJI, "Best drones that follow you automatically (2024)," https://store.dji.com/content/camera-drone-that-follows-you, 2024, accessed: 2025-08-08.

[5] Skydio, "The best follow me drone in 2022," https://www.skydio.com/blog/10-reasons-skydio-makes-the-best-follow-me-drone, 2022, accessed: 2025-02-25.

[6] Autel, "Autel evo ii drone dynamic track mode full review," https://www.autelpilot.com/blogs/buying-guides/autel-evo-ii-drone-dynamic-tracking-mode, 2020, accessed: 2025-02-25.

[7] H. B. Salameh, M. Alhafnawi, A. Masadeh, and Y. Jararweh, "Federated reinforcement learning approach for detecting uncertain deceptive target using autonomous dual uav system," *Information Processing & Management*, vol. 60, no. 2, p. 103149, 2023.

[8] J. Li, J. Brewington, Q. Zhang, and Z. M. Mao, "Wip: Hijacking attacks on uav follow-me systems in realistic scenarios."

[9] J. Hibberd, "Using drones for peeping, burglaries on rise: "it's gotten dramatically worse"," https://www.hollywoodreporter.com/lifestyle/lifestyle-news/drones-spying-robberies-solutions-hollywood-1236166714/, 2025, accessed: 2025-08-13.

[10] P. Dolan, "Like moths to a false flame: Lethality and protection through deception operations," https://www.army.mil/article/286861/like_moths_to_a_false_flame_lethality_and_protection_through_deception_operations, 2025, accessed: 2025-08-13.

[11] D. Davidson, H. Wu, R. Jellinek, V. Singh, and T. Ristenpart, "Controlling {UAVs} with sensor input spoofing attacks," in *10th USENIX workshop on offensive technologies (WOOT 16)*, 2016.

[12] D. Bereska, K. Daniec, S. Fraś, K. Jedrasiak, M. Malinowski, and A. Nawrat, "System for multi-axial mechanical stabilization of digital camera," in *Vision Based Systems for UAV Applications*. Springer, 2013, pp. 177–189.

[13] A. Altan and R. Hacıoğlu, "Model predictive control of three-axis gimbal system mounted on uav for real-time target tracking under external disturbances," *Mechanical Systems and Signal Processing*, vol. 138, p. 106548, 2020.

[14] T. Trippel, O. Weisse, W. Xu, P. Honeyman, and K. Fu, "Walnut: Waging doubt on the integrity of mems accelerometers with acoustic injection attacks," in *2017 IEEE European symposium on security and privacy (EuroS&P)*. IEEE, 2017, pp. 3–18.

[15] Y. Tu, Z. Lin, I. Lee, and X. Hei, "Injected and delivered: Fabricating implicit control over actuation systems by spoofing inertial sensors," in *27th USENIX Security Symposium (USENIX Security 18)*. Baltimore, MD: USENIX Association, Aug. 2018, pp. 1545–1562. [Online]. Available: https://www.usenix.org/conference/usenixsecurity18/presentation/tu

[16] 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 *24th USENIX security symposium (USENIX Security 15)*, 2015, pp. 881–896.

[17] M. Gao, L. Zhang, L. Shen, X. Zou, J. Han, F. Lin, and K. Ren, "Kite: Exploring the practical threat from acoustic transduction attacks on inertial sensors," in *Proceedings of the 20th ACM conference on embedded networked sensor systems*, 2022, pp. 696–709.

[18] N. Shamsi, K. Chandrasekar, Y. Long, C. Limbach, K. Rebello, and K. Fu, "Wip: Threat modeling laser-induced acoustic interference in computer vision-assisted vehicles."

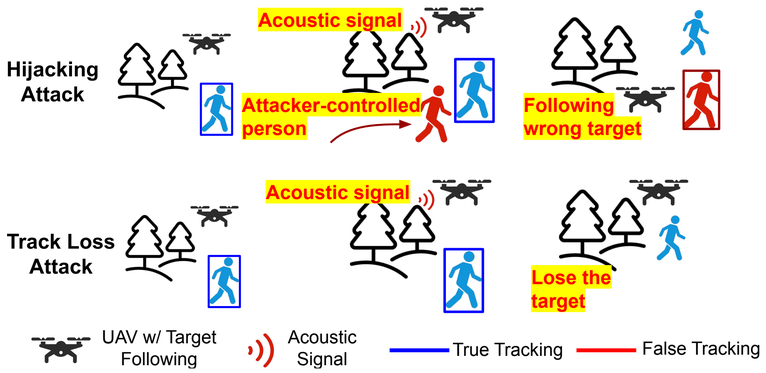
Target Switch Attacks on Gimbal-Stabilized Visual Tracking Systems via Acoustic Injection

Jiarui Li, Joseph Brewington, Qingzhao Zhang, Z. Morley Mao

Computer Science and Engineering, University of Michigan

Overview

Modern UAVs rely on gimbal-stabilized visual tracking to follow targets for applications like surveillance, filming, and navigation. These systems assume that camera motion is smooth and accurately reflects real-world movement. However, we show this assumption can be exploited using sound through **Banshee**, a practical attack that injects carefully crafted acoustic signals to perturb the gimbal. These perturbations create controlled shifts in the camera's viewpoint, breaking the tracking system's assumptions. As a result, the attacker can hijack or deny UAV object tracking. Our attack works in real time and does not require access to the UAV's internal software, demonstrating a new cross-domain vulnerability between acoustics and vision systems.



System Profiling

Our attack leverages an **offline black-box modeling** stage which learns the behavior of gimbal motion through observed response to different injected frequencies. These steps can be automated by an adversary. It is assumed that the adversary can identify and acquire a similar drone as the targeted UAV, which is feasible as commercial UAVs often feature easily identifiable characteristics and may be purchased off the shelf.

Frequency Sweep The attacker sweeps through various frequencies to find resonance peaks, local maxima in motion amplitude vs injected signal frequency. These resonant frequencies are critical for our attack.

Amplitude modeling The amplitude of a driven harmonic oscillator, assuming all other values are fixed, is linear w.r.t the amplitude of the driving signal. As a result, we can learn this relationship via linear regression.

Online control Using a profiled model, the adversary can induce an oscillatory gimbal motion at any desired frequency and amplitude. However, for our attack, an adversary requires additional control. We develop a **feedback switching** methodology which leverages previously explored phase control without requiring access to internal sensor sampling rates.

Attack Design

Threat Model Our threat model assumes the adversary knows the gimbal model, can inject acoustic signals, and can observe the 3D position of UAV and objects in the scene as well as the direction of gimbal motion. We evaluate a **remote attack**, where the adversary uses a speaker and visual detection of gimbal motion direction, and a **direct contact attack**, where the adversary affixes a piezo transducer disc and a motion sensor onto the target gimbal.

Uncertainty To ensure robustness to prediction errors in our algorithm, we apply expectation-over-transformation in our optimization to predicted trajectories.

Attack Algorithm

1. Using UAV body pose and gimbal orientation, construct a 3D-to-2D projection matrix. Project target and attacker detections to 2D space.
2. Update a surrogate tracking model (from the same class as the targeted model) using the estimated 2D bboxes.
3. Predict future locations of the UAV and targets in the scene.
4. Compute a sequence of gimbal motion which which induces a tracking switch in the surrogate tracker over the predicted future positions.
5. Using online directional biasing and the profiled gimbal response, execute the computed gimbal motion.

Testbed We extensively evaluate our attack in a PX4 + Gazebo simulation environment. We implement numerous ablation studies including variation in target positioning, environment, and surrogate tracker to show robustness to such changes. Additional evaluation of our profiling procedure and end-to-end attacks on a physical UAV platform equipped with a gimbal-stabilized camera and standard visual tracking algorithms shows real-world feasibility. We test remote and direct-contact attacks, with different attacker capabilities, to show the flexibility of our attack.

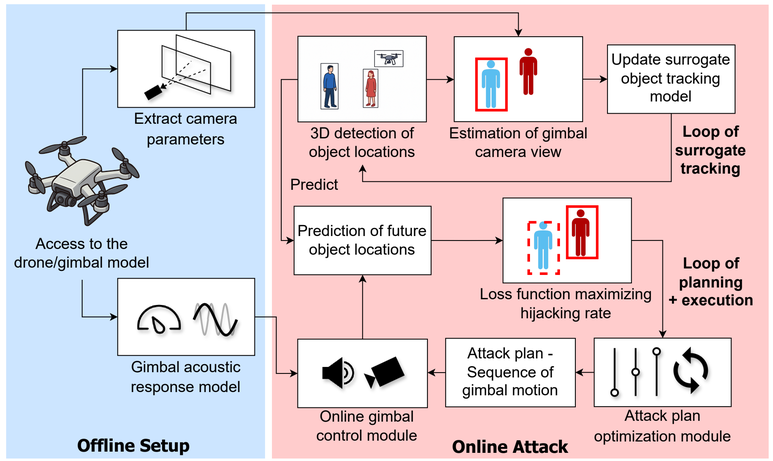


Figure 1. Attack overview

Results

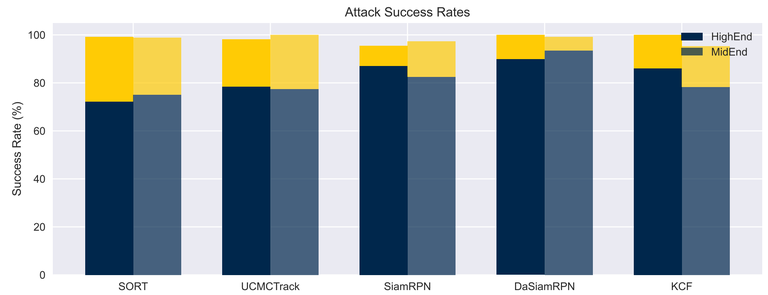


Figure 2. Switch and loss rates of tracking across different tracking algorithms and profiles.

Results Across multiple state-of-the-art trackers, our attack achieves high success rates in both hijacking and disabling tracking. Hijacking succeeds in a large majority of cases, and when combined with disabling our attack approaches near-perfect success across all evaluated trackers and drone platforms. Results are consistent across both high-end and mid-range UAVs, demonstrating that the attack is broadly effective and not tied to a specific implementation.

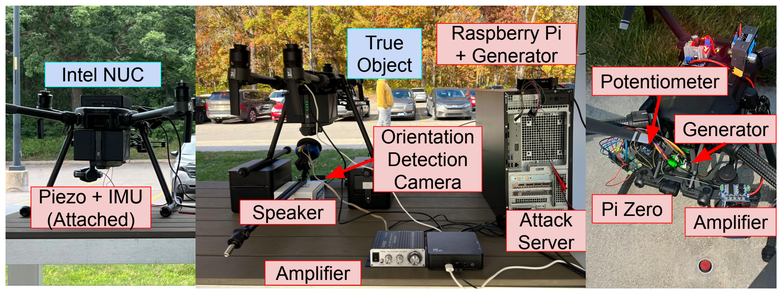


Figure 3. Physical evaluation setups, testing different attack vectors.

Defenses Existing mitigation techniques have notable drawbacks. One approach is to remove the attack vector through physical acoustic isolation of the gyroscope or through more robust signal conditioning and analog to digital converter hardware. These hardware defenses are costly and complex, require changes to off the shelf components, and demand extensive modification of the gimbal, making them practical only for future UAV designs. Software mitigations are easier to deploy on existing platforms, but known techniques are difficult to apply to gimbal systems because Banshee induces large and rapid motions while image stabilization requires a high rate data stream. Existing detection methods for acoustic injection rely on simple motion models, sensor fusion, or specialized hardware, which do not align well with typical UAV gimbal architectures.

Impact Our findings reveal a previously overlooked attack surface in UAV systems: end-to-end acoustic manipulation of inertial sensing enables adversaries to interfere with visual tracking without compromising software or communication channels. The results highlight the need for more robust defenses to secure UAV perception pipelines against physical-layer attacks.