

## Executive Summary

**Project Objective.** Address critical data gaps in event-based high-speed object tracking by constructing a custom ground station to capture a novel, multimodal UAV dataset.

### Key Contributions.

- **3D UAV Tracking Dataset:** The first event-based multimodal benchmark to provide centimeter-level 3D localization via integrated RTK and PPK GNSS for target UAVs.
- **Novel Hardware Integration:** The first hardware-level synchronization for Event/RGB UAV tracking.
- **Dual Stereo Integration:** Pair high-resolution stereo event sensors with stereo RGB for UAV tracking tasks.

## Background



Figure 1. Event cameras pose a paradigm shift in tracking fast moving targets such as UAVs. The left figures are the RGB images, while the right images are the event frames. [1]

- The development of robust, passive early warning systems for national security relies on the ability to track and localize high-speed, elusive targets like Unmanned Aerial Vehicles (UAVs) in real-time.
- **Bottleneck.** Traditional frame-based cameras fail in high-speed scenarios due to **motion blur** caused by fixed frame rates and **low dynamic range** from high-contrast outdoor lighting.
- **Event Cameras.** Bio-inspired sensors that sense light asynchronously at the pixel level, reporting only brightness changes.
- **Microsecond Latency:** Near-instantaneous response to movement.
- **High Dynamic Range:** Superior performance in extreme lighting like the sun or at night.
- **Zero Motion Blur:** Ideal for tracking high-speed targets.

## Experimental Setup

Overcoming the latency of software based synchronization seen in existing datasets, we engineer a custom **dual stereo ground station** featuring strict hardware-level sensor alignment. The optical configuration was optimized for tracking UAVs at extended ranges (**20–150m**). We achieved closely matched Fields of View ( $\sim 40^\circ$ ), maximizing Pixels on Target while minimizing depth error. Centralized high-bandwidth recording is managed by ROS2 running on a NVIDIA Jetson Orin, while a NovAtel SMART7 GNSS receiver provides the necessary framework for centimeter-level 3D localization.

Table 1. Hardware Specifications

| Component       | Model and Details                                                           |
|-----------------|-----------------------------------------------------------------------------|
| Event Cameras   | 2x Prophesee EVK4 (Sony IMX636) w/ Soyo 8mm Lenses<br>1280 x 720 Resolution |
| RGB Cameras     | 2x FLIR Blackfly S USB3 (Sony IMX250) @30FPS w/ 12mm Lenses<br>2448 x 2048  |
| 3D Localization | NovAtel SMART7 (OEM7 + VEXXIS Antenna)                                      |
| GPS Precision   | Centimeter-level Accuracy (via RTK and PPK)                                 |
| Control Unit    | NVIDIA Jetson Orin                                                          |
| UAV             | Skydio Enterprise 2                                                         |

## Experimental Test Plan

### Data Acquisition & Flight Profiles.

- **Target Specifications:** Skydio Enterprise 2 tracked at ranges of **20–150m**.
- **Flight Trajectories:** Diverse and including profiles like high-speed attack-like trajectories and low-altitude sea-skimming profiles.
- **Environmental Diversity:** Challenging captures featuring **severe glare**, tree occlusions, and varying lighting (midday, sunset, evening, night). Objective underexposure and overexposure scores based on pixel saturation. [2]
- **Distractors:** Dynamic non-UAV objects such as **birds and insects** to test tracking robustness.

### Hardware Synchronization & Calibration.

- **Temporal:** STM32-generated hardware pulses provide microsecond-level triggering across all sensors.
- **Spatial:** Kalibr-based calibration for intrinsic parameters and lens distortion correction. [3]

### Ground Truth Annotation Strategy.

- **Event-First Labeling:** Bounding box annotations will be generated primarily on the event accumulated streams rather than the RGB frames. These boxes are stretched across the time axis to filter out non-target events, ensuring highly accurate event-level segmentation. [4]
- **Extended Metadata:** Ground truth continuous 3D trajectories and unique track IDs for both the UAV and non-UAV anomalies.
- **Evaluation Metrics:** Standard evaluation metrics via **mAP@50:95** (Detection), **MOTA/IDF1** (Tracking), and **ADE/FDE** (Forecasting).

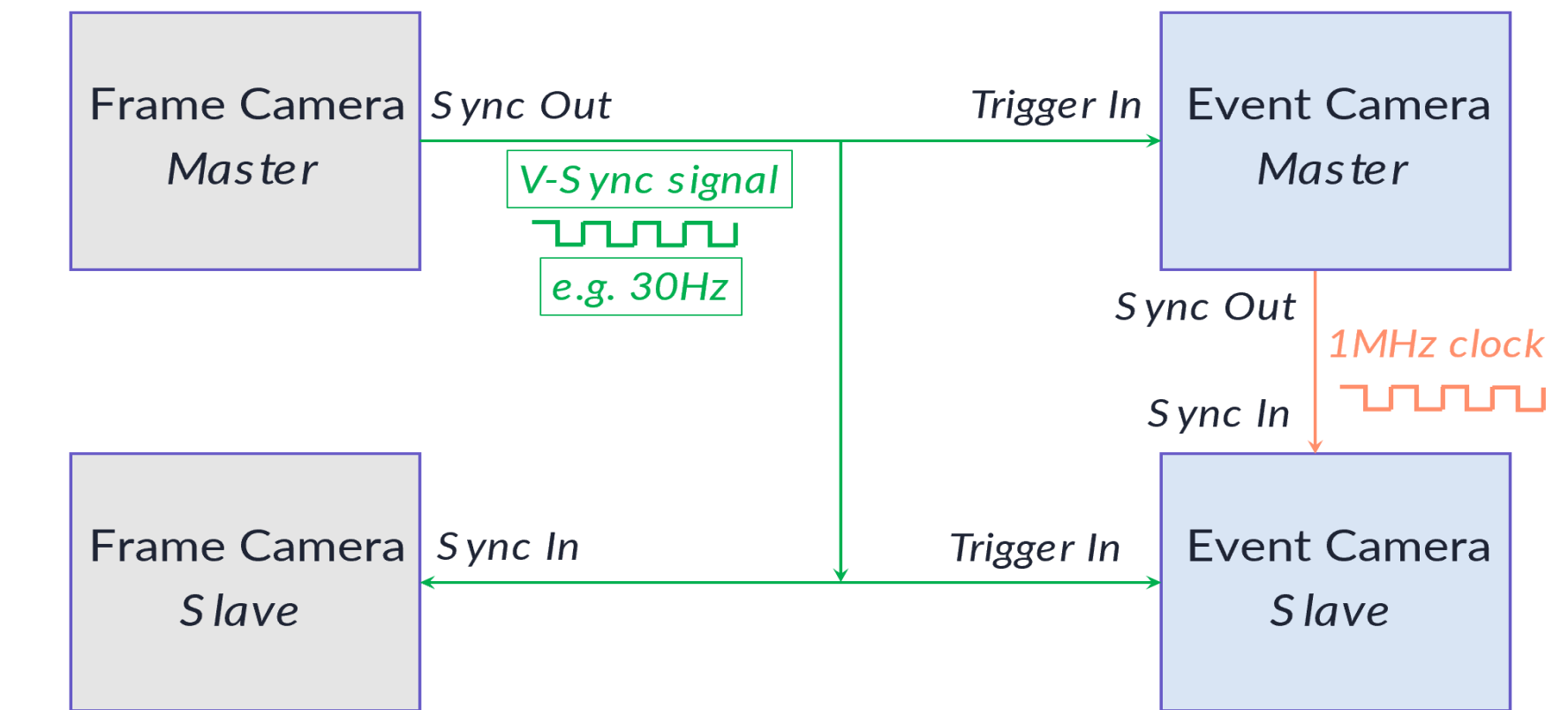


Figure 2. Our Presented Dual Stereo Camera Synchronization Setup

## Preliminary Results & System Validation

- **Temporal Synchronization:** The STM32-based and ROS2 hardware synchronization architecture has been successfully implemented and tested with sub-microsecond alignment.
- **Annotation Pipeline Initialization:** Begun on the "event-first" labeling pipeline, utilizing a strategy of stretching 2D bounding boxes across the temporal axis to generate precise event-level segmentation.
- **Current Development:** The system is currently undergoing integration of the RTK and PPK GNSS receivers to finalize the automated projection of absolute 3D trajectories into the camera frame.

## Conclusions

We present our in-progress work on the creation of a novel multimodal UAV tracking dataset. In the scene of existing event-based UAV tracking literature we are the:

- **First 3D Localized Benchmark:** The first UAV dataset to feature absolute, centimeter-level 3D localization via integrated RTK and PPK GNSS.
- **First Hardware-Synchronized Dataset:** Unlike prior datasets that rely on imperfect software alignment, this is the first to achieve guaranteed sub-millisecond temporal accuracy through strict hardware triggering.
- **First Dual-Stereo Multimodal Rig:** The dataset is first to successfully pair stereo event with stereo RGB for UAV tracking.
- **First to Use Objective Illumination Metrics:** Replace subjective day/night labels with rigorous, pixel-saturation-based underexposure and overexposure metrics to accurately benchmark sensor limits. [2]

## Future Work

- **Further Field Deployment:** Capture high-speed flight data at the UTARI Maverick Autonomous Vehicle Research Center featuring diverse threat profiles and more UAVs.
- **Open-Source Release:** Post-process and organize the data into a public benchmark to facilitate the development of advanced event-based object tracking and 3D localization algorithms.

## References

- [1] G. Magrini et al., "Neuromorphic Drone Detection: an Event-RGB Multimodal Approach," in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2024.
- [2] L. Cui et al., "PEOD: A Pixel-Aligned Event-RGB Benchmark for Object Detection under Challenging Conditions," arXiv preprint arXiv:2511.08140, 2025.
- [3] M. Muglikar, M. Gehrig, D. Gehrig, and D. Scaramuzza, "How to Calibrate Your Event Camera," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2021.
- [4] X.-F. Zhu et al., "Evidential Detection and Tracking Collaboration: New Problem, Benchmark and Algorithm for Robust Anti-UAV System," *IEEE Transactions on Multimedia*, vol. 26, pp. 10643–10656, 2024.