Fast and Robust UAV to UAV Detection and Tracking From Video
IEEE Transactions on Emerging Topics in Computing, August 2021

JING LI , (Member, IEEE), DONG HYE YE, (Member, IEEE), MATHIAS KOLSCH, (Member, IEEE), JUAN P. WACHS , (Senior Member, IEEE), AND CHARLES A. BOUMAN , (Fellow, IEEE)

Literature Review by Shubhan Anchan
Grade XII Student, Bombay International School
September 2025

Introduction

I identified the paper "Fast and Robust UAV to UAV Detection and Tracking From Video" by Li, Wachs, Bouman et al, August 2021, IEEE Transactions on The Emerging Topics in Computing. in support of my Internship Project Report on Improving the Efficiency of Object Detection, Recognition and Identification by Drones. I am also utilizing my learnings from the course, I attended on Machine Learning (including CNNs) and Image Recognition Models conducted by Prof. Pavlos Protopapas, Harvard SEAS.

I present a literature review of the said paper below.

Context

Unmanned aerial vehicles (UAVs) provide functionality around remote sensing, delivery, and security, requiring accurate real-time detection and tracking for collision avoidance and coordinated operation. Camera-based "see-and-avoid" systems use lightweight, high-definition optical sensors but demand robust, computationally efficient video analysis algorithms for autonomous detection and tracking of rapidly moving UAVs in cluttered environments. The unique challenge lies in simultaneous motion of both observer and target UAV, background complexity, small target appearance, and the scarcity of multi-UAV video datasets for training and evaluation.

Classical Methodologies for Detection

I understood, from the paper, that early video-based detection utilized handcrafted features (such as Scale-Invariant Feature Transform - SIFT, Histogram of Oriented Gradients - HOG, Haar, Local Binary Pattern - LBP) and motion models (optical flow, background subtraction), but these have been limited in robustness to scale changes, motion blur, and noisy backgrounds. The paper cites recent research adopting deep convolutional neural networks (CNNs), region-based and single-shot detectors (You Only Look Once - YOLOv3, R-CNN), and hybrid models—significantly improving detection accuracy for general objects in cluttered scenes but often at the cost of high computational loads and difficulties with small, distant targets.

I learnt from the paper, that specialized algorithms for airborne object detection have been proposed, such as motion compensation followed by CNN-based appearance classification, and hybrid systems combining appearance-based detectors and correlation filter trackers. These methods often assume large target appearance regions or significant computational resources, limiting their deployment on lightweight UAV platforms.

Innovative Two-Stage Methodology Proposed by the Authors

In their paper, the Authors propose a modular, highly accurate, and efficient pipeline for UAV-to-UAV detection (U2U-DT). The system divides into two key stages:

- **Target Detector**: Employs background subtraction (via global perspective transform) and sparse optical flow estimation to generate candidate regions. Salient points are classified using a hybrid device—AdaBoost fusing appearance (CNN) and motion classifiers for robust candidate detection. The appearance classifier processes RGB patches from both original and subtracted images, while the motion classifier analyzes seven-dimensional motion feature vectors derived from optical flow and perspective parameters.
- **Target Tracker**: Initiates tracks using candidate detections. Propagation uses flow point-based tracking (Lucas-Kanade method) and a second hybrid classifier to prune inaccurate points. Temporal consistency is achieved with Kalman filtering, greatly improving detection of faint and intermittent targets missed by single-frame methods. Precision and recall are maximized with minimal computational cost.

Comparison of Performance, with Datasets

The authors, through this paper, introduce a unique, publicly available UAV-to-UAV video dataset (U2U-DTD) featuring 50 sequences with outdoor, real-world challenges and manual ground-truth annotations for up to eight UAVs per frame. Experimental results demonstrate superior recall (0.89), precision (0.88), and F-score (0.89) compared to state-of-the-art approaches (EPFL algorithm), with the proposed U2U-DT pipeline sustaining real-time performance on commodity hardware at 30 fps.

Ablation studies highlight the critical impact of combining motion and appearance features—removal of either dramatically reduces performance; Kalman tracking further mitigates missed detections in clutter or occlusion scenarios.

Limitations and Research Directions

The paper recognizes persistent challenges with the need for more extensive annotated datasets, robustness to high occlusion, small distant UAVs, and further reduction in onboard computational requirements for deployment on embedded processors. The paper suggests future expansion to ground vehicle detection and additional hardware optimization.

Conclusion

I submit that this technical paper advances the state-of-the-art in UAV-to-UAV detection by innovating a highly modular hybrid framework, validated on a unique multi-UAV video dataset. I also submit that its real-world applicability is supported by robust experimental results, setting the foundation for continued development in autonomous see-and-avoid, multi-agent UAV coordination, and embedded visual analytics.

Citations:

Li, Jing, et al. “Fast and Robust UAV to UAV Detection and Tracking from Video.” *IEEE Transactions on Emerging Topics in Computing*, vol. 10, no. 3, 1 July 2022, pp. 1519–1531, https://doi.org/10.1109/tetc.2021.3104555.

engineering.purdue.edu/~bouman/publications/pdf/tetc01.pdf