UDC 004.93
DOI: https://doi.org/10.17721/AIT.2025.1.03

Iryna YURCHUK, PhD (Phys. & Math.), Assoc. Prof.
ORCID ID: 0000-0001-8206-3395
e-mail: i.a.yurchuk@gmail.com
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Taras SEMENCHENKO, PhD Student
ORCID ID: 0009-0007-3259-7007
e-mail: taras.semenchenko@knu.ua
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

MOTION-CONSISTENT TEMPORAL FUSION FOR UAV DETECTION AND TRACKING

Background. *Detecting and tracking Unmanned Aerial Vehicles (UAVs) in video streams is essential for modern air-space monitoring yet remains challenging because UAVs are small, fast and easily confused with birds or background clutter. Conventional detectors produce noisy, frame-wise boxes, while standard trackers still suffer from false positives and identity switches. The purpose of this study is to stabilize UAV detections by adding a motion-aware temporal fusion method to a mainstream detector-tracker pipeline.*

Methods. *A detection-tracking pipeline was constructed using an RT-DETR (Real-Time Detection Transformer) and ByteTrack baseline, extended with a lightweight, training-free motion-consistent fusion (MCF) method. The method (i) aggregates bounding-box history over five frames, (ii) averages spatial and confidence values, and (iii) penalizes tracks whose short-term velocity or angular change exceeds empirically chosen thresholds. No appearance features or additional learning are required, so the solution runs in real time on a single GPU.*

Results. *Experiments on a labelled UAV-video dataset show that the proposed method increases Multiple Object Tracking Accuracy (MOTA) from 0.533 to 0.591, precision from 73 % to 84 %, and reduces identity switches from 60 to 28 (a 53 % improvement in ID stability). Recall decreases slightly from 90 % to 76 %, reflecting a deliberate trade-off: the system filters unstable or non-UAV motion to improve track consistency and suppress false positives. The evaluation was performed on more than 1,000 video sequences, ensuring robustness across diverse flight environments.*

Conclusions. *The motion-consistent fusion method significantly enhances both accuracy and temporal coherence while adding minor computational cost. It can be added into existing detection–tracking systems and is particularly suited for real-time UAV surveillance applications, though performance may degrade if drones execute extremely abrupt maneuvers outside the predefined motion thresholds.*

Keywords: *UAV, Object Tracking, Object Detection, Fusion Algorithm, Motion Filter, RT-DETR, ByteTrack.*

Background

Unmanned aerial vehicles (UAVs), commonly named as drones, are increasingly used in both civilian and military applications. As the number of UAV operating in public and private airspace continues to grow, the need for accurate, reliable, and real-time UAV detection systems becomes critical, particularly in video-based monitoring applications.

However, detecting and tracking UAVs in real-world videos is very challenging task. UAVs are typically small, fast-moving, and often visually similar to birds and background objects. Even modern object detectors can produce noisy bounding boxes with volatile confidence values for different frames. Multi-object tracking methods such as ByteTrack can associate detections over time, but they are still vulnerable to false positives, identity switches, and lack semantic understanding of object motion.

This research focuses on the development of a lightweight, real-time fusion method to improve the temporal stability of UAV detections in video. The object of study is a vision-based UAV monitoring system, and the subject is the refinement of frame-level detections through fusion of temporal and motion information.

The goal of this paper is to improve UAV detection accuracy and tracking reliability by introducing a motion-aware temporal fusion module. Special attention is paid to the Multiple Object Tracking Accuracy (MOTA) metric, which combines false positives, missed targets, and identity switches into a single measure of tracking performance.

To achieve this, the following tasks were completed:

- Construction of a detection-tracking pipeline using RT-DETR and ByteTrack;
- Implementation of a temporal fusion module that smooths detection outputs;
- Integration of a motion consistency filter to penalize erratic movement;
- Evaluation of detection and tracking performance.

The scientific novelty of this article lies in the proposed motion-consistent fusion (MCF) mechanism, which extends standard temporal fusion with motion smoothness analysis, which is a method particularly suited for UAVs, that typically follow smooth, linear trajectories. Unlike appearance-based methods, MCF does not require additional training or feature extraction, making it highly practical for real-time UAV tracking.

Experimental results show that the proposed approach improves overall tracking accuracy, indicated by a higher MOTA score, while increasing precision, and significantly lowering identity switches. Recall remains strong, though the method may underperform in cases of highly dynamic or intentionally evasive flight patterns, where the assumption of smooth, UAV-like motion no longer holds.

Related work. The real-time tracking of Unmanned Aerial Vehicles (UAVs) faces bigger challenges than detection alone: extremely small targets, quick moves, frequent disappearances, and camera motion. So, recent research in UAV tracking focuses on two goals: building high-quality benchmarks that reveal these pitfalls and designing trackers that remain robust while still running at video rate on edge hardware.

Jiang et al. introduce Anti-UAV, a large-scale RGB benchmark containing more than 300 video pairs and about 580k manually annotated bounding boxes (Jiang et al., 2021). Beyond its size, the dataset pairs each UAV sequence with an

© Yurchuk Iryna, Semenchenko Taras, 2025

ISSN 2788-6603

empty-background counterpart, enabling dual-stream evaluation, and the authors supply a baseline tracker (DFSC) that exploits semantic flow to stabilize appearance drift, showing a significant margin over earlier Siamese architectures.

To support this, Zhao et al. created DUT Anti-UAV, which combines 10 000 labelled still images for detection with 20 RGB videos for tracking (Zhao et al., 2022). The authors demonstrate that fusing a YOLO-style detector with a lightweight correlation-filter tracker markedly boosts precision, underlining the value of joint "detect-then-track" pipelines when the target is tiny or often disappears.

While both datasets rely on visible light, Zhang, P. et al. extend the problem to visible–thermal UAV tracking (VTUAV) by collecting 500 RGB-T sequences totaling 1.7 million 1080 p frame pairs (Zhang, P. et al., 2022). They also created a new tracker called the Hierarchical Multi-modal Fusion Tracker (HMFT), which combines thermal information at different levels. This layered approach reduces tracking failures by about 20% in their tests, showing that heat data can help when lighting is poor.

Transformer backbones increasingly dominate state-of-the-art trackers. Yu et al. present UTTracker, whose unified Transformer feeds four specialized heads: (i) multi-region local search, (ii) global detection for re-entry, (iii) background correction to cope with camera motion, and (iv) a dynamic head tuned for very small targets (Yu et al., 2023). This holistic design secured the 2nd place in the 3rd Anti-UAV Challenge, highlighting that integrating detection, localization, and motion compensation is more effective than alternating them.

Robustness can also be improved through representation learning. Fu et al. propose PRL-Track, which first learns a coarse appearance-aware representation, then progressively refines it via a hierarchical generator (Fu et al., 2024). Despite the two-stage pipeline, PRL-Track maintains 42.6 FPS on an Nvidia Xavier, illustrating that adaptive feature refinement can balance accuracy with real-time constraints on embedded hardware.

Finally, Do et al. tackle deployment at scale with RAMOTS, a real-time multi-object tracker that couples YOLOv8 (Reis et al., 2023) and YOLOv10 (Wang et al., 2024) (versions of the You Only Look Once real-time object detection algorithm) detectors and BYTETrack (Zhang, Y. et al., 2022) / BOTSort (Aharon, Orfaig, & Bobrovsky, 2022) trackers for fault-tolerant processing (Do et al., 2024). Although aimed at "from-UAV" rather than "against-UAV" scenarios, RAMOTS demonstrates that big-data tooling sustains 28 FPS while managing large video streams–an architectural insight relevant to surveillance networks.

However, existing methods often overlook the specific motion characteristics of UAVs, leading to false positives in cluttered environments. To address these limitations, the method proposed in this paper consists of three main components: a transformer-based detector, an association tracker, and a novel motion-consistent fusion mechanism designed to stabilize trajectories based on flight dynamics.

Methods

The proposed approach comprises three stages: an object detector RT-DETR (Wang et al., 2025), a multi-object tracker ByteTrack (Zhang, Y. et al., 2022), and a custom temporal fusion module that filters and stabilizes detections based on motion consistency.

Detection Stage. The first component of the proposed system is the object detection model, which is responsible for identifying UAV instances in video frames. The RT-DETR (Real-Time Detection Transformer) is used as the backbone detector due to its balance between speed and accuracy, particularly suitable for real-time UAV monitoring applications.

RT-DETR is a transformer-based object detector that directly predicts bounding boxes and class probabilities without the need for anchor generation or post-processing steps such as non-maximum suppression (NMS). It uses a combination of convolutional and attention-based layers to extract rich spatial and contextual information from the input image. The architecture produces a fixed set of object queries, each corresponding to a potential object instance.

At inference time, RT-DETR outputs a set of detections per frame, where each detection includes:

- A bounding box (x, y, w, h);
- A class label (in this case, primarily "UAV");
- A confidence score $c \in [0,1]$.

These frame-wise detections are passed to the tracking and fusion stages for temporal association and refinement.

Tracking stage. To maintain object identity across video frames and enable temporal reasoning, a real-time multi-object tracking method based on ByteTrack is integrated. ByteTrack is a high-performance, association-based tracker that links object detections over time without relying on deep appearance features, making it lightweight and well-suited for UAV tracking in real-time scenarios.

At each frame *t*, the detector (RT-DETR) produces a set of bounding boxes $\{B_j^t\}$ with associated confidence scores. ByteTrack divides these detections into high-confidence and low-confidence sets using a fixed threshold (e.g., 0.5). It then performs data association between the new detections and previously tracked objects.

Each object track maintains a motion model using a Kalman filter. The position of the object is predicted in the current frame using the following linear state transition model:

$$\dot{x_t}|_{t-1} = A \cdot x_{t-1},$$

where x_{t-1} – is the previous state vector of the object (e.g., center position and velocity), *A* – is the state transition matrix (typically constant), $\dot{x_t}|_{t-1}$ – is the predicted state before updating with new observations.

To match detections with predicted tracks, ByteTrack computes the Intersection over Union (IoU) between the predicted bounding box B_i and the detected box B_j:

$$IoU(B_i, B_j) = \frac{|B_i \cap B_j|}{|B_i \cup B_j|}.$$

A cost matrix is constructed based on this metric, and the Hungarian algorithm is used to solve the assignment problem, pairing detections with existing tracks by maximizing overall IoU.

Each matched track is updated with the new detection. Unmatched tracks are temporarily retained and predicted forward using the Kalman filter, allowing the system to recover from short-term occlusions. Tracks that remain unmatched for a predefined number of frames are terminated. Importantly, ByteTrack does not modify detection confidence or bounding box geometry, it only propagates and assigns object identities.

While ByteTrack is highly effective for frame-to-frame association, it is sensitive to detection noise and does not enforce global consistency across time. To deal with these challenges, a temporal fusion method is introduced in the next section that smooths detection results and penalizes unusual motion patterns, which are rare in typical UAV trajectories.

Temporal Fusion Method. Although the integration of ByteTrack provides temporal identity assignment and object association across frames, it does not directly improve the quality or stability of detection results. In real-world UAV tracking scenarios, especially under challenging conditions (e.g., occlusions, fast motion, or cluttered backgrounds), frame-by-frame detections can have high variance in both confidence scores and bounding box locations. To address these limitations, in this paper a lightweight temporal fusion method is introduced that aggregates detection information over time for each tracked object.

The core idea of the proposed fusion method is to keep a fixed-length history buffer for each object track, indexed by its unique track ID. This buffer stores the most recent N detections associated with the object, including:

- The bounding box coordinates $B_i^t = (x, y, w, h)$,
- The detection confidence c_i^t.

At each frame, the fusion method performs the following steps for every active track:

Step 1. Update the buffer with the current detection (bounding box and confidence).

Step 2. Remove the oldest entry if the buffer exceeds the maximum length *N*.

Step 3. Compute a fused bounding box by averaging the coordinates across all buffer entries:

$$\overline{B_i} = \frac{1}{N} \sum t = t - N + B_i^t.$$

Step 4. Compute a fused confidence score as the mean of the stored scores:

$$\overline{c_i} = \frac{1}{N} \sum t = t - N + c_i^t.$$

Step 5. Estimate the detection stability by calculating the standard deviation of the bounding box center positions across the buffer. If the standard deviation exceeds a defined threshold, the detection is flagged as unstable and may be suppressed.

This temporal fusion approach reduces jitter in bounding box placement and filters out short-lived, noisy detections. The averaging strategy also increases the reliability of confidence scores, allowing the system to more robustly handle volatile detections that may otherwise be discarded or mismatched.

The fusion process is non-parametric and computationally cheap, requiring only arithmetic operations over a small buffer. It can be applied in real time and is fully compatible with online tracking systems.

In the next section, this fusion mechanism is extended by incorporating motion consistency analysis to further penalize erratic movement patterns that are unlikely to originate from UAVs.

Motion Consistency Filtering. To further improve the robustness of UAV tracking, a motion consistency filtering mechanism is proposed to enhances the temporal fusion method by incorporating dynamic constraints based on typical UAV movement patterns. While the fusion module stabilizes detections over time, it does not consider the physical plausibility of an object's trajectory. In contrast, motion consistency filtering evaluates whether an object is moving in a smooth, UAV-like manner and penalizes tracks that have unnatural behavior.

UAVs typically have controlled, smooth motion, often with steady direction or hovering behavior. In contrast, false positives such as birds, background clutter, or detection noise tend to result in abrupt position changes or inconsistent movement between frames. The proposed method exploits this behavioral distinction to assess and filter object tracks.

For each object track, the short-term trajectory is analyzed over the last N frames using its center point positions $p_i^t = (x_t, y_t)$.

The following parameters are computed:

- Velocity change (Δ*v*):

$$\Delta v = \left\| p_i^t - 2p_i^{t-1} + p_i^{t-2} \right\|;$$

- Angular deviation (Δθ):

$$\Delta\theta = \arccos\left(\frac{(p_i^{t-1} - p_i^{t-2}) \cdot (p_i^t - p_i^{t-1})}{\left\| p_i^{t-1} - p_i^{t-2} \right\| \cdot \left\| p_i^t - p_i^{t-1} \right\|} \right);$$

- Motion consistency score (MCS):

$$MCS = \alpha \cdot \Delta v + \beta \cdot \Delta\theta.$$

If the motion consistency score exceeds a predefined threshold τ, the corresponding detection is flagged as inconsistent and either penalized (by reducing its confidence) or temporarily ignored by the tracking system.

This filtering mechanism helps suppress false positives caused by fast-moving distractors, reduce identity switches in noisy scenes, and prioritize detections that exhibit UAV-like motion patterns. Importantly, this enhancement is model-free and fully compatible with real-time deployment.

Fusion Algorithm. The final stage of the system combines the results of detection, tracking, temporal fusion, and motion consistency filtering to produce reliable, temporally stable UAV detections with unique identifiers. This stage determines which object tracks are retained and presented as output at each frame.

For each frame, the system performs the following steps:

1. Detection and tracking: The RT-DETR model generates bounding boxes and confidence scores, which are associated across frames using ByteTrack. Each detection is linked to a persistent object track with a unique ID.
2. Temporal fusion: For each active track, a fused bounding box and smoothed confidence score are computed by aggregating the track's detection history over the last N frames. This reduces jitter and stabilizes detection output.

ISSN 2788-6603

3. Motion consistency filtering: The short-term motion pattern of each track is analyzed to compute a motion consistency score. Tracks with erratic or physically implausible motion are flagged and penalized by lowering their confidence or suppressing their output altogether.
4. Decision thresholding: Tracks that meet all criteria – including minimum confidence, temporal stability, and motion consistency – are passed to the final output stage. Each retained track is represented as:

$$Output = (B_i, \bar{c}_i, track\_id),$$

where B_i is the fused bounding box, $\bar{c}_i$ is the smoothed confidence score, and *track_id* is the assigned identity.

To formalize the decision-making process, we define a valid output set O_{final} at frame *t*. A track T_i is included in the final output only if it satisfies both spatial stability and motion consistency constraints. This can be expressed using an indicator function:

$$O_{final} = \{ (\hat{B}_i, \hat{C}_i, id_i) \mid i \in \mathcal{T}_t,\ \sigma(H_i^{bbox}) \leq \sigma_{th} \wedge S_{motion}(i) \leq \tau \}.$$

where O_{final} – is the set of validated output tracks; $\mathcal{T}_t$ – represents the set of all candidate tracks at frame; $\hat{B}_i$ and $\hat{C}_i$ – denote the fused bounding box and confidence score; $\sigma(H_i^{bbox})$ – is the spatial stability metric (standard deviation of the history buffer); $S_{motion}(i)$ – is the motion consistency score; σ_{th} and τ are the stability and motion thresholds, respectively.

The full logic of the fusion and filtering process is showed in Algorithm 1 below.

Algorithm 1. Fusion strategy of tracker

Input:
T_k – set of tracks at frame k
H – history buffer per track ID
N – history length (e.g., 5 frames)
$\sigma_{threshold}$ – stability threshold
θ_{max} – max angle deviation
v_{max} – max velocity change
Output:
T_k^fused – filtered, fused tracks for frame k

```
for each track t in T_k do
    id ← t.id
    bbox ← t.bbox
    conf ← t.confidence
    if H[id] does not exist then
        initialize H[id] ← empty list
    end if
    append (bbox, conf) to H[id]
    if length(H[id]) > N then
        remove the oldest entry from H[id]
    end if
    bboxes ← list of all b from (b,) in H[id]
    confs ← list of all c from (, c) in H[id]
    fused_bbox ← mean(bboxes)
    fused_conf ← mean(confs)
    bbox_std ← standard_deviation(bboxes)
    compute motion_score from recent bboxes
        – velocity ← center_distance(b_n, b_n−1)
        – angle_change ← angle(center_n−2 → center_n−1 → center_n)
        – velocity ← center_distance(b_n, b_n−1)
    if bbox_std < σ_threshold and motion_score < threshold then
        mark track t as "stable"
        output (fused_bbox, fused_conf, id)
    else
        penalize fused_conf
    end if
end for
```

The proposed fusion algorithm introduces a novel enhancement over classical temporal fusion techniques by incorporating a lightweight motion consistency filter. Traditional fusion methods typically rely on statistical averaging (e.g., exponential moving averages or buffer smoothing) across historical detections to reduce jitter and noise. While effective in improving spatial stability, these approaches are blind to the dynamics of object motion and may average out valid but fast movements, or fail to suppress outliers caused by false positives.

In contrast, the proposed method evaluates each fused detection not only based on temporal coherence but also on motion plausibility, derived from short-term velocity and angular trajectory consistency. By explicitly penalizing tracks with erratic movement patterns, such as sudden jumps, sharp turns, or non-UAV-like acceleration, the proposed approach suppresses false positives (e.g., birds, clutter, background detections) without compromising on real-time performance.

This is considered particularly critical in UAV surveillance applications, where:

- Most UAV have smooth, predictable motion patterns.
- False positives often have inconsistent or unnatural motion.
- The system must be efficient and non-reliant on deep appearance models (e.g., no Re-ID or embeddings).

This fusion strategy improves detection precision while maintaining tracking integrity, leading to fewer identity switches, lower false positives, and more consistent trajectories.

Results

The system was evaluated on the (Jiang et al., 2021; Zhang, P. et al., 2022; Zhao et al., 2022) UAV video dataset, which contains more than 1000 video sequences annotated with bounding boxes and object identities. The dataset includes a diverse range of scenes such as urban areas, vegetation, and open skies, with variable lighting and UAV altitudes.

The detection method uses a pretrained RT-DETR model fine-tuned on UAV-specific data. Tracking is performed with ByteTrack, configured with a high-confidence threshold of 0.5 and a Kalman filter for state propagation.

Our temporal fusion method maintains a fixed-length history buffer $N = 5$ and applies motion filtering using two features: center velocity change and angle deviation. Tracks are penalized if their standard deviation exceeds $\sigma_{thresh} = 20$ pixels or if their motion score exceeds a dynamic threshold. All experiments were run on an NVIDIA RTX 3070 ti.

We evaluate performance using:

- MOTA as an overall indicator of tracking accuracy,
- Precision and Recall for detection quality,
- Identity Switches (ID Switches) for measure identity consistency during tracking.

Table 1

Result comparison of fusion configurations

Approach	MOTA	Precision	Recall	ID Switches
RT-DETR only	0.431	0.682	0.824	-
+ ByteTrack	0.480	0.736	0.806	57
+ Fusion (baseline)	0.533	0.733	0.900	60
+ MCF (our approach)	0.591	0.838	0.755	28

Table 1 presents a comparison of detection and tracking performance across different system variants with MOTA used as the primary measure of overall tracking accuracy. The baseline RT-DETR detector achieves a MOTA of 0.431, along with precision of 68% and recall of 82%, but lacks any temporal consistency or object ID tracking.

Adding ByteTrack improves tracking performance to a MOTA of 0.480, driven by gains in precision (up to 74%) and tracking consistency, though 57 identity switches still occur.

The baseline fusion approach further raises MOTA to 0.533, with recall peaking at 90%, confirming that temporal smoothing helps recover missed detections, but identity stability remains limited (60 ID switches).

The proposed method, Motion-Consistent Fusion (MCF), achieves the best overall results:

- MOTA increases to 0.591, the highest score among all tested variants,
- Precision improves to 84%,
- Recall decreases slightly to 76%,
- ID switches are reduced to just 28.

This trade-off, higher MOTA and precision at the cost of a small recall drop-reflects the effect of motion consistency filtering, which suppresses detections with erratic or non-UAV motion. While a few true but noisy detections may be excluded, the result is a more stable and reliable tracking system, with improved resistance to clutter and false positives.

Figure 1 illustrates typical frames where the proposed method smooths detections and eliminates false tracks.

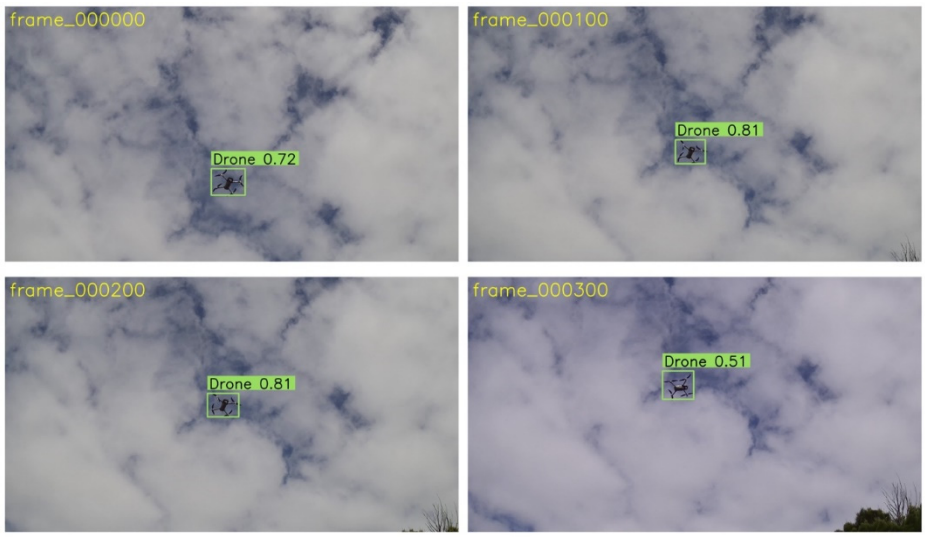

Fig. 1. Examples of the motion-consistent fusion (MCF) results

ISSN 2788-6603

Discussion and conclusions

In this paper, a lightweight and effective temporal fusion framework for UAV detection and tracking in video streams is presented. Based on a standard detection-tracking pipeline with RT-DETR and ByteTrack, a novel motion-consistent fusion (MCF) method is proposed, which improves detection stability by combining temporal smoothing with motion plausibility analysis.

The proposed method introduces two key ideas:

1. **Temporal aggregation** of bounding boxes and confidence scores over a fixed-length buffer, reducing frame-to-frame jitter.
2. **Motion consistency filtering**, which penalizes tracks exhibiting erratic movement patterns unlikely to represent UAV behavior.

Experimental results on a UAV video dataset demonstrate that the proposed approach improves overall tracking performance, achieving a higher MOTA score, increased precision, and significantly fewer identity switches. While recall decreases slightly, this trade-off helps suppress false positives and ensures more stable, UAV-like tracking behavior. MCF increases MOTA from 0.533 to 0.591 – the highest among all tested variants, while maintaining real-time operation, requiring no additional training, and integrating seamlessly with existing detection and tracking systems.

Overall, the proposed method offers a practical and generalizable improvement for UAV detection systems, particularly in real-world conditions where false positives and unstable detections are common.

Authors' contribution: Iryna Yurchuk – general supervision, methodology verification, review and editing of the manuscript, and final approval of the manuscript; Taras Semenchenko – conceptualization of the study, literature review, design and implementation of the motion-consistent fusion algorithm, experimental set-up and evaluation, analysis of results, drafting and final approval of the manuscript.

Sources of funding. This study did not receive any grant from a funding institution in the public, commercial, or non-commercial sectors.

References

Aharon, N., Orfaig, R., & Bobrovsky, B.-Z. (2022). *BoT-SORT: Robust associations multi-pedestrian tracking*. arXiv. https://doi.org/10.48550/arXiv.2206.14651

Do, N.-T., Nguyen, N. N.-Y., Nguyen, D.-P., & Do, T.-H. (2024). Ramots: A real-time system for aerial multi-object tracking based on deep learning and big data technology. In *2024 16th International Conference on Knowledge and System Engineering (KSE)* (pp. 1–6). VNU University of Engineering and Technology. https://doi.org/10.1109/KSE63888.2024.11063545

Fu, C., Lei, X., Zuo, H., Yao, L., Zheng, G., & Pan, J. (2024). Progressive representation learning for real-time UAV tracking. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 5072–5079). School of Electrical and Electronic Engineering, Nanyang Technological University. https://doi.org/10.1109/IROS58592.2024.10803050

Jiang, N., Wang, K., Peng, X., Yu, X., Wang, Q., Xing, J., Li, G., Zhao, J., Guo, G., & Han, Z. (2021). *Anti-UAV: A large multi-modal benchmark for UAV tracking*. arXiv. https://doi.org/10.48550/arXiv.2101.08466

Reis, D., Kupec, J., Hong, J., & Daoudi, A. (2023). *Real-time flying object detection with yolov8*. arXiv. https://doi.org/10.48550/arXiv.2305.09972

Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). Yolov10: Real-time end-to-end object detection. *Advances in Neural Information Processing Systems*, 37, 107984–108011. https://proceedings.neurips.cc/paper_files/paper/2024/hash/c34ddd05eb089991f06f3c5dc36836e0-Abstract-Conference.html

Wang, S., Xia, C., Lv, F., & Shi, Y. (2025). Rt-detrv3: Real-time end-to-end object detection with hierarchical dense positive supervision. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (pp. 1628–1636). Johns Hopkins University. https://doi.org/10.1109/WACV61041.2025.00166

Yu, Q., Ma, Y., He, J., Yang, D., & Zhang, T. (2023). A unified transformer based tracker for anti-UAV tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops* (pp. 3036–3046). IEEE Computer Society; Computer Vision Foundation. https://openaccess.thecvf.com/content/CVPR2023W/Anti-UAV/html/Yu_A_Unified_Transformer_Based_Tracker_for_Anti-UAV_Tracking_CVPRW_2023_paper.html

Zhang, P., Zhao, J., Wang, D., Lu, H., & Ruan, X. (2022). Visible-thermal UAV tracking: A large-scale benchmark and new baseline. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 8886–8895). IEEE Computer Society; Computer Vision Foundation. https://openaccess.thecvf.com/content/CVPR2022/html/Zhang_Visible-Thermal_UAV_Tracking_A_Large-Scale_Benchmark_and_New_Baseline_CVPR_2022_paper.html

Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2022). Bytetrack: Multi-object tracking by associating every detection box. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, & T. Hassner (Eds.), *Computer Vision – ECCV 2022* (pp. 1–21). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-20047-2_1

Zhao, J., Zhang, J., Li, D., & Wang, D. (2022). Vision-based anti-uav detection and tracking. *IEEE Transactions on Intelligent Transportation Systems*, 23(12), 25323–25334. https://doi.org/10.1109/TITS.2022.3177627

Отримано редакцією журналу / Received: 20.07.25
Прорецензовано / Revised: 20.09.25
Схвалено до друку / Accepted: 20.09.25

Ірина ЮРЧУК, канд. фіз.-мат. наук, доц.
ORCID ID: 0000-0001-8206-3395
e-mail: i.a.yurchuk@gmail.com
Київський національний університет імені Тараса Шевченка, Київ, Україна

Тарас СЕМЕНЧЕНКО, асп.
ORCID ID: 0009-0007-3259-7007
e-mail: taras.semenchenko@knu.ua
Київський національний університет імені Тараса Шевченка, Київ, Україна

МЕТОД УЗГОДЖЕНОГО ЗА РУХОМ ЧАСОВОГО Ф'ЮЖНА ДЛЯ ВИЯВЛЕННЯ ТА ВІДСТЕЖЕННЯ БПЛА

В с т у п . ***Виявлення та відстеження безпілотних літальних апаратів (БПЛА) у відеопотоках є критично важливим завданням сучасного моніторингу повітряного простору. Водночас воно залишається складним через малі розміри, швидкий рух БПЛА та їхню схожість із птахами чи фоновими об'єктами. Типові детектори формують нестабільні результати, які залежать від кадрів, а класичні трекери часто дають хибні спрацьовування та помилки детекції. Метою цієї роботи є стабілізація детекцій БПЛА додаванням до стандартного детекції-трекінг пайплайну спеціального методу часово-просторового ф'южна, чутливого до характеру руху об'єкта.***

М е т о д и . ***Базову систему RT-DETR + ByteTrack розширено за допомогою легкого методу узгодженого з рухом згладжування (MCF), який не залежить від навчання. Цей метод: (i) агрегує історію обмежувальних рамок за останні п'ять кадрів, (ii) усереднює просторові координати та рівні довіри, (iii) штрафує знайдені об'єкти, у яких короткочасні зміни швидкості або кута перевищують емпірично обрані пороги. Жодні ознаки зовнішності чи додаткове навчання не потрібні, тож рішення працює в реальному часі на одному GPU.***

Р е з у л ь т а т и . ***Експерименти на розміченому наборі відео з БПЛА показують, що запропонований метод підвищує MOTA з 0.533 до 0.591, Precision – із 73 % до 84 %, а кількість помилок ідентифікації зменшується із 60 до 28 (покращення на 53 % у стабільності ідентифікації). Recall трохи знижується з 90 % до 76 %, що відображає свідомий компроміс: система відфільтровує нестабільні або нехарактерні для БПЛА траєкторії, щоб покращити точність і зменшити кількість хибних спрацьовувань. Оцінювання проведено на понад 1 000 відеозаписах, що забезпечує надійність результатів у різноманітних умовах польоту.***

В и с н о в к и . ***Запропонований метод ф'южна суттєво покращує як точність, так і стабільність результатів у послідовності кадрів відеовідстеження, практично не збільшуючи обчислювальні витрати. Його можна легко інтегрувати у вже наявні системи детекції та трекінгу. Метод особливо ефективний для застосування в реальному часі, хоча його продуктивність може знижуватися у випадках різких або непередбачуваних маневрів БПЛА поза межами заздалегідь визначених параметрів руху.***

К л ю ч о в і с л о в а : ***БПЛА, трекінг об'єктів, детекція об'єктів, алгоритм ф'южна, фільтр руху, RT-DETR, ByteTrack.***

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.

ISSN 2788-6603