What specific temporal features (e.g., motion trajectories, frame-to-frame consistency) most effectively improve object detection and tracking accuracy in drone video streams?

The combination of inter-frame consistency features and motion trajectory modeling provides the most effective improvement in drone video object detection and tracking accuracy.

Abstract

Drone video analysis benefits from temporal features that combine motion trajectory modeling with frame-to-frame consistency. Transformer-based temporal fusion techniques (for example, the SACME module) yielded a 5.1% gain in mean Average Precision, and methods incorporating temporal aggregation (such as temporally adaptive convolution with context queues) reported a precision of 0.786, a success rate of 0.582, and a 5% gain in Area Under the Curve—all at processing speeds up to 125.6 frames per second on standard hardware. In addition, studies using explicit motion trajectory approaches—such as GAO trajectory matching and phase correlation—demonstrated Multiple Object Tracking Accuracy figures ranging from 38.8% to 61.7% and maintained improved detection by reducing false or missed detections.

Together, these findings indicate that temporal features that enforce inter-frame consistency, when combined with accurate motion trajectory modeling, most effectively improve object detection and tracking accuracy in drone video streams.

Paper search

Using your research question "What specific temporal features (e.g., motion trajectories, frame-to-frame consistency) most effectively improve object detection and tracking accuracy in drone video streams?", we searched across over 126 million academic papers from the Semantic Scholar corpus. We retrieved the 50 papers most relevant to the query.

Screening

We screened in papers that met these criteria:

- **Video Stream Type**: Does the study focus on drone/UAV video streams (not static images)?
- **Temporal Analysis**: Does the study include analysis of temporal features or sequences in the video processing?
- **Detection Algorithms**: Does the research include object detection and/or tracking algorithms for video analysis?
- **Empirical Validation**: Does the study include quantitative performance metrics based on empirical testing?
- **Data Source**: Does the study use real-world drone footage or realistic simulated drone scenarios?
- **Implementation**: Does the research include practical implementation and testing of the proposed approach?
- **Context Specificity**: Is the research specifically focused on drone/UAV contexts rather than general video processing?
- **Algorithm Focus**: Does the study focus on algorithmic approaches to video analysis rather than purely hardware solutions?

We considered all screening questions together and made a holistic judgement about whether to screen in each paper.

Data extraction

We asked a large language model to extract each data column below from each paper. We gave the model the extraction instructions shown below for each column.

- **Temporal Feature Extraction Approach**:

  Identify and describe the specific temporal feature extraction methods used in the study:

- What techniques were used to capture temporal information?
- How were motion trajectories or frame-to-frame features analyzed?
- Specify the exact computational or algorithmic approach for temporal feature extraction.

If multiple approaches are described, list all of them. If no explicit temporal feature extraction method is used, note "No specific temporal feature extraction method reported".

Look primarily in the Methods or Experimental Design sections of the paper. Be as precise as possible, including specific algorithms, mathematical models, or computational techniques used.

- **Object Detection and Tracking Methodology**:

  Describe the primary object detection and tracking methodology:

- What type of detection algorithm was used (e.g., neural network, traditional computer vision)?
- What tracking approach was employed (e.g., Kalman filter, multi-object tracking)?
- Specify any unique or novel aspects of the detection/tracking approach.

Extract from the Methods section, focusing on the technical details of how objects were detected and tracked. If multiple methods were used, list them in order of importance or implementation.

- **Performance Metrics for Object Detection/Tracking**:

  List all quantitative performance metrics used to evaluate the object detection and tracking approach:

- Specific accuracy metrics (e.g., precision, recall, F1 score)
- Tracking-specific metrics (e.g., CLEAR MOT, MOTA, MOTP)
- Computational performance metrics (e.g., frames per second, processing time)

Extract from Results or Evaluation sections. Include numerical values if reported. If no specific metrics are used, note "No performance metrics reported".

Ensure to capture the exact metric name, its value, and the context of measurement.

- **Experimental Dataset Characteristics**:

  Describe the characteristics of the dataset used for evaluation:

- Type of video source (e.g., drone, aerial, traffic monitoring)
- Dataset name (if using a standard benchmark)
- Number of video sequences
- Total number of frames
- Types of objects tracked

Extract from Methods or Dataset Description sections. If multiple datasets were used, list them separately. If dataset details are sparse, note the available information and mark any missing details.

- **Computational Environment and Implementation Details**:

  Capture the technical implementation details:

- Hardware used (e.g., GPU type, computational platform)
- Software frameworks or libraries
- Programming languages
- Computational efficiency metrics (processing speed, resource requirements)

Look in Methods, Implementation, or Hardware sections. If multiple environments were used, list them. If no specific details are provided, note "No implementation details reported".

Results

Characteristics of Included Studies

Study	Study Focus	Temporal Features Used	Implementation Method	Performance Metrics	Full text retrieved
Yuan et al., 2024	Multi-object tracking in drone aerial videos	Visual Gaussian mixture probability hypothesis density for trajectory prediction; GAO (Gaussian, Appearance, Optimal) trajectory matching	Holistic transformer; multi-feature trajectory matching; neural network-based	Multiple Object Tracking Accuracy (MOTA): 38.8% (VisDrone), 61.7% (UAVDT)	No
Cheng et al., 2023	Multi-object tracking in UAV videos with motion imperfection	We didn't find mention of a specific temporal feature extraction method	Hybrid deblurring module; motion compensation via feature matching	We didn't find mention of performance metrics	No

Study	Study Focus	Temporal Features Used	Implementation Method	Performance Metrics	Full text retrieved
Zhou et al., 2023	Video object detection in drone videos	Transformer layers; novel temporal fusion network (SACME)	YOLOv7 with SACME addon; transformer-based temporal fusion	mean Average Precision (mAP): +5.1% (VisDrone2019-VID); 31 frames per second (fps)	No
Zhang et al., 2020	Drone video object detection with motion features	We didn't find mention of a specific temporal feature extraction method	Convolutional neural networks (CNNs) with time domain motion features	We didn't find mention of performance metrics	No
Passos et al., 2020	Detection of mosquito breeding grounds in drone aerial videos	Phase correlation for spatial alignment across frames	Deep neural networks (ResNet-101-FPN); phase correlation for tracking	F1-score: 0.78	No
Cao et al., 2022	Aerial tracking with temporal context exploitation	Online temporally adaptive convolution; temporal context queue; convolutions over temporal context	Neural network (AlexNet backbone); temporally adaptive convolution; adaptive temporal transformer	Precision: 0.786; Success: 0.582; frames per second (FPS): 125.6 (personal computer), >27 (NVIDIA Jetson AGX Xavier); Area Under the Curve (AUC): +5%; Center Location Error (CLE): threshold 20	Yes

Study	Study Focus	Temporal Features Used	Implementation Method	Performance Metrics	Full text retrieved
Xiao et al., 2023	Online UAV multi-object tracking with temporal and spatial modules	Temporal feature aggregation module (TFAM)	Neural network-based; TFAM and topology-integrated embedding module (TIEM); one-stage online multi-object tracking	Multiple Object Tracking Accuracy (MOTA): +2.2% (VisDrone2019), +2.5% (UAVDT)	No
García-Fernández and Xiao, 2023	Multi-object tracking for drone-based traffic monitoring	We didn't find mention of a specific temporal feature extraction method	Neural network for detection; Bayesian Trajectory Poisson Multi-Bernoulli Mixture (TPMBM) filter for tracking	We didn't find mention of performance metrics	No
Cohen and Medioni, 1998	Detection/tracking of moving objects in airborne video	Dynamic template from temporal coherence; graph representation of trajectories	Traditional computer vision; motion compensation; dynamic templates; graph-based tracking	We didn't find mention of performance metrics	No
Cohen and Medioni, 2003	Detection/tracking of objects in airborne video imagery	Hierarchical feature-based stabilization; normal component of residual flow; dynamic templates; graph search	Traditional computer vision; hierarchical stabilization; dynamic templates; graph search	We didn't find mention of performance metrics	No

Summary of Study Characteristics:

- Temporal feature extraction or modeling:Explicitly described in 7 out of 10 studies. Common ap-

proaches included temporal fusion networks, phase correlation, temporally adaptive convolutions, dynamic templates, and trajectory-based methods. We didn't find mention of explicit temporal feature extraction in 3 studies.

- Implementation methods:
  - Transformer-based architectures (including hybrids): 3 studies
  - Convolutional neural network-based approaches: 3 studies
  - Hybrid or other neural network-based methods: 3 studies
  - Traditional computer vision methods: 2 studies (some studies used more than one method)
- Performance metrics:
  - Reported in 5 studies
  - Most common metrics: Multiple Object Tracking Accuracy (2 studies), mean Average Precision (1), F1-score (1), Precision (1), Success (1), frames per second (2), Area Under the Curve (1), Center Location Error (1)
  - We didn't find mention of performance metrics in the other 5 studies

---

Effects

Motion Trajectory Enhancement

Study	Feature Type	Detection Improvement	Tracking Improvement	Processing Speed
Yuan et al., 2024	Visual Gaussian mixture probability hypothesis density; GAO (Gaussian, Appearance, Optimal) trajectory matching	No mention found	Multiple Object Tracking Accuracy: 38.8% (VisDrone), 61.7% (UAVDT)	No mention found
Cheng et al., 2023	Motion compensation via feature matching	No mention found	State-of-the-art (no metrics)	No mention found
Zhou et al., 2023	Transformer-based temporal fusion (SACME)	mean Average Precision: +5.1%	No mention found	31 frames per second
Zhang et al., 2020	Motion features between frames	Reduced false/missed detections (qualitative)	No mention found	No mention found
Passos et al., 2020	Phase correlation for alignment	F1-score: 0.78	No mention found	No mention found

Study	Feature Type	Detection Improvement	Tracking Improvement	Processing Speed
Cao et al., 2022	Temporally adaptive convolution; temporal context	Precision: 0.786	Success: 0.582; Area Under the Curve: +5%	125.6 frames per second (personal computer), >27 frames per second (NVIDIA Jetson AGX Xavier)
Xiao et al., 2023	Temporal feature aggregation (TFAM)	No mention found	Multiple Object Tracking Accuracy: +2.2% (VisDrone2019), +2.5% (UAVDT)	No mention found
García-Fernández and Xiao, 2023	Bayesian Trajectory Poisson Multi-Bernoulli Mixture filter	No mention found	No mention found	No mention found
Cohen and Medioni, 1998	Dynamic template, graph-based trajectory	No mention found	No mention found	No mention found
Cohen and Medioni, 2003	Hierarchical stabilization, dynamic template	No mention found	No mention found	No mention found

Summary of Feature Types and Effects:

- Feature types used:
  - Motion-based features: 2 studies
  - Transformer or temporal fusion: 1 study
  - Temporal aggregation or convolution: 2 studies
  - Probabilistic or Bayesian methods: 2 studies
  - Phase correlation: 1 study
  - Template, graph-based, or hierarchical methods: 2 studies
- Detection improvement:
  - Quantitative improvement (mean Average Precision, F1-score, precision): 3 studies
  - Qualitative improvement: 1 study
  - We didn't find mention of detection improvement in 6 studies
- Tracking improvement:
  - Quantitative improvement (Multiple Object Tracking Accuracy, Area Under the Curve, Success): 3 studies
  - Qualitative improvement: 1 study
  - We didn't find mention of tracking improvement in 6 studies
- Processing speed:
  - Reported in 2 studies (31 frames per second; 125.6 frames per second on personal computer and >27 frames per second on NVIDIA Jetson AGX Xavier)

– We didn't find mention of processing speed in 8 studies

Frame Consistency Methods

Study	Method Type	Accuracy Gain	Real-time Performance	Hardware Requirements
Yuan et al., 2024	Holistic transformer for local/global consistency	Multiple Object Tracking Accuracy: 38.8%/61.7%	No mention found	No mention found
Cheng et al., 2023	Hybrid deblurring; motion compensation	State-of-the-art (no metrics)	No mention found	No mention found
Zhou et al., 2023	Transformer layers; temporal fusion (SACME)	mean Average Precision: +5.1%	31 frames per second	No mention found
Zhang et al., 2020	Frame-to-frame motion features	Reduced false/missed detections	No mention found	No mention found
Passos et al., 2020	Phase correlation for registration	F1-score: 0.78	No mention found	No mention found
Cao et al., 2022	Temporally adaptive convolution; context queue	Precision: 0.786; Success: 0.582	125.6 frames per second (personal computer), >27 frames per second (NVIDIA Jetson AGX Xavier)	NVIDIA Jetson AGX Xavier, TITAN RTX graphics processing units
Xiao et al., 2023	Temporal feature aggregation (TFAM)	Multiple Object Tracking Accuracy: +2.2%/+2.5%	No mention found	No mention found
García-Fernández and Xiao, 2023	Bayesian filter for trajectory consistency	No mention found	No mention found	No mention found
Cohen and Medioni, 1998	Temporal coherence, dynamic template	No mention found	No mention found	No mention found
Cohen and Medioni, 2003	Hierarchical stabilization, dynamic template	No mention found	No mention found	No mention found

Summary of Frame Consistency Methods and Effects:

- Temporal and motion-based methods:
  - Transformer-based approaches (holistic, layers, fusion): 2 studies

  - Temporal fusion, aggregation, or convolution: 3 studies
  - Motion features or compensation: 2 studies
  - Phase correlation, Bayesian filtering, and temporal coherence/dynamic template: 1-2 studies each
- Accuracy gain:
  - Multiple Object Tracking Accuracy reported in 2 studies
  - mean Average Precision reported in 1 study
  - F1-score reported in 1 study
  - Precision and success metrics reported in 1 study
  - Qualitative improvements (such as "state-of-the-art" or "reduced false/missed detections") described in 2 studies
  - We didn't find mention of accuracy gain in 3 studies
- Real-time performance and hardware requirements:
  - Real-time performance reported in 2 studies (31 frames per second and 125.6 frames per second on personal computer, >27 frames per second on NVIDIA Jetson AGX Xavier)
  - We didn't find mention of real-time performance in 8 studies
  - Hardware requirements reported in 1 study (NVIDIA Jetson AGX Xavier, TITAN RTX graphics processing units)
  - We didn't find mention of hardware requirements in 9 studies

---

Combined Approaches

Several studies combine motion trajectory modeling with frame consistency methods:

- Yuan et al., 2024:Integrates holistic transformers (for local and global consistency) with GAO (Gaussian, Appearance, Optimal) trajectory matching.
- Cao et al., 2022:Combines temporally adaptive convolution with temporal transformers for both feature extraction and similarity map refinement.

These combined approaches are associated with the highest reported improvements in tracking accuracy and processing speed among the included studies, based on the available abstracts and full texts.

Implementation Considerations

Computational Efficiency

- Cao et al., 2022:Demonstrates high processing speeds (125.6 frames per second on personal computer, >27 frames per second on NVIDIA Jetson AGX Xavier) and provides detailed resource usage statistics (random access memory, graphics processing unit, central processing unit utilization).
- Zhou et al., 2023:Reports 31 frames per second for the SACME module.
- Other studies:We didn't find mention of hardware, software, or efficiency details in most other studies, which limits assessment of practical deployability.

Environmental Factors

- Benchmark datasets:The majority of studies use standard drone or aerial video benchmarks (VisDrone, UAVDT, UAV123, DTB70).
- Dataset details:We didn't find mention of the number of sequences, frames, and object types in most studies, which hinders assessment of external validity and generalizability.
- Cao et al., 2022:Provides comprehensive dataset characteristics, including sequence and frame counts.

References

Á. F. García-Fernández, and Jimin Xiao. "Trajectory Poisson Multi-Bernoulli Mixture Filter for Traffic Monitoring Using a Drone." *IEEE Transactions on Vehicular Technology*, 2023.

Changcheng Xiao, Qiong Cao, Yujie Zhong, L. Lan, Xiang Zhang, Huayue Cai, and Zhigang Luo. "Enhancing Online UAV Multi-Object Tracking with Temporal Context and Spatial Topological Relationships." *Drones*, 2023.

I. Cohen, and G. Medioni. "Detecting and Tracking Moving Objects in Video from an Airborne Observer," 1998.

———. "Detection and Tracking of Objects in Airborne Video Imagery," 2003.

Song Cheng, Meibao Yao, and Xueming Xiao. "DC-MOT: Motion Deblurring and Compensation for Multi-Object Tracking in UAV Videos." *IEEE International Conference on Robotics and Automation*, 2023.

Wesley L. Passos, E. Silva, S. L. Netto, G. Araujo, and A. Lima. "Spatio-Temporal Consistency to Detect Potential Aedes Aegypti Breeding Grounds in Aerial Video Sequences." *arXiv.org*, 2020.

Yubin Yuan, Yiquan Wu, Langyue Zhao, Yaxuan Pang, and Yuqi Liu. "Multiple Object Tracking in Drone Aerial Videos by a Holistic Transformer and Multiple Feature Trajectory Matching Pattern." *Drones*, 2024.

Yugui Zhang, Liuqing Shen, Xiaoyan Wang, and Hai-Miao Hu. "Drone Video Object Detection Using Convolutional Neural Networks with Time Domain Motion Features." *Conference on Multimedia Information Processing and Retrieval*, 2020.

Ziang Cao, Ziyuan Huang, Liang Pan, Shiwei Zhang, Ziwei Liu, and Changhong Fu. "TCTrack: Temporal Contexts for Aerial Tracking." *Computer Vision and Pattern Recognition*, 2022.

Zihao Zhou, Xianguo Yu, Xiangcheng Chen, and Yuke Li. "Self and Cross Motion Extracted in Drone Video to Improve Object Detection." *2023 IEEE International Conference on Unmanned Systems (ICUS)*, 2023.