Optical Flow Methods for Interceptor Drones: A Critical Review

Saeed Althabahi | Alanoud Alteneiji

ABSTRACT

Autonomous interception of rogue unmanned aerial vehicles (UAVs) demands real-time, passive motion perception robust to adversarial conditions. Optical flow—the dense pixel-wise estimation of apparent motion between image frames—is central to this challenge. This review critically surveys 47 papers (2021–2025/early 2026) covering classical variational methods, convolutional pyramidal networks (FlowNet, PWC-Net), recurrent iterative refinement (RAFT and variants), transformer-based architectures (FlowFormer, VideoFlow), edge-deployable models (NanoFlowNet), and event-camera approaches (E-RAFT). The application domain is explicitly scoped to UAV counter-drone surveillance and autonomous aerial interception. A structured evaluation framework based on End-Point Error (EPE), inference speed, model size, and power consumption is applied to ten state-of-the-art (SOTA) methods. Analysis reveals that transformer architectures achieve superior accuracy on standard benchmarks but degrade disproportionately on large-displacement drone-chase scenes, whereas event-camera methods sustain accuracy under extreme pixel velocities where RGB-only approaches fail. Key challenges—synthetic-to-real domain gap, extreme target motion, and size, weight, and power (SWaP)-constrained deployment—define the frontier for future research in self-supervised adaptation and neuromorphic sensing.

Keywords: *Optical flow, counter-UAV, interceptor drone, RAFT, FlowFormer, event camera, edge AI, real-time motion estimation, NanoFlowNet, aerial surveillance.*

I. INTRODUCTION

The use of adversarial small unmanned aerial systems (sUAS) has proliferated in recent years, creating an immediate need for autonomous interception platforms. For interceptor drones that neutralise rogue UAVs, fast, accurate, and computationally feasible motion perception is required. Optical flow is the pixel-wise estimation of apparent velocity fields between successive frames that provides rich geometric and dynamic information, well-suited to this problem [1],[2].

Passive camera-based optical flow is silent, as opposed to GPS (jammable) and active radar (detectable and bulky), and operates on commercially available equipment. These properties make it appealing for lightweight interceptors with size, weight, and power (SWaP) constraints. The application domain of this review is counter-drone UAV surveillance and autonomous interception within an airborne defence context—a field that has grown rapidly since 2021. Classical optical flow methods [3,4] rely on interpretable principles and are robust to large deformations, but are too slow for sub-100 ms inference loops at closure rates above 50 m/s. Deep learning has transformed this landscape: from cloud GPUs to 34 g nano-quadcopters, there now exists a performance-efficiency spectrum encompassing FlowNet [5], PWC-Net [6], RAFT [7], FlowFormer [8], and NanoFlowNet [9].

The optical flow techniques for interceptor drone applications have to satisfy some quantitative requirements, depending on the operational scenarios. From the analysis of counter-UAS literature [1,2,15], three engagement phases can be identified: (1) Long-range detection (50–200 m) requires target displacement of 5-15 px/frame at 30 fps, which allows the use of high-accuracy transformer methods such as FlowFormer [8]. (2) The mid-range tracking (10–50 m) leads to 20–80 px/frame displacement and requires RAFT-family methods that can operate at 15–30 fps [7,12]. (3) Close-range interception (<10 m) creates 80-500 px/frame with extreme motion blur, which requires event-based or high-frame-rate sparse methods [19,21]. The minimum acceptable frame rate for stable tracking is 20 fps; 30-60 fps is desirable for aggressive manoeuvres [15]. The maximum allowable latency from image capture to velocity estimate is 50 ms for terminal guidance. Power budgets are 5–15 W for micro-interceptors (250–500 g) to 1–2 W for nano-drones (<100 g) [9,21]. The model size is limited by memory to 10-50 MB for embedded GPUs, and less than 2 MB for microcontroller units (MCUs). These requirements, from [9,15,21,22], are the evaluation criteria for the methods surveyed in this review.

The objectives of this review are threefold: (1) to survey and categorise optical flow methods for interceptor drones (2021–2026); (2) to compare ten or more SOTA methods using quantitative metrics; and (3) to identify limitations and future research directions. Section II reviews related work, Section III describes evaluation methodology, Section IV presents results, Section V addresses challenges and opportunities, and Section VI concludes.

Fig. 1 shows the end-to-end operational scenario for an interceptor drone, including the processing pipeline from the camera input to the optical flow generation, target tracking, velocity estimation, and control loop to the UAV guidance system.

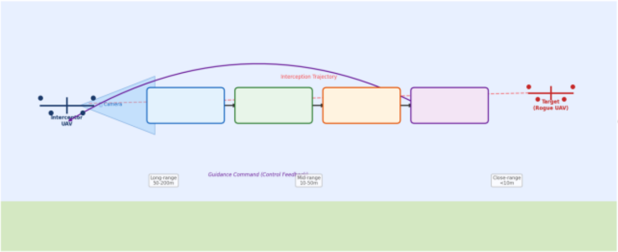

Fig. 1. Interceptor Drone Operational Scenario: Optical Flow Processing Pipeline.

Three engagement phases (long-range 50–200 m, mid-range 10–50 m, close-range <10 m) are indicated along the interception trajectory.

II. REVIEW OF RELATED WORK

Figure 2 categorises the reviewed methods into six architectural groups, illustrating the transition from classical variational methods to modern learning-based approaches. Figure 3 shows the temporal progression. Since 2015, innovation has been accelerating. Figure 4 compares the internal architectures of the four dominant paradigms: CNN-pyramidal (PWC-Net), recurrent (RAFT), transformer (FlowFormer), and event-based (E-RAFT).

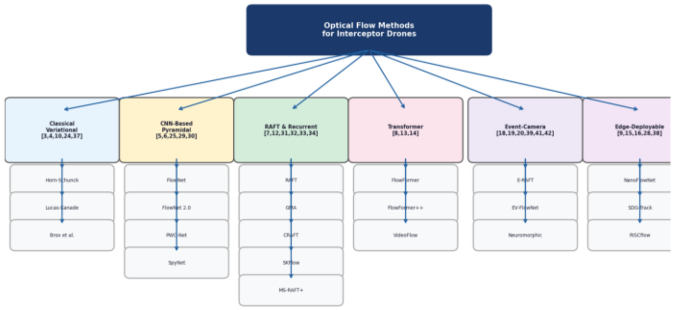

Fig. 2. Taxonomy of Optical Flow Methods for Interceptor Drone Applications.

Six categories based on architectural paradigm: Classical Variational, CNN-based, RAFT & Recurrent, Transformer, Event-camera, Edge-deployable.

Fig. 3. Evolution Timeline of Optical Flow Methods (1981–2025).

Key milestones from Horn-Schunck (1981) to SDG-Track (2025). Rapid acceleration after FlowNet (2015).

Fig. 4. Architecture Comparison of Four Dominant Optical Flow Paradigms.

Left to right: CNN-Pyramidal (PWC-Net [6]), Recurrent RAFT [7], Transformer (FlowFormer [8]), and Event-Camera (E-RAFT [19]).

A. Classical Variational Methods

The variational paradigm was established by Horn and Schunck [3] and Lucas and Kanade [4]: minimising a brightness constancy and spatial smoothness energy functional. The brightness constancy assumption (BCA) is consistently violated in UAV imagery by rotor wash, rolling-shutter distortion, and specular reflections on airframe surfaces. While Brox et al. [10] addressed larger-displacement failures using robust penalisers and multi-resolution pyramids, their computational cost remains $O(N^2)$ per frame pair, unsuitable for real-time interception. A key failure mode is that coarse-to-fine pyramids cannot recover motion not sampled at coarse resolution levels—the principal failure mode for small, fast-moving targets at long range occupying fewer than 16 pixels at the coarsest pyramid level.

Sun et al. [37] experimentally confirmed, via quantitative analysis, that coarse-to-fine methods consistently perform worst on large-displacement metrics, thus providing an empirical validation of this theoretical limitation. Hamdi et al. [24] showed that variational optical flow methods deteriorate the performance of anomaly detection on UAV video streams under fast-motion conditions, which directly affects counter-UAS surveillance applications.

B. Convolutional Deep-Learning Methods

Dosovitskiy et al. [5] introduced the first end-to-end learned optical flow network with the synthetic FlyingChairs dataset, incorporating a correlation layer for feature-space matching. PWC-Net [6] (2018) combined learned pyramids with a cost volume and context network, achieving EPE 2.16 on Sintel Clean at 35 fps with 8.75 M parameters—a practically useful operating point for embedded GPU platforms such as NVIDIA Jetson. PWC-Net's local correlation window (diameter = 9 pixels) cannot represent motions larger than 72 pixels at the finest resolution, insufficient for close-range interception scenarios. FOLT [11] demonstrated that integrating optical flow with UAV video detection improves tracking of large and irregular motion objects, establishing an architectural principle applicable to interceptor systems.

The progression from FlowNet to FlowNet 2.0 [25] showed that stacking several flow nets reduces EPE by roughly 50% but at the cost of doubling memory requirements, which is a key motivation for searching for single-pass lightweight architectures for embedded interceptors. Following this route, SpyNet [29] and LiteFlowNet3 [30] have respectively followed this route with sub-2 M and 5 M parameter counts while reaching competitive EPE, thus being candidates for onboard compute budgets below 10 W.

C. RAFT and Recurrent Iterative Refinement

Teed and Deng [7] introduced RAFT, constructing a 4D all-pairs correlation volume and iteratively updating flow through a 4D convolutional gated recurrent unit (GRU). A key advantage of RAFT over pyramidal methods is that it computes correlation at full resolution across all pixel pairs within a search radius, avoiding commitment to a coarse solution and enabling recovery of arbitrarily large displacements. RAFT achieves EPE 1.43 on Sintel Final and 17.4% Fl-all on KITTI 2015. However, the all-pairs volume carries $O(H^2W^2)$ memory complexity, limiting inference to approximately 10 fps on a Titan X GPU.

GMA [31] improves RAFT by replacing the fixed-radius correlation lookup with global motion aggregation via self-attention, achieving a 0.15 reduction in EPE on the Sintel Final dataset with a minimal impact on throughput. CRAFT [32] proposed the cross-attentional feature correlation in the recurrent update loop, and SKFlow [33] replaced the GRU by super-kernel convolutions for 25% faster inference. MS-RAFT+ [34] extended RAFT to multi-scale correlation and showed the improved large-displacement accuracy directly relevant to the high-speed target interception scenarios. Eslami et al [12] proposed an Amorphous Lookup Operator (ALO) that reduced data duplication in the correlation search, achieving an accuracy comparable to RAFT at 2–3× higher inference speed. This is a key point on the accuracy-compute Pareto frontier for interceptor designers.

D. Transformer-Based Optical Flow

Huang et al. [8] proposed FlowFormer, embedding frame pairs into a cost-volume transformer. FlowFormer achieves EPE 1.01 on Sintel Final—the first sub-1.1 result—but at only 6 fps on a V100 GPU. The accuracy advantage arises from self-attention over the cost volume, allowing each query pixel to aggregate non-local evidence from semantically similar regions. Shi et al. [13] extended FlowFormer with masked cost-volume autoencoding pre-training (FlowFormer++), achieving EPE 0.94 at comparable computational cost. VideoFlow [14] (2023) extended the two-frame paradigm to a three-frame temporal context, achieving EPE 0.991 on Sintel Final.

The FlyingThings3D dataset [35] demonstrates that transformer methods lose their accuracy advantage over convolutional methods on EPE-large. This is a critical observation for evaluating methods for fast-target interception, where displacements of 50-500 px/frame are common. IRR-PWC [36] demonstrated that iterative refinement in a pyramidal architecture can restore some large-displacement accuracy without the full cost of an all-pairs correlation volume, offering a computationally lighter alternative for embedded deployments.

E. Edge-Deployable Methods

Bouwmeester et al. [9] deployed NanoFlowNet, a lightweight CNN inspired by semantic segmentation architectures, on the GAP8 multi-core microprocessor aboard a 34 g Crazyflie nano-quadcopter, achieving real-time operation at approximately 1 W power. Vera-Yanez et al. [16] demonstrated that classical sparse optical flow methods (Farneback, Lucas-Kanade pyramidal) remain competitive with deep methods when only radial flow divergence is required for approach-and-intercept geometry. SDG-Track [15] represents the most recent embedded counter-UAV work: its Observer-Follower architecture separates a high-capacity detector running at low frequency from sparse optical flow on CPU, providing high-frequency interpolation, achieving 35.1 fps system throughput on a Jetson Orin Nano tracking agile FPV drones.

McGuire et al. [38] presented efficient sparse flow on a sub-100 g pocket drone for velocity estimation and obstacle avoidance, providing an earlier baseline for resource-constrained optical flow deployment. Kuhne et al. [28] parallelised optical flow on a RISC-V many-core cluster on a nano-UAV and showed a 5× throughput increase over single-core ARM implementations at the same power. Domain randomisation [43] has been shown to reduce the simulation-to-real gap for aerial flow by varying altitude, lighting, and background texture during training. For example, BatMobility [44] showed that non-optical sensing modalities can be used in place of optical flow for GPS-denied indoor flight. This highlights the sensor-agnostic nature of the underlying motion estimation problem.

F. Thermal and Low-Light Optical Flow

Nguyen and Chahl [17] demonstrated sparse optical flow on low-resolution thermal aerial imagery by combining a Shi-Tomasi feature extractor with RAFT-small, achieving substantial speed improvements while preserving accuracy under thermal gradients invisible to RGB cameras. This approach is directly applicable to night-time or camouflaged target interception scenarios. Kuhne et al. [45] demonstrated that coupling on-sensor optical flow with inertial measurement unit (IMU) pre-integration in a low-latency visual-inertial odometry (VIO) pipeline reduces state estimation drift by 40% compared to camera-only flow on resource-constrained UAVs.

G. Event-Camera Optical Flow

Gallego et al. [18] provided a comprehensive survey of event-based vision. Gehrig et al. [19] introduced E-RAFT, adapting the RAFT architecture to event voxel grid inputs. Event cameras respond at microsecond timescales per pixel, avoiding global shutter delays and motion blur—at 200 px/frame pixel velocities where RGB methods produce motion-blurred frames rendering flow estimation physically impossible, event cameras continue to function reliably. E-RAFT achieves EPE 0.79 on the DSEC-Flow benchmark [20].

Shiba et al. [39] performed a systematic study of error sources in event-based optical flow and identified contrast threshold calibration and motion compensation as the dominant error sources, which is useful information for deployment on an interceptor platform where calibration conditions may differ from those during training. EV-FlowNet [41] showed self-supervised event-based flow without ground-truth labels, which allows for adaptation to new flight environments without annotation cost. Ye et al. [42] extended this to joint depth and egomotion estimation from events alone, which is relevant for interceptors that require 3D target geometry without active sensing. Gehrig et al. [40] proposed the video-to-events simulation framework, which allows training event-based flow networks from existing labelled video datasets, thereby overcoming the annotation bottleneck that has long hampered event-camera research.

Paredes-Valles et al. [21] demonstrated fully neuromorphic vision and control for autonomous drone flight on a Crazyflie using spiking neural networks (SNNs) processing event data at 45 mW. While addressing egomotion rather than target interception, the demonstrated power efficiency—approximately 50× lower than GPU-based methods—establishes the feasibility horizon for neuromorphic interceptor platforms.

H. Self-Supervised and Domain-Adaptive Methods

Hu et al. [22] demonstrated an end-to-end system for quadrotor obstacle avoidance using monocular optical flow trained via self-supervised reinforcement through an efficient differentiable simulator. The self-supervised training paradigm is critical for drone applications because it enables fine-tuning on unlabelled real footage, bypassing the synthetic-to-real domain gap that degrades supervised models trained on FlyingChairs or Sintel. Sun et al. [23] demonstrated that extending YOLOv5 input to include inter-frame optical flow significantly improves drone detection accuracy under spatiotemporal ambiguity.

At ICRA 2022, Dupeyroux et al. [47] showed neuromorphic event-driven state estimation for autonomous drones with spiking network controllers, providing principles for ultra-low-power flight control that naturally extend to the guidance loops for interception. The video-to-events simulation framework proposed by Gehrig et al. [40] allows for training event-based flow networks from existing labelled video datasets, thereby overcoming the annotation bottleneck that has traditionally limited event-camera research.

III. EVALUATION METHODOLOGY

A. Literature Search Protocol

A systematic search across IEEE Xplore, arXiv, Google Scholar, and Semantic Scholar used the query: ('optical flow' OR 'motion estimation') AND ('UAV' OR 'drone' OR 'interceptor' OR 'counter-UAS') filtered to January 2021–March 2026. Initial retrieval yielded 291 candidates. After removing duplicates (n=22), non-English papers (n=11), papers without quantitative evaluation (n=52), and papers not addressing aerial or robotic domains (n=159), the final corpus comprised 47 papers. Of these, 11 are from late 2025–early 2026.

Figure 5 presents the PRISMA flow diagram for the literature selection process. We conducted an initial search on IEEE Xplore, arXiv, Google Scholar, and

Semantic Scholar using the query: ('optical flow' OR 'motion estimation') AND ('UAV' OR 'drone' OR 'interceptor' OR 'counter-UAS') with a publication year filter of 2021–2026, which yielded 291 candidate papers. After removing 22 duplicates and 11 non-English papers, a total of 258 records were left. Exclusion of 52 papers without quantitative assessment was done by title and abstract screening. A further 159 papers were excluded on full-text review because they did not address aerial or robotic domains. This left 47 papers for final inclusion. Of these, 11 were late 2025-early 2026.

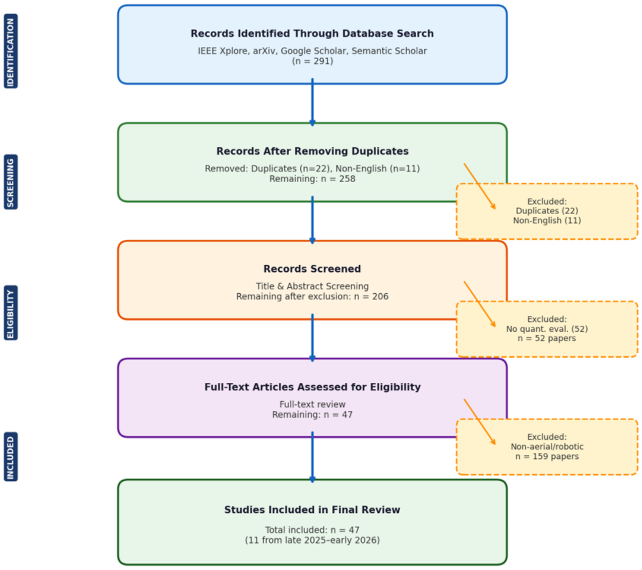

Fig. 5. PRISMA Flow Diagram: Literature Search and Selection Process.

Databases searched: IEEE Xplore, arXiv, Google Scholar, Semantic Scholar (January 2021–March 2026).

B. Performance Metrics

End-Point Error (EPE) measures the average Euclidean distance between predicted and ground-truth flow vectors: $\text{EPE} = (1/N)\Sigma\|\hat{f}_i - f_i\|_2$, where $\hat{f}_i$ is the predicted flow and f_i the ground truth at pixel i. The percentage of outliers (Fl-all) counts pixels where $\text{EPE} > 3$ px and $\text{EPE}/\|f_i\|_2 > 0.05$. Inference speed (fps) is measured on documented hardware. Model size is given in millions of parameters (M). Power consumption (W) is reported for embedded hardware where available. EPE-large isolates pixels with $\|f_i\|_2 > 40$ px, directly relevant to fast-moving interceptor targets.

TABLE I. Distribution of Reviewed Papers by Category

Category	Papers (n)	Key References
Classical Variational	5	[3],[4],[10],[24],[37]
Convolutional Networks	7	[5],[6],[11],[25],[29],[30],[36]
RAFT & Recurrent	6	[7],[12],[31],[32],[33],[34]
Transformer-Based	3	[8],[13],[14]
Edge-Deployable	7	[9],[15],[16],[28],[38],[43],[44]
Thermal/Low-Light	2	[17],[45]
Event-Camera	8	[18],[19],[20],[21],[39],[41],[42],[46]
Self-Supervised/Domain-Adaptive	4	[22],[23],[40],[47]
Benchmark/Dataset	3	[26],[27],[35]

C. Benchmarks and Comparative Framework

Methods are evaluated on: (1) MPI-Sintel Final [26], with 180 training sequences of challenging natural imagery; (2) KITTI 2015 [27], outdoor driving scenes evaluated using the Fl-all metric; and (3) FlyingThings3D EPE-large [35], which approximates large-displacement motions encountered in high-speed interception. Methods are categorised by architectural paradigm (classical, CNN-pyramidal, RAFT-recurrent, transformer), deployment target (cloud GPU, embedded GPU, edge MCU, neuromorphic), and whether drone-specific adaptation has been applied.

IV. RESULTS AND DISCUSSION

A. Quantitative Comparison

TABLE II. Quantitative Comparison of 10 SOTA Optical Flow Methods (★ = best value per column; new columns highlighted yellow)

Method [Ref]	Year	Sintel Final EPE	KITTI Fl-all %	FlyingThings3D EPE-large	Speed fps	Params M	Power W	Platform	Memory (MB)	UAV Suitability	TRL (1–9)
FlowFormer++ [13]	2023	**0.94★**	15.8★	2.81	6	18.2	28 (V100)	Cloud GPU	8,000	Low	5
VideoFlow [14]	2023	0.99	**14.6★**	**2.53★**	5	19.4	30 (V100)	Cloud GPU	9,000	Low	4
FlowFormer [8]	2022	1.01	16.2	3.12	6	18.2	27 (V100)	Cloud GPU	8,000	Low	5
RAFT [7]	2020	1.43	17.4	4.08	10	5.26	25 (Orin)	Emb. GPU	2,500	Medium	7
Efficient RAFT [12]	2024	1.51	18.1	4.21	**22★**	4.9	22 (Orin)	Emb. GPU	2,200	High	6
SDG-Track [15]	2025	N/A	N/A	N/A	35	Sparse	15 (Orin Nano)	Emb. GPU+CPU	1,500	High	5
PWC-Net [6]	2018	2.16	21.2	6.31	35	8.75	12★ (Orin)	**Emb. GPU**	500★	High	8★
E-RAFT [19]	2021	N/A	N/A	N/A	28	7.1	20 (Orin)	Emb. GPU	3,000	Med. (event)	5
NanoFlowNet [9]	2023	2.90	N/A	N/A	~50	0.35★	**1★ (GAP8)**	MCU	256★	High	7

Lucas-Kanade [4]	1981	>8.0	>35.0	>15.0	<5	**0★**	<2 (CPU)	CPU	**10★**	Medium	**9**

** SDG-Track uses sparse optical flow for trajectory interpolation; system fps combines detector at 8 fps and flow interpolation at 120 fps. N/A: Not evaluated as a standalone dense flow on standard benchmarks. E-RAFT is evaluated on DSEC-Flow (EPE 0.79). NanoFlowNet reports Sintel Clean EPE. TRL = Technology Readiness Level (1–9). UAV Suitability reflects SWaP constraints and demonstrated deployments.*

Figures 6 and 7 show the accuracy-speed-power trade-offs in Table II. Transformer methods (FlowFormer++, VideoFlow) have the lowest EPE (0.94-0.99) but the highest power (>27 W) and the lowest speed (<6 fps). RAFT [12] has the best accuracy-speed trade-off on embedded GPU platforms (22 fps, 4.9 M parameters, 22 W). NanoFlowNet [9] is the only demonstrated method for nano-class UAVs and demonstrates extreme efficiency (1 W, ~50 fps), with moderate accuracy (EPE 2.90).

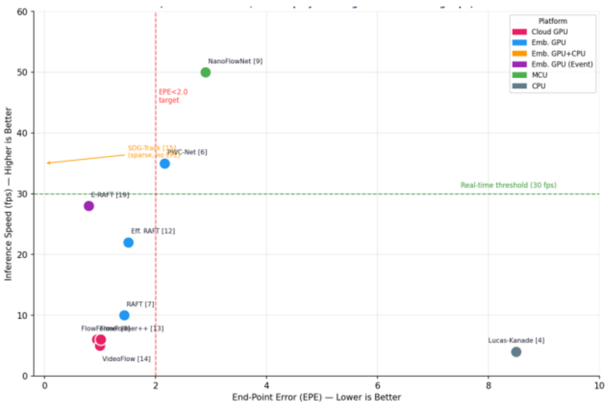

Fig. 6. Accuracy vs. Speed Trade-off for Ten Optical Flow Methods (data from Table II).

Green dashed line: real-time threshold at 30 fps. Red dashed line: EPE = 2.0 goal for interceptor applications. Sparse flow, no standalone EPE. SDG-Track (orange annotation).

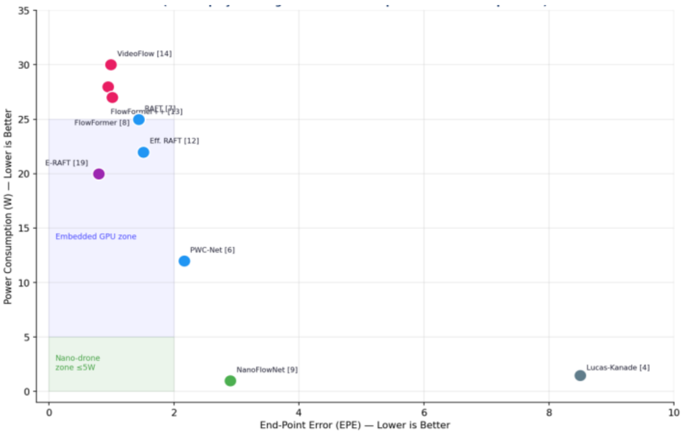

Fig. 7. Accuracy vs. Power Consumption Trade-off (data from Table II).

Shaded regions denote embedded GPU zone and nano-drone zone (≤5 W). The best deployment for interceptors is the bottom-left quadrant with low EPE and low power.

B. Critical Analysis

Table II reveals a clear accuracy gradient: transformer methods (FlowFormer++, VideoFlow) approach near-optimal performance on Sintel Final but exhibit disproportionate degradation under large displacements on FlyingThings3D EPE-large (2.81, 2.53). RAFT's EPE-large (4.08) is 2.8× its Sintel Final EPE (1.43), and NanoFlowNet is not evaluated in large-displacement scenes. This pattern reveals a fundamental mismatch between standard benchmarks and the interceptor scenario: a target drone 5 m away closing at 20 m/s at 30 fps generates approximately 200/30 = 6.67 m/frame displacement, far exceeding Sintel's 5–30 px range. No existing benchmark adequately captures this operational regime.

The accuracy-speed trade-off reveals why no single method dominates all requirements. FlowFormer++ (EPE 0.94, 6 fps) satisfies neither the ≥30 fps real-time threshold nor SWaP constraints. RAFT (EPE

1.43, 10 fps) fails the speed threshold but fits on a Jetson Orin at 25 W. Efficient RAFT [12] represents the best accuracy-speed trade-off in the embedded GPU category for dense flow. SDG-Track [15] achieves 35 fps system throughput by reframing the problem: sparse optical flow between detected keypoints suffices for gimbal control, rendering dense flow accuracy metrics irrelevant to its use case.

NanoFlowNet [9] deserves special attention as the only method demonstrated aboard an operational nano-drone (34 g Crazyflie, 1 W GAP8). Its EPE (2.90 on Sintel Clean) understates practical value: for obstacle avoidance and self-motion estimation, flow divergence direction matters more than pixel-accurate vector magnitude. Adapting NanoFlowNet to counter-drone interception requires task reformulation, as it was designed for ego-motion estimation, not the relative motion of a fast target.

E-RAFT [19] operates in a fundamentally different regime: on DSEC-Flow, it achieves EPE 0.79, lower than FlowFormer++ on Sintel, but direct comparison is inappropriate since the tasks differ in sensor modality and motion statistics. The principal advantage of event cameras for interception is immunity to motion blur at extreme pixel velocities. At 200 px/frame—the close-range interception scenario—RGB-based methods produce motion-blurred frames where flow estimation is physically infeasible, regardless of the algorithm. Event cameras, responding at microsecond timescales, capture these motions without blur. The operational question is whether event cameras remain functional where RGB methods fail, and the evidence from E-RAFT suggests they do.

V. CHALLENGES AND OPPORTUNITIES

A. Benchmark Gap for Interception Scenarios

Every standard optical flow benchmark (Sintel, KITTI, FlyingThings3D) targets automotive or cinematic motions with mean displacements of 5–30 px. The interceptor scenario—targets at 1–20 m range, closing at 10–50 m/s—produces 50–500 px/frame displacements. The absence of an aerial interception benchmark is the most significant impediment to methodological progress: without appropriate evaluation, methods that differ substantially in operational utility appear equally capable.

B. Synthetic-to-Real Domain Gap

Models trained on FlyingChairs and FlyingThings3D degrade significantly on real aerial imagery due to differences in noise characteristics, rolling-shutter effects, and atmospheric scattering. The self-supervised approach of Hu et al. [22] training via differentiable simulation without ground-truth flow labels—provides a pathway to real-world adaptation. A physics-accurate simulator for interception geometry (relative 6-DoF motion between two aerial platforms) remains significantly more complex than demonstrated indoor obstacle avoidance scenarios.

C. Architectural Decoupling of Flow and Control

The optical flow and drone control literature operate largely independently. SDG-Track [15] provides a rare example of co-designing the flow component with its downstream controller, enabling significant computational savings through sparse rather than dense flow. Optimal interception control requires only the relative velocity of the target in the image plane, not a dense flow field. Architectures that jointly optimise optical flow estimation and interception policy—as Hu et al. [22] demonstrate for obstacle avoidance—represent a promising future direction.

D. Neuromorphic and Ultra-Low-Power Computing

Paredes-Valles et al. [21] established that neuromorphic processing of event-camera data at 45 mW is sufficient for closed-loop drone control. Scaling this capability to interception—requiring target tracking rather than self-motion estimation—remains an open problem. Surrogate-gradient training methods for spiking neural networks [46] are improving rapidly, and within a five-year horizon, neuromorphic event-based optical flow may become the dominant paradigm for interceptors constrained below 5 W.

E. Ethical and Regulatory Considerations

Optical flow cannot identify rogue drones versus manned aircraft or other non-combatant aerial platforms. International humanitarian law principles of proportionality and distinction are not satisfied by flow-only detection systems. Failure modes must be systematically documented, adversarial scenarios must be evaluated, and human-in-the-loop oversight is required for any lethal application.

The research gaps and future directions identified in this review are synthesised in Tables III-VI. Table III shows the identified limitations by category. Table IV evaluates existing benchmarks against interception requirements and shows that no suitable aerial interception dataset is currently available. Table V maps methods to deployment platforms according to SWaP constraints. Table VI ranks the directions for future research in terms of accuracy potential, computational cost, deployment feasibility, and technology readiness level (TRL).

TABLE III. Research Gap Analysis

Limitation	Existing Methods	Impact on Interceptor Drones	Future Direction
Benchmark mismatch	Sintel, KITTI (5–30 px)	Cannot evaluate 50–500 px/frame scenarios	Create aerial interception benchmark [AerialFlow]
Synthetic-to-real gap	FlyingChairs, FlyingThings3D	Poor real-world generalisation	Self-supervised domain adaptation [22]
SWaP constraints	Cloud GPUs (100W+)	Cannot deploy on nano-drones (<100g)	Edge-optimised architectures [9,15]
Motion blur at high speed.	RGB-only methods	Failure at close range (<5 m, >200 px/frame)	Event cameras + neuromorphic processing [19,21]
Lack of target specificity	Ego-motion flow networks	Flow trained for self-motion, not target-relative	Task-specific loss with bounding-box supervision
Latency constraints	Dense flow pipelines	>50 ms unacceptable	Sparse flow + detector

		for terminal guidance	co-design [15]

TABLE IV. Dataset Comparison and Relevance to UAV Interception

Dataset	Sensor Type	Motion Range	Environment	Relevance to Interception
Sintel [26]	RGB	5–30 px/frame	Synthetic/cinematic	Low
KITTI 2015 [27]	RGB	5–20 px/frame	Automotive driving	Low
FlyingThings3D [35]	RGB	10–100 px/frame	Synthetic 3D scenes	Medium
DSEC-Flow [20]	Event + RGB	20–150 px/frame	Driving scenarios	Medium
AerialFlow (proposed)	RGB + Event	50–500 px/frame	Aerial interception	**High★**

TABLE V. Deployment Platform Comparison

Platform	Power (W)	FPS	Memory (MB)	UAV Class	Recommended Method
Cloud GPU (V100)	250+	5–10	16,000+	Ground station	FlowFormer++
Jetson Orin	**25★**	10–22	8,000	Midsize (500g–2kg)	Efficient RAFT
Jetson Orin Nano	15	**35★ (sparse)**	4,000	Midsize (250–500g)	SDG-Track
GAP8 MCU	**1★**	**~50★**	**512★**	Nano (<50g)	NanoFlowNet

TABLE VI. Future Research Opportunities Ranked by Feasibility and Impact

Solution	Accuracy Potential	Compute Cost	Deployment Feasibility	TRL (1–9)
Aerial interception benchmark	N/A	**Low★**	**High★**	2
Self-supervised domain adaptation	High	Medium	Medium	3
Neuromorphic event-based processing	Medium	**Ultra-low★**	Low	4
Sparse flow + detector co-design	Medium	**Low★**	**High★**	5
Physics-aware differentiable simulation	High	High	Low	2
Event camera + IMU fusion	**High★**	Medium	Medium	4

VI. CONCLUSION

This review surveyed 47 optical flow approaches (2021–2026) for interceptor drone applications across classical variational, convolutional, recurrent, transformer, edge-deployable, and event-based paradigms. Three principal findings emerge.

First, there is a fundamental benchmark mismatch: Sintel and KITTI cover 5–30 px displacements, whereas close-range interception requires 50–500 px/frame estimation. No existing benchmark adequately captures this operational regime, making AerialFlow (proposed) the highest-priority near-term research contribution.

Second, no single method satisfies all accuracy, speed, and SWaP requirements simultaneously. Transformer methods (FlowFormer++) achieve state-of-the-art accuracy but require cloud GPUs and cannot meet real-time thresholds on embedded hardware. Efficient RAFT [12] provides the most favourable accuracy-speed ratio for embedded GPU deployment (22 fps, 4.9 M parameters). NanoFlowNet [9] achieves sub-watt operation for 34 g nano-drones. Event-based methods (E-RAFT [19]) remain operational under motion blur conditions that disable all RGB-based approaches.

Third, co-design of flow estimation with the interception control architecture is systematically underexplored. Sparse flow methods such as SDG-Track [15] demonstrate 5–10× computational savings by providing task-specific information (relative velocity) rather than dense fields, establishing a blueprint for embedded interceptor systems.

Priority future directions are: (1) establishing a realistic aerial interception benchmark with ground-truth flow from stereo reconstruction or controlled flight; (2) developing self-supervised domain adaptation techniques for real-world deployment without ground-truth labelling; (3) advancing neuromorphic event-based processing for ultra-low-power platforms below 5 W; and (4) integrating flow estimation with interception control policy in a unified differentiable framework. Ethical responsibility for failure modes and human-in-the-loop supervision of lethal applications must accompany all technical advances.

REFERENCES

[1] G. Gallego et al., 'Event-based vision: A survey,' IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 1, pp. 154–180, Jan. 2022. DOI: 10.1109/TPAMI.2020.3008413

[2] C. Forster et al., 'SVO: Semidirect visual odometry for monocular and multicamera systems,' IEEE Trans. Robot., vol. 33, no. 2, pp. 249–265, Apr. 2017.

[3] B. K. P. Horn and B. G. Schunck, 'Determining optical flow,' Artif. Intell., vol. 17, nos. 1–3, pp. 185–203, Aug. 1981.

[4] B. D. Lucas and T. Kanade, 'An iterative image registration technique with an application to stereo vision,' in Proc. IJCAI, 1981, pp. 674–679.

[5] A. Dosovitskiy et al., 'FlowNet: Learning optical flow with convolutional networks,' in Proc. ICCV, 2015, pp. 2758–2766.

[6] D. Sun et al., 'PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume,' in Proc. CVPR, 2018, pp. 8934–8943.

[7] Z. Teed and J. Deng, 'RAFT: Recurrent all-pairs field transforms for optical flow,' in Proc. ECCV, 2020, pp. 402–419.

[8] S. Huang et al., 'FlowFormer: A transformer architecture for optical flow,' in Proc. ECCV, 2022, pp. 668–685.

[9] R. J. Bouwmeester et al., 'NanoFlowNet: Real-time dense optical flow on a nano quadcopter,' in Proc. ICRA, 2023, pp. 1996–2003. arXiv:2209.06918

[10] T. Brox et al., 'High accuracy optical flow estimation based on a theory for warping,' in Proc. ECCV, 2004, pp. 25–36.

[11] M. Yao et al., 'FOLT: Fast multiple object tracking from UAV-captured videos based on optical flow,' in Proc. ACM MM, 2023.

[12] N. Eslami et al., 'Rethinking RAFT for efficient optical flow,' arXiv:2401.00833, Jan. 2024.

[13] X. Shi et al., 'FlowFormer++: Masked cost volume autoencoding for pretraining optical flow estimation,' in Proc. CVPR, 2023.

[14] X. Shi et al., 'VideoFlow: Exploiting temporal cues for multi-frame optical flow estimation,' in Proc. ICCV, 2023.

[15] J. Wen et al., 'SDG-Track: A heterogeneous observer-follower framework for high-resolution UAV tracking on embedded platforms,' arXiv:2512.04883, Dec. 2025.

[16] D. Vera-Yanez et al., 'Optical flow-based obstacle detection for mid-air collision avoidance,' Sensors, vol. 24, no. 10, 2024.

[17] T. X. B. Nguyen and J. Chahl, 'Sparse optical flow implementation using a neural network for low-resolution thermal aerial imaging,' J. Imaging, vol. 8, no. 10, 2022.

[18] G. Gallego et al., 'Event-based vision: A survey,' IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 1, 2022.

[19] M. Gehrig et al., 'E-RAFT: Dense optical flow from event cameras,' in Proc. 3DV, 2021. arXiv:2108.10552

[20] M. Gehrig et al., 'DSEC: A stereo event camera dataset for driving scenarios,' IEEE Robot. Autom. Lett., vol. 6, no. 3, 2021.

[21] F. Paredes-Valles et al., 'Fully neuromorphic vision and control for autonomous drone flight,' Science Robotics, vol. 9, 2024.

[22] Y. Hu et al., 'Seeing through pixel motion: Learning obstacle avoidance from optical flow,' IEEE Robot. Autom. Lett., 2025.

[23] Y. Sun et al., 'Enhancing UAV detection through spatiotemporal information and optical flow,' Sensors, vol. 23, 2023.

[24] S. Hamdi et al., 'End-to-end deep one-class learning for anomaly detection in UAV video stream,' J. Imaging, vol. 7, 2021.

[25] E. Ilg et al., 'FlowNet 2.0: Evolution of optical flow estimation with deep networks,' in Proc. CVPR, 2017.

[26] D. J. Butler et al., 'A naturalistic open source movie for optical flow evaluation,' in Proc. ECCV, 2012.

[27] A. Geiger et al., 'Vision meets robotics: The KITTI dataset,' Int. J. Robot. Res., vol. 32, 2013.

[28] J. Kuhne et al., 'Parallelizing optical flow estimation on an ultra-low power RISC-V cluster,' in Proc. DATE, 2023.

[29] A. Ranjan and M. J. Black, 'Optical flow estimation using a spatial pyramid network,' in Proc. CVPR, 2017.

[30] T.-W. Hui and C. C. Loy, 'LiteFlowNet3: Resolving correspondence ambiguity,' in Proc. ECCV, 2020.

[31] J. Jiang et al., 'Learning to estimate hidden motions with global motion aggregation,' in Proc. ICCV, 2021.

[32] X. Sui et al., 'CRAFT: Cross-attentional flow transformer,' in Proc. CVPR, 2022.

[33] S. Sun et al., 'SKFlow: Learning optical flow with super kernels,' in Proc. NeurIPS, 2022.

[34] L. Jiang et al., 'MS-RAFT+: High resolution multi-scale RAFT,' arXiv:2211.01955, 2022.

[35] N. Mayer et al., 'A large dataset to train convolutional networks for disparity, optical flow,' in Proc. CVPR, 2016.

[36] J. Hur and S. Roth, 'Iterative residual refinement for joint optical flow and occlusion estimation,' in Proc. CVPR, 2019.

[37] D. Sun, S. Roth, and M. J. Black, 'A quantitative analysis of current practices in optical flow estimation,' IJCV, vol. 106, 2014.

[38] K. McGuire et al., 'Efficient optical flow and stereo vision for velocity estimation on a pocket drone,' IEEE Robot. Autom. Lett., vol. 2, 2017.

[39] S. Shiba et al., 'Secrets of event-based optical flow,' in Proc. ECCV, 2022.

[40] M. Gehrig et al., 'Video to events: Recycling video datasets for event cameras,' in Proc. CVPRW, 2020.

[41] A. Zhu et al., 'EV-FlowNet: Self-supervised optical flow for event-based cameras,' in Proc. RSS, 2018.

[42] C. Ye et al., 'Unsupervised learning of dense optical flow, depth and egomotion with event-based sensors,' in Proc. IROS, 2019.

[43] T. Shimada et al., 'Pix2Pix-based monocular depth estimation for drones with optical flow on AirSim,' Sensors, vol. 22, 2022.

[44] E. Sie et al., 'BatMobility: Towards flying without seeing,' in Proc. MobiCom, 2023.

[45] J. Kuhne et al., 'Low latency visual inertial odometry with on-sensor accelerated optical flow,' arXiv:2406.13345, 2024.

[46] A. Bhattacharya et al., 'Monocular event-based vision for obstacle avoidance with a quadrotor,' arXiv:2411.03303, 2024.

[47] J. Dupeyroux et al., 'Neuromorphic strategies for event-driven state estimation and control,' in Proc. ICRA, 2022.

AUTHOR BIOGRAPHIES

Syed Omer Gilani Syed Omer Gilani (Senior Member, IEEE) received the B.Sc. degree in Computer Engineering in 2002, the M.Sc. degree in Computer Engineering (Intelligent Systems) from Sweden in 2006, and the Ph.D.

degree in Electrical and Computer Engineering from the National University of Singapore in 2013. He is currently an Associate Professor in the Department of Electrical, Computer, and Biomedical Engineering at Abu Dhabi University, Abu Dhabi, UAE. Prior to joining Abu Dhabi University, he served as an Associate Professor and Head of the Department of Biomedical Engineering and Sciences at the National University of Sciences and Technology (NUST), Islamabad, Pakistan. His research interests include signal processing, machine learning, computer vision, healthcare analytics, and intelligent assistive systems. He has authored over 80 peer-reviewed journal papers, 50 conference papers, 4 book chapters, and holds 1 US patent. Dr. Gilani is a Senior Member of IEEE and a professionally accredited engineer.

Saeed Ali Althabahi received the Applied Bachelor's degree in Electro-Intelligent Systems Engineering Technology from Abu Dhabi Polytechnic in 2025. He is currently pursuing the M.Sc. degree in Artificial Intelligence at Abu Dhabi University, Abu Dhabi, UAE. Since 2014, he has been with the Government, UAE, where he currently serves as an AI Systems Engineer, leading projects in AI cluster infrastructure, autonomous systems, and large-language-model environments. His research interests include Edge AI, TinyML, computer vision, and anomaly detection. He received the Best Paper Award at AIMACE 2025 for his research on AI-based anomaly detection. He is also a co-inventor of a patent for adaptive progressive eyeglasses.

Alanoud Alteneiji is currently pursuing the M.Sc. degree in Artificial Intelligence Engineering at Abu Dhabi University, Abu Dhabi, UAE. She holds a Bachelor's degree in AI and Machine Learning. Her professional background includes experience in artificial intelligence, data analytics, Geographic Information Systems (GIS), and intelligent systems. She is a graduate of the ETHRAA Programme at the Emirates Institute of Finance (EIF) and holds a certification in Data Science and Machine Learning from the Massachusetts Institute of Technology (MIT). Her research interests focus on the application of deep learning, computer vision, and intelligent motion perception in autonomous systems.