iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** III **Month of publication:** March 2026

DOI: https://doi.org/10.22214/ijraset.2026.78581

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

Emerging Frontiers in Real-Time Moving Object Detection

Leela B. Kamkhede[1], Kavita G. Giri[2]
[1]Lecturer, Electronics Department, Government Polytechnic, Ch.Sambhajinagar
[2]Lecturer, Electronics Department, Government Residential Womens Polytechnic, Latur

Abstract: Moving Object Detection (MOD) is the fundamental backbone of autonomous systems, urban surveillance, and industrial robotics. This paper explores the transition from traditional background subtraction to the current "Edge-First" era dominated by YOLO26 and Real-Time Detection Transformers (RT-DETR). We analyze key innovations including NMS-free inference, temporal context modeling via Vision Transformers (ViTs), and the integration of Small-Target-Aware Label Assignment (STAL) to address long-standing challenges in dynamic environments.

I. INTRODUCTION

Traditional MOD relied heavily on pixel-level change detection (e.g., Frame Differencing, GMM). However, these methods frequently fail in "non-ideal" scenarios—dynamic backgrounds (waving trees), illumination shifts, and camera jitter. In 2026, the paradigm has shifted toward Unified End-to-End Learning, where motion and identity are processed simultaneously within a single neural pipeline.

II. CURRENT STATE-OF-THE-ART ARCHITECTURES

A. The YOLO26 Evolution

The release of **YOLO26** in late 2025 marked a significant architectural simplification aimed at edge deployment. Key enhancements include:

- NMS-Free Inference: By removing Non-Maximum Suppression (NMS), YOLO26 eliminates a traditional post-processing bottleneck, reducing latency by approximately 43% on CPU-side execution.
- ProgLoss & STAL: *Progressive Loss Balancing* prevents large objects from dominating the loss function, while *Small-Target-Aware Label Assignment* ensures tiny or occluded moving objects (critical in drone surveillance) are detected with higher recall.
- MuSGD Optimizer: A hybrid optimizer that stabilizes training convergence, particularly for multi-task models handling detection and segmentation simultaneously.

B. 2.2 Real-Time Vision Transformers (RT-ViT)

While CNNs excel at local feature extraction, Vision Transformers (ViTs) like **RF-DETR** and **DINOv3** provide superior global context.

- Temporal Attention: Unlike single-frame detectors, 2026 models utilize "Multi-frame Context" (as seen in recent SPIE 2026 findings), allowing the model to "remember" an object’s trajectory, significantly reducing flickering in low-light conditions.
- Quantization (2-4 bit): Modern ViTs now support ultra-low-bit quantization, allowing complex attention mechanisms to run on battery-powered IoT devices.

III. KEY TECHNICAL COMPARISON

Feature	Traditional (GMM/Optical Flow)	Deep Learning (YOLOv8/11)	Next-Gen (YOLO26 / RT-DETR)
Speed	Very High (Simple)	High (GPU dependent)	Ultra-High (NMS-free / Edge-optimized)
Accuracy	Low (Noise sensitive)	High	SOTA (Context-aware)
Small Objects	Poor	Moderate	Excellent (via STAL/Multi-scale)
Robustness	Fails in dynamic BG	Good	Exceptional (Adaptive modeling)

IV. CRITICAL CHALLENGES & SOLUTIONS

A. Dynamic Background Suppression

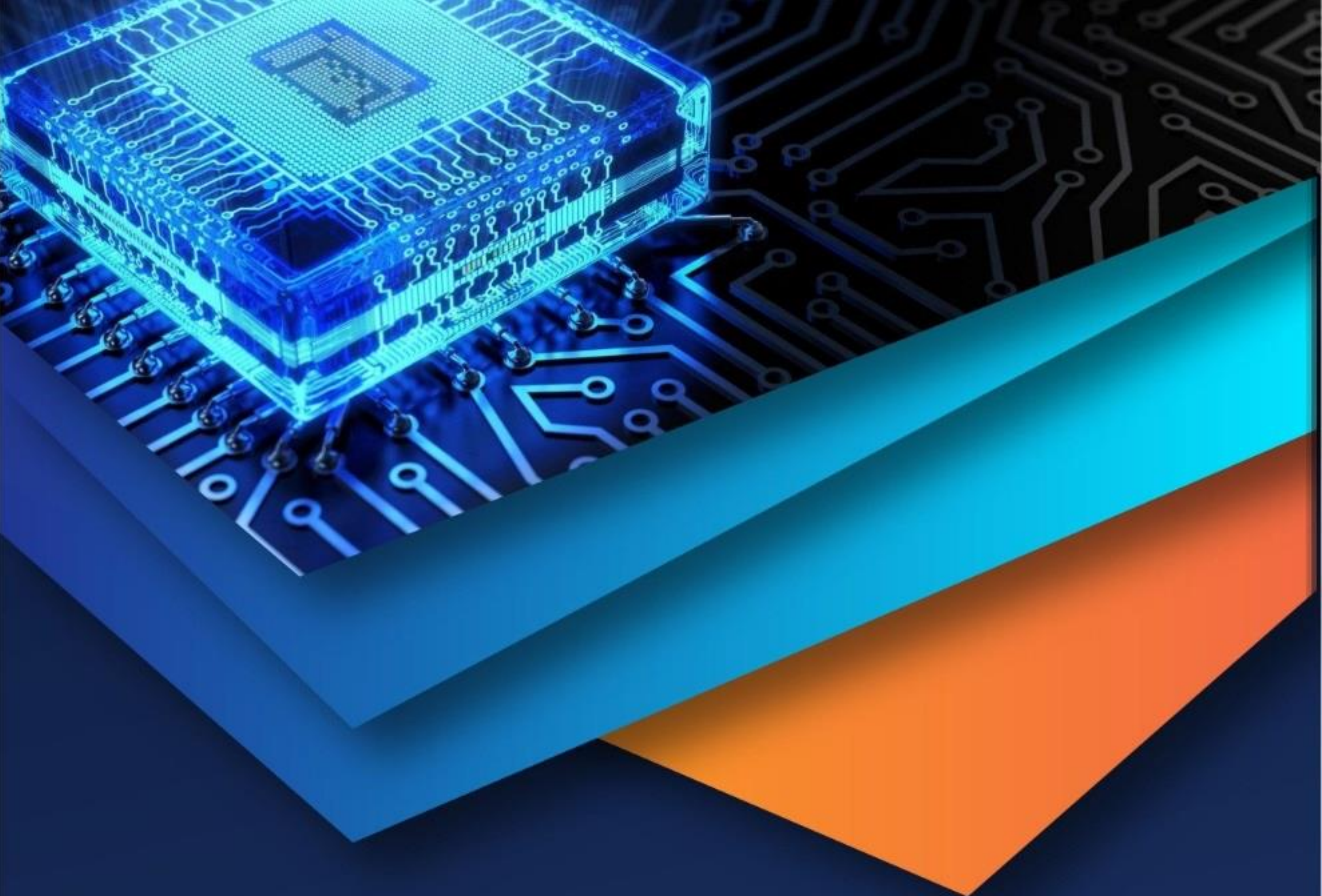
1) Problem: In 2026, the "Waving Tree" problem remains the bane of surveillance.
2) Solution: Modern models incorporate Hybrid Background Modeling (HBM), which combines traditional ViBe algorithms with Faster-RCNN or YOLO heads. This "confines" the deep learning search space only to regions where motion is statistically probable, filtering out static noise.

B. Small Object Scarcity

1) Problem: Moving objects at a distance (e.g., a drone 2km away) often occupy fewer than 10×10 pixels.
2) Solution: High-resolution feature fusion and ProgLoss scheduling ensure the network maintains high sensitivity to these sparse signals without triggering false positives from sensor noise.

V. CONCLUSION & FUTURE OUTLOOK

Moving object detection in 2026 has moved beyond simple "blob tracking." The synergy of NMS-free architectures and Self-Supervised Learning allows systems to adapt to new environments without manual re-labeling. The next frontier involves Multimodal Visual Reasoning, where detectors don't just "see" motion but "understand" the intent behind it (e.g., identifying a "suspicious" gait vs. normal walking).



doi crossref
10.22214/IJRASET

INDEX COPERNICUS
45.98

ISRA JIF
IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY