Proceedings of CONF-SPML 2026 Symposium: The 2nd Neural Computing and Applications Workshop 2025
DOI: 10.54254/2755-2721/2026.TJ30603

Occlusion-Aware Motion Modeling for Robust Multi-Object Tracking: A Comparative Analysis

Haoyu Zhang

Phillips Academy Andover, Andover, USA
haoyu.zhang2008@gmail.com

Abstract. The basis of modern computer vision applications is multi-object tracking, which applies to fields as diverse as autonomous driving, video surveillance, and sports analytics. However, occlusion remains a fundamental problem in how it disrupts tracking performance by causing identity switches and trajectory fragmentation. This paper comparatively analyzes six innovative occlusion-aware motion modeling approaches, revealing that these approaches can be categorized into two main paradigms: methods that combat physical manifestations of occlusion through geometric reasoning and hierarchical processing, and methods that address perceptual and computational consequences through global optimization and attention mechanisms. The analysis demonstrates the evolution of the field from simple occlusion handling to sophisticated modeling. This, in turn, provides valuable insight into selecting the appropriate methods for applications in specific scenarios. The research concludes that in robust multi-object tracking, handling object disappearance is dependent on numerous factors and diverse strategies, and future innovations will come from adaptive systems that can intelligently select the right strategies given the circumstances.

Keywords: multi-object tracking, occlusion handling, motion modeling, computer vision

1. Introduction

Tracking multiple objects is crucial for modern computer vision systems that are used in things like self-driving cars, surveillance, and even sports analysis. However, it is important to remember that these systems face a major and persistent problem: objects frequently get hidden behind other objects or elements, which, in turn, causes trackers to lose them quite often. So, the challenge is not about following visible objects—it is more about maintaining track of objects when they disappear. Therefore, this paper carefully examines six recent approaches that specifically address this hiding problem through better motion modelling. For instance, while methods like PD-SORT and MSPNET improve basic tracking mechanics, others like MPMOT and ORT are instead designed to handle longer disappearances. And so, although all of them aim to solve tracking confusion, they approach it in quite diverse ways: some primarily tackle the physical reality of objects blocking each other, whereas others address the mental and computational challenges that this creates. As a result, this analysis clearly shows the field's evolution from just applying quick fixes to developing much more sophisticated methods that can reliably maintain continuous object paths. This research provides

DOI: 10.54254/2755-2721/2026.TJ30603

significant practical value for both industry applications and academic research by offering guidance in selecting appropriate tracking methods for specific scenarios while identifying current limitations and suggesting future pathways for development.

2. Types of occlusion-aware motion models

2.1. Step-by-step matching (MSPNet)

MSPNet operates on a hierarchical strategy where it matches obvious object pars first before moving on to more difficult cases involving objects that are partially hidden or crowded together [1]. This matching strategy includes multiple association stages, starting with high-confidence motion-based matches. Then it moves on to address more challenging cases using an occlusion-aware head that dynamically adjusts matching thresholds based on occlusion likelihood [1]. This approach efficiently allocates computational resources by processing straightforward associations first and in doing so, preserves computing power for more sophisticated tracking scenarios [1].

2.2. Depth-based methods (PD-SORT)

PD-SORT works on the basis of determining depth to separate objects, essentially working in a 3D field rather than being limited to 2D coordinates [2]. Its function to determine how close an object is to the camera is significantly beneficial to reducing identity switches in crowded scenes. It achieves this through pseudo-depth estimation integrated into the traditional Kalman filter framework. PD-SORT also employs a novel Depth Volume IoU (DVIoU) metric that takes into account both the spatial overlap of an object and its depth consistency to offer reasoning about object relationships in 3D spaces.

2.3. Whole-sequence optimization

Whole-Sequence Optimization, as its name suggests, considers the entire video sequence. Rather than making decisions every single frame, it prioritizes smooth and continuous object path over time, even when there may be individual detections that are not continuous [3]. This approach utilizes cost functions to penalize identity discontinuation across extended sequences, thereby formulating a global trajectory optimization problem. To achieve this effect, a Priority Association Module (PAM) is employed to resolve matches that prioritize global consistency constraints [3]. This consideration of long-term trajectory smoothness over immediate association confidence attains superior performance in complex occlusion scenarios where temporary tracking ambiguities can potentially cause permanent identity switches [3].

2.4. Adaptive feature combining (MPMOT)

MPMOT balances motion and appearance cues based on the duration an object has been visible, or likewise invisible, with an emphasis on visual features for shorter durations of occlusions and an emphasis on motion predictions for longer disappearances [4]. The framework is built on an Adaptive Cost Matrix (ACM) and a Global Connection Model (GCM). The former automatically adjusts weighting between motion and appearance features based on the duration of occlusion, while the latter maintains track hypotheses in the case of long-term disappearances [4]. By selecting the most reliable cues according to the specific contexts and occlusion chrematistics, MPMOT can ensure the most optimal utilization across varying scenarios.

DOI: 10.54254/2755-2721/2026.TJ30603

2.5. Attention-based methods (Transformers)

Transformer-based methods focus only on relevant scene context when objects disappear, using special attention mechanisms to predict their locations while considering all available information. These include the movements of other objects as well as the layout of the scene [5]. These trackers employ a self-attention mechanism to model the relationship between all the objects in a scene. Consequently, this enables contextual reasoning that focuses on relevant spatial and temporal elements while ignoring excess information [5]. The weight of attention an object should receive is learned through the training process, allowing the model to automatically determine the most relevant elements for tracking hidden objects in certain contexts and scenarios [5].

2.6. Long-term disappearance handling (ORT)

ORT's specialty lies in its ability to track objects that have disappeared for an extended period. This is accomplished by maintaining the object's identities using memory networks and motion patterns that were previously learned, effectively tracking the object throughout its disappearance [6]. A dedicated memory module that stores trajectory information and motion patterns for objects that experience extended occlusions is the key to this method. By combining with learned motion priors, it can predict potential reappearance locations and maintain tentative tracks throughout the entire disappearance span [6]. The magic of preserving an object's identity information without direct visual evidence is what establishes its specialization in prolonged occlusion situations [6].

3. Comparison framework

3.1. Evaluation methods

This analysis examines the experimental frameworks and validation approaches employed by each method, focusing specifically on how they demonstrate effectiveness in occlusion handling. The assessment considers the methodological rigor of their testing procedures rather than specific numerical outcomes, examining how each approach validates its claims regarding occlusion robustness.

3.2. Testing setup

The experimental configurations reveal significant variation in testing methodologies across different approaches. MSPNet primarily utilizes standardized benchmarks like MOTChallenge for validation while employing consistent detector outputs to ensure comparability [1]. PD-SORT incorporates depth-aware evaluation metrics to validate its geometric approach, using specialized datasets that emphasize 3D spatial relationships between objects [2]. Optimization-based methods employ longer sequence evaluations to demonstrate its global consistency advantages, focusing on trajectory completeness rather than frame-level accuracy [3]. MPMOT validates its adaptive capabilities through varied occlusion scenarios, testing its dynamic feature balancing across different disappearance durations [4]. Attention-based methods utilize complex scene datasets to demonstrate their contextual reasoning capabilities, particularly in scenarios with multiple simultaneous occlusions [5]. ORT focuses validation on long-term disappearance scenarios, using specialized datasets with extended occlusion periods to verify its memory-based tracking capabilities [6].

DOI: 10.54254/2755-2721/2026.TJ30603

3.3. Performance in different situations

3.3.1. Short-term vs. long-term occlusions

MSPNet demonstrates strong performance in short-term occlusion scenarios, with its hierarchical matching strategy effectively handling brief disappearances through rapid re-identification mechanisms [1]. PD-SORT maintains consistent performance across varying occlusion durations, though its depth-based approach shows strength in medium-length occlusions where geometric consistency provides reliable cues [2]. The optimization method excels in long-term occlusion handling, with its global trajectory optimization maintaining identity consistency through extended disappearance periods [3]. MPMOT shows adaptive performance across occlusion durations, adjusting its feature weighting based on disappearance length to maintain optimal tracking performance [4]. Attention-based methods handle various occlusion lengths competently, though their performance advantage becomes most apparent in medium to long-term scenarios where contextual information proves most valuable [5]. ORT specializes exclusively in long-term occlusions, demonstrating superior performance in extended disappearance scenarios where other methods typically fail [6].

3.3.2. Inter-object vs. scene occlusions

PD-SORT shows superior performance in inter-object occlusion scenarios, where its depth estimation effectively disentangles overlapping objects in crowded environments [2]. The attention-based methods handle both inter-object and scene occlusions competently, using contextual reasoning to maintain tracking through various occlusion types [5]. ORT specializes in occlusions occurring in complex scenarios, particularly with static environmental elements that cause prolonged disappearances [6]. MSPNet performs reliably in inter-object occlusions but shows limitations in complex scene occlusion scenarios where simple hierarchical matching proves insufficient [1]. The optimization method maintains consistent performance across both occlusion types, though its global approach is advantageous in scene occlusions where long-term trajectory consistency is crucial [3]. MPMOT adapts effectively to both occlusion types through its dynamic feature balancing, though it requires sufficient training data to recognize different occlusion characteristics [4].

3.3.3. Crowded vs. sparse scenes

In crowded scenes, MSPNet's hierarchical approach efficiently manages the high density of objects by prioritizing straightforward matches before addressing complex cases [1]. PD-SORT also demonstrates remarkable robustness in crowded environments, where its 3D spatial reasoning prevents identity switches that commonly occur [2]. The optimization method maintains consistent performance across both crowded and sparse scenes, though its computational demands increase significantly in high-density scenarios [3]. MPMOT shows adaptability in crowded scenes, adjusting its association strategies based on scene density and occlusion complexity [4]. Attention-based methods show excellent scalability in crowded scenes, leveraging their context-aware processing to manage complex object interactions and maintain tracking integrity despite frequent occlusions [5]. ORT performs reliably in both crowded and sparse environments, though its specialized long-term capabilities are most valuable in scenarios where objects disappear completely for extended periods rather than briefly occluding in dense crowds [6].

DOI: 10.54254/2755-2721/2026.TJ30603

4. Critical analysis

4.1. Strengths and weaknesses

MSPNet's primary strength lies in its computational efficiency and reliable performance in moderate occlusion scenarios, though it struggles with long-term disappearances and complex scene occlusions [1]. PD-SORT offers robust geometric reasoning but depends heavily on accurate depth estimation, which can be challenging in monocular camera setups and may introduce additional computational costs [2]. The optimization method provides excellent long-term consistency but requires substantial computational resources and may struggle with real-time applications due to its global processing requirements [3]. MPMOT's adaptive feature balancing represents a significant advantage across varying occlusion conditions, though its complex architecture increases implementation complexity and requires careful parameter tuning [4]. Attention-based methods deliver superior contextual reasoning and handle complex occlusion scenarios effectively, but their high computational demands and substantial memory requirements limit practical deployment in resource-constrained environments [5]. ORT provides unmatched performance in long-term disappearance scenarios through its sophisticated memory mechanisms, though this specialization comes at the cost of reduced effectiveness in short-term occlusion handling and increased system complexity [6].

4.2. Speed and efficiency

MSPNet demonstrates favorable computational characteristics due to its hierarchical processing approach, making it suitable for real-time applications with limited resources while maintaining reasonable tracking accuracy [1]. PD-SORT maintains reasonable efficiency despite its geometric computations, though depth estimation introduces additional processing costs compared to traditional 2D approaches, particularly in dense scenes with numerous objects [2]. The optimization method's global processing demands significant computational resources, particularly for long video sequences, limiting its practical applicability [3]. MPMOT exhibits moderate computational requirements that vary based on scene complexity, with its adaptive mechanisms introducing additional processing costs that scale with the number of simultaneous occlusions [4]. Attention-based methods exhibit the highest computational demands among the evaluated approaches, though their performance advantages in complex scenarios may justify these costs in applications where accuracy takes precedence over efficiency [5]. ORT demonstrates variable efficiency depending on occlusion scenarios, with its memory mechanisms operating efficiently during normal tracking but requiring substantial processing during long-term disappearance management [6].

4.3. Adaptability to new environments

MSPNet shows good generalization capabilities due to its reliance on fundamental motion characteristics rather than environment-specific features, though performance may degrade in scenes with unusual motion patterns [1]. PD-SORT's geometric approach provides reasonable cross-environment performance, though its depth estimation accuracy may vary across different camera setups and conditions, requiring calibration for optimal performance [2]. The optimization method demonstrates strong adaptability to new environments when sufficient training data is available, though performance may degrade in significantly different scenarios and requires retraining for optimal cross-domain performance [3]. MPMOT exhibits excellent adaptability to environments

DOI: 10.54254/2755-2721/2026.TJ30603

with similar occlusion characteristics to its training data, though its performance depends on the diversity of scenarios encountered during training and may require additional fine-tuning for novel environments [4]. Attention-based methods show strong generalization capabilities when trained on diverse datasets, leveraging their contextual learning to adapt to new environments, though performance depends heavily on training data comprehensiveness [5]. ORT exhibits excellent adaptability to environments with similar motion patterns to its training data but may require retraining for different scenarios and shows some sensitivity to environmental changes that affect object motion characteristics [6].

4.4. Key findings

The comparative analysis reveals that MSPNet provides the most balanced approach for general-purpose applications, combining reasonable performance with computational efficiency across various occlusion scenarios [1]. PD-SORT offers specialized advantages in scenarios where depth information is reliable and inter-object occlusions dominate, particularly in crowded environments with frequent object interactions [2]. The optimization method represents the optimal choice for applications requiring maximum long-term identity consistency, particularly when computational resources are not a primary constraint and trajectory completeness takes priority over real-time performance [3]. MPMOT's adaptive approach makes it suitable for dynamic environments with varying occlusion characteristics, providing robust performance across different disappearance durations and scene conditions [4]. Attention-based methods excel in complex scenarios that require sophisticated contextual reasoning, though their practical deployment is limited to applications where computational resources are abundant and real-time requirements are relaxed [5]. ORT provides unique capabilities for handling extended disappearances, making it invaluable for specialized applications where objects frequently vanish for prolonged periods and reappear in unpredictable locations [6].

5. Conclusion

This analysis demonstrates the diverse strategies necessary for handling multi-object tracking and occlusions. Each method contains its own strengths and weaknesses, offering unique advantages to specific problems through distinct technical approaches. This study reveals that selecting the right method is dependent on specific application constraints, and that environmental conditions should take precedence over a universal solution. However, this study lacked quantitative performance metrics, real-world deployment analysis, and the testing of hybrid approaches. It also does not feature emerging directions like neuromorphic computing or the ethical implications of robust tracking systems. In the future, methods should become increasingly more hybridized to effectively utilize the strengths from multiple methods while covering for their shortcomings. This could eventually lead to an adaptive system that can select appropriate strategies based on live analysis and computational availability. Additionally, standardized evolution protocols encompassing more diverse scenarios would facilitate more distinctive comparisons and push for more progress in this field. Accounting for both algorithmic performance and practical deployment in the development of a comprehensive benchmarking framework will be essential for translating research advances into real-world applications across domains such as autonomous systems, surveillance, and behavioral analysis.

References

[1] Zheng, Y., Qi, H., Li, L., Li, S., Huang, Y., He, C., & Wang, D. (2024). Motion-guided and occlusion-aware multi-object tracking with hierarchical matching. Pattern Recognition, 151, 110369. https: //doi.org/10.1016/j.patcog.2024.110369

[2] Wang, Y., Zhang, D., Li, R., Zheng, Z., & Li, M. (n.d.). PD-SORT: Occlusion-Robust multi-object tracking using pseudo-depth cues. PD-SORT: Occlusion-Robust Multi-Object Tracking Using Pseudo-Depth Cues. https: //doi.org/10.48550/ARXIV.2501.11288

[3] Li, M., Zhang, Y., Jia, Y., & Yang, Y. (2025). Advancing multi-object tracking through occlusion-awareness and trajectory optimization. Knowledge-Based Systems, 310, 112930. https: //doi.org/10.1016/j.knosys.2024.112930

[4] Meng, W., Duan, S., Ma, S., & Hu, B. (2025). Motion-Perception multi-object tracking (MPMOT): Enhancing multi-object tracking performance via motion-aware data association and trajectory connection. Motion-Perception Multi-Object Tracking (MPMOT): Enhancing Multi-Object Tracking Performance via Motion-Aware Data Association and Trajectory Connection, 11(5), 144. https: //doi.org/10.3390/jimaging11050144

[5] Wang, S., Fang, G., Liu, L., Wang, J., Zhu, K., & Melo, S. N. (2023). Transformer-Based visual object tracking with global feature enhancement. Applied Sciences, 13(23), 12712. https: //doi.org/10.3390/app132312712

[6] Han, S., Wang, H., Yu, E., & Hu, Z. (2025). ORT: Occlusion-robust for multi-object tracking. Fundamental Research, 5(3), 1214-1220. https: //doi.org/10.1016/j.fmre.2023.02.003