Thermal Pedestrian Multiple Object Tracking Challenge (TP-MOT)

Wassim El Ahmar
University of Ottawa
Ontario, Canada
welahmar@uottawa.ca

Angel Sappa
Computer Vision Center, Spain
ESPOL Polytechnic University, Ecuador
sappa@ieee.org

Riad Hammoud
PlusAi Inc.
Santa Clara, California, USA
riad.hammoud@plus.ai

Abstract

Multiple Object Tracking (MOT) has seen significant advancements in the RGB domain, yet remains underexplored in thermal imaging, despite its advantages in low-light and adverse weather conditions. The Thermal Pedestrian Multiple Object Tracking (TP-MOT) Challenge addresses this gap by introducing a large-scale thermal dataset and a standardized evaluation framework. This challenge provides a benchmark for tracking algorithms designed specifically for thermal data, emphasizing robust detection, motion modeling, and identity association in infrared imagery. Participants were required to use a tracking-by-detection pipeline with standardized YOLO-based detectors, ensuring a fair comparison of tracking methodologies. The top-performing approaches leveraged adaptive hyperparameter tuning, motion-based association, and infrared-specific feature extraction to enhance tracking accuracy while maintaining computational efficiency. The results demonstrate that thermal MOT can achieve high performance with dedicated methodologies, offering new insights into tracking pedestrians in challenging conditions. In this first edition a total of 11 teams have been registered for participation. This challenge serves as a catalyst for future research, paving the way for improved thermal tracking solutions in surveillance, autonomous navigation, and security applications.

1. Introduction

Recent progress in multiple object tracking (MOT) research has been primarily driven by advances in RGB-based approaches, which have benefited from abundant training data and well-established evaluation metrics. However, real-world applications—such as surveillance in low-light conditions, nighttime monitoring, and operations in adverse weather—reveal significant limitations of visible-light sensors. These challenging scenarios necessitate robust tracking methods that can operate effectively under compromised lighting and environmental conditions. The Thermal Pedestrian Multiple Object Tracking Challenge (TP-MOT) directly addresses this need by leveraging thermal imaging technology. Through capturing long-wavelength infrared (LWIR) data, thermal sensors provide a complementary modality that excels in conditions where traditional RGB imaging fails.

The TP-MOT challenge distinguishes itself by introducing the first large-scale thermal dataset specifically annotated for MOT [1]. This dataset comprises 30 sequences (9000 frames) captured at five urban intersections using a FLIR ADK thermal sensor. The diversity of the collected data, which spans various public spaces and encompasses a wide range of pedestrian appearances, poses a significant challenge to existing tracking algorithms (e.g., [14], [3], [9]). Inherent characteristics of thermal imagery, such as low resolution, sensor noise, and reduced contrast, require participants to devise innovative detection and association strategies tailored for the thermal domain. This challenge thus fosters a focused exploration of algorithmic techniques that are uniquely suited to exploit the advantages of thermal data while mitigating its limitations.

Unlike multi-modal challenges that leverage data translation between different sensor types [10], TP-MOT adopts a single-modality focus, ensuring that all efforts are concentrated on harnessing the full potential of thermal imaging. Participants are mandated to use a tracking-by-detection framework, with standardized detectors (YOLOV5s [6] or

YOLOV8s [13]) ensuring uniformity across submissions. This requirement guarantees a fair comparison of the tracking algorithms and emphasizes the significance of the subsequent association step. By restricting the experimental setup to thermal imagery, the challenge encourages the development of novel methodologies for box association, motion modeling, and trajectory refinement that are robust against the specific distortions and noise patterns of thermal data.

Building on the insights gleaned from previous tracking competitions, the TP-MOT challenge serves as a catalyst for pushing the boundaries of MOT research into a new modality. It demonstrates that while RGB-based methods have dominated the field, dedicated thermal tracking can achieve comparable—and in some cases superior—performance in challenging environments. The methods and innovations presented by the challenge participants advance the state-of-the-art in thermal MOT and provide valuable benchmarks and insights that can inform future research directions. This paper collates and reviews the key contributions from the challenge, detailing a spectrum of approaches that span from novel deep learning architectures to advanced data association techniques. Through systematically analyzing these contributions, we highlight the transformative impact of thermal imaging on MOT and outline promising avenues for future exploration in robust object tracking.

The manuscript is organized as follows. Section 2 describes previous work on the tracking and thermal imagery domains. Then, Section 3 introduces the dataset and evaluation criteria of this first Thermal Pedestrian Multiple Object Tracking Challenge (TP-MOT). Section 4 presents a summary of this year's participation and outlines the methodologies of the top-performing teams, emphasizing their contributions and illustrating the proposed architectures. Finally, Section 5 presents concluding remarks, while supplementary information regarding the teams and their affiliations can be found in the appendix.

2. Previous Work

In the realm of Multiple Object Tracking (MOT), significant advancements have been made, particularly in the context of thermal imaging. A notable contribution is the work by Ahmar et al. [1]. This study introduces an innovative approach that integrates thermal object identity with motion similarity to improve box association in thermal imagery. The authors also present a comprehensive dataset comprising thermal and RGB images captured in diverse urban settings, serving as a valuable benchmark for future research in thermal MOT.

Beyond thermal imaging, several algorithms have been pivotal in advancing MOT across various modalities. ByteTrack, introduced by Zhang et al. [15], emphasizes the association of every detection box, including low-score detections, to enhance tracking performance. This methodology has demonstrated state-of-the-art results on benchmarks like MOT17 [11]. Similarly, DeepSORT [14] builds upon the original SORT algorithm by incorporating deep learning techniques to create a more robust appearance-based model, effectively reducing identity switches and improving tracking accuracy. Another noteworthy algorithm is OC-SORT [3], which adapts to sudden bounding box changes, often resulting from occlusions, by handling non-linear motion patterns more effectively .

Specific to thermal imaging, Muresan et al. [12] develop a multi-object tracking and segmentation framework tailored for thermal images, incorporating an object validation module to address challenges such as misclassifications and false detections. Their work underscores the importance of integrating domain-specific enhancements to improve the reliability of thermal tracking.

Kumar et al. [9] propose a person tracking framework that integrates a low-resolution thermal infrared sensor with a wide-angle RGB camera. This system aligns signals from both sensors spatially and temporally, enabling the thermal sensor to correct errors from the RGB-based tracker. The combined approach effectively rejects false positives, improves segmentation, and addresses missed detections, demonstrating the potential of low-cost thermal sensors in enhancing person tracking systems.

Despite these advancements, the research community still faces a scarcity of tracking methods that fully exploit the unique properties of thermal imagery. This challenge aims to address this gap by encouraging the development of algorithms tailored to thermal data. By providing a dedicated platform and dataset, the challenge fosters innovation in thermal MOT, contributing to a more robust understanding and application of thermal characteristics in tracking methodologies.

3. Challenge

3.1. Dataset

The dataset used in this challenge is the Thermal MOT Dataset, introduced by Ahmar et al. [1]. This dataset is the first large-scale thermal imaging dataset specifically designed for multiple object tracking (MOT). It was collected using a FLIR ADK thermal sensor across five urban intersections, resulting in 30 sequences, each approximately one minute long. Figure 1 shows sample images from three sequences from the dataset.

Each sequence was captured at a frame rate of 5 frames per second (FPS), producing a total of 300 frames per sequence. In total, the dataset contains 58,590 annotations in the thermal domain, with an average of 6.51 annotations per image. These annotations were manually labeled to provide high-quality ground truth data.

Figure 1. Sample thermal images from the RGB-Thermal MOT dataset.

To ensure proper benchmarking and generalization, the dataset is split into 24 training sequences and 6 validation sequences. Ground-truth annotations are provided only for the training sequences, while participants must evaluate their models on the validation set without additional training on it.

Additionally, the dataset features 313 unique object tracks in the training set and 126 in the test set, allowing researchers to study the persistence and movement patterns of individuals in real-world scenarios. The dataset's structure makes it an ideal benchmark for developing and evaluating thermal object tracking algorithms, particularly those that aim to improve tracking performance in low-visibility environments where RGB-based methods may fail.

The dataset presents unique challenges inherent to thermal imagery, such as low contrast, sensor noise, and variable object appearance based on heat signatures. Unlike RGB-based MOT datasets, where object features are rich and detailed, thermal images rely on infrared radiation, making object discrimination and tracking more dependent on motion cues and robust association techniques.

This dataset provides a standardized benchmark for the thermal MOT community, offering a valuable resource for advancing tracking algorithms that operate beyond the visible spectrum. The inclusion of diverse urban environments further enhances its applicability to real-world scenarios, making it an essential dataset for research in thermal object tracking.

3.2. Evaluation

The Thermal MOT Challenge is designed to advance object tracking research in the thermal domain by providing a structured evaluation framework that ensures fair and reproducible comparisons of tracking algorithms. Given the unique challenges of thermal imagery, including lower resolution, noise, and reduced contrast compared to RGB data, the evaluation methodology prioritizes both tracking accuracy and consistency.

3.2.1. Evaluation Metrics

Submissions are assessed using well-established Multiple Object Tracking evaluation metrics, which provide a comprehensive measure of tracking performance:

- **Multiple Object Tracking Accuracy (MOTA):** Measures the overall tracking accuracy by accounting for false positives, false negatives, and identity switches. Higher MOTA scores indicate more accurate tracking.
- **Multiple Object Tracking Precision (MOTP):** Evaluates how precisely the predicted bounding boxes align with ground truth annotations. This metric provides insight into the spatial accuracy of the detected objects.
- **ID F1 Score (IDF1):** Quantifies the ability of the tracking algorithm to maintain consistent object identities across frames by balancing identity recall and precision.

To ensure objective and fair ranking of participants, the final score is computed using a weighted combination of the **normalized** values of these metrics:

$$\text{Final Score} = 0.5 \times \text{MOTA} + 0.25 \times \text{MOTP} + 0.25 \times \text{IDF1}.$$

This weighting scheme emphasizes tracking accuracy (MOTA) as the most critical factor while still valuing the precision of object localization (MOTP) and the identity consistency of tracked objects (IDF1).

3.2.2. Constraints

To ensure fair comparisons, all submissions must adhere to a tracking-by-detection paradigm, where object detection is performed separately from tracking. Participants must use either YOLOv5s or YOLOv8s as the detection model, ensuring a standardized baseline for comparison.

Each submission must include:

- A ZIP file containing six text files, one for each validation sequence, formatted according to the standard MOT format.
- A GitHub repository with the source code and trained models, allowing challenge administrators to verify the results.

The same tracking parameters must be used across all validation sequences, preventing sequence-specific tuning that could bias the results.

4. Methods

This section briefly summarizes the approaches used by the top three participating teams. Table 1 shows the results of the top 5 teams.

4.1. Rank 1: AutoSKKU

The *AutoSKKU* team from Sungkyunkwan University, South Korea, developed an adaptive hyperparameter tuning framework to enhance real-time multi-object tracking (MOT) in thermal imagery. Their approach systematically optimizes detection and tracking stages to improve tracking accuracy and robustness while maintaining computational efficiency. Unlike conventional MOT pipelines that rely on static parameter settings or complex deep learning-based re-identification (ReID) models, *AutoSKKU* introduces a stage-wise tuning method that dynamically adjusts key hyperparameters to achieve optimal tracking performance across different scenarios. Figure 2 summarizes the framework introduced by *AutoSKKU*.

One of the main challenges in thermal MOT is the lack of rich feature representations compared to RGB-based tracking. Thermal images rely on infrared radiation rather than color or texture information, making object association difficult, especially in low-contrast environments or when multiple objects have similar heat signatures. *AutoSKKU* addresses this by dividing the MOT pipeline into two key stages—detection and tracking—and optimizing each stage independently. The team adopts YOLOv8s [13] as the object detector, fine-tuning hyperparameters such as training and inference image resolution, confidence threshold, and non-maximum suppression settings to improve detection reliability. By carefully adjusting these parameters, their method ensures that the detector provides high-quality bounding boxes while minimizing false positives and missed detections.

The tracking stage is enhanced through adaptive tuning of motion modeling and object association strategies. Instead of relying on pre-set parameters, the framework dynamically updates tracking settings such as motion prediction models (Kalman filter-based), association cost functions, and minimum hits required for track initialization. This ensures that the tracker can handle rapid motion changes, occlusions, and variations in pedestrian trajectories, common challenges in thermal imagery. By optimizing these aspects, the method achieves more stable object tracking with fewer identity switches and fragmented tracks, crucial for real-world deployment in surveillance, security, and autonomous systems.

AutoSKKU's hyperparameter tuning strategy is designed for real-time execution, avoiding computationally expensive processes that would hinder deployment on edge devices or embedded systems. Unlike deep learning-based trackers that require large-scale computations and extensive training, this approach dynamically fine-tunes tracking parameters based on evaluation results, ensuring adaptability to diverse thermal imaging conditions. This makes the solution highly efficient, capable of operating in scenarios where computational resources are limited but high tracking performance is required.

Experimental results on the PBVS Thermal MOT dataset demonstrate the superiority of *AutoSKKU*'s method. Their tracking-by-detection framework achieves high MOTA (Multiple Object Tracking Accuracy) and IDF1 scores, surpassing standard MOT baselines while maintaining real-time performance. The framework performs particularly well in crowded scenes, nighttime surveillance, and urban environments, where thermal tracking often struggles due to occlusions and fluctuating background heat sources. The results confirm that their tuning methodology significantly improves tracking accuracy and identity preservation while reducing computational overhead.

By introducing a lightweight and adaptable tuning-based approach, *AutoSKKU* provides a scalable and efficient solution for thermal MOT. Their method is particularly suited for security monitoring, military surveillance, industrial automation, and smart city applications, where real-time tracking of pedestrians and moving objects in infrared imagery is essential. Their contribution represents a step forward in improving tracking accuracy in thermal imaging without requiring additional deep learning models, making it an important development in the field of thermal multi-object tracking research.

4.2. Rank 2: Fh-IOSB

The *Fh-IOSB* from Fraunhofer IOSB, Karlsruhe, Germany, developed a real-time hyperparameter tuning framework for multi-object tracking (MOT) in thermal imagery, focusing on optimizing both detection and tracking stages. Their approach enhances tracking accuracy by dynamically fine-tuning key parameters, eliminating the need for computationally expensive re-identification (ReID) models.

The workflow is shown in Fig. 3. Images are upscaled for improved small person detection with YOLOv8-s [7]. A ReID model extracts appearance features from detected persons. The association is performed in two stages, prioritizing detections with high confidence ($s \geq s_{\text{track}}$). Detections with confidence below s_{min} are removed. Distance IoU (DIoU) as motion information is combined with Euclidean distance of normalized ReID features as appearance information via a weighted sum: $d_{\text{comb}} = d_{\text{DIoU}} + w\, d_{\text{app}}$. The noise scale adaptive (NSA) Kalman filter based on [8]

Rank	Team	MOTA	MOTP	IDF1	MOTA_NORM	MOTP_NORM	IDF1_NORM	Weighted Result
1	AutoSKKU	98.44	12.63	81.30	1.00	0.00	0.83	0.71
2	Fh-IOSB	82.40	14.43	86.38	0.51	0.18	1.00	0.55
3	HNU-VPAI	74.84	18.89	71.30	0.27	0.64	0.49	0.42
4	GyeongTiger	65.90	21.13	65.31	0.00	0.87	0.29	0.29
5	FFI BASED	65.99	22.45	56.58	0.00	1.00	0.00	0.25

Table 1. Performance Metrics of Teams

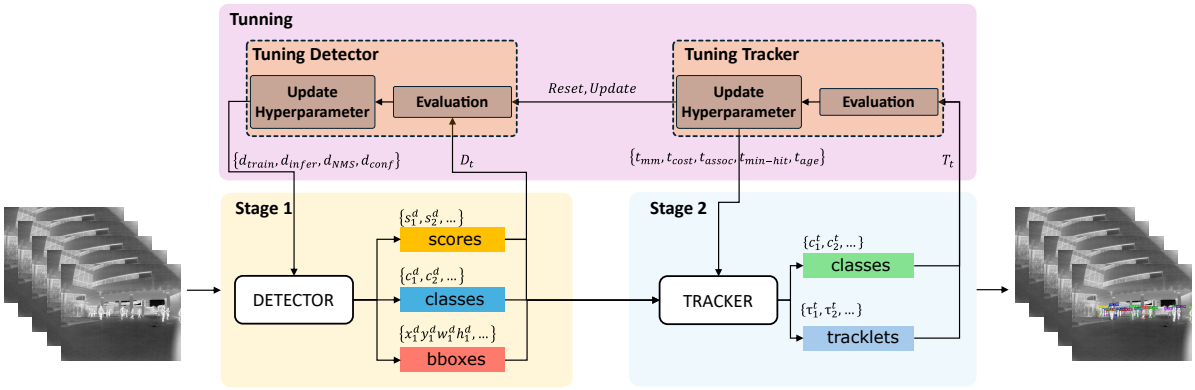

Figure 2. Using the thermal sensor input, Stage 1 involves running the detector to identify each object's class, location, and confidence score. In Stage 2, the tracker refines object locations and predicts their movements. Throughout both stages, the detector and tracker adjust hyperparameters based on evaluation results, either updating or resetting them for improved accuracy.

is enhanced by a new amplifying factor in the confidence-weighted update for matched track–detection pairs. Track features $\mathbf{f}^{\mathrm{T}}$ are updated with detection features $\mathbf{f}^{\mathrm{D}}$ by an exponential moving average (EMA): $\mathbf{f}_k^{\mathrm{T}} = \gamma\, \mathbf{f}_{k-1}^{\mathrm{T}} + (1-\gamma)\, \mathbf{f}_k^{\mathrm{D}}$. Standard motion prediction, initialization of tracks from unmatched detections, and inactivation of unmatched tracks complete the pipeline.

Implementation details:

Detection. The YOLOv8-s model, initialized with weights from a COCO pre-training, is trained with MMYOLO (*github.com/open-mmlab/mmyolo*) on TMOT [2] and FLIR (*flir.com/oem/adas/adas-dataset-form*) datasets for 12 epochs. The initial learning rate of $5{\cdot}10^{-5}$ is divided by 10 after epochs 8 and 11. The classification loss weight is set to 0.75. For data augmentation, random HSV, blurring, random grayscale, and CLAHE are used. During inference, multi-scale (MS) testing with five scales is performed. The IoU threshold of the non-maximum suppression is set to 0.5.

ReID. SOLIDER [4] is employed with original implementation and hyperparameters. The *Tiny* model is used. Subsampled variants of TMOT [2] and VTMOT [5] serve as datasets for training.

Tracker. The minimum detection confidence $s_{\min}$ is set to 0.2. The threshold s_{track} to split detections into high-confidence and low-confidence ones is 0.6. In the combined distance, $w = 20$ is applied, due to much smaller appearance distances compared to motion distances. The maximum distance threshold of the association is set to 3.2 and 2.4 in the first and second stage, respectively. An EMA weighting factor of $\gamma = 0.9$ is used. Tracks are kept inactive for a maximum of $i_{\max} = 40$ frames before termination.

General settings. All code is written in Python and runs on an AMD EPYC 9554 CPU. The detection and ReID models are executed on an NVIDIA L40 GPU.

Runtime. Besides striving for the best accuracy, a runtime-improved version is proposed. In this approach, MS testing is omitted and detector and ReID model are converted to TensorRT. In addition, the SOLIDER [4] model is exchanged by the lightweight OSNet-x1.0 AIN [16]. On the specified hardware, FPS increase from 6.1 to 80.6 by only a small decrease in accuracy.

4.3. Rank 3: HNU-VPAI

The *HNU-VPAI* team introduced a robust data association method for infrared multi-object tracking, enhancing the ByteTrack framework by integrating motion-based association with an infrared texture similarity metric.

They propose a robust and efficient data association method tailored for infrared image multi-object tracking, building upon the fundamental framework of ByteTrack.Unlike conventional methods that typically discard low-confidence detection results, our method utilizes both high-confidence and lower-confidence detections systematically, facilitating improved tracking robustness particu-

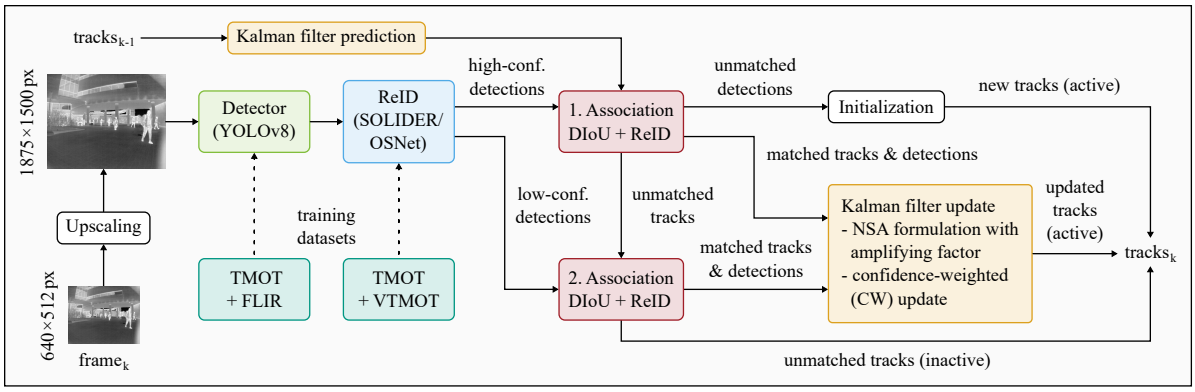

Figure 3. *Fh-IOSB* tracking pipeline. Highlights include an advanced association distance and an improved Kalman filter.

larly in challenging infrared imagery conditions. Specifically, our approach introduces a two-component cost calculation strategy, comprising a motion cost and an infrared texture similarity cost. The motion cost innovatively integrates detection confidence scores with Intersection-over-Union (IOU) metrics, aiming to achieve a more reliable association under moderate scene dynamics. Moreover, to effectively leverage the distinct characteristics of infrared images, they introduce a novel infrared texture feature derived from local gradient orientation histograms (HOG). This infrared-specific feature plays a pivotal role in accurately associating targets under complex scenarios such as thermal signature variations, background clutter, and occlusions. Consequently, our adaptive strategy dynamically selects and combines these two cost components based on pre-defined thresholding conditions, significantly enhancing tracking performance across diverse infrared tracking scenarios.

To effectively address the specific challenges posed by infrared imagery—such as low contrast, indistinct boundaries, and variable thermal signatures, they introduce an innovative infrared feature representation using Histogram of Oriented Gradients (HOG). Infrared targets typically exhibit unique yet subtle texture patterns rather than distinct color or intensity differences found in visible imagery, making traditional appearance models less effective. HOG descriptors excel at capturing local gradient information, which is particularly suitable for encoding infrared signatures characterized by subtle temperature-driven intensity variations and texture distributions. By computing the HOG-based texture histogram within each detected bounding box, they derive a robust thermal identity feature that encapsulates local structural characteristics. Subsequently, they calculate the similarity between predicted and candidate target regions based on their HOG feature histograms, providing a reliable measurement that accounts for thermal variations, low contrast conditions, and ambiguous object boundaries common in infrared data. This targeted use of texture information significantly enhances association accuracy, particularly under complex tracking conditions such as occlusions or overlapping thermal signatures.

Our proposed method lies a robust tracking framework that seamlessly integrates thermal and motion data, as shown below:

Specifically, let the set of predicted tracking boxes be denoted as $T = \{t_1, t_2, ..., t_m\}$ and detection boxes as $D = \{d_1, d_2, ..., d_n\}$. The total association cost $C_{total}(t_i, d_j)$ between predicted target t_i and detection candidate d_j is computed using an adaptive two-component formulation, consisting of a motion cost C_{motion} and an infrared similarity cost $C_{thermal}$. The motion cost integrates Intersection-over-Union (IOU) and the detection confidence score as follows:

$$C_{motion}(t_i, d_j) = 1 - (IOU(t_i, d_j) \times Confidence_{d_j}) \quad (1)$$

For the infrared similarity component, given the histograms H_{t_i} and H_{d_j} calculated from the HOG descriptors within each Region of Interest (ROI), the infrared cost $C_{thermal}$ between each predicted and detected bounding box is computed through histogram intersection:

$$C_{thermal}(t_i, d_j) = 1 - \frac{\sum_{k=1}^{b} \min(H_{t_i}(k), H_{d_j}(k))}{\sum_{k=1}^{b} H_{d_j}(k)} \quad (2)$$

The final cost $C_{total}(t_i, d_j)$ is computed adaptively based on two empirically selected thresholds A and B ($A < B$):

$$C_{total}(t_i, d_j) = \begin{cases} C_{motion}(t_i, d_j), & C_{motion}(t_i, d_j) < A \\ \alpha \cdot C_{motion}(t_i, d_j) + (1 - \alpha) \cdot C_{thermal}(t_i, d_j), & A \leq C_{motion}(t_i, d_j) < B \\ C_{thermal}(t_i, d_j), & C_{motion}(t_i, d_j) \geq B \end{cases} \quad (3)$$

Here, A and B are empirically determined thresholds satisfying $A < B$, and α is an experimentally optimized weighting factor balancing the relative contributions of motion and infrared texture similarity metrics.

5. Conclusion

The Thermal Pedestrian Multiple Object Tracking Challenge has demonstrated the potential of thermal imaging for multi-object tracking in real-world scenarios where RGB-based methods struggle. By introducing a large-scale thermal dataset and standardizing evaluation metrics, this challenge has fostered innovation in tracking algorithms tailored for thermal data. The results highlight the effectiveness of adaptive parameter tuning, motion-based association, and domain-specific enhancements in improving tracking accuracy and robustness.

The top-performing approaches showcased diverse strategies, from hyperparameter optimization to advanced data association techniques, underscoring the unique challenges posed by thermal imagery, such as low contrast and sensor noise. Notably, the best solutions balanced tracking accuracy with computational efficiency, making them suitable for deployment in resource-constrained environments.

Moving forward, the TP-MOT Challenge lays a strong foundation for further research in thermal-based tracking. Future work could explore integrating additional sensor modalities, improving feature extraction for identity preservation, and enhancing tracking robustness under extreme environmental conditions. By continuing to refine methodologies and expanding datasets, the research community can unlock new possibilities for thermal MOT applications in surveillance, autonomous systems, and beyond.

Appendix A. Teams Information

The organization team acknowledge the participants and utilize edited versions of top-performing team submissions to provide additional method explanations.

TP-MOT 2025 organization team:
Members: Wassim El Ahmar, Angel Sappa, Riad Hammoud
Affiliation: University of Ottawa, Computer Vision Center, ESPOL Polytechnic University, PlusAi Inc

Top Participating Teams:

AutoSKKU
Members: Duong Nguyen-Ngoc Tran, Long Hoang Pham, Chi Dai Tran, Quoc Pham-Nam Ho, Huy-Hung Nguyen, Jae Wook Jeon
Affiliation: Sungkyunkwan University

Fh-IOSB
Members: Daniel Stadler, Andreas Specker
Affiliation: Fraunhofer IOSB

wangzhiyu918
Members: Zhiyu Wang, Weiqing Lu, Puhong Duan, Bin Sun, Xudong Kang, Shutao Li
Affiliation: Hunan University

GyeongTiger
Members: Gyeongho Cho
Affiliation: Pusan National University,

FFI BASED
Members: Sigmund Rolfsjord, Jon Andre Ottensen, Jorgen Ahlberg
Affiliation: Norwegian Defence Research Establishment, University of Oslo, Linkoping University

References

[1] Wassim El Ahmar, Dhanvin Kolhatkar, Farzan Nowruzi, and Robert Laganiere. Enhancing thermal mot: A novel box association method leveraging thermal identity and motion similarity. *arXiv preprint arXiv:2411.12943*, 2024. 1, 2

[2] Wassim El Ahmar et al. Enhancing thermal mot: A novel box association method leveraging thermal identity and motion similarity. *arXiv:2411.12943*, 2024. 5

[3] Jinkun Cao, Jiangmiao Pang, Xinshuo Weng, Rawal Khirodkar, and Kris Kitani. Observation-centric sort: Rethinking sort for robust multi-object tracking. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9686–9696, 2023. 1, 2

[4] Weihua Chen et al. Beyond appearance: A semantic controllable self-supervised learning framework for human-centric visual tasks. In *CVPR*, 2023. 5

[5] Yabin Zhu et al. Visible–thermal multiple object tracking: Large-scale video dataset and progressive fusion approach. *Pattern Recognition*, 2025. 5

[6] Glenn Jocher, Alex Stoken, Jirka Borovec, Liu Changyu, Adam Hogan, Laurentiu Diaconu, Jake Poznanski, Lijun Yu, Prashant Rai, Russ Ferriday, et al. ultralytics/yolov5: v3. 0. *Zenodo*, 2020. 1

[7] Glenn Jocher et al. Ultralytics YOLO. https://github.com/ultralytics/ultralytics, 2023. 4

[8] Hyeonchul Jung et al. Conftrack: Kalman filter-based multi-person tracking by utilizing confidence score of detection box. In *WACV*, 2024. 4

[9] Suren Kumar, Tim K Marks, and Michael Jones. Improving person tracking using an inexpensive thermal infrared sensor. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 217–224, 2014. 1, 2

[10] Spencer Low, Oliver Nina, Angel D. Sappa, Erik Blasch, and Nathan Inkawhich. Multi-modal aerial view image challenge: Translation from synthetic aperture radar to electro-optical domain results - pbvs 2023. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 515–523, 2023. 1

[11] Anton Milan, Laura Leal-Taixé, Ian Reid, Stefan Roth, and Konrad Schindler. Mot16: A benchmark for multi-object tracking. *arXiv preprint arXiv:1603.00831*, 2016. 2

[12] Mircea Paul Muresan, Radu Danescu, and Sergiu Nedevschi. Multi-object tracking, segmentation and validation in thermal images. In *2023 IEEE Intelligent Vehicles Symposium (IV)*, pages 1–8. IEEE, 2023. 2

[13] Rejin Varghese and M Sambath. Yolov8: A novel object detection algorithm with enhanced performance and robustness. In *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, pages 1–6. IEEE, 2024. 2, 4

[14] Nicolai Wojke, Alex Bewley, and Dietrich Paulus. Simple online and realtime tracking with a deep association metric. In *2017 IEEE international conference on image processing (ICIP)*, pages 3645–3649. IEEE, 2017. 1, 2

[15] Yifu Zhang, Peize Sun, Yi Jiang, Dongdong Yu, Fucheng Weng, Zehuan Yuan, Ping Luo, Wenyu Liu, and Xinggang Wang. Bytetrack: Multi-object tracking by associating every detection box. In *European conference on computer vision*, pages 1–21. Springer, 2022. 2

[16] Kaiyang Zhou et al. Omni-scale feature learning for person re-identification. In *ICCV*, 2019. 5