

Supplementary: Motion Tracking with Rotated Bounding Boxes on Overhead Fisheye Imagery

Jordan Lam
jordanlam@zju.edu.cn

Zhejiang University
Zhejiang, China

In this document, supplementary materials are provided that additionally supports the claims of the manuscript. This document is structured as follows:

- Appendix **A** provides details on our empirical analysis which motivates this study.
- Appendix **B** describes the dataset for our experiments in further detail.
- Appendix **C** provides further experimental results with oracle detections and visual results.

A Empirical Analysis

For this study, we first conducted an empirical analysis of the challenges associated with tracking OBB on fisheye imagery. Considering the significant radial distortion that alters object appearance and shape, capturing these changes consistently poses a non-trivial task. To validate our analysis, experiments are performed with oracle detections, which are also ground truth detections. In our first hypothesis, we evaluate the suitability of motion trackers for tracking rotated bounding boxes (OBB). For the second hypothesis, we demonstrate that an object's aspect ratio, size, and speed are not consistent across the entire frame.

A.1 Analysis 1: Motion tracking

Through analysing motion trackers, specifically Bytetrack [8], for tracking OBB on fish-eye imagery, several insights are observed. Firstly, we observed that motion trackers remain effective for OBB, even without incorporating the rotation parameter. Secondly, due to the continuous change in aspect ratio, we noticed that tracking the width and height of the bounding box directly yields superior results compared to tracking the aspect ratio or scale, as observed in most motion trackers. Despite the effectiveness of motion tracking, there are still limitations compared to tracking on rectilinear imagery. Specifically, challenges arise in scenarios involving crowded scenes and nonlinear object movements. Within crowded scenes, different detection box angles will affect the Intersection over Union (IoU) scores differently. Furthermore, when a rotation parameter is added to the association process, the computation of IoU is increased due to needing the exact corner coordinates and frame size.

Dist from center	Δ Area	Δ Aspectratio	Δ Angle
< 20%	16.38	0.21	74.83
20-40%	10.20	0.13	39.50
40-60%	7.51	0.08	10.21
60-80%	3.79	0.06	15.72
80+%	1.37	0.07	8.77

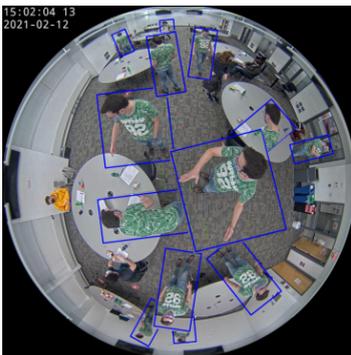
Table 1: The average change in the area, aspect ratio, and angle per frame dependent on the distance from the centre.

A.2 Analysis 2: Region Changes

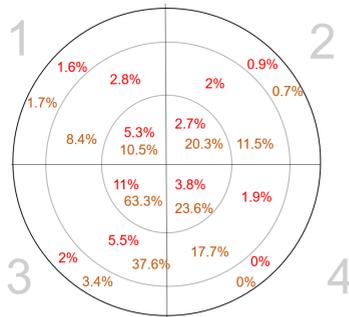
In our second analysis, we performed a region-based investigation with a total of 16 regions, assigning each detection based on its centre. We analysed how an object’s area, aspect ratio, and angle changed as it moved through the video.

Our analysis confirmed that the aspect ratio and area differ significantly across regions, particularly as most cameras are not directly positioned at the centre of the room. In Fig 1a, the central zones tend to exhibit larger detections and aspect ratios are closer to a square shape, while the peripheral zones adhere to the standard aspect ratio for a person. This indicates that our method should cover these possible issues where lost detections might appear larger within the central zone of the frame. Additionally, with the change in aspect ratio, tracking the aspect ratio might not be the optimal method to handle this.

To further understand the speed of changes, we selected and recorded objects with high levels of movement activity. The results presented in Table 1 support our hypothesis that aspect ratio and area are associated with the proximity to the centre. Therefore, the closer the object is to the centre, the more noise is likely to appear during tracking. Whereas movement in the peripheral regions tends to be less affected. These findings provide valuable insights into the dynamic characteristics of objects in fisheye videos.



(a) Multi-exposure of a single observed person.



(b) Percentage difference from the bottom left outer quadrant.

Figure 1: Analysis Visualisation: Highlighting the difference in size and aspect ratio in different locations of the frame.

B Dataset Setup

Our tracking method is evaluated on a custom evaluation dataset combining CEPTDOF [14] and WEPDToF [15]. The dataset consists of 20 videos with a total of 28k frames that tackle a wide range of challenges which is shown in Table 2. This final evaluation dataset includes numerous real-world obstacles such as crowded spaces and severe occlusions, making it a suitable dataset to assess our model.

Video Clip	Original Dataset	Frames	Description
Convenience Store	WEPDToF	1075	Cropped View
Exhibition	WEPDToF	690	Occlusions, Crowded, Tiny people
Exhibition Setup	WEPDToF	1800	Camouflage, Tiny people
High Activity	CEPDToF	7202	Constant walking activity
IRfilter	CEPDToF	3000	Low light - worse detections
IRill	CEPDToF	3000	Low light - worse detections
It Office	WEPDToF	500	Tiny people
Jewelry Store	WEPDToF	450	Occlusions
Jewelry Store 2	WEPDToF	450	Occlusions
Kindergarten	WEPDToF	549	Children
Large Office	WEPDToF	350	Occlusion, Tiny people
Large Office 2	WEPDToF	350	Occlusion, Tiny people
Lunch1	CEPDToF	1201	Common walking and sitting
Lunch2	CEPDToF	3000	Crowded scenes
Lunch3	CEPDToF	900	Common walking and sitting
Printing Store	WEPDToF	1055	Camouflage
Repair Store	WEPDToF	947	Camouflage, Partly visible
Street Grocery	WEPDToF	231	Distorted aspect ratio, Non-circular view
Tech Store	WEPDToF	633	High camera, Tiny people, Occlusions
Warehouse	WEPDToF	496	High camera, Cropped view

Table 2: Videos within the Evaluation Dataset.

C Further Experimental results

C.1 Results with Oracle Detections

In addition to evaluating estimated detection from RAPiD [14], we also evaluated on the dataset’s oracle detections in Table 3. Through the evaluation, we have observed that our proposed method does outperform but is not as significant as using estimated detections. Additionally, whilst using KLD and BD as the association method, metrics such as AssA and IDF1 are similar to tracking without rotation with minimal improvements. Whereas only GWD shows significant improvements in all metrics.

Tracker	HOTA \uparrow	MOTA \uparrow	DetA \uparrow	AssA \uparrow	IDFI \uparrow	IDSW \downarrow
SORT [10]	89.15	97.70	95.55	83.23	87.56	<u>430</u>
Botsort [11]	77.94	84.87	78.63	78.67	82.29	1075
OC-SORT [9]	90.78	97.64	97.84	84.22	88.04	439
ByteTrack [8]	90.50	98.57	97.55	83.97	88.05	584
Ours (KLD)	90.50	98.51	98.76	82.92	87.14	523
Ours (BD)	91.04	98.42	98.71	83.97	88.07	553
Ours (GWD)	<u>91.42</u>	<u>98.68</u>	<u>98.90</u>	<u>84.52</u>	<u>88.95</u>	475

Table 3: Experimental results on the evaluation dataset with oracle detections. The best results have been underlined.

C.2 Visualisation Results

In this section, we demonstrate some limitations of our proposed method in Fig 2 and some areas where it excels in Fig 3.

There are some highlighted limitations of our proposed tracker which will need to be further addressed. When there are occlusion and rapid aspect ratio changes, the association can not capture this leading to no matches. This indicates that the same target can not be linked again and a new identity is generated once the target appears again. Additionally, due to an additional parameter, our proposed method is a lot more sensitive to targets that have little movement. This is demonstrated in four different clips within Fig 2.

Next in Fig 3, we focus on areas in which our RF-Tracker performs well. We have observed that through the use of a dynamic buffer, we can more comfortably rapid irregular motion. In Fig 3a, where the blue target quickly steps back, it can comfortably capture this change in movement. Additionally, in other clips, our method can account for the angle and capture targets that were occluded and lead to increased track retention.

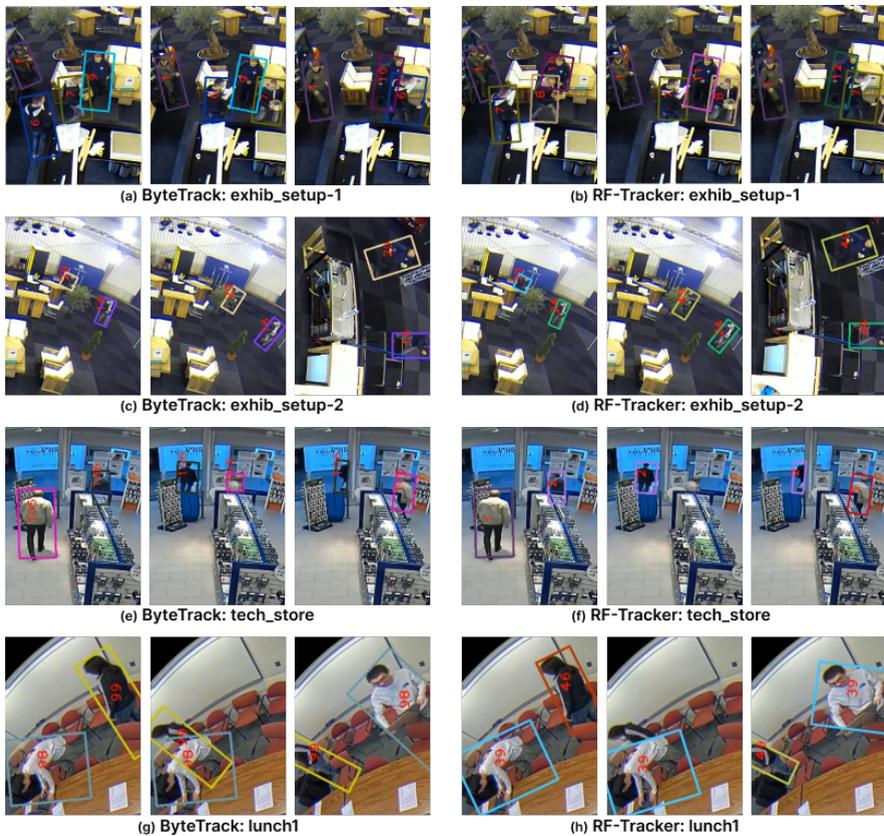


Figure 2: Visualisation results of our proposed method RF-Tracker. Samples of where RF-Tracker performs worse than ByteTrack which ignores the rotation parameter. There are limitations when there are big shape changes which lead to disjointed tracking of the same target.

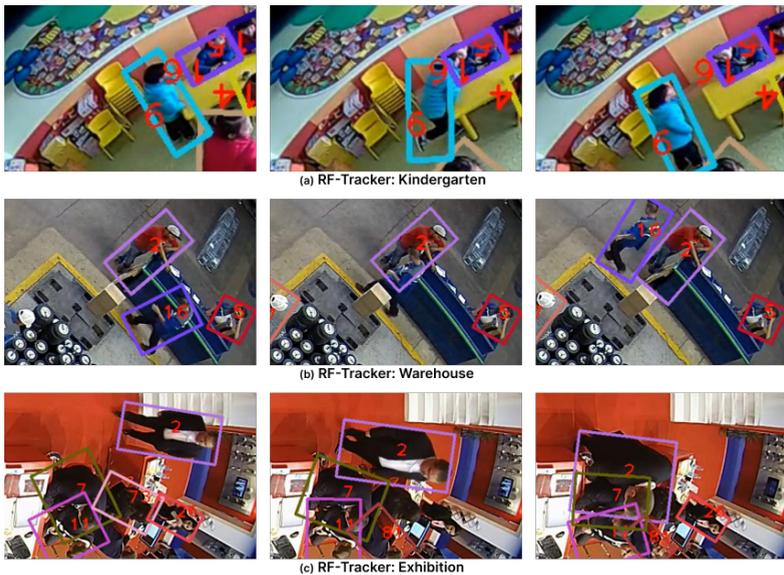


Figure 3: Visualisation results of our proposed method RF-Tracker. Three sequences are selected to demonstrate the effectiveness of the RF-Tracker in handling difficult cases such as occlusion and irregular movement. The same identities are represented through the same box colour and box number.

References

- [1] Nir Aharon, Roy Orfaig, and Ben-Zion Bobrovsky. Bot-sort: Robust associations multi-pedestrian tracking. *ArXiv*, abs/2206.14651, 2022.
- [2] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft. Simple online and realtime tracking. In *IEEE international conference on image processing (ICIP)*, pages 3464–3468, 2016.
- [3] Jinkun Cao, Xinshuo Weng, Rawal Khirodkar, Jiangmiao Pang, and Kris Kitani. Observation-centric sort: Rethinking sort for robust multi-object tracking. *ArXiv*, abs/2203.14360, 2022.
- [4] Zhihao Duan, M. Tezcan, Hayato Nakamura, Prakash Ishwar, and Janusz Konrad. Rapid: Rotation-aware people detection in overhead fisheye images. *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 2700–2709, 2020.
- [5] Ozan Tezcan, Zhihao Duan, Mertcan Cokbas, Prakash Ishwar, and Janusz Konrad. Wepdtof: A dataset and benchmark algorithms for in-the-wild people detection and tracking from overhead fisheye cameras. *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 503–512, 2022.
- [6] Yifu Zhang, Pei Sun, Yi Jiang, Dongdong Yu, Zehuan Yuan, Ping Luo, Wenyu Liu, and Xinggang Wang. Bytetrack: Multi-object tracking by associating every detection box. *IEEE/CVF European conference on Computer Vision (ECCV)*, 2021.