Motion Tracking with Rotated Bounding Boxes on Overhead Fisheye Imagery

Jordan Lam
jordanlam@zju.edu.cn

Zhejiang University
Zhejiang, China

Abstract

Although recent progress in multiple object tracking (MOT) has been notable, effectively tracking rotating bounding boxes and views from an overhead angle remains a considerable challenge. Previous methods typically ignore the rotation parameter such as using the centre point, or focus more on the appearance cues. This paper introduces a simple motion-based tracker that is effective for fisheye imagery, addressing these specific challenges. Our proposed method focuses on motion estimation and detection associations. The approach is composed of (1) transforming rotated bounding box detections into 2D Gaussian distributions, (2) distribution distances that can replicate Intersection over Union (IoU) to associate detections, and (3) a dynamic buffer during association to alleviate irregular movement in overhead views. In this paper, we have experimented with three different distribution distances which have been shown to replicate the IoU behaviour during association. Through these distribution distances, we can effectively track rotated bounding boxes and be applied on a linear Kalman Filter. Experimental results show that our method achieves promising performance on multi-object tracking on overhead fisheye surveillance datasets and demonstrates comparable results on the MOT datasets.

1 Introduction

Multi-object tracking (MOT) has been critical for security and retail analytics. For example, in smart retail, businesses can efficiently gain valuable insights about customer interactions, dwell time, and hot spot identification [3, 10, 14]. Whilst wireless signal tracking has been a popular choice, it has drawbacks, especially for scenarios involving privacy concerns and individuals who do not use mobile devices [13, 17]. As a result, visual tracking remains in high demand, and we believe fisheye lenses is a great option in many situations due to their wide field-of-view (FOV), which provides a 180-degree omnidirectional view with a single camera. The FOV allows for extended tracking duration without the need for multiple conventional cameras and re-identification techniques. In recent years, fisheye lenses have been effectively applied multiple domains including autonomous vehicles, body mounts and surveillance, offering an omnidirectional view and reducing the requirement for multiple lenses [6, 8, 11, 12, 15, 20, 26].

Despite significant advancements in MOT [1, 4, 7, 24, 33], challenges such as rotated/orientated bounding boxes (OBB), appearance change, and irregular motion have yet to be fully resolved. Related research has shown OBB to outperform the standard horizontal

bounding boxes in capturing effective space on overhead fisheye [15, 20]. However, from our research, we have noticed limited work on tracking OBB and previous works will typically ignore the rotation parameter. Through understanding the importance of including the rotation parameter, there are several tracking challenges with OBB on fisheye lenses that must be tackled. Firstly, is the introduction of an additional periodic rotation parameter, which complicates the tracking process when using a non-linear Kalman Filter. Secondly, fisheye lens distortions affect the appearance and geometric consistency assumptions for tracking, as objects' size and appearance change with their movement within the frame. Thirdly, with the indoor overhead omnidirectional view, irregular motion patterns are often noticed. As objects can be observed moving in any direction which is a lot more complicated than simple linear motions such as a straight path down an aisle.

To address these challenges, we first conducted an empirical analysis of the challenges associated with rotated bounding boxes and fisheye lenses. Our empirical findings show that popular motion trackers can effectively handle OBB whilst not utilising any appearance cues. However, if the tracker ignores the rotation parameter of the BB, limitations can be seen in slightly more crowded areas and when there sudden movements of objects. Next, the aspect ratio and area of detections are shown to change significantly under fisheye images in different zones and distances. This analysis further supports our proposed method to tackle these challenges and our empirical analysis can be found in our supplementary material.

In this paper, we propose RF-tracker (Fisheye-Rotated Tracker), which is a multiple-object tracker under fisheye imagery with rotated bounding box detections. Our approach focuses on motion tracking to optimise computational efficiency, which can compensate for the additional complexity when detecting objects with OBB representations. We transform detection to a 2D Gaussian distribution representation and leverage geometric motion matching to mitigate ambiguity resulting from appearance distortions caused by the fisheye lens. In our experimental findings, we report promising gains in metrics like HOTA [16] and IDF1 [18] compared to other state-of-the-art MOT methods(ByteTrack [32] and OC-SORT [4]) which ignore the rotation parameter.

2 Related Works

2.1 SORT-based Tracking

Current MOT algorithms typically follow one-stage and two-stage methods. One-stage trackers typically handle both detection and tracking within a single model, whereas two-stage approaches prioritise the association aspect within a tracking-by-detection paradigm. In this paper, we focus on two-stage tracking-by-detection because our method is centred on the association method. This also provides for greater freedom in detector selection since detectors in this field are still relatively young. Additionally, motion tracking has been attracting great attention again, with ByteTrack [32] and OC-SORT [4] demonstrating improved performance with new techniques and higher-quality detectors. These methods are based on SORT [2] which uses location and motion cues to achieve real-time online tracking. DeepSORT [24] and StrongSort [7] are SORT adaptions that incorporate appearance cues to enhance tracking performance. However, for this study, we see appearance cues as a challenge due to the constant change from the distortion, therefore, we do not utilise appearance cues.

2.2 Rotated/Orientated Bounding Boxes

Orientated or rotated bounding boxes (OBB) were motivated by the need to better capture objects with irregular shapes and large empty spaces that cannot be efficiently represented by standard horizontal bounding boxes. OBBs have proven to be particularly useful in various scenarios for object detection, including satellite imagery and text detection [22, 28, 29]. Similarly, this applies to overhead views, where objects can appear at any angle, making OBBs a suitable choice for more accurate detection [8, 15, 20].

However, tracking OBB has seen limited works, where the majority focused on segmentation-based or centre-based approaches rather than directly working with orientated bounding boxes [23, 25]. While segmentation methods may offer greater accuracy in many cases, they often come with significant computational requirements, which makes achieving real-time tracking challenging.

2.3 Tracking on Fisheye

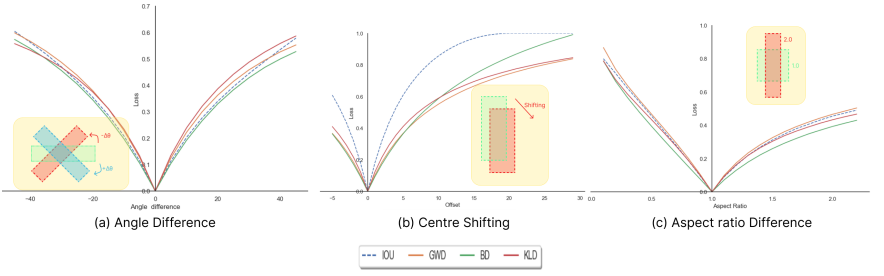
Many current state-of-the-art tracking methods developed for rectilinear videos have yet to be adapted to OBB and fisheye. Especially, with the increasing use of OBBs in modern detection methods, which presents new challenges compared to the conventional horizontal bounding boxes.

Sagastiberri et al. [19] proposed an online MOT on fisheye inspired by FairMOT [33]. Their work identified the primary challenge being the changing appearance of objects as they transition from the peripheral to the central regions of the field of view. To address this, they employed a convolutional LSTM network to track long-term appearance while leveraging temporal information to capture appearance variations over time. Cokbas et al. [5] focuses on re-identification on fisheye which consists of the advantage of overlapping FOV between cameras. Their method fuses multiple features consisting of deep-learning appearance cues, colour histograms and detection centres. Haggui et al. [9] utilised a centroid tracker [34] on tracking people within a scene based on only centre point detections.

In contrast to previous methods, our research focuses solely on motion-based tracking in fisheye videos without rectification or extensive pre-processing. To the best of our knowledge, this is the first work that considers pure motion tracking for OBB or fisheye imagery.

3 Rotated-Fisheye Tracker

In this section, we propose our tracking method, RF-Tracker, which is a motion tracker that tackles rotated bounding boxes on fisheye imagery. Drawing inspiration from SORT-based approaches, our method overcomes the challenges while ensuring simplicity and effectiveness. To initiate the tracking process, we begin by transforming the current frame's detected rotated bounding boxes into 2D Gaussian distributions. Subsequently, we utilise the Kalman Filter to predict the locations of the tracked objects in the subsequent frames. We then introduce our Gaussian distribution distances for the association task that allows efficient target matching between frames. Due to the nature of the different distances, three different distances that can replicate IoU behaviour are introduced. Finally, to capture the irregular motion patterns, we incorporated a dynamic buffer based on the detection's centre location.



(a) Angle Difference (b) Centre Shifting (c) Aspect ratio Difference

Figure 1: A comparison between our proposed association algorithms which can replicate the behaviour of IoU. Dotted blue is IoU, then orange, green, and red are GWD, BD, and KLD respectively. (a) shows the behaviours on different angles, (b) on the distance when shifting from the top left to bottom right, and (c) different aspect ratios with the same centre and area.

3.1 2D Gaussian Transformation

Rotated bounding boxes have witnessed remarkable advancements in the field of object detection for satellite imagery and text detection. Existing research has revealed that the introduction of rotation with IoU can lead to challenges such as boundary discontinuity and square-like problems in detection [30]. Therefore, our initial step transforms detections into their Gaussian distribution representation, which is then tracked by the Kalman Filter. This removes the limitations of tracking the periodic rotation parameter within the linear Kalman Filter. This study also shown that tracking the rotation parameter degrades performance while utilising this Gaussian representation improves it.

When dealing with rotated bounding boxes, the commonly stored parameters are the centre coordinates, width, height, and rotation angle. The transformation process, as outlined by [30] transforms $B(x,y,w,h,\theta)$ to $\mathcal{N}(\mu,\Sigma)$. μ is expressed as $\mu=(x,y)$ which are the centre point of the bounding box. Then we get the covariance with the following equation:

$$\begin{aligned}\Sigma^{\frac{1}{2}} &= RSR^T \\ &= \begin{pmatrix}\cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}\frac{w}{2} & 0 \\ 0 & \frac{h}{2}\end{pmatrix}\begin{pmatrix}\cos\theta & \sin\theta \\ -\sin\theta & \cos\theta\end{pmatrix} \\ &= \begin{pmatrix}\frac{w}{2}\cos^2\theta+\frac{h}{2}\sin^2\theta & -\frac{w}{2}\sin\theta\cos\theta+\frac{h}{2}\sin\theta\cos\theta \\ \frac{w}{2}\sin\theta\cos\theta+\frac{h}{2}\cos^2\theta & -\frac{w}{2}\sin^2\theta+\frac{h}{2}\cos^2\theta\end{pmatrix} \\ &= \begin{pmatrix}\sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy}\end{pmatrix},\end{aligned} \tag{1}$$

where R represents the rotation matrix and S represents the diagonal matrix. This covariance transforms the way we can store the width, height, and angle.

Within the Kalman Filter, we store the five parameters x, y, σ_{xx}, σ_{yy}, and σ_{yx}. Since the parameter σ_{yx}, which captures the correlation between the width and height of the bounding box, is not subject to periodic variations like the rotation angle r is. Therefore, tracking the Gaussian distribution provides a more comprehensive representation of the object's state, resulting in improved tracking accuracy.

3.2 Association with Distribution Distances

For the association, we proposed three different Gaussian distributions which can replicate the behaviour of IoU for bounding box association. We have noticed that using these distances performs better than using IoU with a rotation parameter. The three distances introduced in this section are Gaussian Wasserstein Distance (GWD), Kullback-Leibler divergence (KLD), and Bhattacharyya distance (BD).

GWD [30] quantifies the dissimilarity between two Gaussian distributions by measuring the minimum amount of work required to transform one distribution into the other, considering both the difference in means and the disparity in their spread. GWD is shown as,

$$\begin{aligned} dist_{xy} &= \|\mu_1 - \mu_2\|_2^2, \\ dist_M &= \mathrm{Tr}(\Sigma_1 + \Sigma_2 - 2(\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2})^{1/2}). \end{aligned} \tag{2}$$

where μ represents the mean, and Σ denotes the covariance matrix.

Our next association algorithm is KLD [31] which similarly quantifies the additional average amount of information needed to encode data from one distribution, measuring their dissimilarity.

$$\begin{aligned} dist_{xy} &= \frac{1}{2}(\mu_1 - \mu_2)^\top \Sigma_2^{-1}(\mu_1 - \mu_2), \\ dist_M &= \frac{1}{2}\mathrm{Tr}(\Sigma_1^{-1}\Sigma_2) + \frac{1}{2}\ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) - 1. \end{aligned} \tag{3}$$

The third is the Bhattacharyya distance shown in Equation 4, BD quantifies the dissimilarity between two Gaussian distributions, by assessing the overlap between them, incorporating both their means and covariances.

$$\begin{aligned} \Sigma_{comb} &= \frac{1}{2}(\Sigma_1 + \Sigma_2), \\ dist_{xy} &= \frac{1}{8}(\mu_2 - \mu_1)^T \Sigma_{comb}^{-1}(\mu_2 - \mu_1), \\ dist_M &= \frac{1}{2}\log\frac{|\Sigma_{comb}|}{\sqrt{|\Sigma_1| \cdot |\Sigma_2|}}. \end{aligned} \tag{4}$$

These distance metrics provide a meaningful measure of how much the shape and location of one Gaussian distribution needs to be adjusted to align with another, offering a comprehensive understanding of their divergence. To exhibit the behaviour similar to IoU with our Gaussian distances, with the following equation:

$$\begin{aligned} Dist_{ori} &= dist_{xy} + dist_M, \\ Dist_{final} &= (1 - \frac{1}{1 - \log(1 + \sqrt{Dist_{ori}})}) * \beta. \end{aligned} \tag{5}$$

GWD and KLD are already losses that have been successfully implemented in [35] for object detection in a similar manner. However, as the outputted similarity distance is lower than IoU, a β is applied in Equation 5 as a scaler, to further replicate the IoU behaviour. Each association algorithm we use has a different β to make up the difference with IoU.

3.3 Buffered Matching Mechanism

To address the changes in shape and irregular motion, we proposed a buffering mechanism during the association process. This was inspired by the work of [27], to similarly account for irregular motion patterns. The purpose of the buffer is to compensate for estimation errors that arise when objects undergo irregular motion, resulting in deviations from their actual locations. By expanding the distribution, the estimated locations are brought closer to the true positions.

Our empirical analysis revealed that a significant portion of lost associations occurs in the central region. Consequently, we adopted a dynamic buffer that adjusts based on the distance between the current track prediction's centre and the centre of the frame. Compared to a fixed buffer, a dynamic buffer is more suitable for this scenario due to the smaller peripheral zones, where increasing the buffer size could lead to further mismatches. We introduced a buffer parameter, denoted as λ, representing the maximum percentage increase of the buffer. The buffer increase is calculated as the percentage of the distance from the centre, such that if an object is $d\%$ away from the centre, the buffer ratio will be $d\lambda$. To incorporate the buffer and loosen the distribution, the tracking representation's covariance matrix is multiplied by the buffer ratio.

3.4 Tracking Process

The tracking process is shown in Algorithm 1 with information on the input and output from each step. For detections, we utilised RAPiD [8], which is a detector that outputs detections with rotations based on fisheye imagery. However, other detectors can also be used where each detection output consists of the inputs of (x, y, w, h, r). For our experiments, all the detections within the video are first pre-gathered, and then each frame is treated independently in chronological order keeping to online tracking principles.

Algorithm 1 Association with new detections $\mathcal{N}(m, \Sigma)$ (at time t)

Input : A set of track $T_{t-1} = \{T_1, ..., T_N\}$,
A set of dets that has rotations $D_t = \{D_1, ..., D_M\}$
Output: Update set of Tracks

// Transformation
$D_{gaus} \leftarrow convertGaus(D_t)$ // equation 1

// Get latest estimations from each track's Kalman Filter

// Associate
$temp_D \leftarrow buffer(D_{gaus})$

$matched, T_u, D_u = Matching(temp_D, temp_T)$

$T_t \leftarrow update(matched)$

// remove unmatched tracks and initialise new tracks

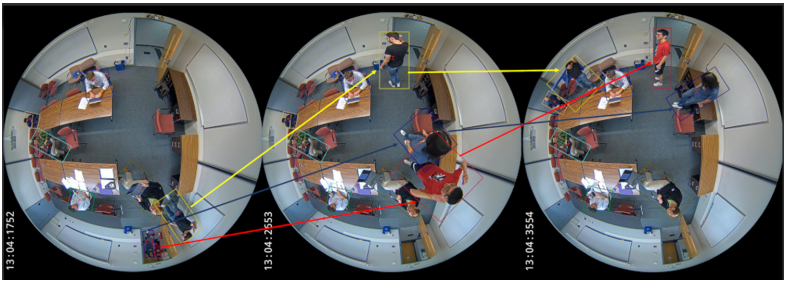


Figure 2: An example of tracking with our proposed method RF-Tracker on CEPDTOF with oracle detections.

4 Experiments

4.1 Datasets and Evaluation metrics

Dataset and Implementation. In this area of research, we encountered limited availability of public datasets tailored for tracking OBB on fisheye imagery. Therefore, to construct our comprehensive evaluation dataset, 20 videos are selected from both CEPDOF [8] and WEPDTOF[21] datasets, totalling over 27K frames. This final evaluation dataset offers considerable tracking challenges due to the presence of overlapping objects and irregular motion across multiple frames in diverse scenarios. Our experimental setup consists of estimated detections from RAPiD [8], with publicly available weights, and the oracle detections that are the ground truth detections. Furthermore, we used the MOT16 training dataset as one of our evaluation datasets to evaluate the performance of standard horizontal detections.

For our implementation, we maintained similar hyperparameters to ByteTrack [32]. To account for overhead scenes, potential occlusions, and association method, we expanded the track duration of our tracker to 60 frames and the matching threshold to 60%. As our method only focuses on the association step, it retains simple and real-time, with low computation requirements, and can be run with a CPU

Metrics. Following the recent works, we adopted HOTA[16] and IDF1[18] as the primary metrics to evaluate our tracking performance. HOTA is a metric that aims to be a more holistic assessment by considering accurate detection, association, and localisation, thereby balancing their effects. IDF1 assesses the capability of preserving identities and emphasises the performance of associations. As, MOTA places greater emphasis on the performance of object detections, using the same detections, the difference in performance is not as significant. To provide a more detailed analysis, we also include additional metrics such as AssA which evaluates the consistency and track fragments focusing on the stability and continuity of tracks, and ID switches (IDSW) which reflects the accuracy of identity maintenance throughout the tracking process.

4.2 Main Results

4.2.1 Evaluation on the Fisheye Dataset with RAPiD Detections

Table 1 compares our proposed RF-tracker with the state-of-the-art MOT methods on our evaluation dataset. To ensure fairness, all methods were evaluated using the same detections from the RAPiD detector [8]. For the horizontal trackers, the rotation parameter is ignored

Tracker	HOTA ↑	MOTA ↑	DetA ↑	AssA ↑	IDFI ↑	IDSW ↓
DeepSORT[24]	28.4	48.5	38.8	12.8	16.6	3965
SORT[2]	38.3	49.4	38.8	38.2	46.8	566
Botsort[1]	41.5	49.2	40.8	43.1	51.4	968
OC-SORT[4]	40.1	49.2	38.8	41.9	50.2	532
ByteTrack[32]	42.8	52.7	42.1	44.0	52.5	645
Ours (KLD)	44.8	56.3	44.8	45.1	55.3	640
Ours (BD)	45.1	56.5	44.9	45.6	56.3	593
Ours (GWD)	45.0	56.6	45.0	45.4	56.1	556

Table 1: Experimental results on the evaluation dataset with estimated detections from RAPiD. The best results have been underlined.

Tracker	HOTA ↑	MOTA ↑	DetA ↑	AssA ↑	IDF1↑	FPS↑
SORT [2]	19.1	12.6	10.5	34.9	18.5	64.5
BotSORT [1]	21.6	15.3	12.7	36.7	22.3	66.2
ByteTrack [32]	22.1	15.5	12.9	37.8	22.9	66.1
OC-SORT [4]	18.8	12.6	10.3	34.4	18.6	63.8
Ours (KLD)	21.4	15.3	12.7	35.6	22.2	53.2
Ours (BD)	21.5	15.3	12.7	36.3	22.3	55.8
Ours (GWD)	21.7	15.4	12.7	37.1	22.8	55.5

Table 2: Experimental Results on MOT 2016 dataset with MOT public detections on the training set. The best results have been underlined.

in the tracking process, as they were originally only based on standard horizontal bounding box detections.

Our proposed method, all three association methods, demonstrates promising performance results, outperforming the state-of-the-art results with an increase of over 2 in HOTA and 3.6 in MOTA. In this experiment, our proposed method with BD obtains the best results with 45.1 HOTA and our method with GWD does not differ too much with 45 HOTA. Additionally, our findings align with our initial hypothesis that appearance-based tracking approaches, such as DeepSORT [24], encounter significant challenges when applied to fisheye.

4.2.2 Evaluation on the MOT Dataset

To understand our tracker ability on non-rotated detections, our proposed tracker was evaluated on the MOT16 dataset. Although our proposed method does not surpass current trackers, it does not differ significantly. Our results differ from ByteTrack by only 0.5 HOTA and less than 0.2 MOTA. Upon analysing the data, it became evident that the dataset's viewing angle and occlusion were contributing factors. We found that tracking with the top left corner of the MOT dataset's viewing angle is more consistent than tracking using its centre coordinates. Next, from looking at speed (tracker only), despite the additional computation, our method maintains a satisfactory speed suitable for real-time, thanks to its inherent simplicity.

Association	Inc Rotation	Transformation	HOTA	IDF1
IoU	X	X	0	0
IoU	✓	X	-2.31	+0.03
KLD	✓	X	-5.22	-3.33
BD	✓	X	-5.03	-3.31
GWD	✓	X	-5.64	-3.10
KLD	✓	✓	+0.44	-0.58
BD	✓	✓	+1.04	+0.48
GWD	✓	✓	+1.47	+1.48

Table 3: Ablation study on association and transformation, this table demonstrates the percentage difference to the baseline of ByteTrack. To highlight the improvements, the colour red signifies a performance increase.

Association	Dynamic	Max Buffer Size		
		0%	30%	50%
GWD	X	0	-2.49	-6.52
GWD	✓	-	+0.29	-0.13

Table 4: Ablation study on the dynamic buffer with only the GWD association on oracle detections with HOTA results. To highlight the improvements, the colour red signifies a performance increase.

4.3 Ablation Study

To further evaluate our proposed method, we conducted an ablation study focusing on each individual component with oracle detections on the evaluation dataset. By employing oracle detections, we can isolate and evaluate the tracking performance independently of the detection aspect, thereby mitigating the impact of detector limitations. This ablation analysis allowed us to assess the key contributions: Gaussian transformation, different association algorithms, and dynamic buffer.

In this experimentation, we split our results into three parts: association method, rotation inclusion, and transform detection for KF. For the association, we have standard IoU as our baseline and the three proposed distances. Next, in our baseline, we do not include the rotation parameter in the tracking process, which yields the same results as ByteTrack. The third is transformation, which is the transformation from OBB to 2D Gaussian for the Kalman Filter estimations. Otherwise, the transformation only happens during the association process and the rotation parameter is estimated. Additionally, we measure the importance of our dynamic buffer by analysing the use of dynamic ratio and the maximum buffer expansion.

The results, as presented in Table 3 and Table 4, demonstrate our ablation results. In the first row of Table 3, our baseline ignores the rotation parameter during the tracking process and uses IoU for association. First, our results demonstrate that the performance will be negatively impacted by using a linear Kalman filter to track the rotation parameter naively. The performance further degrades if we use Gaussian distances to associate detection which contains the estimated rotation parameter. Finally, we demonstrate that transforming the detections prior to the Kalman Filter leads to superior HOTA performance which outperforms associating with IoU.

In Table 4, it is observed that a dynamic buffer expansion outperforms static buffer expansion. However, the buffer size has to be controlled to be not too big or the tracking performance will worsen. Our optimal results are around a 30% buffer expansion.

4.4 Evaluation Summary

Based on the results obtained from the experiments, our proposed method has demonstrated significant advancements compared to standard horizontal trackers for tracking OBB on fisheye imagery. Moreover, we demonstrated that our tracker remains usable for real-time tracking not only with OBB but with standard horizontal bounding boxes as well. Overall, among the three association approaches, GWD demonstrates great performance and is the most consistent for both high and low quality detections. Finally, as our proposed method operates as a two-stage tracker, it allows for easy integration of improved detectors in future iterations, which can further enhance overall performance.

4.5 Discussion

Although our method brings novelty in tracking detections with rotations, there are still limitations which remain unresolved or this method additionally introduces. Firstly, occlusion, one of the commonly addressed in tracking, our work slightly alters the problem as the top-down view removes a lot of the common occlusion seen in other works. However, on peripheral edges, it faces similar challenges as rectilinear views but with even smaller objects. The proposed method has not directly addressed this issue and currently uses both high and low quality detections proposed by ByteTrack [52] to tackle this problem. Next, as our method uses a geometric association approach, there are further limitations compared to using IoU. From our study, there has been noticeable increase in noise sensitivity and struggles on objects with no motion. Additionally, in this research, as we have only focused on SORT-based MOT with limited computations to remain efficient, deeper approaches and comparisons to other current SotA deep approaches should be addressed in further work. Finally, in our supplementary material, we have included further experiments to show the effectiveness of our method.

5 Conclusion

In this paper, we addressed the tracking challenges of rotated bounding boxes and overhead fisheye imagery through the utilisation of a pure-motion tracking algorithm. The task of tracking OBB in overhead fisheye imagery is challenging due to the periodic rotation parameter, and the constant change of appearance and irregular motion. The proposed approach involves the transformation of OBB detections into 2D Gaussian distribution representation and the utilisation of distribution distances that can replicate association with IoU. Moreover, to handle irregular motion patterns, we introduced a dynamic buffer expansion during the association process. Significant improvements are observed in our fisheye evaluation dataset compared to our SORT-based method baselines where we ignore the rotation within the tracking process. Furthermore, our ablation study further dives into our individual components and highlights the efficiency of our proposed method. In conclusion, we proposed RF-Tracker, which provides promising tracking performance while maintaining simplicity compared to state-of-the-art trackers for this task.

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] Lorena Bourg, Thomas Chatzidimitris, Ioannis Chatzigiannakis, Damianos Gavalas, Kalliopi Giannakopoulou, Vlasios Kasapakis, Charalampos Konstantopoulos, Damianos Kypriadis, Grammati Pantziou, and Christos Zaroliagis. Enhancing shopping experiences in smart retailing. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–19, 2021.

[4] 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.

[5] Mertcan Cokbas, Prakash Ishwar, and Janusz Konrad. Spatio-visual fusion-based person re-identification for overhead fisheye images. *ArXiv*, abs/2212.11477, 2022.

[6] Liuyuan Deng, Ming Yang, Hao Li, Tianyi Li, Bing Hu, and Chunxiang Wang. Restricted deformable convolution-based road scene semantic segmentation using surround view cameras. *IEEE Transactions on Intelligent Transportation Systems*, 21: 4350–4362, 2018.

[7] Yunhao Du, Zhicheng Zhao, Yang Song, Yanyun Zhao, Fei Su, Tao Gong, and Hongying Meng. Strongsort: Make deepsort great again. *IEEE Transactions on Multimedia*, 2023.

[8] 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.

[9] Olfa Haggui, Hamza Bayd, and Baptiste Magnier. Centroid human tracking via oriented detection in overhead fisheye sequences. *The Visual Computer*, 2023.

[10] Loh Li Har, Umi Kartini Rashid, Lee Te Chuan, Seah Choon Sen, and Loh Yin Xia. Revolution of retail industry: from perspective of retail 1.0 to 4.0. *Procedia Computer Science*, 200:1615–1625, 2022.

[11] Dong-Hyun Hwang, Kohei Aso, Ye Yuan, Kris Kitani, and Hideki Koike. Monoeye: Multimodal human motion capture system using a single ultra-wide fisheye camera. *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, 2020.

[12] Varun Ravi Kumar, Sandesh Athni Hiremath, Stefan Milz, Christian Witt, Clément Pinard, Senthil Kumar Yogamani, and Patrick Mäder. Fisheyedistancenet: Self-supervised scale-aware distance estimation using monocular fisheye camera for autonomous driving. *IEEE International Conference on Robotics and Automation (ICRA)*, pages 574–581, 2019.

[13] Andreas D Landmark and Børge Sjøbakk. Tracking customer behaviour in fashion retail using rfid. *International journal of retail & distribution management*, 45(7/8): 844–858, 2017.

[14] Nils Magne Larsen, Valdimar Sigurdsson, and Jørgen Breivik. The use of observational technology to study in-store behavior: Consumer choice, video surveillance, and retail analytics. *The Behavior Analyst*, 40:343–371, 2017.

[15] Shengye Li, M. Tezcan, Prakash Ishwar, and Janusz Konrad. Supervised people counting using an overhead fisheye camera. *IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–8, 2019.

[16] Jonathon Luiten, Aljosa Osep, Patrick Dendorfer, Philip Torr, Andreas Geiger, Laura Leal-Taixé, and Bastian Leibe. Hota: A higher order metric for evaluating multi-object tracking. *International Journal of Computer Vision (IJCV)*, 129:548–578, 2021.

[17] Meera Radhakrishnan, Sharanya Eswaran, Archan Misra, Deepthi Chander, and Koustuv Dasgupta. Iris: Tapping wearable sensing to capture in-store retail insights on shoppers. *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 1–8, 2016.

[18] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. *IEEE/CVF European Conference on Computer Vision (ECCV) Workshops*, pages 17–35, 2016.

[19] Itziar Sagastiberri, Noud van de Gevel, Jorge García, and Oihana Otaegui. Learning sequential visual appearance transformation for online multi-object tracking. *IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–7, 2021.

[20] Masato Tamura, Shota Horiguchi, and Tomokazu Murakami. Omnidirectional pedestrian detection by rotation invariant training. *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1989–1998, 2019.

[21] 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.

[22] Jinwang Wang, Jian Ding, Haowen Guo, Wensheng Cheng, Ting Pan, and Wen Yang. Mask obb: A semantic attention-based mask oriented bounding box representation for multi-category object detection in aerial images. *Remote. Sens.*, 11:2930, 2019.

[23] Qiang Wang, Li Zhang, Luca Bertinetto, Weiming Hu, and Philip H.S. Torr. Fast online object tracking and segmentation: A unifying approach. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

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

[25] Bao Xin Chen and John Tsotsos. Fast visual object tracking using ellipse fitting for rotated bounding boxes. *IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, 2019.

[26] Weipeng Xu, Avishek Chatterjee, Michael Zollhöfer, Helge Rhodin, P. Fua, Hans-Peter Seidel, and Christian Theobalt. Mo2cap2: Real-time mobile 3d motion capture with a cap-mounted fisheye camera. *IEEE Transactions on Visualization and Computer Graphics*, 25:2093–2101, 2018.

[27] F. Yang, Shigeyuki Odashima, Shoichi Masui, and Shan Jiang. Hard to track objects with irregular motions and similar appearances? make it easier by buffering the matching space. *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 4788–4797, 2022.

[28] Xue Yang, Jirui Yang, Junchi Yan, Yue Zhang, Tengfei Zhang, Zhi Guo, Xian Sun, and Kun Fu. Scrdet: Towards more robust detection for small, cluttered and rotated objects. *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8231–8240, 2018.

[29] Xue Yang, Qingqing Liu, Junchi Yan, and Ang Li. R3det: Refined single-stage detector with feature refinement for rotating object. *ArXiv*, abs/1908.05612, 2019.

[30] Xue Yang, Junchi Yan, Qi Ming, Wentao Wang, Xiaopeng Zhang, and Qi Tian. Rethinking rotated object detection with gaussian wasserstein distance loss. *International Conference on Machine Learning (ICML)*, pages 11830–11841, 2021.

[31] Xue Yang, Xiaojiang Yang, Jirui Yang, Qi Ming, Wentao Wang, Qi Tian, and Junchi Yan. Learning high-precision bounding box for rotated object detection via kullback-leibler divergence. *Advances in Neural Information Processing Systems*, 34:18381–18394, 2021.

[32] 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.

[33] Yifu Zhang, Chunyu Wang, Xinggang Wang, Wenjun Zeng, and Wenyu Liu. Fairmot: On the fairness of detection and re-identification in multiple object tracking. *International Journal of Computer Vision (IJCV)*, 129:3069–3087, 2021.

[34] Xingyi Zhou, Vladlen Koltun, and Philipp Krähenbühl. Tracking objects as points. *IEEE/CVF European Conference on Computer Vision (ECCV)*, pages 474–490, 2020.

[35] Yue Zhou, Xue Yang, Gefan Zhang, Jiabao Wang, Yanyi Liu, Liping Hou, Xue Jiang, Xingzhao Liu, Junchi Yan, Chengqi Lyu, Wenwei Zhang, and Kai Chen. Mmrotate: A rotated object detection benchmark using pytorch. *30th ACM International Conference on Multimedia*, 2022.