Distortion-Aware Multi-Object Tracking via Virtual Plane Projection in Overhead Fisheye Cameras

https://github.com/PanithiVanasirikul/DAFTrack

Panithi Vanasirikul[1], Piyanon Charoenpoonpanich[1], Ekapol Chuangsuwanich[2]
[1]OxygenAI, [2]Department of Computer Engineering Chulalongkorn University

Overview

- Two main **challenges in fisheye Multi-Object Tracking**:
  - **Nonlinear Motion** breaking the Kalman-based tracker's constant velocity assumption.
  - **Highly Varying Appearances** challenging appearance-based ReID models.

  Therefore, **combining motion and appearance-based methods also does not overcome these limitations.**
- We proposed **DAFTrack**:
  - **Tracking in a distortion-aware plane** to linearize motion.
  - **Unscented Kalman Filter-based nonlinear updates** linking virtual and image planes.
  - **Camera Height Incorporation** for consistent real-world association.
- **Results:** DAFTrack outperforms SOTA by **+6.1 HOTA on CEPDOF** and **+3.3 HOTA on WEPDTOF datasets**.

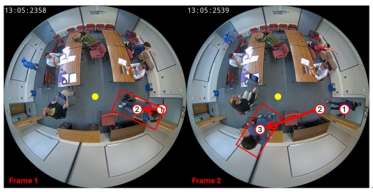

Figure 1: Illustration of fisheye-induced nonlinear motion.

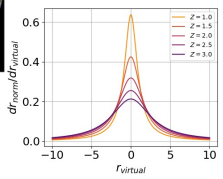

Figure 2: $dr_{norm}/dr_{virtual}$ across camera heights Z, showing stronger apparent acceleration at lower heights near the optical center.

Method

- **Input:** High and low confidence bounding boxes.
- **Step 1:** Convert all bounding boxes to image plane representation.
- **Step 2:** Predict next state of active tracks.
- **Step 3:** Project high confidence image plane representation to virtual plane representation.
- **Step 4:** Incorporate camera height to get real-world camera coordinates.
- **Step 5:** Associate high confidence detections from step 2 with active tracks' predicted state from step 3 using Euclidean distance on virtual plane.
- **Step 6:** Associate low confidence from step 1 with remaining active tracks' using Gaussian Wasserstein Distance on image plane.
- **Step 7:** Update the Kalman filter using true measurement from image plane by using the unscented transform.

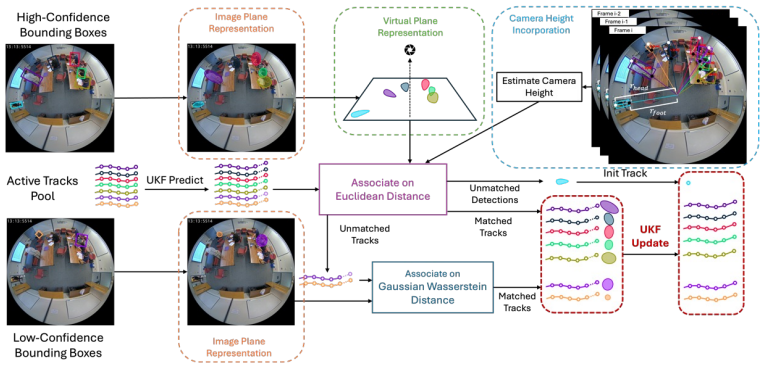

Figure 3: System diagram of our proposed method.

Experiments

- **Main Results**
  - **CEPDOF** DAFTrack achieves state-of-the-art performance, surpassing the previous best HOTA by 6.1 points. Its automatic camera height estimation (che) performs comparably to the version using ground truth heights (gt).

Tracker	HOTA↑	DetA↑	MOTA↑	IDF1↑	AssA↑	IDS↓
General-purpose Trackers						
BoT-SORT [3]	46.9	70.8	68.1	52.6	33.2	1392
OC-SORT [4]	47.5	75.1	77.6	47.5	31.8	11654
ByteTrack [5]	49.5	77.1	78.3	49.0	33.6	1030
Fisheye Trackers						
HDA [6]	51.9	79.5	79.4	51.8	36.0	1565
RF-Tracker [7]	52.9	80.4	80.7	54.4	36.8	796
DAFTrack with gt (Ours)	**59.0**	**83.0**	**83.2**	58.5	**44.0**	**700**
DAFTrack with che (Ours)	58.0	82.2	82.5	**58.8**	43.3	725

  - **WEPDTOF** DAFTrack achieves state-of-the-art performance, surpassing the previous best HOTA by 3.3 points using automatic camera height estimation (che).

Tracker	HOTA↑	DetA↑	MOTA↑	IDF1↑	AssA↑	IDS↓
General-purpose Trackers						
BoT-SORT [3]	51.9	49.7	42.0	64.5	57.8	213
OC-SORT [4]	57.9	57.2	**50.6**	65.2	59.9	329
ByteTrack [5]	56.7	56.2	49.6	64.0	59.2	273
Fisheye Trackers						
HDA [6]	58.1	53.2	46.5	64.7	66.2	183
RF-Tracker [7]	59.3	55.3	48.1	66.4	65.9	**173**
DAFTrack with che (Ours)	**62.6**	**57.8**	50.4	**69.9**	**69.1**	205

- **Datasets:** CEPDOF[1], WEPDTOF[2]
- **Metrics:** HOTA, DetA ,MOTA, IDF1, AssA, IDS
- **Baseline:** BoT-SORT[3], OC-SORT[4], ByteTrack[5], HDA[6], RF-Tracker[7]
- **Ablation Studies** Tracking on the virtual plane consistently improves all metrics, while true measurements (tm) further enhance HOTA and AssA. Incorporating both ground truth (gt) and estimated (che) camera height yields higher HOTA and AssA, indicating more accurate tracking and associations.

Tracker	HOTA↑	IDF1↑	AssA↑	IDS↓
Image Plane + KF	46.2	43.0	30.4	1869
Virtual Plane + KF	50.9	59.7	34.8	927
Virtual Plane + UKF (tm)	55.4	**59.8**	40.0	**672**
Virtual Plane + UKF (tm, gt)	**59.0**	58.5	**44.0**	700
Virtual Plane + UKF (tm, che)	58.0	58.8	43.3	725

References

[1] Zhe Duan, M.O. Tezcan, H. Nakamura, P. Ishwar, and J. Konrad. Rapid: Rotation-aware people detection in overhead fisheye images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pages 636–637. IEEE, 2020.

[2] M. 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. In 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pages 1381–1390, 2022. doi: 10.1109/WACV51458.2022.00145.

[3] Nir Aharon, Roy Orfaig, and Ben-Zion Bobrovsky. Bot-sort: Robust associations multi-pedestrian tracking. arXiv preprint arXiv:2206.14651, 2022.

[4] 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.

[5] 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 Proceedings of the European Conference on Computer Vision (ECCV), 2022.

[6] Ping-Yang Chen, Jun-Wei Hsieh, Ming-Ching Chang, Munkhjargal Gochoo, Fang-Pang Lin, and Yong-Sheng Chen. Fisheye multiple object tracking by learning distortions without dewarping. In 2023 IEEE International Conference on Image Processing (ICIP), pages 1855–1859, 2023. doi: 10.1109/ICIP49359.2023.10222872.

[7] Jordan Lam. Motion tracking with rotated bounding boxes on overhead fisheye imagery. In 35th British Machine Vision Conference 2024, BMVC 2024, Glasgow, UK, November 25-28, 2024. BMVA, 2024. URL https://papers.bmvc2024.org/0257.pdf.