

Cross-View Multi-Object Tracking (MOT)

Computer Vision Spring 2024 Project Proposal

Bhav Beri¹, Harshit Aggarwal²

I. INTRODUCTION

Our project is based on the paper - "DIVOTrack: A Novel Dataset and Baseline Method for Cross-View Multi-Object Tracking in DIVERse Open Scenes" [1]

In today's world, accurately tracking multiple objects across different viewpoints and scenes is tough. Cross-view tracking is a specialized field within computer vision that focuses on tracking objects across multiple camera viewpoints simultaneously.

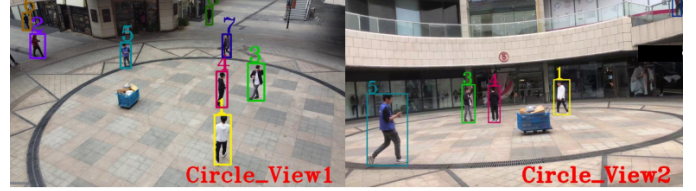
Unlike single-view tracking, which operates within the constraints of a single camera's perspective, cross-view tracking integrates information from multiple cameras to provide a more comprehensive understanding of object movements and interactions within a scene. Specifically, given multiple synchronized videos capturing the same scene from different viewpoints, there is a high probability that an object obscured in one view is visible in another. Cross-view settings can compensate for the occlusion of information from single-view monitoring with their complementary information. By aggregating information from multiple viewpoints, cross-view tracking provides a more comprehensive understanding of object trajectories, leading to improved tracking accuracy and reliability.

Multiple cross-view tracking datasets have appeared in the past ([2], [3], [4], [5], [6]), the DIVO-Track dataset aims to fill the gaps and shortcomings with the older datasets. This paper also proposed an end-to-end cross-view multi-object tracking baseline framework named CrossMOT to learn object embeddings from multiple views, extended from the single-view tracker FairMOT [7]. CrossMOT [1] is a unified joint detection and cross-view tracking framework that uses an integrated embedding model for object detection, single-view tracking, and cross-view tracking.

II. PROJECT PLAN

A. General Project Plan

We aim to replicate the results of this paper. We would train the whole model by ourselves and then get the results. An important point to note is that the paper's code is available online on GitHub¹, but none of the checkpoint files or models are available openly online, and they require a special license for getting those and thus are inaccessible. We would be doing the experiments on five datasets, as done by the paper and shown results, namely - DIVOTrack [1],



WILDTRACK [2], EPFL [3], MvMHAT [4], and CAMPUS [6].

We also plan to run some ablations and experiments to check further the cases where the model fails to identify the objects correctly, which will allow us to get a better understanding of the working of the model and ways to improve it further.

B. New Things Planned

As part of our project proposal, we aim to extend the applicability of our model by evaluating its performance on several other generic multi-view Datasets - MultiviewX [8], BBC Old-School Dataset [9], MCMOD [10].

Through thoroughly examining the paper, we have pinpointed specific scenarios where our model might encounter challenges, such as cases involving motion dynamics, crowded scenes, among others. One notable dataset that exemplifies these challenges is the BBC Old-School Dataset [9]. We intend to assess our model's performance on this dataset to identify areas for improvement and try to bring up certain ideas for improvement.

An extension to the project (which may or may not be feasible in the course timeline) would be introducing an innovative method for cross-view multi-person/multi-object tracking, leveraging adaptive multi-scale feature representation learning. This method would dynamically adjust feature extraction to accommodate variations in object characteristics across diverse viewpoints.

III. COMPUTE RESOURCES AVAILABLE

We both have access to the ADA Cluster of IIIT Hyderabad [11]. Both of us are Dual-Degree Students, and thus, by the research account privileges, we have access to unlimited minutes of usage with up to 4 GPUs allowed. We also have access to a desktop with RTX 3080ti [12] and a laptop with RTX 3070ti mobile [12], which would allow us to run the model locally as well in most cases.

¹ 2021111013 ² 2021111015

¹<https://github.com/shengyuhao/DIVOTrack>

REFERENCES

- [1] S. Hao, P. Liu, Y. Zhan, K. Jin, Z. Liu, M. Song, J.-N. Hwang, and G. Wang, "Divotrack: A novel dataset and baseline method for cross-view multi-object tracking in diverse open scenes," *International Journal of Computer Vision*, pp. 1–16, 2023.
- [2] T. Chavdarova, P. Baqué, S. Bouquet, A. Maksai, C. Jose, T. Bagautdinov, L. Lettry, P. Fua, L. Van Gool, and F. Fleuret, "Wildtrack: A multi-camera hd dataset for dense unscripted pedestrian detection," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5030–5039.
- [3] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua, "Multicamera people tracking with a probabilistic occupancy map," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 267–282, 2008.
- [4] L. Y. W. F. S. W. Yiyang Gan, Ruize Han, "Self-supervised multi-view multi-human association and tracking," in *ACM MM*, 2021.
- [5] Y. Xu, X. Liu, L. Qin, and S.-C. Zhu, "Cross-view people tracking by scene-centered spatio-temporal parsing," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, Feb. 2017. [Online]. Available: <https://ojs.aaai.org/index.php/AAAI/article/view/11190>
- [6] Y. Xu, X. Liu, Y. Liu, and S.-C. Zhu, "Multi-view people tracking via hierarchical trajectory composition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 4256–4265.
- [7] Y. Zhang, C. Wang, X. Wang, W. Zeng, and W. Liu, "Fairmot: On the fairness of detection and re-identification in multiple object tracking," *International Journal of Computer Vision*, vol. 129, no. 11, p. 3069–3087, Sept. 2021. [Online]. Available: <http://dx.doi.org/10.1007/s11263-021-01513-4>
- [8] Y. Hou, L. Zheng, and S. Gould, "Multiview detection with feature perspective transformation," 2021.
- [9] "Old school dataset," 4 2022. [Online]. Available: <https://www.bbc.co.uk/rd/projects/old-school-dataset>
- [10] Y. Q. Y. H. S. L. Zihao Sun, Xiao Hu, "Mcmmod: The multi-category large-scale dataset for maritime object detection," *Computers, Materials & Continua*, vol. 75, no. 1, pp. 1657–1669, 2023. [Online]. Available: <http://www.techscience.com/cmc/v75n1/51540>
- [11] "IIITH high performance computing (HPC) cluster." [Online]. Available: <https://www.iiit.ac.in/hpc/>
- [12] "Nvidia graphics cards - 30 series." [Online]. Available: <https://www.nvidia.com/en-in/geforce/graphics-cards/30-series/>