Dynamic Clustering for Efficient Trajectory Prediction in Dense Crowd Environments

Antonius Bima Murti Wijaya[1] Paul Henderson[2] Marwa Mahmoud[3]

I. INTRODUCTION

Dense crowd trajectory prediction is an important problem for public safety systems, and particularly critical in high-risk environments, including transportation hubs, large-scale public events, and densely populated urban spaces. In such settings, pedestrians continuously adapt their movements in response to surrounding individuals, often aligning their behaviour with nearby pedestrians to avoid collisions and maintain smooth crowd flow [1], [2], [3].

Most existing trajectory prediction methods have been developed for relatively low-density pedestrian scenarios and rely on manually annotated tracking data that provides reliable individual trajectories [4], [5], [6]. However, these approaches face significant challenges when applied to dense crowd environments. Modelling interactions between large numbers of pedestrians increases computational cost, particularly for approaches that explicitly model neighbourhood or group relationships [7], [3]. In addition, in highly crowded scenes, pedestrians frequently occlude one another, and the head becomes the only consistently visible body part across individuals. This limited visual information reduces tracking reliability, leading to missing trajectories, identity switches, and re-identification failures. These tracking errors propagate to downstream trajectory-prediction models, degrading prediction accuracy.

To address these limitations, we propose a clustering-based representation for dense crowd trajectory prediction inspired by the collective movement patterns observed in human crowds. Instead of modelling each pedestrian individually, we group nearby pedestrians into clusters during the tracking stage and represent each cluster using its centroid. The trajectory prediction model then operates on these cluster-level representations rather than individual pedestrian trajectories.

Our contributions focus on the proposed online trajectory clustering-based representation, cluster evaluation metrics, and the evaluation of trajectory predictors on cluster outputs in a crowd dataset.

II. METHODS

The proposed dynamic clustering method takes time-varying pedestrian location data as input, typically obtained from a tracking system. In each frame, the clustering algorithm calculates a centroid for each cluster. These centroid positions replace individual pedestrian locations as inputs to a trajectory prediction model that forecasts future paths. This design seeks to decrease training and inference time, lower memory requirements, and enhance robustness to noisy tracking data.

[1,2,3]All of the authors are with School of Computing Science, University of Glasgow, G12 8QQ, United Kingdom [1] `a.wijaya.1@research.gla.ac.uk` [2]`paul.henderson@glasgow.ac.uk`, [3]`marwa.mahmoud@glasgow.ac.uk`

The proposed dynamic clustering approach for dense crowds is illustrated in Figure 1. The process begins with initialization using a nested agglomerative clustering method, followed by direction evaluation every 10 frames using LOF (Local Outlier Factor), and centroid computation. When LOF detects outliers, each outlier is either reassigned to a nearby cluster or placed in an unassigned list. The nested clustering procedure is triggered again once the number of elements in the unassigned list reaches a predefined threshold. The Centroid Trajectory Calculation component describes how each cluster computes the acceleration of its centroid trajectory based on the average deviation value derived from its members. The Cluster Assignment component demonstrates how the algorithm dynamically performs the clustering process.

Our evaluation function consists of three main metrics: Cluster Trajectory Errors Occurrence (CTEO): It counts the number of trajectory errors in each cluster and divides that by the total number of clusters, as shown in equation **??**.

$$CTEO = Average(\frac{1}{f_i}\sum_{t=0}^{f} 1[dist_t > T]) \tag{1}$$

$dist$ is the distance between the location or direction of a centroid between its frames, T is a threshold for noticeable distance, n is the cluster number and f is the frame number on each centroid.

Cluster Trajectory Errors Length (CTEL).It counts every cluster trajectory error length and divides it by the number of clusters, as shown in equation 2.

$$\text{CTEL} = Average(\sum_{t=1}^{f} \begin{bmatrix} dist_t & \text{if } distl_t > T \\ 0 & \text{otherwise} \end{bmatrix}) \tag{2}$$

Cluster Member Distance Deviations (CMDD): This metric only considers the centroids that have more than two and shows how far the member is from its centroid, as shown in equation 3.

$$\text{CMDD} = Average\left(\frac{1}{z_i}\sum_{j=0}^{z_i} |ped_t - C_t|\right) \tag{3}$$

where z_i is the number of pedestrians in cluster i, ped is the member locations or direction, and C_t is the cluster locations or direction. The direction took a more crucial part since

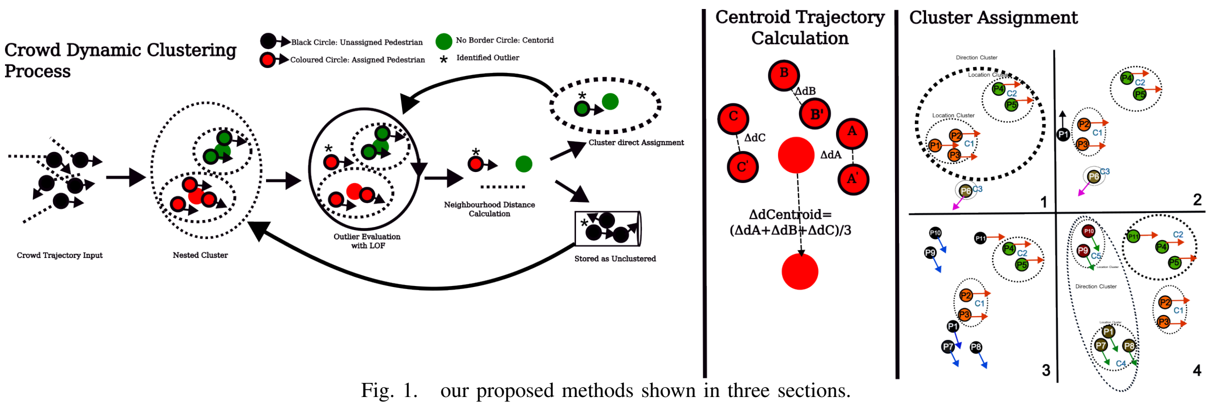

Fig. 1. our proposed methods shown in three sections.

TABLE I

CLUSTERING EVALUATION FOR CROWDHD DATASET

Data	CMDD ↓	CTEO ↓	CTEL ↓	ncluster	npeds
HT-02	7.93	0.0	0.0	308/460	1264
GT-02	6.77	0.26	0.01	267/378	602
HT-03	7.04	0.23	0.01	443/882	2577
GT-03	5.60	0.98	0.02	353/482	811
HT-04	7.99	0.31	0.01	281/653	1387
GT-04	8.084	0	0.0	270/367	580

pedestrians can not be in a cluster if they move in a different direction.

III. Brief Experiment and Discussion

Our experiment data are based on the CrowdHD dataset[7]. We make two parts of the experiment, the first part is to evaluate the clustering performance, and the second is to evaluate how this clustering approach can assist the existing trajectory prediction. In this paper, we have only shown the experiment, which is based on the MART algorithm[8]; meanwhile, the other algorithms are shown in our main paper.

Table I presents a cluster evaluation based on automated tracking outputs (HT) and ground-truth trajectories (GT). All HT scenarios demonstrate consistently low CMDD values, indicating a minimal average distance across all frames. The ncluster column represents the number of clusters formed, ranging from clusters with a single member to those containing more than two members, while npeds indicates the total number of pedestrians.

Table II illustrates a substantial reduction in execution time and demonstrates that the clustering approach outperforms both the tracking-based results and traditional random selection methods in terms of execution time and memory usage, while showing only a slight reduction in ADE and FDE.

IV. Conclusion and future Research

This research lays the groundwork for future investigations into real-time prediction by simplifying dense crowd trajectory data on preprocessed steps. This research lays the groundwork for future investigations into real-time prediction by simplifying dense crowd trajectory data. Crowd is a challenging scenario for trajectory prediction, not only of

TABLE II

TRAJECTORY PREDICTION PERFORMANCE WITH MART

Scene	Source	Exec(s)↓	ADE 50 ↓	FDE 50 ↓	Mem↓(MiB)
02	GT	44905	5.56	8.61	34731
02	Cluster	**14179.78**	6.22	9.78	**16167**
02	Tracking	18849.30	**5.99**	**9.32**	25912
02	Random	8191	6.37	10.16	15167
03	GT	89113.05	6.59	2.72	55976
03	Cluster	**10040.14**	9.77	16.18	**21189**
03	Tracking	20547.49	**8.89**	**15.00**	26864
03	Random	8564	10.66	17.49	20414
04	GT	22673.73	6.24	9.12	23457
04	Cluster	**9356.21**	7.96	12.51	**8476**
04	Tracking	12582.06	**7.18**	**11.03**	12802
04	Random	7634.674	7.47	11.67	8077

the massive data that comes with noise but also dynamic behaviour, which means it has evolved pattern depend on certain conditions. In order to solve this, we currently work on a crowd dataset that has real-world coordinates, longer duration, and pattern varieties. Therefore then can enable work on trajectory prediction in dynamic dense crowd scenarios.

References

[1] A. Corbetta and F. Toschi, "Physics of human crowds," *Annual Review of Condensed Matter Physics*, vol. 14, pp. 311–333, 3 2023.

[2] B. F. Warren William, K. Yoshida, T. Wirth, and B. A. Free, "Human crowds as social networks: Collective dynamics of consensus and polarization," *Perspectives on Psychological Science*, vol. 19, 2024.

[3] A. Dehghan and M. Shah, "Binary quadratic programing for online tracking of hundreds of people in extremely crowded scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, 2018.

[4] T. Salzmann, B. Ivanovic, P. Chakravarty, and M. Pavone, "Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data," in *Computer Vision – European Conference on Computer Vision*, 2020, pp. 683–700.

[5] P. Xu, J.-B. Hayet, and I. Karamouzas, "Socialvae: Human trajectory prediction using timewise latents," pp. 511–528, 2022.

[6] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social lstm: Human trajectory prediction in crowded spaces," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 6 2016, pp. 961–971.

[7] R. Sundararaman, C. D. A. Braga, E. Marchand, and J. Pettre, "Tracking pedestrian heads in dense crowd," in *Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2021.

[8] S. Lee, J. Lee, Y. Yu, T. Kim, and K. Lee, "Mart: Multiscale relational transformer networks for multi-agent trajectory prediction," in *ECCV(European Conference of Computer Vision)*. ECVA(Europen Computer Vision Association), 2024, pp. 89–107.