IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 14, 2020, accepted August 4, 2020, date of publication August 10, 2020, date of current version August 19, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3015375

Scene-Independent Motion Pattern Segmentation in Crowded Video Scenes Using Spatio-Angular Density-Based Clustering

ABHILASH K. PAI[1], A. KOTEGAR KARUNAKAR[1], (Senior Member, IEEE), AND U. RAGHAVENDRA[2]

[1]Department of Computer Applications, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India
[2]Department of Instrumentation and Control Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India

Corresponding authors: A. Kotegar Karunakar (karunakar.ak@manipal.edu) and U. Raghavendra (raghavendra.u@manipal.edu)

ABSTRACT Motion pattern segmentation for crowded video scenes is an open problem because of the inability of existing approaches to tackle unpredictable crowd behaviour across varied scenes. To address this problem, we propose a Spatio-Angular Density-based Clustering (SADC) approach, which performs motion pattern segmentation by clustering the spatial and angular information obtained from the input trajectories. The k-nearest neighbours of each trajectory and the angular deviation between trajectories constitute the spatial and angular information, respectively. Effective integration of the spatio-angular information with an improvised density-based clustering algorithm makes this approach scene-independent. The performance of most clustering algorithms in the literature is parameter-driven. Choosing a single parameter value for different types of scenes decreases the overall clustering performance. In this article, we have shown that our approach is robust to scene changes using a single threshold, and, through the analysis of parameters across eight commonly occurring crowded scenarios, we point out the range of thresholds that are suitable for each scene category. We evaluate the proposed approach on the benchmarked CUHK dataset. The experimental results show the superior clustering performance and execution speed of the proposed approach when compared to the state-of-the-art over different scene categories.

INDEX TERMS Clustering, crowd analysis, crowd behaviour analysis, crowd flow segmentation, group detection, motion pattern segmentation.

I. INTRODUCTION

Pedestrians in a crowded scene exhibit interesting patterns of motion over time. Analysing these motion patterns across different types of crowded scenes helps to understand complex crowd behaviours, detect anomalous behaviours and predict unforeseen events which could pose a threat to the safety of the crowd.

Motion pattern segmentation is an automated visual surveillance task that divides a scene into regions of consistent and coherent motion. Grouping the patterns of crowd motion simplifies the process of crowd behaviour understanding/recognition and crowd anomaly detection [1]–[4] (the terms motion pattern segmentation and group detection are used interchangeably and have the same meaning in the context of this article). In spite of various efforts [5]–[10], precise motion pattern segmentation remains a challenging task due to varying crowd dynamics across different types of scenes and intricate interactions between the pedestrians within a scene. In general, a crowded scene can be either structured or unstructured. In a structured crowded scene, the direction of motion of the crowd remains same for most of the time and is easily predictable. Whereas, in an unstructured crowded scene, the direction of motion of the crowd changes chaotically and is unpredictable. Hence, a model which is designed specifically for one type of crowded scene need not be efficient with the other. Most of the attempts to tackle the problem of motion pattern segmentation perform less efficiently when the type of scene changes, which results in wrongly detected segments.

In this article, we propose a scene-independent motion pattern segmentation approach by using the spatial as well

The associate editor coordinating the review of this manuscript and approving it for publication was Yongqiang Zhao.

as angular features of a trajectory and subsequently applying an improvised density-based clustering algorithm to group trajectories exhibiting similar motion patterns. The input trajectories generated by using the generalized KLT (gKLT) based key-point tracker [7] are averaged for a given period of time. The average spatial location and average angular orientation of each trajectory constitute its spatial and angular features, respectively. The spatio-angular features are then used to generate spatial information in terms of k-nearest neighbours of each trajectory and angular information in terms of angular deviation between trajectories. Finally, the spatio-angular information is used by the density-based clustering algorithm to group similar motion patterns. For simplicity, we denote our approach as SADC (Spatio Angular Density-based Clustering).

The proposed SADC algorithm is evaluated on the publicly available standard CUHK dataset [8] containing different types of scenes of varying densities and perspectives with ground truth available for motion pattern segmentation. For analysis of performance across different kinds of scenes, we divide the CUHK dataset into meaningful scene categories. Comparison of our results with the state-of-the-art motion pattern segmentation techniques shows that the proposed SADC approach is efficient and computationally faster than other techniques.

The four main contributions of this article are: (i) a faster, efficient and scene-independent method for grouping similar motion patterns, (ii) an averaging-based approach to extract meaningful features of crowd motion, (iii) an improvised density-based clustering algorithm where cluster membership is decided based on closeness of spatial and angular features of a trajectory when compared to other trajectories and (iv) a detailed analysis across different scene categories, which proposes a range of parameter values applicable for each scene category.

II. RELATED WORK

There have been numerous attempts to improve the efficiency of the process of motion pattern segmentation in crowded scenes starting from the Lagrangian Particle Dynamics based approach proposed by Ali and Shah [5]. A concise review of these approaches can be found in [11]. According to [11], these approaches have been divided into three categories: flow field model-based, similarity-based, and probability model-based approaches. Most of these approaches either use homogeneous datasets (in terms of scene dynamics) or they create ground truth which are in-line to their proposed method. In this article we use a similarity-based clustering approach which is evaluated over the CUHK dataset, containing diverse real-time scenarios along with ground truth for motion pattern segmentation. Hence, the remainder of this section discusses about: (i) the notable and recent efforts towards efficient motion pattern segmentation/group detection using CUHK dataset, (ii) the approaches which are related to similarity-based trajectory clustering.

It was Zhou *et al.* [7] who initially created a Collective motion dataset in order to measure the collective behaviour within the crowd. In their proposed approach, an algorithm called Collective Merging (CM) was used to find coherent groups of trajectories within the crowd. The CM algorithm models (i) the local behaviour of the crowd using a weighted k-Nearest Neighbour (k-NN) graph and (ii) the global behaviour among the non-neighbours by finding similar paths in the k-NN graph. Subsequently Shao *et al.* [8] created the CUHK dataset from the Collective Motion dataset by adding new scenes and defining ground truth for the detection of similar motion patterns/groups within a scene (more details about CUHK can be found in Section IV-A). In their work, a Collective Transition (CT) algorithm was proposed for group detection. The CT algorithm improved an earlier approach for group detection based on Coherent Filtering (CF) [6], by modelling the trajectory data using Markov Chains. Both the CF and CT based algorithms used the concept of Coherent Neighbour Invariance (initially introduced in [6]) to keep track of the persistent members within a group over the course of time.

In another interesting work, Wang *et al.* [9] used a thermal diffusion-based model to generate a strong coherent motion field (known as the thermal energy field) from the noisy optical flow field computed from the input crowd video. Triangulation-based boundary detection, watershed segmentation algorithm and two-step graph-based clustering strategy were applied over the thermal energy field to cluster coherent motion regions. Trojanova *et al.* [12] adopted a weighted k-NN graph-based clustering approach, which used a data-driven threshold as compared to a static threshold in [7]. Fan *et al.* [13] used a Natural Nearest Neighbour algorithm (3N) to adaptively determine the optimal number of the nearest neighbours (the k-value) as compared to k-NN based approach where the k-value needs to be experimentally determined. In their work, the 3N algorithm generates a crowd motion network from which similar motion patterns are detected using the concept of coherent neighbour invariance. A different approach called Hybrid Social Influence Model (HSIM) was proposed by Ullah *et al.* [10], which used a density-independent version of the Social Force Model [14], [15] to model crowd motion and Communal model [16] to group similar motion patterns. The topic models which are popular in language processing, have been used by Chen *et al.* [17] for group detection. In their proposed approach, after dividing the input crowd image into a fixed number of patches (using a Simple Linear Iterative Clustering algorithm), a descriptor was computed for each patch by combining the feature points generated by the gKLT tracker and orientation distribution of each feature points within the patch. The Latent Dirichlet Allocation-based model is then combined with the Markov Random Field to determine the topics. Based on these topics the features are grouped together. In a recent work, Wang *et al.* [18] proposed a self-weighted multiview clustering approach that combines an orientation-based graph and a structural context-based graph

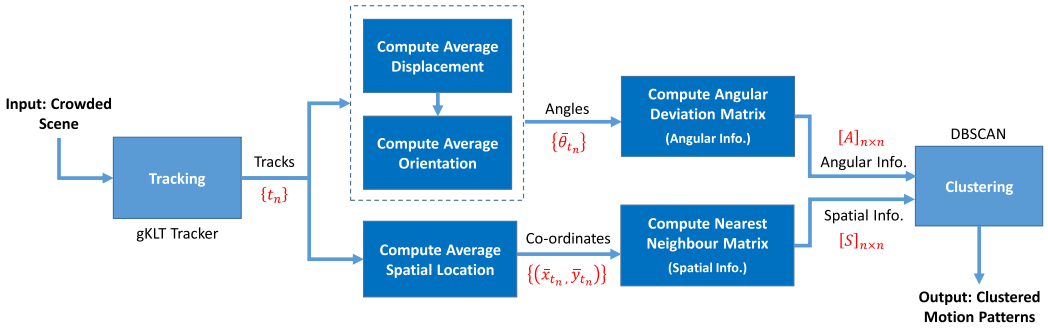


FIGURE 1. Overview of the proposed approach where, n - no. of trajectories, gKLT Tracker - Generalized KLT Tracker, $\{t_n\}$ - Set of trajectories, $\{\bar{\theta}_{t_n}\}$ - Set of averaged angular orientations, $\{(\bar{x}_{t_n}, \bar{y}_{t_n})\}$ - Set of averaged spatial location co-ordinates, $[A]_{n\times n}$ - Angular deviation matrix, $[S]_{n\times n}$ - Nearest Neighbour Matrix.

and apply a tightness-based merging strategy to detect groups within the crowd.

The approaches in [6]–[9] detect groups frame-by-frame which results in group-switching (or cluster-switching) for each frame. Motion pattern segmentation being a prior to crowd behavioural monitoring/ anomaly detection process, regular switching of groups/clusters creates inconsistent results for these processes. Our approach is similar to He and Liu [19] who used a Density-based Clustering algorithm to perform motion pattern analysis in crowded scenes. However, there are a couple of differences in the approaches: (i) to extract and represent the motion information from input crowded video scene, we compute the averaged trajectory generated from an efficient and accurate KLT tracker [6], [20] compared to He *et al.* who used a global optical flow field computed from the traditional Lucas Kanade-based approach [21], (ii) to determine the spatial proximity, we employ a nearest neighbour-based strategy compared to the Euclidean-distance-based thresholding approach which is scene-dependent and fails when the scene-perspective changes.

The task of motion pattern segmentation can also be considered as a trajectory clustering problem. Trajectory clustering have been extensively researched and the recent surveys on moving object trajectory clustering methods presented in [22]–[24] point out the challenges faced. Finding a suitable measure to compute the similarity among trajectories with varied properties and finding a suitable algorithm to cluster the trajectories based on their similarities are the two main challenges in trajectory clustering. Various similarity measures [25]–[28] and clustering algorithms [18], [19], [29]–[32] have been used for trajectory based motion pattern analysis. In our approach we average the trajectories and utilise the averaged vector information to perform clustering.

Most of the motion pattern segmentation approaches discussed so far are either applicable to specific categories of crowded scene, or suffer from excessive cluster switching or are computationally intensive. This article proposes a spatio-angular density-based clustering approach which is efficient and stable for different scene categories, which has minimal cluster switching and is computationally faster.

III. SPATIO-ANGULAR DENSITY-BASED CLUSTERING (SADC) FOR MOTION PATTERN SEGMENTATION

This article considers scenarios involving sparse to dense crowds in shopping malls, train stations, escalators, street/sidewalk/market, crosswalks, road-traffic, marathon, military-parade, public events, other indoor and outdoor scenes which are under surveillance for crowd management using static cameras. According to prior research [7], [33], the social behaviour of people walking in groups is such that all the members within the group tend to move in same direction and are spatially close to each other. Based on these scenarios one can infer that a set of spatially close motion patterns are considered to be in the same group, if they move together in the same direction. This work proposes a spatio-angular density-based clustering for detecting such coherent groups. The block diagram shown in Figure 1 illustrates the three phases of proposed approach, namely, (i) Extraction of motion information, where motion information in the form of trajectories (tracks) is extracted from the input video by tracking key-points using a gKLT Tracker, (ii) Computation of Angular & Spatial Information from the motion features, where the motion features in the form of average angular orientation and average spatial location (computed from each of the trajectories) are used to create angular and spatial information matrices, respectively, (iii) Improvised Density-based clustering, which generate clusters containing similar motion patterns by considering the angular and spatial information. The proposed approach is explained in detail in the following sections.

A. EXTRACTION OF MOTION INFORMATION

Given an input video, the first step is to capture the motion of the crowd. This is usually done by tracking a set of key-point features across the frames in the video which results in the generation of trajectories (or tracks) [34]. The problem of extracting motion information is now a tracking problem.

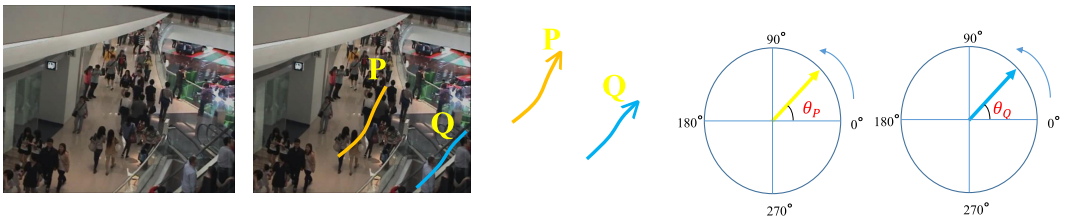


FIGURE 2. (a) A frame from a crowded video scene, (b) Two motion patterns P & Q, (c) P & Q represented in space with arrow pointing towards the direction of motion, (d) The directions of the motion patterns P & Q represented as angles θ_P & θ_Q respectively in the polar co-ordinate system (Best viewed in color).

The proposed approach uses the generalized KLT tracker (gKLT tracker [7] which is derived from the traditional KLT tracker [20]) because of its tracking accuracy and computational efficiency. Each of the generated trajectories is a combination of 2-D spatial co-ordinates $\{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$, where $T = 1 : m$. The key-points which are tracked may lie in the background or generated due to illumination variations resulting in the generation of short, static & noisy trajectories. In our approach, such trajectories are filtered out using the method by Shao *et al.* [8] which improves the overall accuracy of the motion pattern segmentation. The following subsection explains the next phase, where the refined trajectories are used to extract motion features from which angular and spatial information are generated.

B. COMPUTATION OF ANGULAR AND SPATIAL INFORMATION FROM THE MOTION FEATURES

Two types of motion features are extracted from the refined trajectories namely angular orientation and spatial location. The angular orientation feature gives the direction of motion of a crowd. If we consider only the directions of motion patterns P & Q (as shown in Figure 2(a) & 2(b)) as a feature vector, it gets clustered as a same set of motion trajectories. In reality, they should form different set of trajectories. Hence, this work introduces a second motion feature, the spatial location, that enables to identify trajectories that are spatially distant from each other.

Since the length of the trajectories are not same, computing the pair-wise similarities between them requires normalization of trajectory lengths. Averaging the trajectory data is not only an effective way to normalize the length of the trajectories but also enables to capture the behaviour of the trajectories over time. The refined trajectory data is averaged before computing the two motion features as suggested in [12]. If the trajectory data for a trajectory t_i is represented as $t_i = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$, where m is the length of the trajectory, its average spatial location feature ($\bar{s_{t_i}}$) is computed as follows:

$$\bar{s_{t_i}} = \frac{1}{m}\sum_{k=1}^{m}(x_{k_{t_i}}, y_{k_{t_i}}) \tag{1}$$

To obtain the average angular orientation feature, the average displacement ($\overline{V}_{t_i}$) of a trajectory is computed first as follows:

$$\overline{V}_{t_i} = \frac{1}{(m-1)}\sum_{k=2}^{m}\left((x_{k_{t_i}} - x_{(k-1)_{t_i}}), (y_{k_{t_i}} - y_{(k-1)_{t_i}})\right) \tag{2}$$

where, the average displacement vector, $\overline{V} = [\overline{u}, \overline{v}]$, consists of an x-component ($\overline{u}$) and a y-component ($\overline{v}$). The average angular orientation feature ($\overline{\theta}_{t_i}$) is then computed as follows:

$$\overline{\theta}_{t_i} = \begin{cases} cos^{-1}\left(\dfrac{\overline{V}.\hat{V}}{||\overline{V}|| * ||\hat{V}||}\right) * \dfrac{180}{\pi}, & \overline{v} > 0 \\ \left[2\pi - cos^{-1}\left(\dfrac{\overline{V}.\hat{V}}{||\overline{V}|| * ||\hat{V}||}\right)\right] * \dfrac{180}{\pi}, & \overline{u} \neq 0, \overline{v} \leq 0 \\ 0, & \overline{u}, \overline{v} = 0 \end{cases} \tag{3}$$

where $\hat{V} = [0, 1]$ is the unit vector in the horizontal direction. The value of $\overline{\theta}_{t_i} \in (0, 2\pi - 1)$ varies according to the values of the vectors $\overline{u}$ and $\overline{u}$ and is equal to zero when there is no motion. Computing the average angular orientation in this manner i.e., orientation value obtained from averaged displacement vectors rather than the orientation value obtained by averaging the orientations of the individual displacement vectors, enables to implicitly encode the effect of magnitude information as well. In the case of abrupt changes in the direction of a trajectory, computing average angular orientation for the entire path leads to erroneous motion features. By using a fixed set of frames (rather than the entire set of frames containing a trajectory) for computing both the angular and spatial features, the error incurred is reduced. The number of frames considered for averaging is determined empirically.

The obtained motion features ($\bar{s_{t_i}}$ & $\overline{\theta}_{t_i}$) are then used to generate matrices containing angular and spatial information respectively. The angular information matrix ($[A]_{n\times n}$) contains the pair-wise angular deviation between average angular features of two trajectories t_i & t_j. The distance function ('$d_{angular}$') used to compute the angular deviation between the average angular features $\overline{\theta}_{t_i}$ & $\overline{\theta}_{t_j}$ is defined as follows:

$$d_{angular}(\overline{\theta}_{t_i}, \overline{\theta}_{t_j}) = min(|\overline{\theta}_{t_i} - \overline{\theta}_{t_j}|, 2\pi - |\overline{\theta}_{t_i} - \overline{\theta}_{t_j}|) \tag{4}$$

where the function $d_{angular} \in (0, \pi)$ overcomes the issues with computing distance (angular deviation) in circular domain.

On the other hand, the spatial information matrix conveys whether the trajectories are spatially close to each other and is constructed by finding the k-Nearest Neighbours for each trajectory. Pair-wise Euclidean distance between the average spatial location features is used to find the nearest neighbours. The spatial information-based matrix is defined as follows:

$$S(i,j) = \begin{cases} 1, & \text{if } t_j \in N(t_i) \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where $N(t_i)$ consists of a set of k-Nearest Neighbours of the trajectory t_i. The value of 'k' depends on the number of trajectories (n) in a scene and is computed as follows:

$$k = \alpha * n \tag{6}$$

where $\alpha \in [0,1]$ denotes how much percentage of the total number of trajectories must be considered to determine the value of k. Using a density-based clustering algorithm, the angular and spatial information matrices are then combined to form clusters of similar trajectories.

C. IMPROVISED DENSITY-BASED CLUSTERING

The obtained angular and spatial information depicts how the motion patterns are spread across the scene. Finding similar motion patterns and grouping them (using the obtained information) becomes a clustering problem. The crowded scenes can contain motion patterns of arbitrary shape/size and can have motion patterns which does not belong to any group (noisy data). A density-based clustering algorithm would be an obvious choice in such cases. In the proposed approach, we use the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm. As proposed by Ester *et al.* [35], DBSCAN consists of two parameters: (i) the *Eps*-neighbourhood and (ii) *MinPts* parameter. In this approach, the Eps-neighbourhood value, which is essentially a threshold, determines the maximum angular deviation that is allowed between the average angular features of two trajectories. We improvise the Eps-neighbourhood criteria by using an angular threshold (*Eps*-neighbourhood value is called as λ_θ in our work) and then combine it with the spatial information to decide the cluster membership. The inclusion of spatial information is crucial, because it removes the distant trajectories with orientation less than the angular threshold (λ_θ). Further, the inclusion of spatial information makes our the proposed motion pattern segmentation algorithm resistant to scene perspective changes. Therefore, the trajectories which have similar orientation (defined by the λ_θ parameter) and are spatially close to each other (defined by the α parameter) belong to the same cluster. This is consistent with our definition in Section III. The proposed approach along with the improvised section of DBSCAN is shown in Algorithm 1.

Algorithm 1 Spatio-Angular Density-Based Clustering (SADC)

Input: Set of 'n' trajectories, $\{t_n\}$ from a crowded scene.
Output: 'm' clusters, where each cluster contains similar motion patterns

for k = 1 to n **do**
1: Compute $\overline{\theta}_{t_k}$ and $\overline{s}_{t_k}$ using Eq. 3 and Eq. 1, respectively.
2: Compute the matrices $A_{k\times k}$ from $\{\overline{\theta}_{t_k}\}$ and $S_{k\times k}$ from $\{\overline{s}_{t_k}\}$ using the distance functions defined in Eq. 4, Eq. 5 and Eq. 6, respectively.
end for
3: Perform DBSCAN to obtain 'm' clusters of similar motion features:
for each (i, j) in A and S **do**
if $A(i,j) \leq \lambda_\theta$ and $S(i,j) == 1$ **then**
Assign t_j to t_i's cluster.
end if
end for

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. DATASET

CUHK dataset [8] is the only publicly available standard dataset with ground truth for motion pattern segmentation. The dataset contains a total of 474 videos captured from 215 scenes out of which 300 videos are used for the purpose of motion pattern segmentation. The proposed approach is evaluated on these 300 videos captured across different types of scenes. For the purpose of analysis, these videos are divided (manually) into different categories based on (i) the property of crowd movement as (a) structured, (b) unstructured, (ii) the location of the scene as (a) Indoor, (b) Outdoor and (iii) the type of the scene as (a) cross walk, (b) escalator, (c) mass movement, (d) market (e) public walkway, (f) shopping mall, (g) street. Figure 3 and Figure 4 shows example scenes and statistics, respectively for this categorization.

B. EXPERIMENTAL SET-UP

The experiments are performed, over all the 300 scenes of CUHK-dataset, on a personal computer with Intel Core *i5*-8400@2.80 GHz processor and 8 GB RAM. Firstly, we replicate the experimental set-up used by Shao *et al.* [8] to extract motion information. The gKLT tracker is initialized with a set of 3000 key-points which are tracked across the frames in the video. Short/static/noisy trajectories are filtered out by discarding those having their trajectory length less than 10-frames & those with zero-displacement vectors for more than half of its duration. Secondly, for the purpose of grouping, trajectory data from multiple frames (30-frames) are considered in our approach. By doing so, we aim to obtain more information on the history of the path followed by a trajectory. Finally, for the purpose of clustering, the values for angular threshold & the α-value for determining the spatial threshold,

FIGURE 3. Example scenes for (a) cross walk, (b) escalator, (c) market, (d) mass movement, (e) public walkway, (f) shopping mall, (g) station, and (h) street. Scenes (d) and (f) are good examples for structured and unstructured, indoor and outdoor scenes respectively.

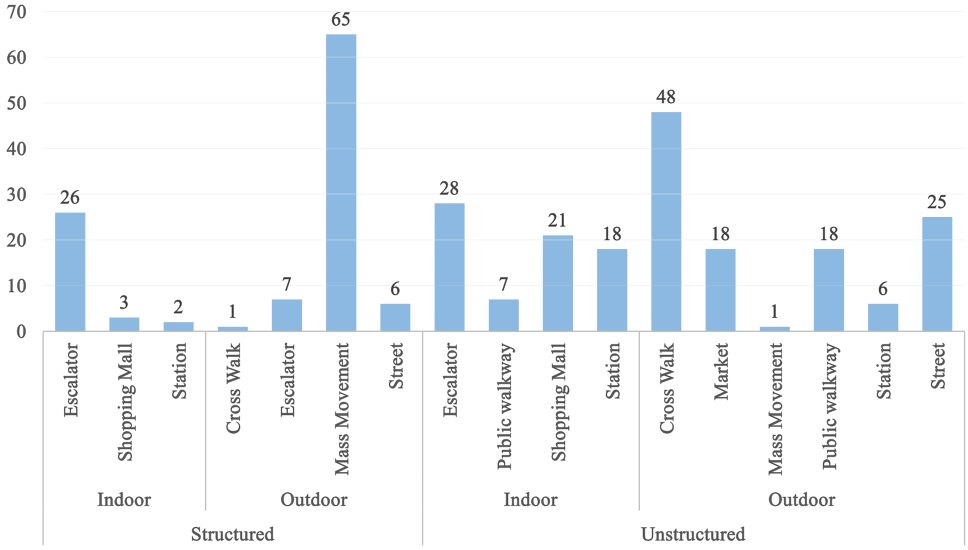


FIGURE 4. CUHK dataset categorization statistics. It can be observed that most of the Mass Movement scenes are Structured. Otherwise, majority of the scenes within the dataset are Unstructured.

are chosen empirically ($\lambda_\theta = 20°, \alpha = 15$), whereas the value for minimum points to form a cluster is set to $MinPts = 5$.

C. RESULTS

To prove the effectiveness of the proposed approach, its performance is compared with the state-of-the-art approaches CM [7] and CT [8], whose binaries are publicly available. Note that, we have not considered CF for comparison because the CT algorithm (which is considered) is an improvement over CF. We have kept the original setting for the parameters of CM and CT. Since both the approaches generate clusters for each frame, for the purpose of evaluation and comparison, we choose the clusters generated at a particular frame (for each scene) as defined in the CUHK-dataset and align the proposed approach accordingly.

1) QUALITATIVE RESULTS

The category-wise qualitative results are shown in the Figure 5. It can be observed that the proposed SADC-algorithm performs well in all categories of the CUHK dataset even in case of complex scenes (Figure 5, columns 5-8). In fact, most of the clusters generated by the SADC-algorithm are more closer to the ground truth clusters when compared to CM [7] and CT [8]. The CM-algorithm uses an graph-based approach which relies on finding similar weighted paths and a connected component-based algorithm which clusters the similar nodes on the graph based on a threshold. While the path-based approach is effective, the less effective connected component-based clustering does not generate accurate clusters when scene-type changes. Also, the CM algorithm is dependent on the K-parameter value,

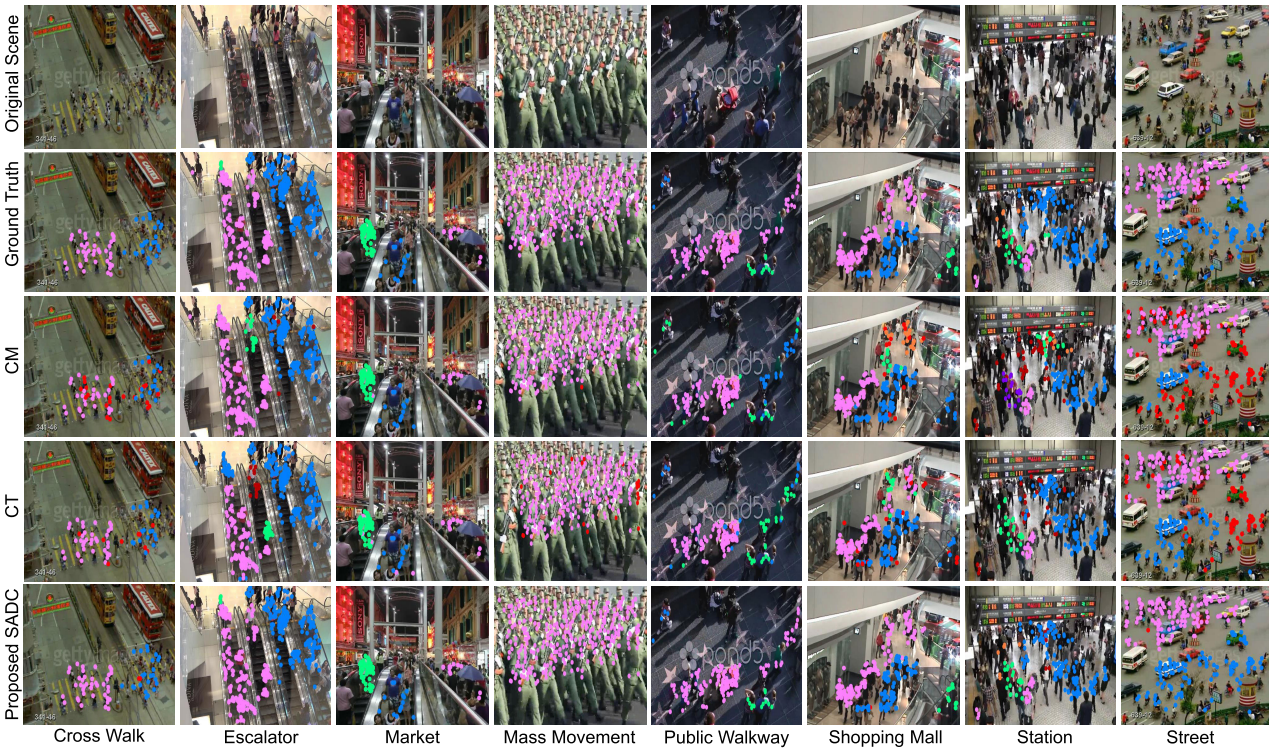


FIGURE 5. Qualitative comparison between the proposed approach (SADC), CM [7] and CT [8] across the eight-categories of CUHK-dataset. The trajectory-points with similar motion patterns have same color. Red-color represents noise points (Best viewed on screen).

which is another reason for the inconsistent results. The CT-algorithm improves the CF-algorithm using a Markov-chain based approach. Since, the CT-algorithm is built on top of the CF-algorithm, it is partly-dependant on the output of the CF-algorithm due to which some trajectory-points belonging to a particular cluster in the ground truth are considered as noise by the CT-algorithm. This is because, the trajectory key-points which are labelled as noise by CF-algorithm (not-included in the clustering result) is not considered for building the Markov-chain and remain as noise (red-color points). Furthermore, CM, CF and CT uses a frame-by-frame approach to generate the clusters. Therefore, for each pair of frames, only the key-points vectors at that instance is considered for clustering. Due to this, the cluster members keep changing and results are inconsistent across the duration of the video, which is another reason for the occurrence of noise key-points in Figure 5. Whereas, the proposed SADC uses an averaging-based approach that extracts the history of the trajectory (for a fixed set of 30-frames) in terms of average angular and spatial features. This results in the reduction of noise key-points (as seen in Figure 5) which leads to consistent clustering with minimal cluster switching.

2) QUANTITATIVE RESULTS

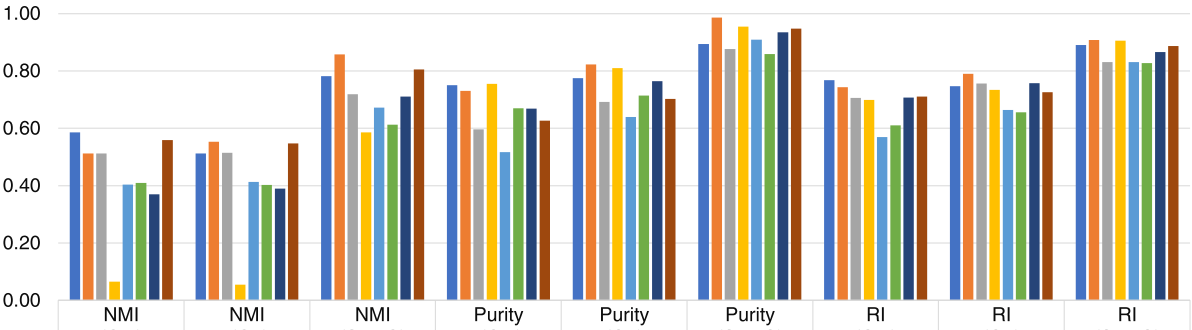
For quantitative evaluation, motion pattern segmentation is considered as a clustering problem and the performance evaluation done using four widely used external cluster evaluation measures such as Purity [36], Normalized Mutual Information (NMI) [37] & Rand Index (RI) [38]. All the performance measures are in the range of [0,1], where a higher value indicates a better clustering performance. The comparison results in Table 1 shows the average performance for 300 scenes of CUHK-dataset and indicates that the proposed approach outperforms the CM and CT-based approaches.

TABLE 1. Quantitative comparison of proposed approach with the state-of-the-art approaches (averaged for 300 scenes).

Approach	NMI	Purity	RI
CM [7]	0.43	0.69	0.70
CT [8]	0.41	0.76	0.73
Proposed SADC	**0.72**	**0.94**	**0.88**

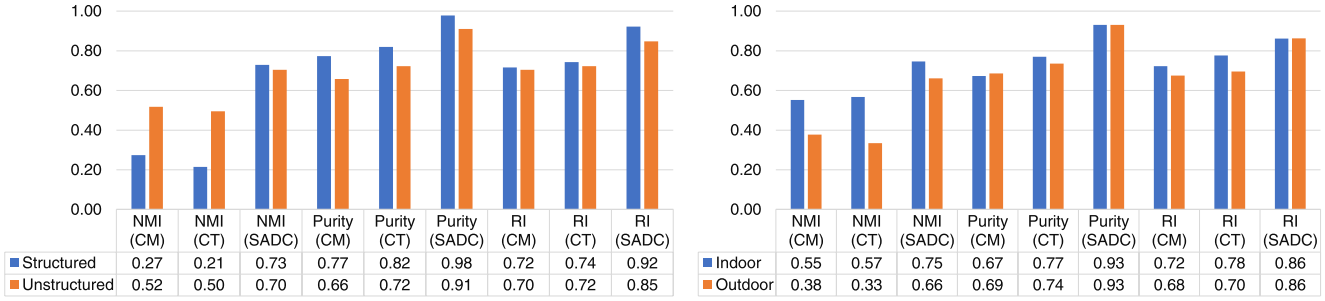
Notice that, the NMI value for CT is lesser than CM, but CT performs better than CM for other metrics. On inspection, we found that the NMI-metric always generates NMI $= 0$ if either one of the two clustering assignment (ground truth or clustering result) has a single cluster but the other clustering assignment has more than one cluster, which is not desirable. This is due to an underlying mathematical issue with the computation of Mutual Information, the discussion of which is beyond the scope of this article. Out of 300 scenes in CUHK dataset, 25% of scenes are having one-cluster ground truth. This is one of the main reason for other literatures reporting a lower value of NMI when evaluated over the CUHK dataset.

Figures 6 & 7 shows the comparison results over different scene categories as defined in Section IV-A. For fair



	NMI (CM)	NMI (CT)	NMI (SADC)	Purity (CM)	Purity (CT)	Purity (SADC)	RI (CM)	RI (CT)	RI (SADC)
Cross Walk	0.59	0.51	0.78	0.75	0.77	0.89	0.77	0.75	0.89
Escalator	0.51	0.55	0.86	0.73	0.82	0.99	0.74	0.79	0.91
Shopping Mall	0.51	0.51	0.72	0.60	0.69	0.88	0.71	0.76	0.83
Mass Movement	0.07	0.05	0.59	0.76	0.81	0.95	0.70	0.73	0.91
Public Walkway	0.40	0.41	0.67	0.52	0.64	0.91	0.57	0.66	0.83
Station	0.41	0.40	0.61	0.67	0.71	0.86	0.61	0.66	0.83
Street	0.37	0.39	0.71	0.67	0.76	0.93	0.71	0.76	0.87
Market	0.56	0.55	0.80	0.63	0.70	0.95	0.71	0.73	0.89

FIGURE 6. Quantitative comparison result based on type of scene (Best viewed in color).



	NMI (CM)	NMI (CT)	NMI (SADC)	Purity (CM)	Purity (CT)	Purity (SADC)	RI (CM)	RI (CT)	RI (SADC)
Structured	0.27	0.21	0.73	0.77	0.82	0.98	0.72	0.74	0.92
Unstructured	0.52	0.50	0.70	0.66	0.72	0.91	0.70	0.72	0.85

	NMI (CM)	NMI (CT)	NMI (SADC)	Purity (CM)	Purity (CT)	Purity (SADC)	RI (CM)	RI (CT)	RI (SADC)
Indoor	0.55	0.57	0.75	0.67	0.77	0.93	0.72	0.78	0.86
Outdoor	0.38	0.33	0.66	0.69	0.74	0.93	0.68	0.70	0.86

FIGURE 7. Quantitative comparison result based on - Left: Property of crowd movement and Right: Scene location (Best viewed in color).

evaluation, we have considered equal number of scenes for each scene-category by randomly sampling the scenes. The proposed SADC algorithm performs not only the best among the three algorithms but also performs above par in all scene categories. This is because of the fact that, (i) SADC considers the history of the trajectory for multiple frames, which plays a crucial role during clustering, (ii) SADC computes the angular deviation between the averaged trajectory vectors and is well complemented by the k-nearest neighbour method, both of which are effective inputs to perform efficient density-based clustering for different scene categories. The NMI issue, as discussed earlier, is clearly evident from Figure 6 & 7, where the Mass Movement scene category has average NMI value close to zero for CM and CT and low for SADC (in this case, the value for RI-metric can be considered). Mass Movement category involves scenes like Marathon, Military Parade, Protest Rallies whose ground truth, in most cases, contains only one motion pattern.

D. PARAMETER ANALYSIS

As discussed in Section III-B & III-C, the proposed approach depends on two parameters, namely λ_θ (works on angular information) & α (works on spatial information). Figure 8 & 9 demonstrates the effect of varying these parameters. Figure 8 shows the effect of combining the spatial information along with angular information. Among the two profiles seen in Figure 8, the one which is generated after including both spatial and angular information is better than the other profile where only angular information is included (discussed briefly in Figure 2). This asserts the need to combine spatial information with angular information. It can also be observed that, the performance of the proposed algorithm gradually decreases with increase in λ_θ value which means that a larger value for λ_θ would result non-similar motion patterns being grouped together.

Figure 9 indicates the performance profiles of SADC algorithm with various combinations of λ_θ & α. The similar λ_θ profiles for varying values of α indicates the dominance of angular component over the spatial component in making decisions regarding clustering. Although the profiles look similar, the one with $\alpha = 15$ is observed to be more stable than the others. Also, it can be noticed that λ_θ value between 10° and 30° performs better on an average.

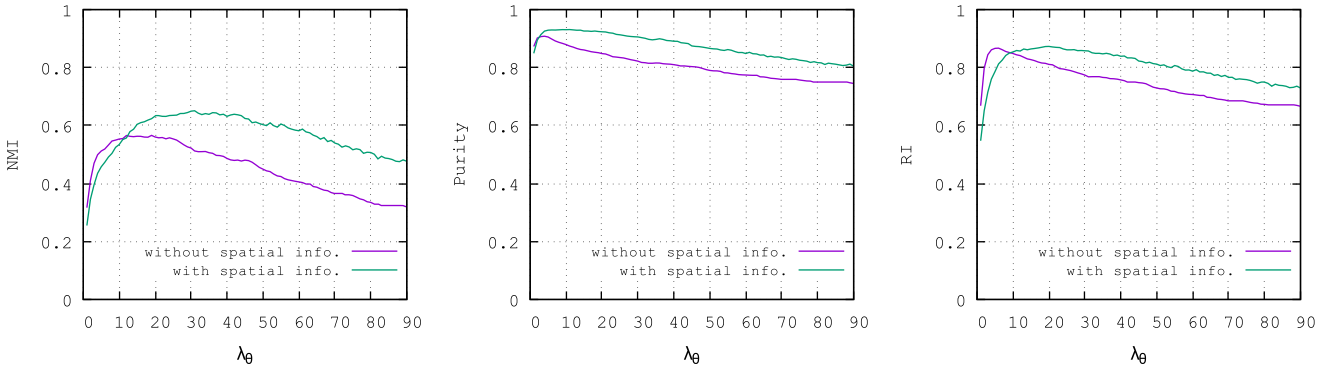


FIGURE 8. Average NMI, Purity and Rand Index profiles respectively (from left to right) with respect to varying λ_θ values for 300 scenes of CUHK dataset (Best viewed in color).

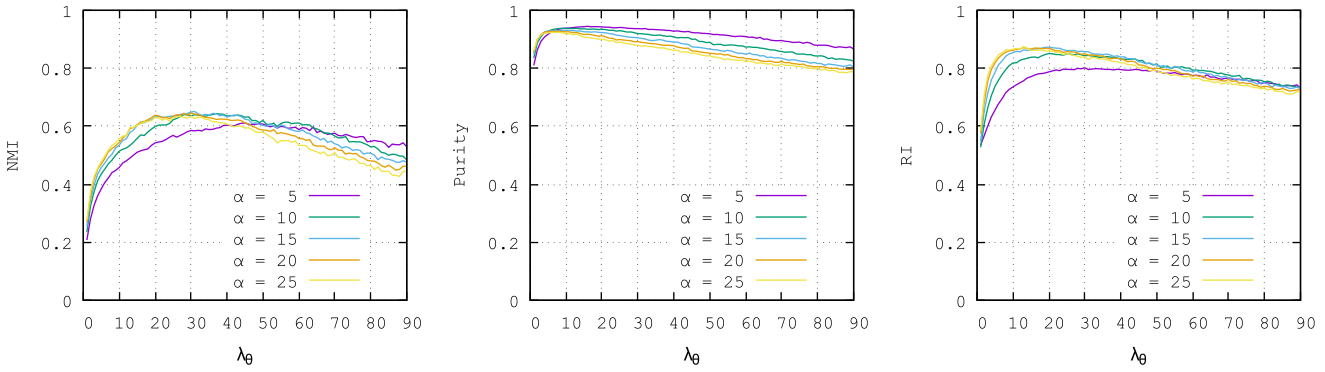


FIGURE 9. Average NMI, Purity and Rand Index profiles respectively (from left to right) for with respect to varying λ_θ and α values for 300 scenes of CUHK dataset (Best viewed in color).

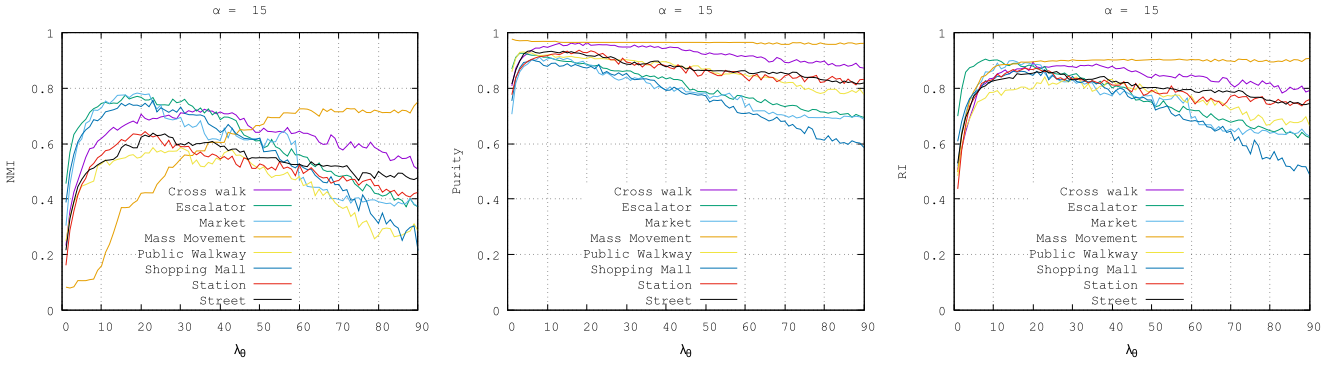


FIGURE 10. Average NMI, Purity and Rand Index profiles respectively (from left to right) for with respect to varying λ_θ value and $\alpha = 15$, for scene categorization based on type of scene (Best viewed on screen).

Further, in Figures 10 & 11 we report the performance profiles for each of the scene categories by keeping alpha constant ($\alpha = 15$) and varying λ_θ value. For this purpose, we consider all the scenes in each category without any sampling. Apart from the NMI-profile for mass-movement, structured and outdoor scene (which contain scenes with one-cluster ground truth) all the other profiles show the actual trend. The experimental results in section IV-C were obtained by applying a single threshold ($\lambda_\theta = 20°, \alpha = 15$) for all the scenes of CUHK dataset. Even though the obtained results for a single threshold are well above par, in real world scenes the threshold must depend on the type of scene. Therefore, with the help of the profiles in Figures 10 & 11 we point out the suitable values for λ_θ across different scene categories (λ_θ plays a major role in deciding cluster membership compared to α). The following inferences can be made about the choice of λ_θ for SADC algorithm when used for analysing crowded scenes: (i) For most cross walk and some mass movement

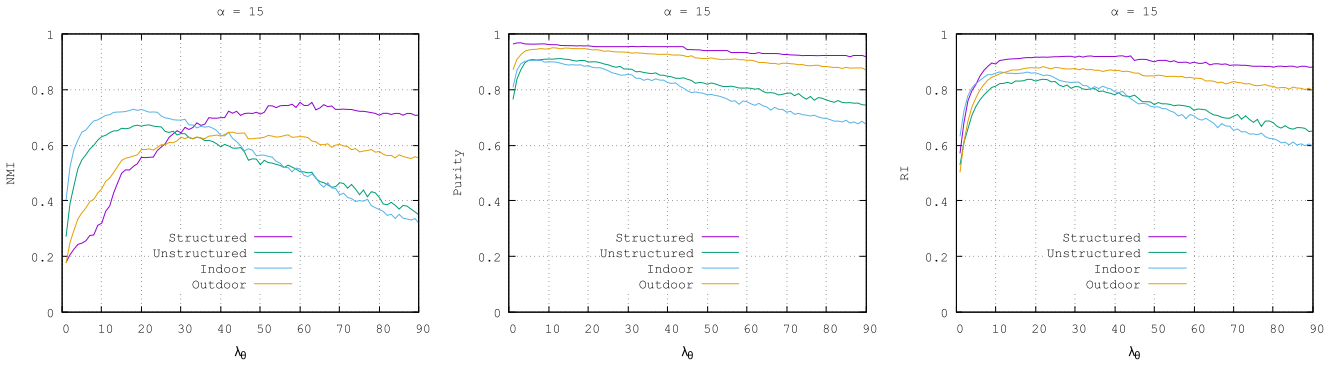


FIGURE 11. Average NMI, Purity and Rand Index profiles respectively (from left to right) for with respect to varying λ_θ value and $\alpha = 15$, for scene categorization based on property of crowd movement and scene location (Best viewed in color).

TABLE 2. Average time taken (in seconds) to generate clustering result for 30-frames.

Approach	Time taken (seconds)
CM [7]	2.52
CT [8]	134.34
Proposed SADC	**0.39**

scenes, λ_θ between 10° and 45° is suitable. But, for most of the single-cluster mass movement scenes which are also structured and outdoor, the angular deviation between the trajectories would be very low and hence a higher λ_θ value greater than 20° is suitable because a lower λ_θ value could result in splitting of similar clusters. (ii) For escalator scenes (indoor), λ_θ between 10° and 35° is desirable. Increasing the λ_θ value will result in the inclusion of non-escalator trajectories to escalator's cluster. (iii) For other complex and unstructured scenes such as market, public walkway, shopping mall (indoor), station (indoor) and street (where the number of motion patterns/groups are high), a lesser λ_θ value will enable the SADC to capture more groups. Hence, for such scenes, a λ_θ value between 10° and 20° is desirable.

E. COMPUTATIONAL COMPLEXITY

Time taken to generate the motion pattern segmentation results play an important role, especially in real-time scenarios. To calculate the time taken, we use the setting as mentioned in Section IV-B and report the average time taken (in seconds) to generate the clustering result for 30-frames over the entire 300 scenes of CUHK dataset. Note that we do not include the time taken to generate the trajectories as they are the same for all three methods (CM, CT, SADC). Both CM and CT need to iterate frame-by-frame to perform clustering, whereas the proposed SADC algorithm generates clusters for every 30-frames. Therefore, the time taken by SADC to generate clusters is significantly faster than the other two methods, which is substantiated in Table 2.

V. CONCLUSION AND FUTURE WORK

This article proposed a spatio-angular density-based clustering approach to group similar motion patterns. Spatial and angular features are obtained by averaging the trajectories across a fixed set of frames to capture better information about the history of the trajectories. These two features are then utilized to facilitate an improvised density-based clustering algorithm where the cluster membership is decided on the basis of two parameters, namely angular deviation threshold and spatial threshold. Qualitative and quantitative analysis through a set of parameters for different scene categories have shown the robustness of the proposed algorithm.

In the future work, we plan to integrate the crowd anomaly detection framework with our proposed approach. Furthermore, we intend to explore more on the inclusion of new features to tackle complex scenarios.

REFERENCES

[1] M. S. Zitouni, H. Bhaskar, J. Dias, and M. E. Al-Mualla, "Advances and trends in visual crowd analysis: A systematic survey and evaluation of crowd modelling techniques," *Neurocomputing*, vol. 186, pp. 139–159, Apr. 2016.

[2] X. Zhang, Q. Yu, and H. Yu, "Physics inspired methods for crowd video surveillance and analysis: A survey," *IEEE Access*, vol. 6, pp. 66816–66830, 2018.

[3] G. Tripathi, K. Singh, and D. K. Vishwakarma, "Convolutional neural networks for crowd behaviour analysis: A survey," *Vis. Comput.*, vol. 35, no. 5, pp. 753–776, May 2019.

[4] M. Ravanbakhsh, E. Sangineto, M. Nabi, and N. Sebe, "Training adversarial discriminators for cross-channel abnormal event detection in crowds," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2019, pp. 1896–1904.

[5] S. Ali and M. Shah, "A Lagrangian particle dynamics approach for crowd flow segmentation and stability analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–6.

[6] B. Zhou, X. Tang, and X. Wang, "Coherent filtering: Detecting coherent motions from crowd clutters," in *Computer Vision—ECCV*. Springer, 2012, pp. 857–871.

[7] B. Zhou, X. Tang, H. Zhang, and X. Wang, "Measuring crowd collectiveness," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1586–1599, Aug. 2014.

[8] J. Shao, C. C. Loy, and X. Wang, "Learning scene-independent group descriptors for crowd understanding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 6, pp. 1290–1303, Jun. 2017.

[9] W. Lin, Y. Mi, W. Wang, J. Wu, J. Wang, and T. Mei, "A diffusion and clustering-based approach for finding coherent motions and understanding crowd scenes," *IEEE Trans. Image Process.*, vol. 25, no. 4, pp. 1674–1687, Apr. 2016.

[10] H. Ullah, M. Ullah, and M. Uzair, "A hybrid social influence model for pedestrian motion segmentation," *Neural Comput. Appl.*, vol. 31, no. 11, pp. 7317–7333, Nov. 2019.

[11] T. Li, H. Chang, M. Wang, B. Ni, R. Hong, and S. Yan, "Crowded scene analysis: A survey," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 3, pp. 367–386, Mar. 2015.
[12] J. Trojanová, K. Křehnáč, and F. Brémond, "Data-driven motion pattern segmentation in a crowded environments," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2016, pp. 760–774.
[13] Z. Fan, J. Jiang, S. Weng, Z. He, and Z. Liu, "Adaptive crowd segmentation based on coherent motion detection," *J. Signal Process. Syst.*, vol. 90, no. 12, pp. 1651–1666, Dec. 2018.
[14] D. Helbing and P. Molnár, "Social force model for pedestrian dynamics," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 51, no. 5, p. 4282, 1995.
[15] N. Rinke, C. Schiermeyer, F. Pascucci, V. Berkhahn, and B. Friedrich, "A multi-layer social force approach to model interactions in shared spaces using collision prediction," *Transp. Res. Procedia*, vol. 25, pp. 1249–1267, 2017.
[16] W. Pan, W. Dong, M. Cebrian, T. Kim, J. H. Fowler, and A. S. Pentland, "Modeling dynamical influence in human interaction: Using data to make better inferences about influence within social systems," *IEEE Signal Process. Mag.*, vol. 29, no. 2, pp. 77–86, Mar. 2012.
[17] M. Chen, Q. Wang, and X. Li, "Patch-based topic model for group detection," *Sci. China Inf. Sci.*, vol. 60, no. 11, Nov. 2017, Art. no. 113101.
[18] Q. Wang, M. Chen, F. Nie, and X. Li, "Detecting coherent groups in crowd scenes by multiview clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 1, pp. 46–58, Jan. 2020.
[19] W. He and Z. Liu, "Motion pattern analysis in crowded scenes by using density based clustering," in *Proc. 9th Int. Conf. Fuzzy Syst. Knowl. Discovery*, May 2012, pp. 1855–1858.
[20] C. Tomasi and T. Kanade, "Detection and tracking of point features," *Int. J. Comput. Vis.*, vol. 9, no. 3, pp. 137–154, Jan. 1991.
[21] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. DARPA Image Understand. Workshop*, 1981, pp. 121–130.
[22] G. Yuan, P. Sun, J. Zhao, D. Li, and C. Wang, "A review of moving object trajectory clustering algorithms," *Artif. Intell. Rev.*, vol. 47, no. 1, pp. 123–144, Jan. 2017.
[23] J. Bian, D. Tian, Y. Tang, and D. Tao, "A survey on trajectory clustering analysis," 2018, *arXiv:1802.06971*. [Online]. Available: http://arxiv.org/abs/1802.06971
[24] Y. Li, H. Liu, X. Zheng, Y. Han, and L. Li, "A top–bottom clustering algorithm based on crowd trajectories for small group classification," *IEEE Access*, vol. 7, pp. 29679–29698, 2019.
[25] A. M. Cheriyadat and R. J. Radke, "Detecting dominant motions in dense crowds," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 4, pp. 568–581, Aug. 2008.
[26] R. Chaker, Z. A. Aghbari, and I. N. Junejo, "Social network model for crowd anomaly detection and localization," *Pattern Recognit.*, vol. 61, pp. 266–281, Jan. 2017.
[27] S. D. Khan, S. Bandini, S. Basalamah, and G. Vizzari, "Analyzing crowd behavior in naturalistic conditions: Identifying sources and sinks and characterizing main flows," *Neurocomputing*, vol. 177, pp. 543–563, Feb. 2016.
[28] J.-G. Lee, J. Han, and K.-Y. Whang, "Trajectory clustering: A partition-and-group framework," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2007, pp. 593–604.
[29] Y. Wu, Y. Ye, and C. Zhao, "Coherent motion detection with collective density clustering," in *Proc. 23rd ACM Int. Conf. Multimedia*, 2015, pp. 361–370.
[30] R. Sharma and T. Guha, "A trajectory clustering approach to crowd flow segmentation in videos," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 1200–1204.
[31] W. Lu, X. Wei, W. Xing, and W. Liu, "Trajectory-based motion pattern analysis of crowds," *Neurocomputing*, vol. 247, pp. 213–223, Jul. 2017.
[32] A. S. Hassanein, M. E. Hussein, W. Gomaa, Y. Makihara, and Y. Yagi, "Identifying motion pathways in highly crowded scenes: A non-parametric tracklet clustering approach," *Comput. Vis. Image Understand.*, vol. 191, Feb. 2020, Art. no. 102710.
[33] Y. Wu, Y. Ye, C. Zhao, and Z. Shi, "Collective density clustering for coherent motion detection," *IEEE Trans. Multimedia*, vol. 20, no. 6, pp. 1418–1431, Jun. 2018.
[34] J. Shi and C. Tomasi, "Good features to track," Cornell Univ., Ithaca, NY, USA, Tech. Rep., 1993.
[35] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, vol. 96, 1996, pp. 226–231.
[36] F. Nie, X. Wang, and H. Huang, "Clustering and projected clustering with adaptive neighbors," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2014, pp. 977–986.
[37] A. Strehl and J. Ghosh, "Cluster ensembles—A knowledge reuse framework for combining multiple partitions," *J. Mach. Learn. Res.*, vol. 3, pp. 583–617, Dec. 2002.
[38] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *J. Amer. Stat. Assoc.*, vol. 66, no. 336, pp. 846–850, Dec. 1971.

ABHILASH K. PAI received the bachelor's and master's degrees from the Cochin University of Science and Technology, Cochin, Kerala, in 2010 and 2014, respectively. He is currently pursuing the Ph.D. Research Scholar with the Department of Computer Applications, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India. His research interests include computer vision and machine learning.

A. KOTEGAR KARUNAKAR (Senior Member, IEEE) received the B.Sc. and M.C.A. degrees from Karnataka University, Karnataka, India, in 1995 and 1998, respectively, and the Ph.D. degree from the Manipal Academy of Higher Education, Karnataka, in 2009.

He is currently a Professor and the Head of the Department of Computer Applications, Manipal Academy of Higher Education. His research interests include image/video processing and communication, scalable video coding, media aware network elements, multi-view video coding, scalable video over peer-to-peer networks, error resilient and concealment for scalable video, stereo vision, and image and video forensics.

U. RAGHAVENDRA received the Ph.D. degree from the Manipal Academy of Higher Education, Manipal, India. He is currently a Faculty with the Department of Instrumentation and Control Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education. He has published several papers in refereed international SCI-IF journals and international conference proceedings. His current research interests include 3D computer vision, image processing, and medical image analysis. He has a patent to his credit and received an Invention Award from Intellectual Ventures, USA, for his innovations, in 2014.