

# Extracting Descriptive Motion Information from Crowd Scenes

Muhammad Saqib  
School of Software,  
University of Technology Sydney,  
NSW 2007, Australia  
muhammad.saqib@student.uts.edu.au

Sultan Daud Khan  
Makkah Technology Valley,  
Saudi Arabia  
sdkhan@gistic.org

Nabin Sharma, Michael Blumenstein  
School of Software,  
University of Technology Sydney,  
NSW 2007, Australia  
nabin.sharma@uts.edu.au  
michael.blumenstein@uts.edu.au

**Abstract**—An important contribution that automated analysis tools can generate for management of pedestrians and crowd safety is the detection of conflicting large pedestrian flows: this kind of movement pattern, in fact, may lead to dangerous situations and potential threats to pedestrian's safety. For this reason, detecting dominant motion patterns and summarizing motion information from the scene are inevitable for crowd management. In this paper, we develop a framework that extracts motion information from the scene by generating point trajectories using particle advection approach. The trajectories obtained are then clustered by using unsupervised hierarchical clustering algorithm, where the similarity is measured by the Longest Common Sub-sequence (LCS) metric. The achieved motions patterns in the scene are summarized and represented by using color-coded arrows, where speeds of the different flows are encoded with colors, the width of an arrow represents the density (number of people belonging to a particular motion pattern) while the arrowhead represents the direction. This novel representation of crowded scene provides a clutter free visualization which helps the crowd managers in understanding the scene. Experimental results show that our method outperforms state-of-the-art methods.

## I. INTRODUCTION

Crowd scene constitutes a significant number of people gathered together in one place. Such places are often holding events such as concerts, political and religious gatherings. Public safety and security are a major concern in the crowded situations. The analysis of crowd has many real-world applications in surveillance, public space design, crowd management and in the behavioral study of animals and marine life. The study of the crowd has therefore attracted multidisciplinary research from physics, psychology, biology, civil engineering and computer science. In computer science, the importance of automatic analysis of crowd in computer vision has received much attention in the last decade. The visual analysis of crowd includes various computer vision tasks such as object detection, tracking and using motion information for understanding the behavior of the crowd. However, the crowd is a complex phenomenon, and the application of particular computer vision technique depends upon the structure and density of the crowd. Detection and tracking might work very well in the low-density crowd because of the clear visibility of pedestrians in the crowd. However, in high-density crowd, detector most often miss the detections or result in false-positive detections.

Moreover, tracking is also not feasible in the high-density crowd. The tracker is lost easily and requires re-initialization at regular interval. The reason for suboptimal performance is that there is less number of pixels per person and inter-object occlusions, which make the detection and tracking unreliable solutions. Therefore in such scenarios, a holistic understanding of the scene is required. The study of crowd behavior includes the study of flows, dominant flows/common pathways, and identification of inflows/source and sinks/outflows. This study of behavior can be beneficial as a preprocessing step in the understanding of crowd behavior in extraordinary situations such as congestion.

It is very critical for a computer vision system to capture accurate motion information for understanding the behavior of crowd. Thus the performance of the system is dependent upon how the motion is represented. The desired motion representation should generate long and most importantly reliable trajectories. These trajectories can then be used to describe the flow for the whole video. Traditionally, optical flow is used for computing the pixel-wise flow between the two frames [4], [12]. However, there are some inherent problems in the direct usage of optical flow for motion representation. Firstly, optical flow produces ambiguous results on the boundaries of conflicting flows and perform poorly in the case of very slow moving objects. Secondly, optical flow does not represent the long-range spatio-temporal motion representation required in many applications. Typically obtaining complete trajectories is a difficult task from crowd motion in the high-density crowd. Therefore to get complete trajectories a notion of tracklet is used to capture short-term motion. A tracklet is a fragment of trajectory obtained by the tracker for the object. As previously mentioned, detection of the object and their tracking does not perform well in a high-density crowd. In most of the previous studies, tracklets are extracted from feature points and subsequently tracked for the very short duration to generate tracklets. After tracklets generation, various approaches such as Markov Random Field (MRF) [22] are used to ensure spatio-temporal dependencies between the tracklets [22], [23]. The tracklets give a robust high-level representation of the crowd and are less likely to drift as compared to complete trajectories [11].

In this paper, we propose a framework for the study of

crowd behavior by analyzing the flows in the crowd and describing the flows quantitatively and qualitatively. Our proposed framework used particle flow computed by moving the grid of particles with the optical flow, initialized by the dynamical system [17]. Numerical integration of the dynamical system over the segment of the video is used to generate tracklets. The first step in the crowd analysis is to generate reliable long trajectories. Firstly, we make use of tracklets to generate short term reliable trajectories for a small segment of video and then use clustering both spatially and temporally to produce flows for the whole video. By using the resultant flows, we can identify source/inflow and outflow/sink of the scene. Once the reliable flows are achieved, we compute basic parameters like speed, density and direction of the dominant flows. These parameters can fully characterize pedestrian movement and measuring these will help the crowd managers in understanding over all motion in the scene. Once dominant flow are describe by these parameters, we can visually analyze speed-density relationship that serve basis for understanding the behavior of complex system. The remainder of the paper is organized as follows. In section II, discusses the most related work in the literature. In section III we discuss the proposed methodology. In section IV, we present and analyze the experimental results. Finally, the paper is concluded in section V.

## II. LITERATURE REVIEW

There has been a body of work on the usage of tracklets for understanding the collective motion of pedestrians in the crowd. Zhou *et al.* [22] used Random Field Topic (RFT) to obtain semantic regions by non-parametric clustering of the tracklets. The RFT model ensures spatio-temporal coherence between the tracklets. The information about source and sink in the scene are given to the models as a highly semantic-level priors during the inference. However, our framework does not assume sources and sinks to be known a priori. In a variation of this approach, Zhou *et al.* formulated the whole crowd with the agent-based model called Mixture model of Dynamic pedestrian Agents (MDA) [23]. In this model, each pedestrian is modeled as an agent with the three factors. The factors included the belief of starting and the destination position of the pedestrian, movement dynamics, and the timing of pedestrian entering the scene. The collective behavior is analyzed using agent after learning from real data. However, this agent-based approach is only suitable for the low-density crowd, where agents can easily model each pedestrian movement. Blei and Frazier [3] used Distance Dependent Chinese Restaurant Process (DD-CRP) to cluster the coherent tracklets in the first stage. In the second stage, the clusters are aggregated temporally to obtain pathways, or dominant motion flows in the scene. Next, a geometric analysis is applied on the pathway shape to determine the sources and sinks in the scene. Hassanein *et al.* [7] considered tracklet as a directed line segment and used line geometry to find similarity between the tracklets. Upon clustering, semantic regions are extracted which include source, sink, and common pathways.

Wang *et al.* analysed motion patterns by clustering the hybrid generative-discriminative feature maps [20]. The extracted feature map associate motion pattern with low-level feature. Finally, hierarchical clustering is used to cluster feature maps into semantic regions. However, all of the above mentioned approaches only describe the crowd flows qualitatively without further details about the motion in the scene. Meta-tracking is another technique employed to discover the scene structure in the crowd [9]. The Orientation Distribution Functions (ODFs) model is used to find out the direction of the flow at each pixel. A particle meta-tracking is used to obtain particle trajectories called meta-tracks. Finally, meta-tracks are clustered to get dominant motion flows in the crowd. This kind of motion representation is complex for extracting dense trajectories.

The raw data about path taken by pedestrian is considered noisy. For instance, Nedrich *et al.* used weak tracking system to identify source and sink through the mean shift clustering algorithm [15]. Manifold is a denoising algorithm which is also used for clustering of the data. Xu *et al.* proposed a shrinkage-based framework for unsupervised trajectory clustering using adaptive kernel-based estimation [21]. The framework can be viewed as the combination of mean-shift clustering and manifold base model. The center of the cluster represent manifold, and the structure of cluster shows motion patterns exhibited by the pedestrians in the crowd. Trajectories in the cluster are considered as noisy traversal of the manifold. Thus suppressing the noise and reconstructing manifold is seen as clustering of the trajectories.

Some approaches have been proposed to extract feature point for extracting trajectories called feature trajectories. The features are extracted using SIFT, HOG, and Histogram of Oriented Flow (HOF) or combination of various feature descriptors [19]. SIFT features are used to find correspondence between two frames for trajectories. Similarly, Sun *et al.* used SIFT descriptor along with KLT tracker to track feature of the sampled points from dense SIFT descriptor for dense trajectories [18]. Nawaz *et al.* proposed an end-to-end approach for trajectory clustering in aerial videos [14]. Firstly, camera motion is compensated before clustering of trajectories obtained from the co-efficient of Discrete Wavelet Transform (DWT). In another approach motion is analyzed and performed by the clustering of the trajectories in the frequency domain [8]. However, these approaches are not able to capture whole motion information because only few feature points are detected which are not sufficient for the extraction of dense trajectories. The most related work to our proposed work used KLT tracker to extract motion trajectories and used hierarchical clustering to obtain dominant flows [5], [6], [13]. However, the approaches are not able to capture the long-term motion required for understanding the behavior of the crowd. Furthermore, these approaches do not identify source and sink locations of the dominant flows. Similar approach is followed by Khan *et al.* [10] for extracting dominant motion flows and identification of sources and sinks. However, the motion flows do not describe scene in terms of speed and density. Together, these studies provide useful insight

into the study of crowd flow regarding motion representation, similarity measures and crowd behavior qualitatively such as the identification of sources, sinks and dominant motion pattern or common pathways. However, previous studies did not quantify the speed, density, and direction of the flows in the crowd. Our proposed framework extract dense motion trajectories using tracklets as a basic motion unit. Furthermore, the flows are described and visualized by the speed and direction of dominant motion patterns.

### III. PROPOSED METHODOLOGY

In this paper, we propose a framework for describing flows exhibited by pedestrians in the crowd. In the first step, a stable, dense trajectories are generated in the crowded scene. Secondly, the generated trajectories are clustered together spatially and temporally to get Long Dense Trajectories (LDT). These flows help at a later stage in the identification of sources and sinks for the behavior analysis. According to the proposed framework, input video is divided into multiple equal length segments as shown in the Figure 1. To generate reliable tracklets, the first frame of each segment is overlaid with the grid of particles. After initialization of the dynamical system, the particles are moved through the segment by optical flow [17]. The integration of the dynamical system gives reliable particle trajectories for particular video segment. The trajectories are then clustered into local tracks using Longest Common Sub-sequence (LCS) metric. Global tracks are obtained by applying hierarchical clustering algorithm to the local tracks. The following are the steps discussed in greater detail to extract descriptive motion information from the dominant flows.

#### A. Motion flow-field computation

Consider a feature point  $i$  in the frame of the video segment at time  $t$  is given by the following equation 1

$$Z_{i(t)} = (X_{i(t)}, V_{i(t)}) \quad (1)$$

Where  $Z_{i(t)}$ ,  $X_{i(t)}$ ,  $V_{i(t)}$  are the flow vector, location and velocity of the  $i^{th}$  feature point at time  $t$  respectively. Furthermore,  $V_{i(t)} = (vx_{i(t)}, vy_{i(t)})$  incorporate both change in x and y directions. Similarly,  $X_{i(t)} = (x_{i(t)}, y_{i(t)})$  is the initial location of the particle in the grid and  $X_{i(t+1)} = (x_{i(t+1)}, y_{i(t+1)})$  is next location at time  $t + 1$  approximated for all particles in the grid by solving the dynamical system given by the equation 3. There is an angle  $\theta_i$  associated with velocity vector  $V_{i(t)}$  where  $\theta$  is  $0^\circ \leq \theta \leq 360^\circ$ . The motion flow-field constitute all the flow vectors given by  $\{Z_1, Z_2, \dots, Z_m\}$  where  $m$  is the number of all foreground feature points in the frame. The dynamical system is initialized in which the velocity of feature point  $i$  is related to its optical flow at time  $t$  given by equation 2.

$$V_{i(t)} = F(X_{i(t)}) \quad (2)$$

#### B. Particle advection

In this step, a grid of particle is overlaid on the first frame and advect it on the optical flow field for each segment of the video. The grid resolution is kept at a lower resolution to make the process less computationally expensive. Let  $X_{i(t)} = (x_{i(t)}, y_{i(t)})$  is the initial location of the particle in the grid. The next location  $X_{i(t+1)} = (x_{i(t+1)}, y_{i(t+1)})$  can be approximated for all the particle in the grid by solving the following equation 3.

$$X_{i(t+1)} = F(X_{i(t)}) + X_{i(t)} \quad (3)$$

Let  $\Omega_i$  is the trajectory of the particle, which is essentially the movement of a particle from one frame to another frame according to the flow map. The pair of the flow maps are computed for each particle in the video segment as in [17]. In the unstructured crowd, the particle may drift from the flow of pedestrian and background noise can also induce its effect in the particle flow. Therefore a threshold term  $B_i$  is included in equation 4 to remove erroneous particles out of the computation. As a result of particle advection short trajectories called tracklets are obtained.

$$X_{i(t+1)} = F(X_{i(t)}) + X_{i(t)} * B_i \quad (4)$$

#### C. Clustering of tracklets

Tracklets only provide information about the local movements. Because the tracklets are short-term, therefore to obtain dense trajectories the tracklets are clustered together to generate dominant motion flows. The source and sink of the dominant flows are identified which help in describing the speed, density, and mean direction of the flows.

The first step in the clustering in getting the dense tracks is by extending shorter tracklets through the segments of the video. The assumption here is that these shorter tracklets are located spatially close together and can have plausible continuation if they can be connected to each other in a head-to-tail manner for a long trajectory construction. For the construction of long trajectory, the first tracklet called a query tracklet (that need to be extended) is compared with the potential candidates tracklets from the neighborhood of query tracklet's position on the grid. The best among the candidate is selected for the continuation of query tracklet. The selected candidate tracklet is now considered to be a query tracklet, and the same process is repeated until there are no more tracklets to be extended. Once the set of these tracklets are achieved, which are then connected by mean of  $K^{th}$  order least square polynomial regression. As shown in the Figure 1 after long trajectory construction, tracks are fed to the hierarchical clustering algorithm which clusters them into local tracks using Longest Common Sub-sequence as a similarity metric (LCS). The LCS is an efficient technique for matching tracks of unequal length to determine the longest common portion between the tracks. Finally, the local tracks are combined into the global tracks using the same hierarchical clustering algorithm. We call these global tracks as *Longest Dense Trajectories (LDT)*.

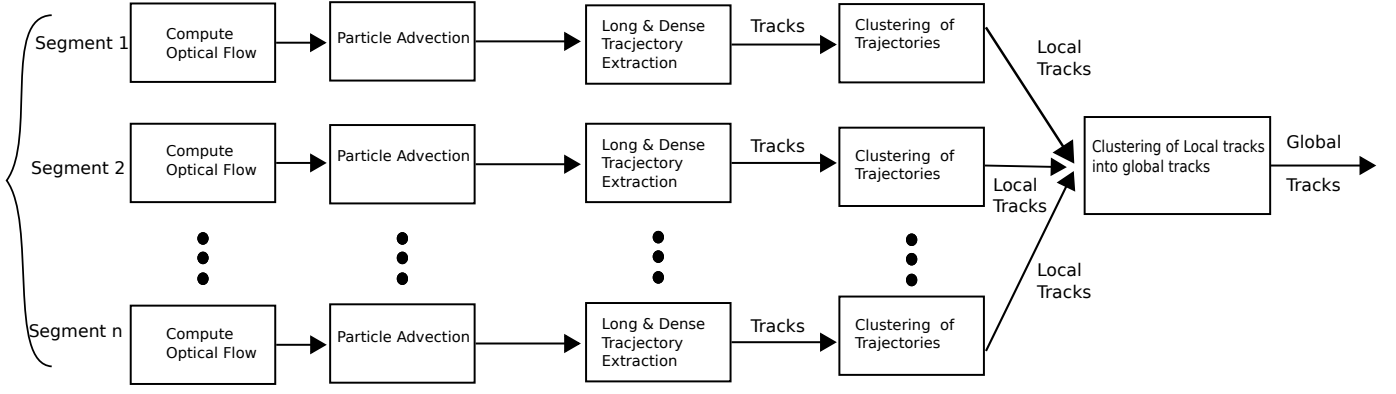


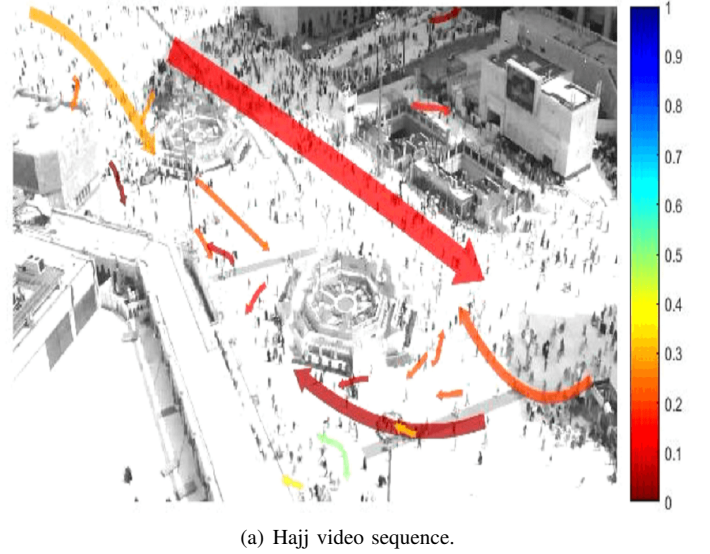
Fig. 1. Block diagram of proposed framework for dense trajectories extraction

Once the LDTs or global tracks (dominant flows) are achieved, the next step is to compute basic parameters that describe the flows. The first parameter that describe the flow is average speed of the flow. Speed together with the density can serve a base for understanding the behavior of complex systems. Let  $G_1$  and  $G_2$  are the two dominant flows obtained after clustering. Let  $G_i = \{t_1, t_2, \dots, t_n\}$ , where  $t_j$  are the tracklets that belong to the dominant flow  $G_i$ . In order to find the average speed of  $G_i$ , we first compute the speed of each tracklet  $t_j$  belongs to  $G_i$ . The instantaneous speed  $v_k$  of each point  $p_i \in t_j$  can be computed as  $v_k = \frac{\Delta_d}{\Delta_t}$ , where  $\Delta_d$  is the displacement among two consecutive points  $p_i$  and  $p_j$  that belong to  $t_j$  and  $\Delta_t$  is the frame rate of video. The overall speed  $v$  of tracklet  $t_j$  can be computed by  $v = \frac{1}{n} \sum_{k=1}^n v_k$  where  $n$  is the total number of points in  $t_j$ . In this way we compute the speed of every tracklet that belongs to  $G_i$  and finally average speed of  $V_i = \frac{1}{N} \sum_{m=1}^N v_m$ , where  $N$  is the number of tracklets belong to  $G_i$ .

In the next step, we compute the mean direction of the flow in the same way as average speed is computed. Here, we compute instantaneous flow  $f_k$  of each point  $p_i \in t_j$ . since  $f_k$  is a vector showing displacement  $u$  in x-dimension and displacement  $v$  in y-dimension, therefore we compute the angle  $\theta_k = \tan^{-1} \frac{u}{v}$ . The mean direction of  $t_j = \frac{1}{n} \sum_{k=1}^n \theta_k$  and finally average direction of  $G_i$  becomes  $\frac{1}{N} \sum_{m=1}^N \theta_m$ . We compute the density  $D_i$  of  $G_i$  by  $\frac{1}{M}$ , where  $N$  is total number of tracklets belong to  $G_i$  and  $M$  is the total number of tracklets belonging to all dominant flows. The summarized representation of dominant flows is shown in Fig 2(a). It is obvious in the figure that multiple dominant flows with their respective descriptions are detected by our algorithm. From the figure, we can easily visualize the inverse relationship of speed-density. The red-thick and long arrow show the main dominant flow which implies that density on this path is high relative to other paths in the scene.

#### IV. EXPERIMENTAL RESULTS

We have conducted several experiments to evaluate the performance of the proposed framework on publically available datasets. All the algorithms are implemented in MATLAB and



(a) Hajj video sequence.

Fig. 2. Summarized representation of dominant flows after the application of the algorithm. More red is the color the low is the speed of the flow while thicker the arrow represents the high density. (Best viewed in color)

the experiments are carried out on PC with the configuration of 2.7GHz (core i7) processor and 8 GB of RAM. The datasets include both indoor/outdoor scenes comprising of structured/unstructured [16] crowd of varying density level. The complete details of the datasets are shown in Table I. The input to our proposed framework is the segments of the video. We have divided every video into the segments of 50 frames. The ground truth for the all the videos is established manually by the human observer so that obtained results can be compared to the ground truth. As we are employing particle advection for extracting dense, long trajectories. The resolution of the grid overlaid on the segment of the videos is kept at 1/10th of the actual resolution determined experimentally. The increase in resolution increases the computation, and there is no significant improvement seen in the extraction of trajectories.

The effectiveness of our proposed framework is compared with the relevant current approaches in the extraction of



TABLE I  
DETAIL OF THE DATASETS USED IN THE EXPERIMENTS

Sequence	Resolution	Location	Crowd Density
Airport [1]	480X360	Indoor	Medium-density
Gallery [10]	1812X1461	outdoor	High-density
Unimib [2]	480X360	Indoor	Low-density
Escalator [6]	480X480	Indoor	Low-density
Hajj [10]	2946X1656	Outdoor	High-density
Station [6]	957X719	Indoor	Low-density

dominant flows/global tracks and also in the identification of sources and sinks of the dominant flows. The extraction of accurate flows is vital to the understanding of crowd scene. Our framework uses the extracted information to describe the speed, density, and mean direction of the dominant flows. Firstly, we compare the global tracks/dominant flows generated by our framework with ground-truth dominant flows. Moreover, a similar comparison is drawn between the current approaches and ground-truth flows using a flow similarity metric. The flow similarity metric is used to measure the similarity between dominant flows generated and the ground-truth flows. The flow similarity (Sim) is given by the equation 5

$$Sim = \frac{\left( \sum_{i=1}^{N_{gf}} \arg \max_{j \in [1, M_{df}]} \frac{LCS(G_j, Gth_i)}{Length(Gth_i)} \right)}{N_{gf}} \quad (5)$$

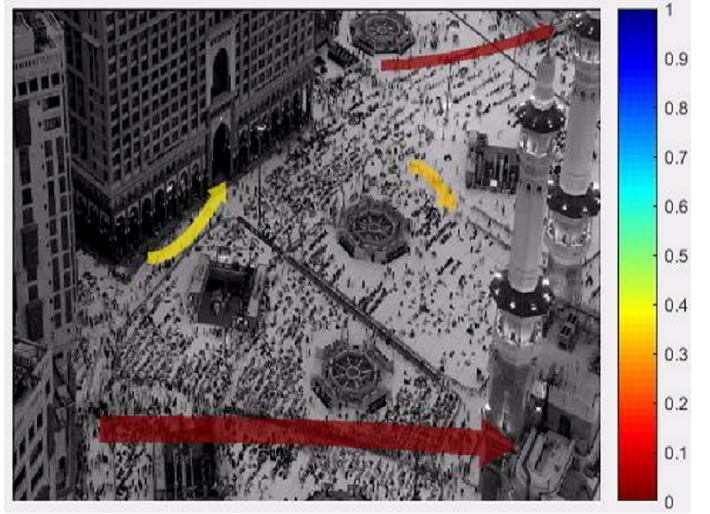
Whereas  $N_{gf}$  and  $M_{df}$  represent the total number of ground truth flows  $Gth$  and detected number of dominant flows  $G$  respectively. The Fig 4 shows the performance of our method encoded with the colors. The darker the color which corresponds to lower value on color bar, the more similar is the extracted global tracks with the ground truth flows. Similarly, a lighter color indicates lower resemblance of the detected global tracks with the detected one.

Secondly, The second metric identifies how closely the locations of sources/sinks of extracted trajectories lie together to the ground truth sources/sinks. For this comparison, merely using distance metric such as euclidean will not work as it is hard to relate changes in the scene to the changes in the real-world.

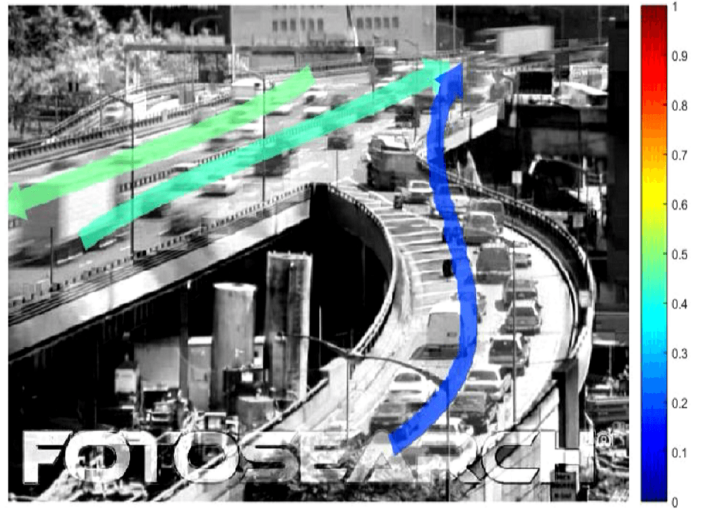
Therefore, we have used an association matrix to relate the locations of sources and sinks by their joint probability distribution. We have constructed an association matrix for sources and sinks identified by the ground-truth, our method, and other current methods. The measure of *Kullback-Leibler (KL) divergence* also referred to as relative entropy is adapted to compute the difference between the association matrices of compared methods with the ground truth matrices. The Kullback-Leibler formula is given below in equation 6

$$D_{KL}(AM_{Gth} || AM_{Gtrack}) = \sum_i AM_{Gth}(i) \ln \frac{AM_{Gth}(i)}{AM_{Gtrack}(i)} \quad (6)$$

Whereas  $AM_{Gth}$  and  $AM_{Gtrack}$  are the association matrices of ground truth flows and dominant flow produced by the



(a) Gallery video sequence.



(b) Traffic video sequence.

Fig. 3. Summarized representation of dominant flows after the application of the algorithm. More red is the color the low is the speed of the flow while thicker the arrow represents the high density. (Best viewed in color)

compared methods respectively. The comparative analysis is shown in the Fig 4. The figure explores the difference between the relative entropies of our method with other methods encoded with colors. The smaller value of the entropies for our method indicates that the sources and sinks locations are closed to the locations established by the ground-truth entropies as compared to the other state-of-the-art methods.

We have computed the average speed and density for three video sequences after successfully extracting the dominant flows with sources and sinks. The arrow overlaid on the sample frames represent the dominant flows as shown in the Fig 2(a) and 3(a) 3(b). The overall speed calculated for the dominant flows is by finding the average speed of the tracklets comprising it. The associated color bar with the frames indicates the speed. The density of the crowd can be visualized by the width of the arrows representing dominant

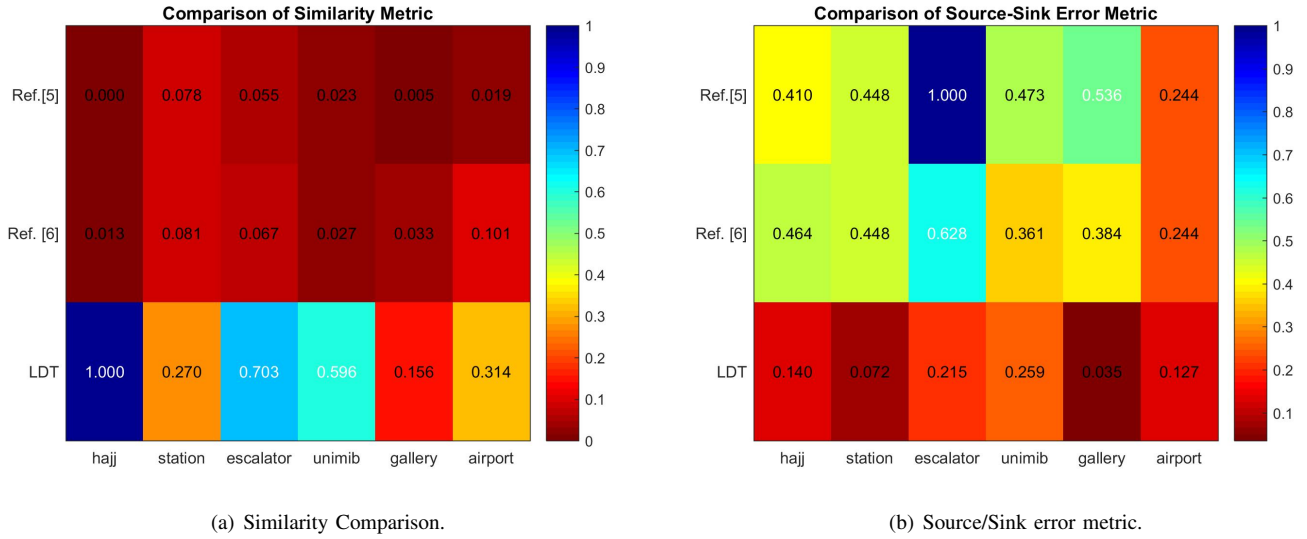


Fig. 4. (a) Similarity metric comparison of our LDT with base line trajectories obtained by other similar methods (Low values on color bar indicate high resemblance and vice versa). (b) Shows comparison error metric of our LDT with other similar methods (Best viewed in color)

flows, and the arrowhead gives the mean direction of the flow. Our proposed framework can extract useful descriptive information which is essential for understanding of the crowd behavior.

## V. CONCLUSION

In this paper, we have proposed a framework for extracting descriptive motion information from the crowd scenes. The framework makes sure that reliable, dense and trajectories are extracted. The descriptive information includes speed, density, and mean direction of dominant flows.

## REFERENCES

- [1] S. Ali and M. Shah. A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–6. IEEE, 2007.
- [2] S. Bandini, A. Gorrini, and G. Vizzari. Towards an integrated approach to crowd analysis and crowd synthesis: A case study and first results. *Pattern Recognition Letters*, 44:16–29, 2014.
- [3] D. M. Blei and P. I. Frazier. Distance dependent chinese restaurant processes. *Journal of Machine Learning Research*, 12(Aug):2461–2488, 2011.
- [4] T. Brox and J. Malik. Object segmentation by long term analysis of point trajectories. *Computer Vision–ECCV 2010*, pages 282–295, 2010.
- [5] A. M. Cheriadat and R. J. Radke. Detecting dominant motions in dense crowds. *Selected Topics in Signal Processing, IEEE Journal of*, 2(4):568–581, 2008.
- [6] W. Chongjing, Z. Xu, Z. Yi, and L. Yuncai. Analyzing motion patterns in crowded scenes via automatic tracklets clustering. *china communications*, 10(4):144–154, 2013.
- [7] A. S. Hassanein, M. E. Hussein, and W. Gomaa. Semantic analysis for crowded scenes based on non-parametric tracklet clustering. In *IJCAI*, pages 3389–3395, 2016.
- [8] W. Hu, X. Li, G. Tian, S. Maybank, and Z. Zhang. An incremental dpm-based method for trajectory clustering, modeling, and retrieval. *IEEE transactions on pattern analysis and machine intelligence*, 35(5):1051–1065, 2013.
- [9] P.-M. Jodoin, Y. Benezeth, and Y. Wang. Meta-tracking for video scene understanding. In *Advanced Video and Signal Based Surveillance (AVSS), 2013 10th IEEE International Conference on*, pages 1–6. IEEE, 2013.
- [10] 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*, 177:543–563, 2016.
- [11] T. Li, H. Chang, M. Wang, B. Ni, R. Hong, and S. Yan. Crowded scene analysis: A survey. *IEEE transactions on circuits and systems for video technology*, 25(3):367–386, 2015.
- [12] W.-C. Lu, Y.-C. F. Wang, and C.-S. Chen. Learning dense optical-flow trajectory patterns for video object extraction. In *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*, pages 315–322. IEEE, 2010.
- [13] M. B. Muhammad Saqib, Sultan Daud Khan. Detecting dominant motion patterns in crowds of pedestrians, 2017.
- [14] T. Nawaz, A. Cavallaro, and B. Rinner. Trajectory clustering for motion pattern extraction in aerial videos. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 1016–1020. IEEE, 2014.
- [15] M. Nedrich and J. W. Davis. Learning scene entries and exits using coherent motion regions. In *International Symposium on Visual Computing*, pages 120–131. Springer, 2010.
- [16] M. Rodriguez, S. Ali, and T. Kanade. Tracking in unstructured crowded scenes. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 1389–1396. IEEE, 2009.
- [17] B. Solmaz, B. E. Moore, and M. Shah. Identifying behaviors in crowd scenes using stability analysis for dynamical systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(10):2064–2070, 2012.
- [18] J. Sun, Y. Mu, S. Yan, and L.-F. Cheong. Activity recognition using dense long-duration trajectories. In *Multimedia and Expo (ICME), 2010 IEEE International Conference on*, pages 322–327. IEEE, 2010.
- [19] J. Sun, X. Wu, S. Yan, L.-F. Cheong, T.-S. Chua, and J. Li. Hierarchical spatio-temporal context modeling for action recognition. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2004–2011. IEEE, 2009.
- [20] C. Wang, X. Zhao, Z. Wu, and Y. Liu. Motion pattern analysis in crowded scenes based on hybrid generative-discriminative feature maps. In *Image Processing (ICIP), 2013 20th IEEE International Conference on*, pages 2837–2841. IEEE, 2013.
- [21] H. Xu, Y. Zhou, W. Lin, and H. Zha. Unsupervised trajectory clustering via adaptive multi-kernel-based shrinkage. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4328–4336, 2015.
- [22] B. Zhou, X. Wang, and X. Tang. Random field topic model for semantic region analysis in crowded scenes from tracklets. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 3441–3448. IEEE, 2011.
- [23] B. Zhou, X. Wang, and X. Tang. Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2871–2878. IEEE, 2012.