BEYOND PARTICLE FLOW: BAG OF TRAJECTORY GRAPHS FOR DENSE CROWD EVENT RECOGNITION

Yanhao Zhang[⋆]*, Lei Qin*[†]*, Hongxun Yao*[⋆]*, Pengfei Xu*[⋆]*, Qingming Huang*[⋆†]

[⋆]Harbin Institute of Technology, Harbin, 150001, China
[†]Inst. of Comput. Tech., Chinese Academy of Sciences, Beijing, 100190, China
{yhzhang,h.yao,pfxu}@hit.edu.cn {lqin,qmhuang}@jdl.ac.cn

ABSTRACT

In this paper, a novel crowd behavior representation, Bag of Trajectory Graphs (BoTG), is presented for dense crowd event recognition. To overcome huge loss of crowd structure and variability of motion in previous particle flow based methods, we design group-level representation beyond particle flow. From the observation that crowd particles are composed of atomic subgroups corresponding to informative behavior patterns, particle trajectories which simulate motion of individuals will be clustered to form groups at the first step. Then we connect nodes in each group as a trajectory graph and discover informative features to depict the graphs. A clip of crowd event can be further described by Bag of Trajectory Graphs (BoTG)-occurrences of behavior patterns, which provides critical clues for categorizing specific crowd event and detecting abnormality. The experimental results of abnormality detection and event recognition on public datasets demonstrate the effectiveness of our proposed BoTG on characterizing the group behaviors in dense crowd.

Index Terms— Bag of Trajectory Graphs, Attributes, Crowd Behavior, Event Recognition.

1. INTRODUCTION

Crowd event recognition plays a significant role in video surveillance domain and has gained more and more attention. Understanding of the crowd behavior, to some extent, faces many challenges like complex interactions, group-level relationships and various semantics. Meanwhile, the efforts for crowd event recognition [1] [2] [3] [4] [5], such as abnormal traffic detection for crowd dynamics or aggressive chaos event involving crowds, have been explored.

Most previous works which focus on the recognition of events involve multiple agents in the crowd scenes. The events are typically defined by the motion patterns of individuals or the entire crowd behaviors. The existing approaches can be divided into three categories depending on the unit granularity of analyzing. The first type is object centric [1] [6] [7], which concentrates on the detected targets and makes recognition by analyzing the trajectories of the object-level targets. The second type is flow centric, in which optical flow [8], particle flow [9] [10] as well as local gradients or appearances space-time sub volumes [11] [12] [13] are popularly utilized to simulate the crowd flow instead of tracking the individuals. However, this type of approaches is always related to a macroscopic model which is not close to the behavior semantics. The third type is group centric, which tries to encode more group semantic information and insightful structure to represent the crowd behavior by utilizing group structures [14], particle trajectory [15] [16], or energy potentials [17]. As we can see, flow strategy ensures that dense complex movements can be captured, which would be beneficial for the dense crowd understanding. Moreover, group-level structure also provides an essential factor which combines macroscopic view of crowd and microscopic dynamics as well as the interaction as a whole.

From above inspirations, in this paper we propose the idea of crowd group-level representation which organizes the crowd motion patterns by graphs and provides meaningful features to infer behavior insight. Specifically, particle trajectory graphs act as a crucial link to the holistic crowd and individuals. The constructed graphs would greatly benefit (1) reflecting the basic properties and context of groups as well as (2) recognizing the holistic crowd behavior patterns.

Therefore, we describe fine grained individual motion as particle trajectories, and further consider the spatial proximity to form trajectory graphs. By applying informative features on representing the structure and motion of graphs, we signify crowd behavior patterns using Bag of Trajectory Graphs (BoTG). BoTG records the temporal co-occurrence of certain behavior patterns appearing in different types of crowd events. In summary, our primary contributions include:

1) Proposing a framework to automatically discover informative trajectory graphs by advection of particles in dense crowd instead of detecting and tracking individuals.

2) Introducing a novel group-level representation, BoTG, for effectively describing the structure and motion of the crowd behavior patterns.

3) Emphasizing various semantic features which affect the performance of BoTG, and demonstrating the effectiveness in abnormality detection and behavior pattern recognition.

2. TRAJECTORY GRAPH CONSTRUCTION

The processing flow of the proposed method is illustrated in Figure 1. Given clips of crowd videos, we first obtain the particle trajectories by a particle advection approach. We then construct trajectory graphs and extract features for the clustered groups, which aims at representing group-level structure and motion clues. Finally, we utilize Bag of Trajectory Graphs representation for dense crowd event detection and recognition tasks with supervised learning approach.

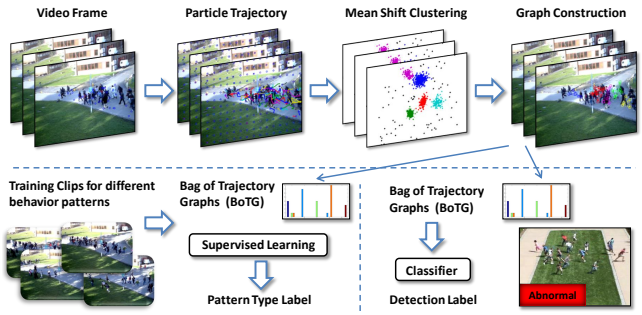

Fig. 1. Processing flow of the proposed BoTG method.

2.1. Particle Advection for Trajectory Extraction

Particle advection scheme [15] [18] is utilized to simulate crowd motion behavior by regarding the individual as particle in dense crowd scene. Meanwhile, particles are moving with the optical flow field, which reflects the property of the continuous evolution in the group motion.

To start with, a homogeneous grid of particles is placed over the frame with the scale of grid mainly depending on the density of crowd. Subsequently, along with the bilinear interpolation of the optical flow field, velocities for each particle are computed using 4th-order Runge-Kutta algorithm. Particles will follow the trajectories in a fluid flow by the guidance of average neighborhood. The trajectory of a particle $P_{t_1}^{t_T}(t)$ in the flow field consists of T tuples:

$$P_{t_1}^{t_T}(t) = (s_i, v_i, t_i)_{t_i=t_1}^{t_T} \tag{1}$$

where s_i and v_i denote the position and velocity vectors of the particle at time t_i which are obtained from the optical flow.

2.2. Mean Shift Based Trajectory Clustering

To identify the particles with similar motion patterns, we cluster particle trajectories into groups for modeling the group-level behavior characteristics by mean shift method [19]. It prompts reliable modes determined by particle density (unlike K-means) and robust to variety of trajectories. For each trajectory, the closest mode of a sample distribution is computed iteratively by mean shift which starting from a hypothesized mode. Specially, given c sample x_i, $i = 1, \ldots, c$, in T-dimensional space Λ, the kernel density estimation of function $f(x)$ can be written as

$$\hat{f}(x) = \frac{c_{K,h}}{c} \sum_{i=1}^{c} K(\frac{d^2(x, x_i)}{h}) \tag{2}$$

where $K(z) > 0 (z \geq 0)$ is a radically symmetric kernel satisfying $\int K(z)dz = 1, c_{K,h} > 0$ represent normalization coefficient. $d(x, x_i)$ and h define distance measurement and bandwidth scale respectively in which the samples are considered for probability density estimation. Then, the random selected point x_j shift to the point x_{j+1} with the highest probability density in current scale by calculating mean shift vector $m_h(x)$ as follows.

$$m_h(x) = c'_{K,h} \frac{\nabla \hat{f}(x)}{\hat{f}(x)} = \frac{\sum_{i=1}^{c} x_i Q(x_i)}{\sum_{i=1}^{c} Q(x_i)} - x \tag{3}$$

where the kernel profile $Q(z) = -K'(z)$, so that the next point $x_{j+1} = x_j + m_h(x_j), j = 1, \ldots, c$. The kernel is recursively moved and converge to the nearest mode as cluster center. Repeat the above iterative process until all the samples finish clustering.

In the context of our case, a sample corresponds to a particle trajectory. The bandwidth h is defined as the scale of group-level size. We set it as the grid size which is related to the density of particles. In practice, we use the simple histogram filter process to remove the background noise trajectories which may be caused by the illumination, distortion and background movement (as shown in Fig.2(a-b)). The filtered trajectory space is the effective clustering space $\Lambda' = \{P^T(s, v, t) | \, \|s_1 - s_T\|_2^2 > \mathrm{D}_{Th}\} \subset \Lambda$, and $D_{Th} > 0$ is the distance threshold (5 pixels in experiment).

In particle trajectory graph construction, trajectories in each group-level cluster are fully connected using the Euclidean distance as the edge weights (Fig.2(c)). Thus, a T-frame video sequence can be represented by particle trajectory graphs with 3-tuple $G = (V, E, W)$ for each, in which V is a set of vertices, $E \subseteq V \times V$ is a set of edges, W is the edge weight assigning for E. We next try to specify the detailed description based on the basic graphs.

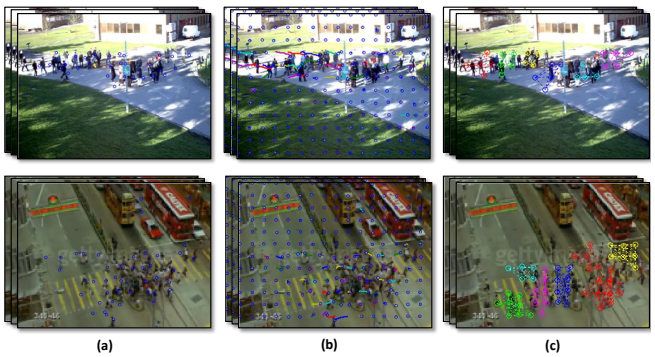

Fig. 2. Examples of graph construction. (a) filtered particles. (b) filtered trajectories. (c) graphs after clustering.

3. BAG-OF-TRAJECTORY-GRAPHS

As for understanding of behavior patterns or group-level types of crowd motion, the occurrence of informative trajectory graphs should be more critical than visual patterns or spatial-temporal motion volumes. Our mid-level representation, BoTG, reflects group-level behavior patterns when considering the graph structure, the group attributes as well as motion dynamics information.

3.1. Feature Representation for Trajectory Graphs

Graph Structure. In order to represent the group structure, utilizing spectrum to reflect the structural characteristics of the graphs is a good choice since Laplacian spectrum achieved a good performance in many recognition and classification problems [20]. Suppose we have N graphs with m trajectories for each in T-frame clip, for each graph $G_k = (V_k, E_k, W_k)$, the Laplace matrix is defined as:

$$A_k = \begin{cases} d_{ij} & \text{if } i \neq j \\ -\sum_{j=1}^{m} d_{ij} & \text{otherwise} \end{cases} \tag{4}$$

where $i, j = 1, 2, \ldots, m, k = 1, 2, \ldots, N$, and $d_{ij} = W_k$ refers to the distances between different vertices. The eigenvalues of Laplacian matrix of G_k can be obtained by the method of singular value decomposition(SVD),

$$A_k = U \Delta U^T \tag{5}$$

where $\Delta = diag\{\tau_1, \tau_2, \cdots, \tau_m\}$,$\tau_1 \geq \tau_2 \geq \cdots \geq \tau_m \geq 0$. We selected the largest 3 eigenvalues as the graph structure feature which are discriminative to distinguish various structure patterns.

Group Attributes. Group attributes express the characteristics of groups including orientation distribution and speed distribution. In each trajectory graph $G = (V, E, W)$, for each node $v_k = \{s_1 \cdots s_T | s_i = (x_i, y_i)\} \in V$, the basic orientation and speed feature channels are computed as follows,

$$\begin{aligned} S(v_k) &= \sqrt{(x_T - x_1)^2 + (y_T - y_1)^2} \\ O(v_k) &= \arctan(\tfrac{x_T - x_1}{y_T - y_1}) \end{aligned} \tag{6}$$

where $s_i = (x_i, y_i), i \in [1, T]$ is the position of trajectory node v_k. The attributes can be regarded as quantitative measurements for the properties of group behaviors. We characterize statistical orientation and speed histogram with n bins $(n = 8)$ to define group attributes as follows:

$$H_{ori} = \{O_i\}_{i=1}^{n}, H_{spd} = \{S_i\}_{i=1}^{n} \tag{7}$$

Motion Dynamics. Besides the inner attributes of the groups, external motion information is also needed to describe the group in the entire crowd. For each trajectory graph G_j, we select the top 3 $S(v_k)$ as trajectory graph speed and treat average nodes position as the graph location to record the dynamic motion.

These features robustly capture the structure and motion of the trajectory graphs and effectively describe typical group-level behavior patterns. We next built our bag of words learning scheme by concatenating these features.

3.2. Vocabulary Building of Trajectory Graphs

Motivated by visual words that describe the local patterns of an image, trajectory graphs represent group behavior patterns for certain sequences, which are applicable for group event recognition. The concatenated feature vectors are clustered using K-means to build a vocabulary of trajectory graph words, in which a word indicates a certain type of group behavior pattern. Therefore, each T-frame crowd video can be represented by a Bag of Trajectory Graphs (BoTG) words. Accordingly, BoTG can capture informative cues of groups by means of preserving occurrence patterns. In this way, we construct BoTG from crowd clips and train SVM to recognize different event types. As a result, BoTG serves as an effective representation for group-level behavior patterns.

4. EXPERIMENTAL EVALUATION

4.1. Abnormality Detection

To validate the effectiveness of our proposed model on abnormal event detection, we conduct it on the UMN dataset [21]. In the experiment, the detection performance of each method is evaluated by event-level measurement as in [22].

UMN dataset. It consists of 11 clips of crowded escape events which are captured in 3 different scenes including indoor and outdoor. Each video begins with normal behaviors and ends with panic escaping. All the video frames are resized to 120×160 pixels for computation cost.

Measurement. In the particle advection scheme, we set a particle every 5 pixels in the optical flow field and the length of the trajectory T is set to be 10 frames. During the mean shift clustering part, the bandwidth h equals the half of amount of particles. For the construction of the visual words, we compute all the trajectory graphs in each 10 frames. The vocabulary contains 10 cluster centers. In the experiment,

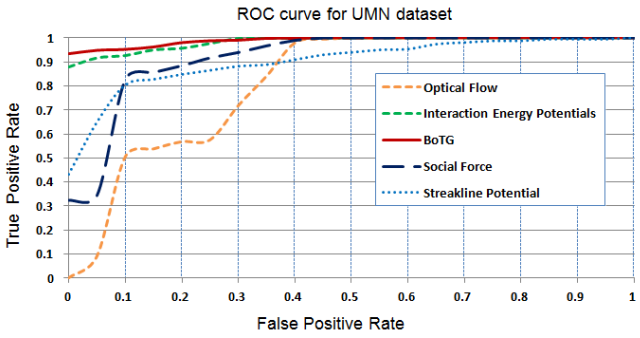

Fig. 3. ROC curves of abnormal detection in UMN dataset.

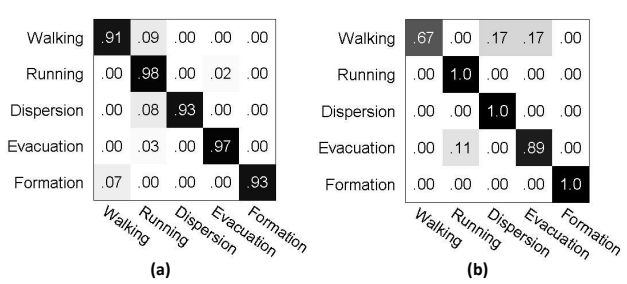

Fig. 4. Confusion matrices for event recognition. (a) Results in UMN. (b) Results in PETS2009 S3. Rows are ground truths and columns are the predictions.

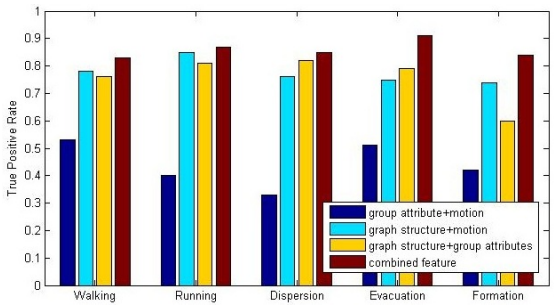

Fig. 5. TPR of UMN recognition performance for different event using different combination of features.

we utilize SVM with RBF kernels to train the model on 10 videos and compute the FPR and TPR on the left one.

Insight. Figure 3 illustrates the ROC curves of the experiments compared with other state-of-the-art high-level modeling methods. Results listed for comparison are directly gained from paper [17] [15] [18]. Table 1 shows that BoTG (AUC = 0.990) can achieve better performance over available high-level methods including Interaction Energy Potentials [17] (IEP), Social Force [18] (SF), Streakline Potential [15] (SP) and Optical Flow (OF). Better effect comes from the fact our graph structure and group attributes are more discriminative for group pattern changes and motion speed feature also perform significantly in the dispersing abnormality. It dedicates BoTG has superiority to improve the performance in detecting the abnormal behavior pattern by considering contextual group-level attributes as well as motion information.

Table 1. Comparison of high-level methods in UMN dataset.

Method	**BoTG**	IEP [17]	SF [18]	SP [15]	OF
AUC	**0.990**	0.985	0.96	0.90	0.86

4.2. Event Recognition

In this experiment, we consider the event recognition for the crowd scene. We perform to classify the video clips into 5 pre-defined event types: group regular walking, group regular running, group local dispersion, group evacuation (rapid dispersion) and group formation (splitting) at different time instances. We conducted the experiment on the UMN [21] and PETS2009 S3 dataset [23].

Dataset. The UMN are manually segmented into 450 clips of 10 frames. PETS 2009 S3 are segmented into 65 clips. All the video clips are labeled with the occurred event mentioned above. Each frame is resized to 240×320 pixels.

Evaluation Protocol. We randomly select 60% of the clips for training and the rest for testing. A one-vs-all SVM with RBF kernels are trained for each type using BoTG. Since the ground truth annotation is given per clip, we evaluate the recognition performance with confusion matrices. To verify the validity of features, we also show the True Positive Rate (TPR) of recognition for all event types on UMN.

Results and Discussion. Figure 4 shows the confusion matrices between 5 types of events on UMN and PETS2009. BoTG effectively recognize different patterns, while confusion only occurs in very similar components. Figure 5 illustrates TPR results of different feature component combination strategies on UMN. Several conclusions can be drawn. First, "graph structure + motion dynamics" stands for the top significant principle for *Walking*, *Running* and *Formation*, demonstrating those behavior patterns are quite related to the motion information. Second, "graph structure + group attributes" outperforms the others, illustrating they are the most critical features to recognize events of global and local dispersion. Graph structures are more discriminative than motion information to distinguish these patterns. Third, attributes information is less significant compared to graph structure when combining with motion shown in dark blue bar. Nevertheless, it also works well for the inter-group event like *Dispersion* and *Evacuation*, since it records the differences of various groups. Finally, the performance of the combined feature is the best for all the event recognition due to the features are complementary to each other. Through the confusion matrices calculation, we achieve an overall above 90% accuracy in UMN (94.4%) and PETS2009 S3 dataset (91.2%).

5. CONCLUSION

In this paper, a Bag of Trajectory Graphs representation is proposed for dense crowd event recognition. We present a efficient graph construction approach with informative group-level graph description, which effectively capture group-level structure and motion dynamic of behavior patterns. Further, experiments indicate effectiveness of our approach is notable on abnormality detection and event recognition work.

6. ACKNOWLEDGEMENT

This work was supported in part by National Basic Research Program of China (973 Program): 2012CB316400, in part by National Natural Science Foundation of China: 61025011, 61035001, 61003165 and 61133003, and in part by Beijing Natural Science Foundation: 4111003.

7. REFERENCES

[1] A. Basharat, A. Gritai, and M. Shah, "Learning object motion patterns for anomaly detection and improved object detection," in *CVPR 2008*. IEEE, 2008, pp. 1–8.

[2] D. Kuettel, M.D. Breitenstein, L. Van Gool, and V. Ferrari, "What's going on? discovering spatio-temporal dependencies in dynamic scenes," in *CVPR 2010*. IEEE, 2010, pp. 1951–1958.

[3] T. Hospedales, S. Gong, and T. Xiang, "A markov clustering topic model for mining behaviour in video," in *ICCV 2009*. IEEE, 2009, pp. 1165–1172.

[4] D. Tran and J. Yuan, "Optimal spatio-temporal path discovery for video event detection," in *CVPR 2011*. IEEE, 2011, pp. 3321–3328.

[5] Yanhao Zhang, Lei Qin, Hongxun Yao, and Qingming Huang, "Abnormal crowd behavior detection based on social attribute-aware force model," in *ICIP 2012*. IEEE, 2012, pp. 2689–2692.

[6] S. Pellegrini, A. Ess, K. Schindler, and L. Van Gool, "You'll never walk alone: Modeling social behavior for multi-target tracking," in *ICCV 2009*. IEEE, 2009, pp. 261–268.

[7] Zhongwei Cheng, Lei Qin, Qingming Huang, Shuqiang Jiang, and Qi Tian, "Group activity recognition by gaussian processes estimation," in *ICPR 2010*. IEEE, 2010, pp. 3228–3231.

[8] N. Ihaddadene and C. Djeraba, "Real-time crowd motion analysis," in *ICPR 2008*. IEEE, 2008, pp. 1–4.

[9] S. Ali and M. Shah, "A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis," in *CVPR 2007*. IEEE, 2007, pp. 1–6.

[10] S. Ali and M. Shah, "Floor fields for tracking in high density crowd scenes," *ECCV 2008*, pp. 1–14, 2008.

[11] L. Kratz and K. Nishino, "Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models," in *CVPR 2009*. IEEE, 2009, pp. 1446–1453.

[12] J. Kim and K. Grauman, "Observe locally, infer globally: a space-time mrf for detecting abnormal activities with incremental updates," in *CVPR 2009*. IEEE, 2009, pp. 2921–2928.

[13] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz, "Robust real-time unusual event detection using multiple fixed-location monitors," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 3, pp. 555–560, 2008.

[14] W. Ge, R.T. Collins, and B. Ruback, "Automatically detecting the small group structure of a crowd," in *WACV 2009*. IEEE, 2009, pp. 1–8.

[15] R. Mehran, B. Moore, and M. Shah, "A streakline representation of flow in crowded scenes," *ECCV 2010*, pp. 439–452, 2010.

[16] S. Wu, O. Oreifej, and M. Shah, "Action recognition in videos acquired by a moving camera using motion decomposition of lagrangian particle trajectories," in *ICCV 2011*. IEEE, 2011, pp. 1419–1426.

[17] X. Cui, Q. Liu, M. Gao, and D.N. Metaxas, "Abnormal detection using interaction energy potentials," in *CVPR 2011*. IEEE, 2011, pp. 3161–3167.

[18] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," in *CVPR 2009*. IEEE, 2009, pp. 935–942.

[19] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 5, pp. 603–619, 2002.

[20] J. Tang, C.Y. Zhang, and B. Luo, "A graph and pnn-based approach to image classification," in *Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on*. IEEE, 2005, vol. 8, pp. 5122–5126.

[21] "Unusual crowd activity dataset of university of minnesota, from http://mha.cs.umn.edu/movies/crowdactivity-all.avi.," .

[22] Y. Cong, J. Yuan, and J. Liu, "Sparse reconstruction cost for abnormal event detection," in *CVPR 2011*. IEEE, 2011, pp. 3449–3456.

[23] "Pets2009 dataset, http://ftp.cs.rdg.ac.uk/pets2009," .