SIViP (2014) 8 (Suppl 1):S173–S181
DOI 10.1007/s11760-014-0669-9

ORIGINAL PAPER

Representing dense crowd patterns using bag of trajectory graphs

Yanhao Zhang · Qingming Huang · Lei Qin · Sicheng Zhao · Hongxun Yao · Pengfei Xu

Received: 13 December 2013 / Revised: 23 April 2014 / Accepted: 28 June 2014 / Published online: 22 July 2014

Abstract The aim of this paper was to address the problem of dense crowd event recognition in the surveillance video. Previous particle flow-based methods efficiently capture the convolutional motion in the crowded scene. However, the group-level description was rarely studied due to huge loss of group structure and intra-class variability. To address these issues, we present a novel crowd behavior representation called bag of trajectory graphs (BoTG). Firstly, we design a group-level representation beyond particle flow. From the observation that crowd particles are composed of atomic subgroups corresponding to informative behavior patterns, particle trajectories that simulate motion of individuals will be clustered to form groups. Secondly, we connect nodes in each group as a trajectory graph and propose 3 informative features to encode the graphs, namely, graph structure, group attribute, and dynamic motion, which characterize the structure, the motion within, and among the trajectory graphs. Finally, each clip of crowd event can be further described by BoTG as the occurrences of behavior patterns, which provides critical clues for categorizing specific crowd event. We conduct extensive experiments on public datasets for abnormality detection and event recognition. The results demonstrate the effectiveness of our BoTG on characterizing the group behaviors in dense crowd.

Keywords Bag of trajectory graphs · Group attributes · Crowd behavior · Event recognition

Y. Zhang (✉) · Q. Huang · S. Zhao · H. Yao · P. Xu
School of Computer Science, Harbin Institute of Technology, Harbin 150001, Heilongjiang, China
e-mail: yhzhang@hit.edu.cn

Q. Huang
e-mail: qmhuang@jdl.ac.cn

S. Zhao
e-mail: sczhao@hit.edu.cn

H. Yao
e-mail: h.yao@hit.edu.cn

P. Xu
e-mail: pfxu@hit.edu.cn

L. Qin
Institue of Computer Techology, Chinese Academy of Science, Beijing 100010, China
e-mail: lqin@jdl.ac.cn

1 Introduction

Crowd event recognition plays a significant role in video surveillance domain and has gained more and more attention in the computer vision literature. Understanding of the crowd behavior, to some extent, faces many challenges like complex interactions, group-level relationships and various semantics.

Previous works on crowd dynamics analysis mainly focused on the relatively object tracks [4], the interactions between moving agents and objects [11,17,28], or the particle interaction among the neighborhood [31]. The efforts for crowd event recognition, such as abnormal traffic detection for crowd dynamics or aggressive chaos event involving crowds have been explored. However, the structures among a group of particles occur frequently as the flow patterns in dense crowd scenarios, and the study of dense crowd behavior patterns based on visual cues has great potentials for smart video surveillance applications.

Considering the motion patterns of particle flow vary inherently, the highly restrictive approximations for the trajectories make the crowd patterns more invariant rather than distinguishable. Such property of particle flow is caused by the dense crowd which is hard to track for every individual and inevitably brought by the unconstrained motion patterns.

Although the trajectories of particles or objects may easily represent the typical group activity, it would be more difficult to capture the group-level characteristics due to two factors: (1) *Huge loss of structure*, the difficulties caused by the dense crowd with varying number of people and occlusions make the group activity lack of specific structures. (2) *Intra-class variability*, large variability of motion patterns comes from the fact that the trajectories do not explicitly encode the information of the interactions with a group. It is difficult and infeasible to robustly track a long trajectory of a individual in the dense crowd with occlusions. In addition, motion flow in crowded scenes involves many variations like the scaling color and light changes.

In this paper, we first investigate how to effectively represent dense crowd event from a group-level perspective. The number of people involved in a crowded scene is dense, which leads to that basic tracking methods are generally of little use for characterizing individual trajectories. Meanwhile, low level visual patches or optical flow patterns are not enough to describe the structure of a group. Thus, the middle level visual cues, e.g., motion trajectories and their interactions are critical for representing group structures. We propose to represent the dense crowd patterns with three types of features based on the clustering of particle trajectories, eg., graph structure, group attribute, and dynamic motion, which encode the structure properties of particle trajectories, the dynamic motion within, and among the trajectories, respectively. These graph-based features describe the structures and attributes of temporally short-term behaviors for the dense crowds. Finally, the crowd patterns can be represented by using bag of trajectory graphs (BoTG). The public database is then used to evaluate the effectiveness of our proposed representation for recognizing dense crowd events.

1.1 Related works

Most previous works which focus on the recognition of events involve multiple agents in the crowded scene. The events are typically defined by 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.

Object centric: The first type is object centric [4,5,24], which concentrates on the detected targets and makes recognition by analyzing the trajectories of the object-level targets. The typical methods are based on the statistical modeling on the trajectories which are obtained by specific trackers. [26] learn the joint co-occurrence of the patterns in the activity using the outdoor tracker. By creating the hierarchical binary tree classification, the moving objects are classified. In [14], the typical object trajectories are learned by active shape model. The distribution of the object trajectories is used to identify the incident and recognize the event. Furthermore, [23] obtain the object tracks by blob tracker and model the interaction by statistical Bayesian approach HMM and CHMM. [24] consider moving entities as agents and model them using social behavior analysis. The motion of each agent is driven by its destination and plans to avoid collisions with other moving objects. [22] apply filters on trajectories and take the responses as features for recognition which shows a promising result based on motion trajectories. [29] propose a semantic region modeling approach using Dual-HDP, which co-clusters both words and documents from data for the region detection.

Flow centric: The second type is flow centric methods, in which optical flow [13], particle flow [2,3] as well as the local gradients or space-time sub volumes [1,15,16] are popularly utilized to simulate the crowd flow instead of tracking the individuals. This type of approaches is always related to a macroscopic model from the flow field by cooperating the behavior semantics. The flow field in the macroscopic model can be viewed as extensions of multiple agents complemented by contextual and social information. Considering the positions as the main source, the interactions can be modeled by psychology, physics and sociology to exploit the social interactions or dynamics system to infer interpersonal relationships. [12] construct a directed neighborhood graph to measure the closeness based on the flow field. The motion patterns are grouped hierarchically based on the flow vectors. [20] place grid of particles over the image which are driven by the social forces created by the space-time optical flow as individuals. The forces are then estimated using the social force model to identify abnormal behaviors. In the dynamic system modeling, [2] reveal the Lagrangian coherent structure in the underlying Finite Time Lyapunov Exponent flow field generated by the particle advection. The streakline [21] is further designed to compute the important aspects such as flow and potential function. [30] extend the model by representing trajectories using chaotic invariants as the maximal Lyapunov Exponent. The maximum likelihood estimation criterion is adopted to identify the abnormal event.

Group centric: The third type is group centric, which tries to encode more group semantic information and the insightful structure to represent the crowd behavior by utilizing group structures [10,32], or energy potentials [9]. Group activities require the detection of a group structure in terms of the social aggregations in which the motion features need to be combined with the sociological and psychological information to infer the relations. The literature usually addresses them as clustering of groups and detection of the structures. [10] seek a deeper analysis of group structures in social behaviors which are discovered by hierarchical clustering and Hausdorff distance. [19] propose a group representative to model the relationships with a number of group members using Asynchronous Hidden Markov

Model (AHMM). [18] infer the contextual group structure by a latent tree structure model to improve activity recognition performance.

Most works analyze the crowd from the social viewpoint, which extend the group concept to imply socially-aware interactions. [8] detect the F-formations in the scene to infer the occurrences of the interactions between two or more persons. The interactions in small groups of people can be encoded in causality features to characterize based on the trajectories of the subjects. Considering the constraints imposed by the groups, [25] introduce the interaction hypothesis using CRF which takes into account social factors to estimate the group memberships. Finally, an interaction energy potential [9] can be extracted to model the relationships among groups of people. The relationships between the current states of the subjects and the corresponding reactions are then used to model the normal and abnormal behaviors.

As we can see, the flow strategy ensures that the dense complex movements can be captured, which would be beneficial for understanding the dense crowd. Moreover, the group-level structure also provides an essential factor which combines macroscopic view of crowd and microscopic dynamics, as well as the interactions as a whole. The conference version of this work was published in [32]. More technical details, theoretic analysis and experimental evaluations are provided in this paper.

1.2 The proposed method

From the above inspirations, in this paper, we propose the idea of constructing the group-level representation of crowds which organizes the crowd motion patterns by graphs and provides meaningful features to infer the behavior insight. Specifically, the particle trajectory graphs act as a crucial link between 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 the fine grained individual motion as the particle trajectories, and further consider the spatial proximity to form trajectory graphs. By applying informative features on representing structures and motion of graphs, we signify crowd behavior patterns using BoTG. BoTG records the temporal co-occurrences of the certain behavior patterns appearing in different types of crowd events, which benefits from the frequencies for the occurrences of the trajectory graphs appearing in each temporal window. It preserves the occurrence patterns of the groups that are effective for analyzing group activities and types. We demonstrate the capability of BoTG to recognize different event types and detect the abnormality. In summary, our primary contributions include the following:

1. We propose a framework to automatically discover the informative trajectory graphs, which is generated by advection of particles in dense crowds instead of detecting and tracking individuals. The combination of particle flow and the clustering operation makes the motion of graphs invariant and strengthens the descriptive power.
2. We properly introduce a novel group-level representation, BoTG, for describing the structures and motion of the crowd behavior patterns. It is capable of representing the dense crowd event which is highly robust and feasible.
3. We emphasize the various semantic features which affect the performances of BoTG, and demonstrate the effectiveness in the abnormality detection and the behavior pattern recognition.

2 The trajectory graphs model

2.1 Trajectory graph construction

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

2.1.1 Particle advection for trajectory extraction

Particle advection scheme [20,21] is utilized to simulate crowd motion behaviors by regarding individuals as particles in dense crowded scenes. Meanwhile, the particles are moving with the optical flow field, which reflects the property of the continuous evolution in the group motion.

Given the video of a crowded scene, it is divided into a series of clips which is represented by $T \times W \times H$ volumes where T denotes the number of frames and $W \times H$ denotes the width and height of the frame. The optical flow of each clip is computed to obtain the local motion velocity in pixel level, which is denoted by

$$\left\{(U_w^t, V_h^t) | w \in [1, W], h \in [1, H], t \in [1, T-1]\right\} \qquad (1)$$

To start with, a homogeneous grid of particles is placed over the frame with the scale of grid mainly depending on the density of crowds. Subsequently, the particle advection is performed to estimate the positions of moving particles along with the bilinear interpolation of the optical flow field. The

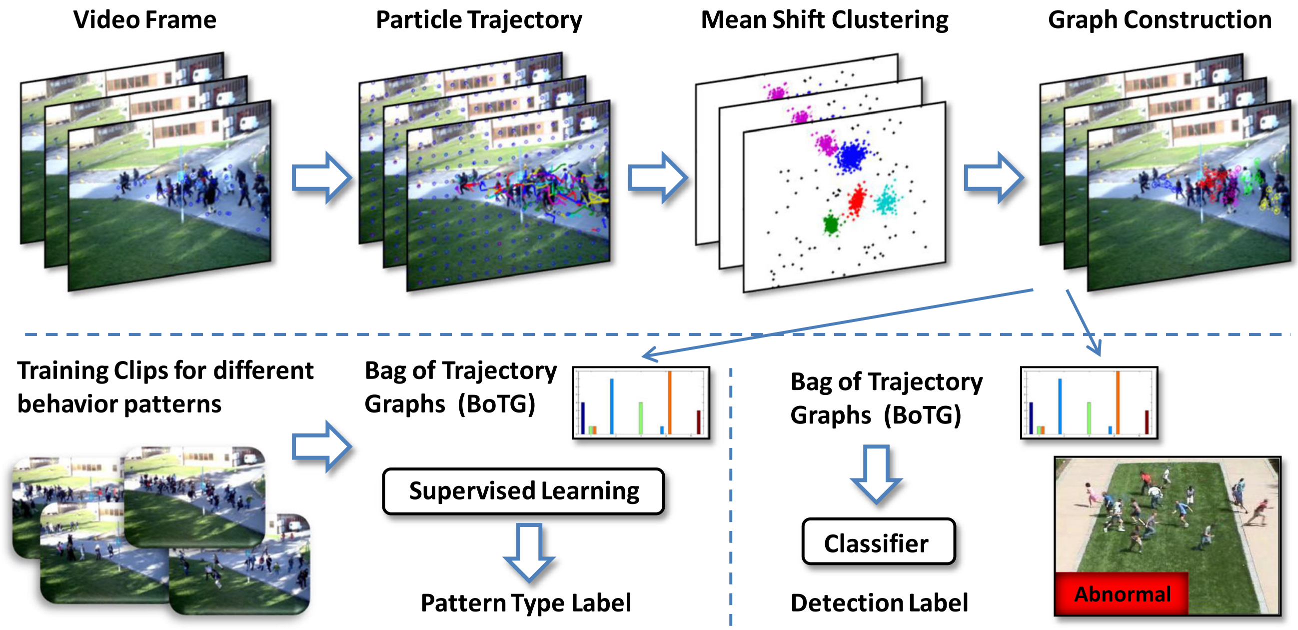

Fig. 1 Processing flow of the proposed BoTG method

velocities for each moving particle are computed using 4th-order Runge–Kutta algorithm.

$$X_w^{t+1} = X_w^t + U_w^t \tag{2}$$
$$Y_h^{t+1} = Y_h^t + V_h^t \tag{3}$$

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:

$$P_{t_1}^{t_T}(t) = (s_i, v_i, t_i)_{t_i=t_1}^{t_T} \tag{4}$$

where s_i and v_i denote the position vectors of particle i in (X_w^{t+1}, Y_h^{t+1}) and velocity vectors of the particle i in (U_w^t, V_h^t) at time t_i which are obtained from the optical flow.

2.1.2 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 [6]. 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 starts 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:

$$\hat{f}(x) = \frac{c_{K,h}}{c} \sum_{i=1}^{c} K\left(\frac{d^2(x, x_i)}{h}\right) \tag{5}$$

where $K(z) > 0 (z \geq 0)$ is a radically symmetric kernel satisfying $\int K(z)dz = 1$, $c_{K,h} > 0$ represents 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. Let the kernel profile $Q(z) = -K'(z)$, and compute the gradient of $\hat{f}(x)$. 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 following:

$$m_h(x) = c'_{K,h} \frac{\nabla \hat{f}(x)}{\hat{f}(x)} = \frac{\sum_{i=1}^{c} x_i Q(x_i)}{\sum_{i=1}^{c} Q(x_i)} - x \tag{6}$$

where $c'_{K,h}$ is the derived coefficient using gradient ascent algorithm. The next point x_{j+1} is shifting from x_j as,

$$x_{j+1} = x_j + m_h(x_j), \quad j = 1, \ldots, c. \tag{7}$$

The kernel is recursively moved and converges to the nearest mode as the 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. 2a, 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).

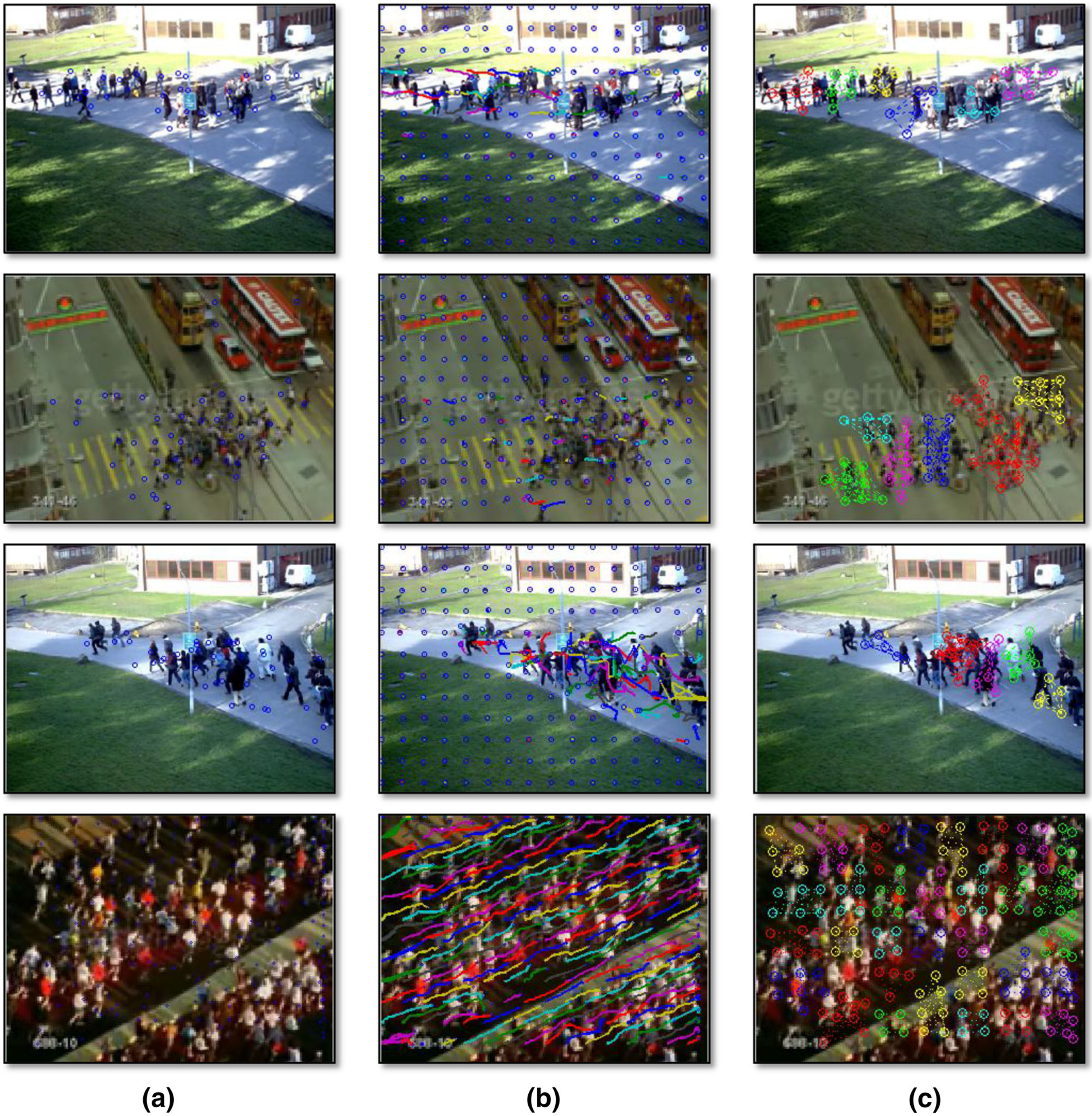

Fig. 2 Examples of graph construction. **a** Filtered particles. **b** Filtered trajectories. **c** Graphs after clustering

In particle trajectory graph construction, trajectories in each group-level cluster are fully connected using the Euclidean distance as the edge weights (Fig. 2c). 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 assigned for E. We next try to specify the detailed description based on the basic graphs.

2.2 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 the motion dynamics information.

2.2.1 Feature representation for trajectory graphs

We represent the properties of trajectory graphs by using the three following features:

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 [27]. Suppose we have N graphs with m trajectories for each in a T-frame clip, for each graph $G_k = (V_k, E_k, W_k)$, the Laplace matrix is defined as:

$$A_k = \begin{cases} d_{ij} & \text{if } i \neq j \\ -\sum_{j=1}^{m} d_{ij} & \text{otherwise} \end{cases} \tag{8}$$

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),

$$A_k = U \Delta U^T \tag{9}$$

where $\Delta = diag\{\tau_1, \tau_2, \ldots, \tau_m\}$,$\tau_1 \geq \tau_2 \geq \cdots \geq \tau_m \geq 0$. 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 $G_k = (V_k, E_k, W_k)$, for each node $v_k = \{s_1 \ldots s_T | s_i = (x_i, y_i)\} \in V_k$, the basic speed $S(.)$ and orientation $O(.)$ feature channels are computed as follows:

$$\begin{aligned} S(v_k) &= \sqrt{(x_T - x_1)^2 + (y_T - y_1)^2} \\ O(v_k) &= \arctan\left(\frac{x_T - x_1}{y_T - y_1}\right) \end{aligned} \tag{10}$$

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:

$$\begin{aligned} H_{ori} &= \{O_i\}_{i=1}^{n} \\ H_{spd} &= \{S_i\}_{i=1}^{n} \end{aligned} \tag{11}$$

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. As above, all the features can be represented by 24D vector (concatenating 3+8+8+3+2 D vector). They describe the group-level structure that depicts the "shape" of groups and the motion of the group that draws the "appearance" of the groups. We next built our bag of words learning scheme by concatenating these features.

2.3 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. The BoTG represent the crowd behavior patterns by a histogram vector h_j,

$$h_j = (f_1, f_2, \ldots, f_i, \ldots, f_d)^T \tag{12}$$

$$f_i = \frac{n_{ij}}{n_j} \tag{13}$$

where d the selected word number which is the size of the vocabulary. n_{ij} is the frequency of the occurrence of the ith trajectory graph appearing in the jth clip. The vector h_j contains the distribution information in a certain crowded scene, which is normalized by the total number of the graphs in the jth clips. Therefore, each T-frame crowd video can be represented by BoTG histogram. 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.

3 Experimental evaluation

3.1 Abnormality detection

To validate the effectiveness of our proposed model on abnormal event detection, we conduct it on the UMN dataset.[1] In the experiment, the detection performance of each method is evaluated by event-level measurement as proposed in [7].

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

[1] http://mha.cs.umn.edu/movies/crowdactivity-all.avi.

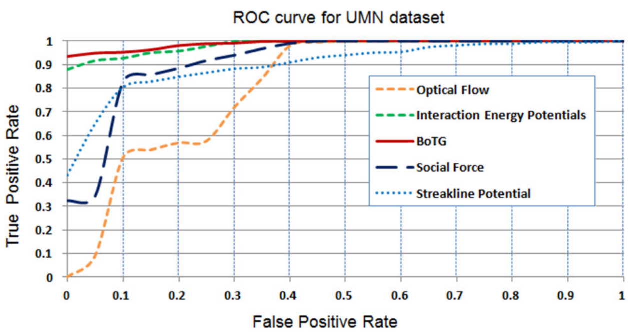

Fig. 3 ROC curves of abnormal detection in UMN dataset

mean shift clustering part, the bandwidth h equals the half of number 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, 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 [9,20,21]. Table 1 shows that BoTG (AUC = 0.990) can achieve better performance over available high-level methods including Interaction Energy Potentials [9] (IEP), Social Force [20] (SF), Streakline Potential [21] (SP) and Optical Flow (OF). Better effect comes from the fact that our graph structure and group attributes are more discriminative for group pattern changes and the motion speed feature also performs significantly in the dispersing abnormality. It indicates BoTG has the 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	AUC
BoTG	0.990
IEP [9]	0.985
SF [20]	0.96
SP [21]	0.90
OF	0.86

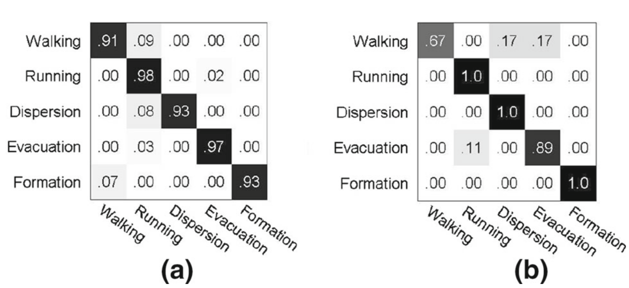

Fig. 4 Confusion matrices for event recognition. **a** Results in UMN. **b** Results in PETS2009 S3. *Rows* are ground truths and *columns* are the predictions

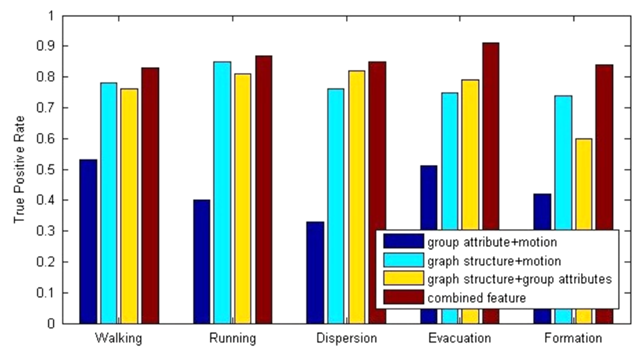

Fig. 5 TPR of UMN recognition performance for different event using different combination of features

3.2 Event recognition

In this experiment, we consider the event recognition for the crowded scenes. 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 dataset and PETS2009 S3 dataset.[2]

Dataset. The UMN dataset is manually segmented into 450 clips of 10 frames. PETS 2009 S3 dataset is segmented into 65 clips. All the video clips are labeled with the events 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 is trained for each type using BoTG. Since the ground truth annotation is given per clip, we evaluate the recognition performance with confusion matrices. Since the UMN contains more complex behavior than PETS2009 does, to verify the validity of features, we also show the true positive rate (TPR) of recognition for all event types on the UMN.

Results and discussion. Figure 4 shows the confusion matrices between 5 types of events on the UMN and PETS2009. BoTG effectively recognizes different patterns, while confusion only occurs in very similar components. Figure 5 illustrates TPR results of different feature component combination strategies on the 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

[2] http://ftp.cs.rdg.ac.uk/PETS2009.

Table 2 Computational complexity cost evaluation of different part

Steps	Time cost (s)
The particle advection	1.411
The mean shift clustering	0.271
The feature representation	0.178

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 the 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 that the features are complementary to each other. Through the confusion matrices calculation, we achieve an overall above 90 % accuracy in the UMN (94.4 %) and PETS2009 S3 dataset (91.2 %).

Computational efficiency analysis. We further give the time cost of our scheme for constructing the BoTG, which contains three following parts. The first part is particle advection for the trajectory computation, which is approximately $O(n \times k^4)$ with n particle locations and an adjust value k for advection. The second part is mean shift clustering process with the time complexity as $O(i \times j)$ for i-weight by j-height frames. The last part is $O(n \times n) + 2O(n)$ for the feature representation. Table 2 further shows the complexity comparisons of the different parts for $T = 10$ frames of video, which is measured by seconds. It is obvious that our proposed method is capable of reaching online efficiency for crowd applications.

4 Conclusion

In this paper, a BoTG representation is proposed for dense crowd event recognition. Different from the previous works, we present an efficient graph construction approach to embed the particle trajectories into a group-level representation. We also propose informative group-level graph descriptors, which effectively capture structures and motion dynamic group-level of behavior patterns. Furthermore, experimental results indicate the effectiveness of our approach is notable on abnormality detection. Our approach is verified with different feature combination and classifiers for event recognition and promising performance is achieved. In the future work, we will focus on the hierarchical model and learn the graph-based crowd patterns to recognize the crowd behavior. The social priors and subgraph mining method will also be considered.

Acknowledgments 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, 61133003, 61332016 and 61035001.

References

1. Adam, A., Rivlin, E., Shimshoni, I., Reinitz, D.: Robust real-time unusual event detection using multiple fixed-location monitors. IEEE Trans. Pattern Anal. Mach. Intell. **30**(3), 555–560 (2008)
2. Ali, S., Shah, M.: A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In: CVPR 2007, IEEE, pp. 1–6 (2007)
3. Ali, S., Shah, M.: Floor fields for tracking in high density crowd scenes. ECCV **2008**, 1–14 (2008)
4. Basharat, A., Gritai, A., Shah, M.: Learning object motion patterns for anomaly detection and improved object detection. In: CVPR 2008, IEEE, pp. 1–8 (2008)
5. Cheng, Z., Qin, L., Huang, Q., Jiang, S., Tian, Q.: Group activity recognition by gaussian processes estimation. In: ICPR 2010, IEEE, pp. 3228–3231 (2010)
6. Comaniciu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Mach. Intell. **24**(5), 603–619 (2002)
7. Cong, Y., Yuan, J., Liu, J.: Sparse reconstruction cost for abnormal event detection. In: CVPR 2011, IEEE, pp. 3449–3456 (2011)
8. Cristani, M., Bazzani, L., Paggetti, G., Fossati, A., Tosato, D., Del Bue, A., Menegaz, G., Murino, V.: Social interaction discovery by statistical analysis of f-formations. In: BMVC, pp. 1–12 (2011)
9. Cui, X., Liu, Q., Gao, M., Metaxas, D.: Abnormal detection using interaction energy potentials. In: CVPR 2011, IEEE, pp. 3161–3167 (2011)
10. Ge, W., Collins, R., Ruback, B.: Automatically detecting the small group structure of a crowd. In: WACV 2009, IEEE, pp. 1–8 (2009)
11. Hospedales, T., Gong, S., Xiang, T.: A markov clustering topic model for mining behaviour in video. In: ICCV 2009, IEEE, pp. 1165–1172 (2009)
12. Hu, M., Ali, S., Shah, M.: Learning motion patterns in crowded scenes using motion flow field In: ICPR 2008, pp. 1–5 (2008)
13. Ihaddadene, N., Djeraba, C.: Real-time crowd motion analysis. In: ICPR 2008, IEEE, pp. 1–4 (2008)
14. Johnson, N., Hogg, D.: Learning the distribution of object trajectories for event recognition. Image Vis. Comput. **14**(8), 609–615 (1996)
15. Kim, J., Grauman, K.: Observe locally, infer globally: a space-time mrf for detecting abnormal activities with incremental updates. In: CVPR 2009, IEEE, pp. 2921–2928 (2009)
16. Kratz, L., Nishino, K.: Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In: CVPR 2009, IEEE, pp. 1446–1453 (2009)
17. Kuettel, D., Breitenstein, M., Van Gool, L., Ferrari, V.: What's going on? Discovering spatio-temporal dependencies in dynamic scenes. In: CVPR 2010, IEEE, pp. 1951–1958 (2010)
18. Lan, T., Wang, Y., Yang, W., Mori, G.: Beyond actions: discriminative models for contextual group activities. NIPS **4321**, 4322–4325 (2010)
19. Lin, W., Sun, M.T., Poovendran, R., Zhang, Z.: Group event detection with a varying number of group members for video surveillance. IEEE Trans. Circuits Syst. Video Technol. **20**(8), 1057–1067 (2010)
20. Mehran, R., Oyama, A., Shah, M.: Abnormal crowd behavior detection using social force model. In: CVPR 2009, IEEE, pp. 935–942 (2009)

21. Mehran, R., Moore, B., Shah, M.: A streakline representation of flow in crowded scenes. ECCV **2010**, 439–452 (2010)
22. Ni, B., Yan, S., Kassim, A.: Recognizing human group activities with localized causalities. In: IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE, pp. 1470–1477 (2009)
23. Oliver, N., Rosario, B., Pentland, A.: Graphical models for recognizing human interactions. In: NIPS, Citeseer, pp. 924–930 (1998)
24. Pellegrini, S., Ess, A., Schindler, K., Van Gool, L.: You'll never walk alone: Modeling social behavior for multi-target tracking. In: ICCV 2009, IEEE, pp. 261–268 (2009)
25. Pellegrini, S., Ess, A., Van Gool, L.: Improving data association by joint modeling of pedestrian trajectories and groupings. In: ECCV (2010)
26. Stauffer, C., Grimson, W.E.L.: Learning patterns of activity using real-time tracking. IEEE Trans. Pattern Anal. Mach. Intell. **22**(8), 747–757 (2000)
27. Tang, J., Zhang, C., Luo, B.: A graph and pnn-based approach to image classification. In: Proceedings of 2005 International Conference on Machine Learning and Cybernetics, 2005. IEEE, vol 8, pp. 5122–5126 (2005)
28. Tran, D., Yuan, J.: Optimal spatio-temporal path discovery for video event detection. In: CVPR 2011, IEEE, pp. 3321–3328 (2011)
29. Wang, X., Ma, K.T., Ng, G.W., Grimson, W.E.L.: Trajectory analysis and semantic region modeling using nonparametric hierarchical bayesian models. Int. J. Comput. Vis. **95**(3), 287–312 (2011)
30. Wu, S.W.S., Moore, B., Shah, M.: Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes. CVPR (2010)
31. Zhang, Y., Qin, L., Yao, H., Huang, Q.: Abnormal crowd behavior detection based on social attribute-aware force model. In: ICIP 2012, IEEE, pp. 2689–2692 (2012)
32. Zhang, Y., Qin, L., Yao, H., Xu, P., Huang, Q.: Beyond particle flow: bag of trajectory graphs for dense crowd event recognition. In: ICIP (2013)