Real-Time Understanding of Abnormal Crowd Behavior on Social Robots

Dekun Hu[1,3], Binghao Meng[1], Shengyi Fan[2], Hong Cheng[1](✉),
Lu Yang[1], and Yanli Ji[1]

[1] Center for Robotics, School of Automation Engineering,
University of Electronic Science and Technology of China, Chengdu, China
{hudekun,hcheng}@uestc.edu.cn
[2] Ricoh Software Research Center of Beijing, Beijing, China
[3] College of Computer Science, University of Chengdu, Chengdu, China

Abstract. Perceiving the crowd behavior is very important for social cloud robots, who serve as guiders at transportation junctions. In this paper, we propose a real-time algorithm based on background modeling to detect collective motions in complex scenes. The proposed algorithm not only avoids unstable foreground extraction, but also has low computational complexity. To detect the abnormal crowd escape, we refer to the definition of the moving energy of patches and use the energy histogram of the patches to effectively and accurately represent the crowd distribution information in the crowd scenes. We have applied the proposed algorithm to the real surveillance videos which contain the aggregation and dispersion events. The experimental results show the significant outperformance of the proposed algorithm in comparison to the-state-of-the-art approach.

Keywords: Crowd behavior analysis · Social robots · Motion energy of patches · Background modeling

1 Introduction

With the development of the robot technology, cloud computing technology and Internet of things technology, cloud service robots have been used in various places, such as airports, bus stops, shopping malls, subway stations, *etc.* They can serve as the guiders for people to take traffic tools or the assistants of shopping. Moreover, they can serve as the safety guards by detecting the abnormal events in real-time. Real-time crowded scene analysis is a very difficult task for social robots due to the inherent complexity and vast diversity such as illumination changes, low resolution, scene depth, camera position, *etc.* With the development of computer vision technologies, various vision based approaches have been proposed to detect abnormal events in surveillance scenes, background modeling [15], sparse representation [4], object tracking [19], face recognition [20] and people counting [10], are considered as the fundamental elements that compose

Y.-S. Ho et al. (Eds.): PCM 2015, Part II, LNCS 9315, pp. 554–563, 2015.
DOI: 10.1007/978-3-319-24078-7_56

an intelligent surveillance system for the anomaly detection. In this paper, we aim to detect the abnormal crowd behaviors [7,9,12,16] based on the computer vision technology in real-time. A variety of algorithms have been proposed to detect abnormal events in scenes, these approaches could be divided into three categories according to the scene representation: a trajectory based approaches [8,18,19,21], patch based approaches [5,6,13,18], sparse coding based approaches [2,7,11,14].

In [8], a normal dictionary set, was constructed by collecting trajectories of normal behaviors and extracting the control point features of cubic B-spline curves, which was further divided into Route sets. Sparse reconstruction coefficients and residuals of a test trajectory to the Route sets could be calculated with the trajectory sparse reconstruction analysis (SRA). A new descriptor named as Social Affinity Maps (SAM) [21] and priors over origin and destination (OD-prior) were proposed to understand the crowd behavior at the scale of million pedestrians for human mobility in crowded spaces such as city centers or train stations. In [18], a robust algorithm was proposed to detect stationary group activities and understand crowd scene. A locally shared foreground codebook and was used to shape the 3D stationary time map.

A collectiveness descriptor for crowd as well as their constituent individuals along with the efficient computation were proposed in [6]. In [5], a novel patch entropy approach to represent the crowd distribution information and the optical flow was introduced to describe the crowd speed. The Gaussian Mixture Model (GMM) over the normal crowd behaviors was used to predict the anomalies in the detecting stage. A hybrid agent system [13] was used to detect abnormal behaviors in crowded scenes, which included static and dynamic agents to observe efficiently the corresponding individual and interactive behaviors in a crowded scene. The crowd behaviors were represented as a bag of words through the integration of static and dynamic agents information.

The Social Force model [14] treated the moving patches as individuals. And their interaction forces are estimated and mapped into the image plane to obtain Force Flow for every pixel in every frame. Randomly selected spatio-temporal volumes of Force Flow are used to model the normal behaviors of the crowd. In [11], based on inherent redundancy of video structures, an efficient sparse combination learning framework was proposed for abnormal behaviors detection. It achieved high detection rates on benchmark datasets at a speed of 140∼150 frames per second on average when computing on an ordinary desktop PC by MATLAB. In [2], unlike existing approaches based on sparse coding , the abnormal events detection model directly sparsely coded the motion features of the center patches with features of its surrounding patches.

In this paper, we take advantage of the distribution of the patches in the frame to simulate the distribution of the individuals in the crowd. The speed of the patches in the frame to simulate the speed of the individuals in the crowd. We proposed patch moving energy to effectively represent the crowd distribution information. As the number of frames with abnormal crowd behaviors is only a small portion of the entire video, it is obvious that abnormal crowd behavior detection is an unbalanced problem. In this paper we simultaneously use the crowd speed and the distribution information to predict abnormal crowd dispersion behaviors. The comparison experiments conducted on surveillance dataset validate the advantages of the proposed algorithm.

The rest of the paper is organized as follows: In Sect. 2, we introduce the procedures of proposed algorithm. In Sect. 3, we present experimental results and the comparisons with the-state-of-art algorithm. Section 4 concludes the paper.

2 The Proposed Algorithm

2.1 Crowd Aggregation Detection

In the field of public security, the massive mass incidents are often from small crowd gathered to evolution. Therefore, moderate scale crowd aggregation detection and its alarming are crucial to social robots for surveillance purpose. The gathered crowd can't be measured using the algorithm based on optical flow due to the relatively static state of crowd in a particular area within a period of time. Hence, a crowd aggregation detection based on real background modeling and hierarchical alarm algorithm is proposed in this paper. The algorithm process is shown in Fig. 1 and described in following subsections.

Background Modeling. In order to model the background in the video scenes, a robust Pixel-to-Model (P2M) background modeling and recovery approach [17] is used in our work. Each pixel is represented by a context feature which consists of local compressive descriptors. The novel P2M distance is employed to classify the potential background pixels. Furthermore, the P2M distances are also utilized to adaptively update the background model in the space of local descriptors in a smooth and efficient way. The P2M based background recovery can robustly reconstruct the clean background and suppress real-world noises.

As shown in Fig. 1, there are two different types of background computing in the crowd aggregation detection: clean background modeling, which is called real background, and the dynamic background. At time t, the dynamic background of scene based on P2M is denoted as $B_{dy}(t)$. The real background B_{real} is computed by

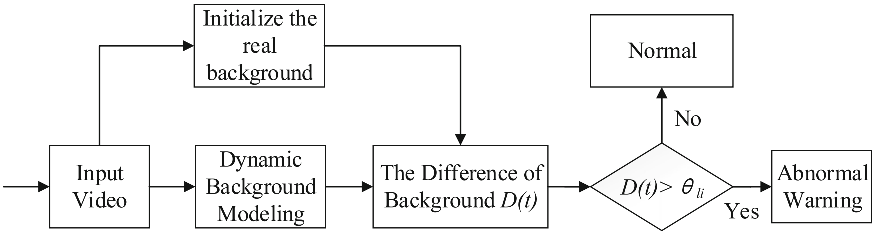

Initialize the real background
Normal
No
Input Video
Dynamic Background Modeling
The Difference of Background D(t)
D(t)> θ_li
Yes
Abnormal Warning

Fig. 1. The framework of the proposed crowd aggregation detection.

$$B_{real} = \frac{1}{N}\sum_{n=1}^{N} b_n, \tag{1}$$

where b_n is a random selected background image from $B_{dy}(t)$ with $t \in [0, T]$ and N is the number of the random selected background images. In practice, the real background can be selected manually for better performance.

Event Detection. One of the main characteristics of the gathered crowd is relatively static within certain areas for a period of time. In order to extract the "static" people from the scene, a $i \times j$ grid of patches is placed over every difference image of clean background and dynamic background, and the size of the patch $P_{(i,j)}$ is $m \times n$. The difference image should be binarized, thus all of pixels with value of 1 presents the "static" foreground. Part of static pixels describes the information of gathered crowd in the scene. As we use the "static" patches to represent the gathered crowd, in order to statistic the "static" patches, we denote the difference image before binarization at time t as $D(t)$, which is calculated by Eq. (2). The value of a patch $V_{P_{(i,j)}}$ is defined by the proportion of non-zero pixels in it as Eq. (3)

$$D(t) = |B_{real} - B_{dy}(t)|, \tag{2}$$

$$V_{P_{(i,j)}} = \begin{cases} 1, & if \sum_{i=1}^{m}\sum_{j=1}^{n} P_{i,j}(x,y) \geq \frac{m \times n}{T_p} \\ 0, & otherwise \end{cases}, \tag{3}$$

where $P_{i,j}(x,y)$ is the value of pixels in the patch $P_{(i,j)}$ of the binary difference image. And T_p is the threshold of a static patch. The value of T_p is 8 in this work.

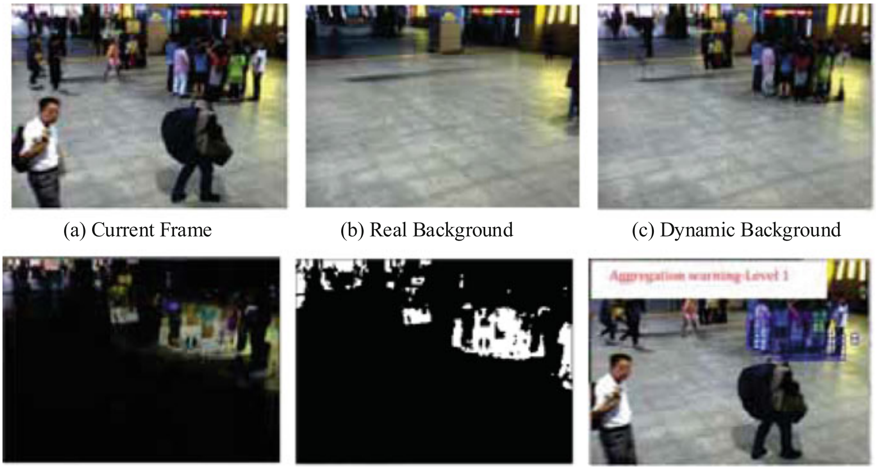
Crowd aggregation area is composed of multiple adjacent static patches. So a weighted sliding window is used to detect the crowd aggregation. The size of the window is integer times of $m \times n$, and its sliding step length is also equal to integer times of the patch size. The value of the window is defined as

$$V_w = \sum_{a=1}^{A}\sum_{b=1}^{B} \lambda_{c1} V_{P_{(i,j)}}, \tag{4}$$

where A is integer times of m and B is integer times of n. λ_{c1} is a compensation parameter of camera calibration, which can improve the effect of the patches far from the camera. A group threshold θ_{l_i} are used to classify the different levels of crowd aggregation according to the value of V_w. An example of crowd aggregation detection is shown in Fig. 2.

2.2 Crowd Escape Detection

In real-life situations, crowd escape occurs by violent movement which is apparent as sudden speeding up, or chaotic movement in a restricted area, or movement contrasting with that of one's neighbors such as in a panic situation. In statistical



(a) Current Frame (b) Real Background (c) Dynamic Background

(d) The Difference of Backgrounds (e) The Binary Image of (d) (f) Detection Result

Fig. 2. An example of crowd aggregation detection.

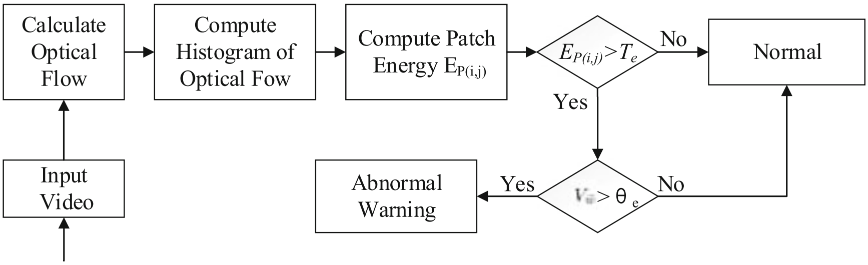


Fig. 3. The framework of the proposed crowd escape detection.

mechanics, entropy is used to measure uncertainty. The greater entropy means the higher disorder, thus the patch entropy approach is proposed to estimate the distribution of the moving patches in [6]. We refer to [6] and propose the patch energy to simulate the distribution of the pedestrians in the crowd. The main steps of the patch energy approach are summarized in Fig. 3 and described in following subsections.

Calculate the Dense Optical Flow. As the moving pedestrians are able to cause the abnormal crowd behaviors, only them need to be concerned about when we detect the crowd escape. We use the moving patches to represent the moving pedestrians in this work. The moving patches extraction stated as the following. Firstly, the velocity of every pixel is calculated by dense optical flow [3]. In order to reduce the influence of illumination change, the average optical flow of continuous several frames is extracted. The map of optical flow is shown in Fig. 4(a). Secondly, every map is divided into $M \times N$ patches. We estimate every patch's velocity with the energy of motion according to the velocity of every pixel in the patch as described in the following subsection.

Patch's Energy of Motion. Assuming the size of every patch in the map of optical flow is $X * Y$, a histogram of the patch is calculated by the different velocities of every pixel(as shown in Fig. 4(b)), every patch in the image has an energy of motion defined by Eq. (5). An example of patch energy change is shown in Fig. 4(c).

$$E_{p_{(i,j)}} = \frac{1}{2}\sum_{r=1}^{H} h_r v_r^2, \ \sum_{r=1}^{H} h_r = X * Y, \tag{5}$$

where v_r is the rth bin in the histogram and h_r is the number of pixels in the rth bin.

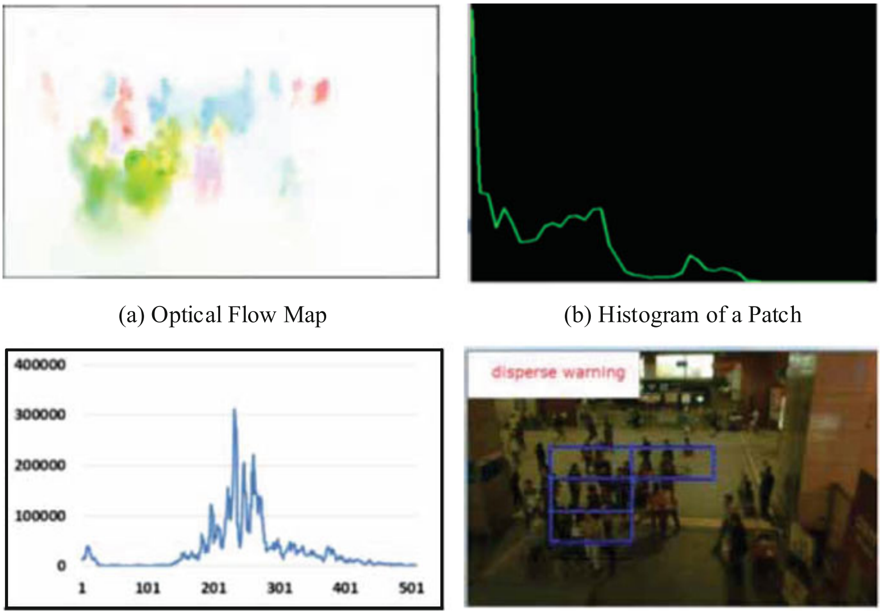
Moving Patches Extraction. We denote a patch as "moving patch" if the energy of motion for it is greater than the threshold T_e. In order to extract the moving patches, we compare every patch's energy of motion with the T_e in turn, the value of T_e is given by experiences in different video scenes. The value of a patch $V_{\widetilde{P}_{(i,j)}}$ is defined as

$$V_{\widetilde{P}_{(i,j)}} = \begin{cases} 1, & if E_{p_{(i,j)}} \geq T_e \\ 0, & otherwise \end{cases}. \tag{6}$$

Event Detection. There will be more patches involving in the escape area if there are more running directions. So a weighted sliding window is used to detect the crowd escape. The size of the window is integer times of the patch, and its sliding step length is also equal to integer times of patch size. The value of the window $V_{\widetilde{w}}$ is defined as

$$V_{\widetilde{w}} = \sum_{i=1}^{A}\sum_{j=1}^{B} \lambda_{c2} V_{\widetilde{P}_{(i,j)}}, \tag{7}$$

where A is integer times of X and B is integer times of Y. λ_{c2}, which can improve the effect of the patches far from the camera, is a compensation parameter of camera calibration, and the value of it will be increased with the increase of the distance between the camera and the real point. If the $V_{\widetilde{w}}$ is greater than the threshold θ_e, the crowd escape behavior happens in this window area. As the crowd escape behavior usually involving plenty of persons, we consider that the crowd escape behavior must happen in more than 2 adjacent window areas. In this way, a lot of false positives have been avoided. An example of crowd escape detection is shown in Fig. 4(d), and the blue boxes in the picture indicate the area where the event happens.



(a) Optical Flow Map

(b) Histogram of a Patch

(c) The Change of Patch Energy

(d) The Result of Detection

Fig. 4. An example of crowd escape detection.

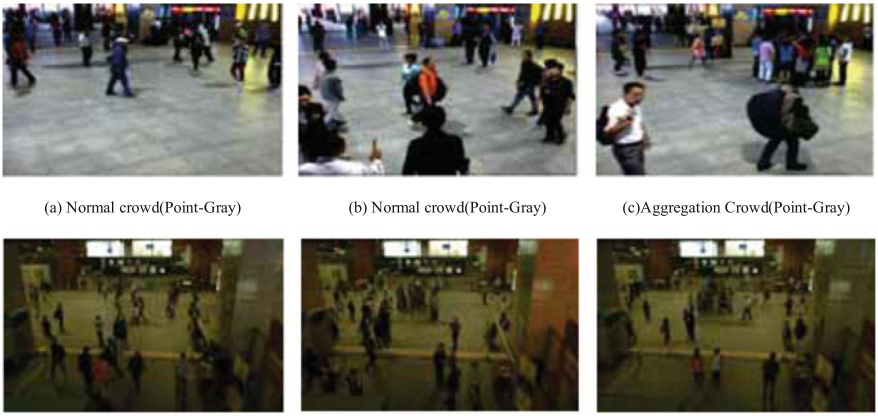
3 Experiment Results and Analysis

3.1 Dataset

To validate the performance of the proposed algorithm, we test it on UMN and RICOH dataset in comparison to the particle entropy algorithm.

The publicly available dataset of the unusual crowd activities from University of Minnesota (UMN) [1] is used to verify the effectiveness of the proposed abnormal crowd event detection algorithm. The dataset consists of 3 different indoor or outdoor scenes with the escape events.

RICOH dataset consists of two kinds of video from two different cameras (TYZX camera and Point-Gray camera). The dataset from TYZX camera includes two dispersing events. It is a low resolution complex scene, involving more moving pedestrians, with illumination changing drastically. So it is very difficult to detection the abnormal crowd behaviors. Another kind of video from Point-Gray camera includes some crowd aggregation in 6 videos, with different view of camera, different direction of aggregation of crowd and different scale of crowd. It is also difficult to detect the event because of the serious occlusion resulting from the low installing location of the camera. The example images of RICOH dataset is shown in Fig. 5.



(a) Normal crowd(Point-Gray)

(b) Normal crowd(Point-Gray)

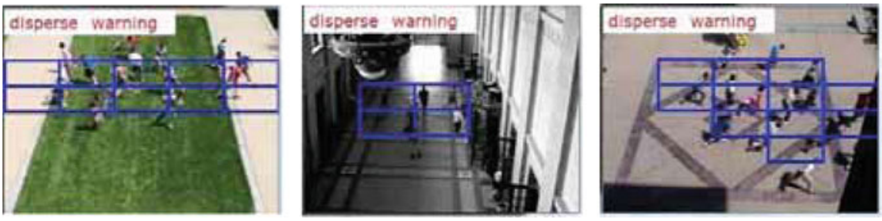
(c)Aggregation Crowd(Point-Gray)

(d) Normal Crowd(TYZX)

(e) Dispersing Crowd(TYZX Scene 1)

(f) Dispersing Crowd(TYZX Scene 2)

Fig. 5. The forward and backward estimation of optical flow.



(a)UMN Scene 1

(b)UMN Scene 2

(c)UMN Scene 3

Fig. 6. The result of crowd escape detection on UMN dataset.

3.2 Experiments

The experiment is conducted as follows. Firstly, the experiment on the crowd aggregation detection is devised on the videos from the Point-Gray camera, the average precision rate is 88 % with 94 % recall rate. In order to compare our algorithm with the state-of-the-art particle entropy algorithm, we conduct the experiments on UMN and TYZX dataset for crowd escape detection. Figure 6 shows some results on UMN Scenes.

The experiment for crowd escape is conducted secondly. Table 1 shows the quantitative comparisons to the particle entropy algorithm in the UMN dataset and TYZX. The precision and recall rate of our algorithm are much better than the particle entropy except the precision on UMN scene2. And the best result even achieves 100 % precision rate and 82 % recall rate. Hence, the proposed algorithm can significant outperform the state-of-the-art particle entropy algorithm on most tested datasets.

Table 1. The result of comparison.

Method Dataset	Our method		The particle entropy [6]	
	precision	recall	precision	recall
UMN Scene 1	99.1%	78.1%	98.1%	75.3%
UMN Scene 2	83.5%	51.1%	96.4%	37.3%
UMN Scene 3	100%	82%	96%	57.5%
TYZX Scene 1	85.3%	35.8%	9.3%	13.6%
TYZX Scene 2	94.3%	58.3%	7.6%	11.4%

4 Conclusions

In the future, robots will play more and more important roles in our life. As one of the crucial role of public security guards, they can sense the abnormal crowd behaviors and activate alarm and evacuate the stream of people, thus reducing the occurrence of public events. In this paper, we propose a novel crowd aggregation detection algorithm based on background modeling firstly. The algorithm can make grading warning to crowd congestion in public security. Secondly, another energy of moving approach is proposed to represent the crowd distribution information. The experimental results on RICOH dataset show the good performance of the proposed approach. Specially, our algorithm is robust to illumination changes, low resolution, scene depth and camera position. The experiments conducted on publicly available dataset showed the effectiveness of the approach and that our algorithm outperforms the state-of-the-art particle entropy algorithm. In the future work, the thresholds in our algorithm will be self-adaptive to avoid the complicated manual modulation for different surveillance scenes.

Acknowledgment. The authors would like to thank all the reviewers for their insightful comments and special thanks to the RICOH (Beijing) for their dataset, financial and technical support. This work was partially supported by NSFC (No.61305033, 61273256, NO.61305043), Fundamental Research Funds for the Central Universities (ZYGX2013J088, ZYGX2014Z009) and SRF for ROCS, SEM. Youth Research Funds of Chengdu university (20805066)

References

1. Unusual crowd activity dataset of university of minnesota. http://mha.cs.umn.edu/proj_events.shtml#crowd
2. Alahi, A., Ramanathan, V., Fei-Fei, L.: Socially-aware large-scale crowd forecasting. In: CVPR. IEEE (2014)
3. Brox, T., Bruhn, A., Papenberg, N., Weickert, J.: High accuracy optical flow estimation based on a theory for warping. In: Pajdla, T., Matas, J.G. (eds.) ECCV 2004. LNCS, vol. 3024, pp. 25–36. Springer, Heidelberg (2004)

4. Cheng, H., Liu, Z., Yang, L., Chen, X.: Sparse representation and learning in visual recognition: theory and applications. Signal Process. **93**(6), 1408–1425 (2013)
5. Cho, S.H., Kang, H.B.: Abnormal behavior detection using hybrid agents in crowded scenes. Pattern Recogn. Lett. **44**, 64–70 (2014)
6. Gu, X., Cui, J., Zhu, Q.: Abnormal crowd behavior detection by using the particle entropy. Optik **125**, 3428–3433 (2014)
7. Kratz, L., Nishino, K.: Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In: CVPR. IEEE (2009)
8. Li, C., Han, Z., Ye, Q., Jiao, J.: Visual abnormal behavior detection based on trajectory sparse reconstruction analysis. Neurocomput. **119**, 94–100 (2013)
9. Liao, Z., Yang, S., Liang, J.: Detection of abnormal crowd distribution. In: Int'l Conference on Green Computing and Communications & Int'l Conference on Cyber, Physical and Social Computing. IEEE/ACM (2010)
10. Liu, X., Tu, P.H., Rittscher, J., Perera, A., Krahnstoever, N.: Detecting and counting people in surveillance applications. In: AVSS. IEEE (2005)
11. Lu, C., Shi, J., Jia, J.: Abnormal event detection at 150 FPS in MATLAB. In: ICCV. IEEE (2013)
12. Mahadevan, V., Li, W., Bhalodia, V., Vasconcelos, N.: Anomaly detection in crowded scenes. In: CVPR. IEEE (2010)
13. Mehran, R., Oyama, A., Shah, M.: Abnormal crowd behavior detection using social force model. In: CVPR. IEEE (2009)
14. Tang, X., Zhang, S., Yao, H.: Sparse coding based motion attention for abnormal event detection. In: ICIP. IEEE (2013)
15. Thijs, G., Lescot, M., Marchal, K., Rombauts, S., De Moor, B., Rouze, P., Moreau, Y.: A higher-order background model improves the detection of promoter regulatory elements by gibbs sampling. Bioinform. **17**(12), 1113–1122 (2001)
16. Wang, B., Ye, M., Li, X., Zhao, F., Ding, J.: Abnormal crowd behavior detection using high-frequency and spatio-temporal features. Mach. Vis. Appl. **23**(3), 501–511 (2012)
17. Yang, L., Cheng, H., Su, J., Li, X.: Pixel-to-model distance for robust background reconstruction. In: TCSVT PP (2015)
18. Yi, S., Wang, X., Lu, C., Jia, J.: L0 regularized stationary time estimation for crowd group analysis. In: CVPR. IEEE (2014)
19. Yilmaz, A., Javed, O., Shah, M.: Object tracking: a survey. CSUR **38**(4), 13 (2006)
20. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: a literature survey. CSUR **35**(4), 399–458 (2003)
21. Zhou, B., Tang, X., Zhang, H., Wang, X.: Measuring crowd collectiveness. Pattern Anal. Mach. Intell. **36**(8), 1586–1599 (2014)