Abnormal Crowd Behavior Detection Using Novel Optical Flow-Based Features

Cem Direkoglu
Middle East Technical University – Northern Cyprus Campus, Kalkanli, Guzelyurt, North Cyprus, via Mersin 10 Turkey
cemdir@metu.edu.tr

Melike Sah
Near East University, Near East Boulevard, 99138, Nicosia, North Cyprus, via Mersin 10 Turkey
melike.sah@neu.edu.tr

Noel E. O'Connor
The Insight Centre for Data Analytics, Dublin City University, Dublin, Ireland
noel.oconnor@dcu.ie

Abstract

In this paper, we propose a novel optical flow based features for abnormal crowd behaviour detection. The proposed feature is mainly based on the angle difference computed between the optical flow vectors in the current frame and in the previous frame at each pixel location. The angle difference information is also combined with the optical flow magnitude to produce new, effective and direction invariant event features. A one-class SVM is utilized to learn normal crowd behavior. If a test sample deviates significantly from the normal behavior, it is detected as abnormal crowd behavior. Although there are many optical flow based features for crowd behaviour analysis, this is the first time the angle difference between optical flow vectors in the current frame and in the previous frame is considered as a anomaly feature. Evaluations on UMN and PETS2009 datasets show that the proposed method performs competitive results compared to the state-of-the-art methods.

1. Introduction

In recent years, many computer vision approaches have been proposed for detection, tracking and recognition of individual objects in surveillance videos, such as a person, car, animal etc. for behavior understanding. On the other hand, crowd (i.e. a group of people) behavior analysis is a new research area in computer vision with a range of important applications such as automatic detection of panic and escape behavior because of violent events, natural disasters, riots or chaotic acts in crowds. Despite the fact that there is much research on vision-based activity analysis for individuals [1], crowd activity analysis remains a challenging problem [2]. In a crowd, there are usually many people located at different positions and moving in different directions making it difficult to find effective features for higher level analysis.

Abnormal event detection can be categorized into two classes: Local and Global abnormal event [16]. In local abnormal event, the behavior of an individual is different from the other individuals in a crowded scene. On the other hand, in global abnormal event, the group behavior of the global scene is abnormal, for example, where the pedestrians suddenly scattered due to an explosion. In our work, we particularly focus on global abnormal crowd behaviour detection.

In general, there are two main approaches for understanding global crowd behaviors: object-based and holistic approaches. In object-based methods, a crowd is considered as a collection of individuals [3][4]. It is necessary to detect and track each individual to understand the crowd behavior [5]. This approach faces considerable complexity in detection of objects, tracking trajectories, and recognizing activities in dense crowds where the whole process is affected by occlusions. On the other hand, holistic approaches [6][7][8] consider the crowd as a global entity and extract features such as using optical flow to represent the state of motion in the whole frame for higher level analysis.

In our research, we particularly focus on global abnormal crowd behavior detection in surveillance videos such as people suddenly start to run around in the same or different directions. Anomaly detection, also named as outlier detection, refers to detecting patterns in a given dataset that do not conform to an established normal behavior. To achieve this, two key issues need to be addressed: event representation and anomaly measurement. For abnormal event representation, some methods consider the spatial-temporal information, such as Histogram of Optical Flow (HOF) [9], Histogram of Motion Direction (HMD) [10], spatial-temporal gradient [11], social force model [12], chaotic invariant [13], mixtures of dynamic textures [14], force field [15] and sparse representation [16]. On the other hand, for anomaly measurement, generally a one-class learning method is used to learn normal samples. For example, Gaussian Mixture Model, Hidden Markov Model [17], one-class Support Vector Machine (SVM), Replicator Neural Networks [18] and Convolutional Neural Network [19][20], and Bayesian model [21]. Then, in the testing phase, if there is any test sample that significantly deviates from the normal type, it is labelled to be abnormal. Extensive Surveys on crowd behavior analysis are provided in [2][22].

In this paper, we propose a new holistic approcah for event feature extraction. In our approach, an abnormal crowd event is defined to be a panic or escape situation

978-1-5386-2939-0/17/$31.00 ©2017 IEEE

where each individual in the group runs around. The main contribution of this paper is an event feature extraction method that is based on optical flow. In an abnormal situation people panic and start to run around, and according to our observation, this abnormal behavior increases not only the optical flow magnitude but also the angle difference between the optical flow vectors computed in the previous frame and in the current frame at each pixel location (especially in motion regions). We propose a mathematical model and algorithm to produce event features that are effective, noise free and invariant to the direction of motion. First, at each pixel location, we evaluate the angle difference between optical flow vectors computed in the current frame and in the previous frame. However there are also small noisy optical flow measurements and their angle difference would affect the observation. To remove these noisy observations, we multiply the angle difference with the optical flow magnitude in the current frame. Finally, the obtained values at each pixel location are sorted in ascending order and a set of the maximum values, i.e. the first 101 values, are selected to represent the feature vector of the current frame. A one-class SVM is used to learn normal crowd behavior. In the testing phase, if there is any test frame that is significantly deviates from the normal behavior, it is labeled as abnormal behavior. It is important to note that although there are many optical flow based features have been proposed for crowd behaviour analysis, this is the first time the angle difference between optical flow vectors in the current frame and in the previous frame is considered as a anomaly feature. Our experiments are conducted on two well-known publicly available surveillance datasets, namely UMN [23] and PETS2009 [24]. Results show that our method is competitive with the state-of-the-art methods in both datasets.

The paper is organized as follows. Section 2 introduces the proposed features based on optical flow. Section 3 presents the detection using the proposed features. Section 4 is the evaluations on UMN and PETS2009 datasets. Conclusions and future work is given in Section 5.

2. Proposed Features based on Optical Flow

The proposed event feature extraction is based on optical flow a well known important image motion description technique. We compute the optical flow at each frame using the Lucas-Kanade algorithm [26]. However, the raw optical flow data cannot be directly used for event representation since it is susceptible to background noise and direction of movement. For example, in an escape situation, each individual in the group may run in similar directions, or in opposite directions. Thus we need features, based on optical flow, that are invariant to the direction of motion and that can discriminate the usual and unusual event at every time instant. Since in an abnormal situation people panic and start to run around, we noticed that, especially in motion regions, the angle difference between the optical flow vectors in the current frame and in the previous frame increases at each pixel location. The angle difference between two vectors, at each pixel location, is computed as follows:

$$\theta_t = \arccos\left(\frac{(u_{t-1}.u_t + v_{t-1}.v_t)}{\left(\sqrt{u_{t-1}^2+v_{t-1}^2}\;.\;\sqrt{u_t^2+v_t^2}\right)}\right) \tag{1}$$

where $\overrightarrow{o_{t-1}}=(u_{t-1}, v_{t-1})$ and $\overrightarrow{o_t}=(u_t, v_t)$ are optical flow vectors, respectively, in the previous frame $(t-1)$ and in the current frame (t) at each pixel location. θ_t is the angle difference at the current frame. However there are also small noisy optical flow measurements and their angle difference would affect the observation. To remove these noisy observations, we multiply the angle difference with the optical flow magnitude in the current frame as given below,

$$I = \sqrt{u_t^2 + v_t^2}\;.\,\theta_t \tag{2}$$

Finally the computed values at each pixel location in the current frame are sorted in ascending order and a set of maximum values, i.e. the first 101 values that is determined experimentally, are selected to form the feature vector representing the event in the current frame.

3. Detection using the Proposed Features

We investigate the use of the proposed features with a one-class Support Vector Machine (SVM) classification. A one-class SVM is used to learn normal events. The test frame is classified using the neighbourhood frames. In UMN [23] dataset, the test frame is classified using the 70 by 70 neighbourhood frames (70 from past and 70 from future neighbourhoods), which is determined experimentally as shown in Figure 6. This means that the window size is 141 (including the test frame). On the other hand, in PETS2009 [24] dataset, the test frame is classified using the 35 by 35 neighbourhood frames, which is also determined experimentally as shown in Figure 6. The window size is 71 including the test frame. Each of the frames in the window is labelled with the one-class SVM classifier. If there is any frame that is significantly deviating from the normal type, it is labeled to be abnormal. Otherwise it labeled to be normal. Then the most frequent class is selected to represent the behaviour of the test frame. In one-class SVM, a Polynomial kernel with parameters $nu = 0.86$ and $cost = 0.1$ is used.

4. Experimental Results

Our experiments are conducted on two different publicly available surveillance datasets, namely UMN and

PETS2009, which are widely used for evaluation. In addition, we compare our method with the state-of-the-art methods introduced for global abnormal event detection such as the method based on Bayesian model (BM) [21], sparse reconstruction cost (SRC) [16], chaotic invariants (CI) [13], the social force model (SF) [12], and the force field model (FF) [15]. We follow the same evaluation settings, and we use the same accuracy measurement as outlined in [21]. The accuracy is defined to be the percentage of correctly identified frames that is calculated by comparing with the ground truth. Although some of the compared methods perform evaluations on videos that aee gathered from the Internet, these videos are not available online for comparison. Therefore, we focus on evaluations on PETS2009 and UMN benchmark datasets, which are publicly available.

4.1. Evaluations on UMN Dataset

The UMN crowd dataset [23] contains normal and abnormal crowd behaviors which are captured at indoor and outdoor scenes of University of Minnesota. Each video starts with a normal crowd behavior and ends with sequences of abnormal crowd behavior. The dataset contains 11 videos with a total of 7736 frames that is captured at three different scenes. Scene 1 is an outdoor scene, which contains two scenarios. Scene 2 consists of six scenarios and it was captured in an indoor location. Scene 3 contains three scenarios that was captured in an outdoor scene. We use the same evaluation settings as described in [21]. We select 50 nonescape (i.e. normal behaviour) frames randomly in each scenario where the features extracted in these frames represent the training data. Resulting in 100, 300, and 150 frames training for scene 1, 2 and 3 respectively.

Table 1: Accuracy (%) comparison of the proposed method with BM, FF, CI, SF and SRC for abnormal crowd behavior detection in the UMN dataset

	Proposed Method	BM	FF	CI	SF	SRC
Scene 1	99.10	99.03	88.69	90.62	84.41	90.52
Scene 2	94.85	95.36	80.00	85.06	82.35	78.48
Scene 3	97.76	96.63	77.92	91.58	90.83	92.70
Overall Accuracy	96.46	96.40	81.04	87.91	85.09	84.70

Table 1 shows accuracy comparison of six methods for three different scenes. Overall, the proposed method acvieves the best accuracy with an average of 96.46%, which is higher than the accuracy of BM (96.40%), FF (81.04%), CI (87.91%), SF (85.09%) and SRC (84.70%). The accuracy of the other methods under the same evaluation settings was taken from [21]. We observe that both the proposed method and the BM performs superior compared to other methods. Overall, the proposed method performs slightly better than the BM. Some illustrative accuracy comparison of the proposed method in UMN dataset is also shown in Figure 1 and 2. The proposed method achieves very accurate anormaly detection compared to other methods.

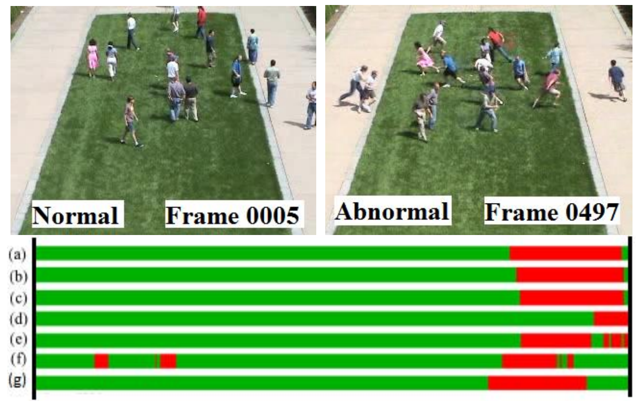


Figure 1: Crowd escape behavior detection for scenario 1 in the UMN dataset. (a) Ground truth, (b) result of the proposed method, (c) result of BM, (d) result of FF, (e) result of CI, (f) result of SF, and (g) result of SRC.

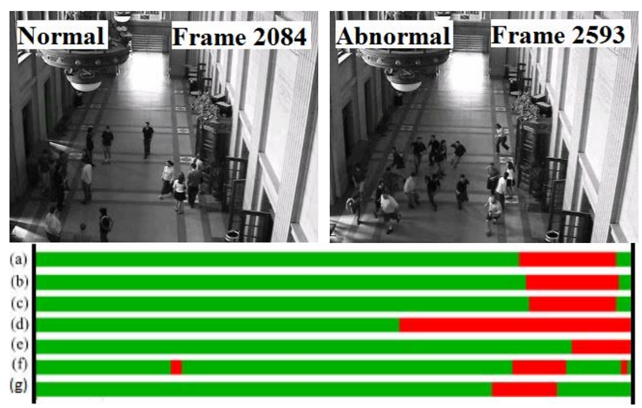


Figure 2: Crowd escape behavior detection for scenario 6 in the UMN dataset. (a) Ground truth, (b) result of the proposed method, (c) result of BM, (d) result of FF, (e) result of CI, (f) result of SF, and (g) result of SRC.

4.2. Evaluations on PETS2009 Dataset

The PETS2009 dataset [24] consists of different crowd activities and contains five parts: (i) calibration data, (ii) training data, (iii) person count and density estimation data, (iv) person tracking data and (v) flow analysis and event recognition data. Each part contains several sequences and each sequence contains different views. There are two scenarios in this dataset related to abnormal crowd behaviour. Each of these scenarios was captured from four different camera views, resulting in 8 videos. The same crowd escape scenario is captured from four different camera views, resulting in significant differences in the field of view and illumunation. As a result of variations in view and illumunations, detecting abnormal crowd behavior for the same scenario becomes a difficult task. The first scenario contains 107 frames, and in this scenario crowd is walking from right to left, and suddenly they start to run in the same direction. The second scenario consists of 374 frames. In this scenario, three different group of people walks to the center and stay there for a while. Then, suddenly crowd starts to run away in different directions randomly.

We compare the proposed method with BM [21], FF [24], CI [13] and SF [12]. We follow the same evaluation settings as outlined in [21]. In the first scenario, for each view, we prepare the training data using the features of the randomly selected 30 nonescape frames.

Table 2 shows the accuracy comparison of five methods for each view in the first scenario. In overall, the proposed method outperforms other methods. In this dataset, especially for views 3 and 4, videos were captured from an angle where the field of view is occluded by a tree. In addition, from these camera angles, as soon as the crowd appear, the crowd escape event starts very quickly, which means there is less number of training frames for normal crowd behavior. Despite this, the proposed features performs good, which is better than the performance of BM, FF, SF and CI (for view 4). In view 1 and 2 the proposed features achieves very good results with an average of 92.24% and 89.50% (which is the best of all methods).

Table 2: Accuracy (%) comparison of the proposed method with BM, FF, CI and SF for crowd escape behavior detection for the first scenario in the PETS2009 dataset

	Proposed Method	**BM**	**FF**	**CI**	**SF**
View 1	99.07	92.45	37.74	56.60	63.21
View 2	98.13	83.02	37.74	83.02	70.76
View 3	62.62	89.62	37.74	81.13	52.83
View 4	97.20	90.57	37.74	52.83	48.11
Overall Accuracy	89.25	88.92	37.74	68.40	58.73

In the second scenario, we again follow the same evaluation settings as described in [21]. In the second scenario, for each view, we prepare the training data using the features of the randomly selected 100 nonescape (i.e. normal behaviour) frames. Table 3 illustrates accuracy evaluation of five methods for the second scenario. Overall, the proposed method performs better than the other methods. Only for the view 4 is the proposed method ranked behind the BM and FF methods. This mainly because there is a significant resolution problem on this view and the video is suddenly cut and started with another activity. As a result, calculation of angle difference between optical flow vectors is affected. Despite this, for the second secanario, the proposed method achieves the best overall accuracy with an average of 96.72%, compared to the overall results of BM (94.22%), FF (87.66%), CI (92.62%) and SF (84.97%) methods. The accuracy of the other methods under the same evaluation settings was taken from [21].

Table 3: Accuracy (%) comparison of the proposed method with BM, FF, CI and SF for crowd escape behavior detection for the second scenario in the PETS2009 dataset

	Proposed Method	**BM**	**FF**	**CI**	**SF**
View 1	98.66	96.01	94.50	94.95	91.22
View 2	99.20	94.15	63.83	92.02	89.36
View 3	99.47	95.21	95.48	94.15	94.68
View 4	89.57	91.49	96.81	89.36	64.63
Overall Accuracy	96.72	94.22	87.66	92.62	84.97

Some representative accuracy comparison of the proposed method in the PETS2009 dataset is also illustrated in Figure 3 and 4. The proposed method performs more accurate anomaly detection compared to the other methods. Furthermore, in Figure 5, we illustrate the accuracy comparisons in graphical form for all datasets and videos.

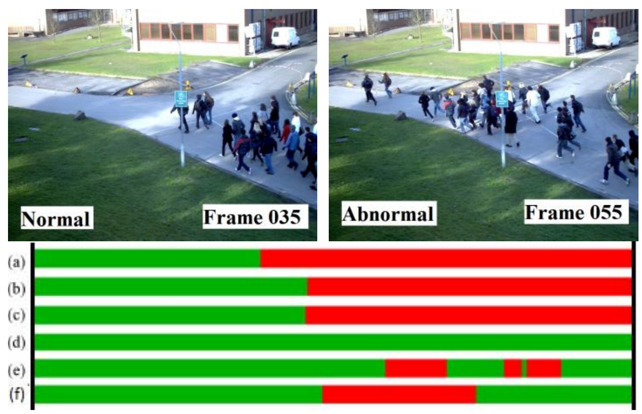


Figure 3: Crowd escape behavior detection for view 1 in the first scenario of PETS2009 dataset. (a) Ground truth, (b) result of the proposed method, (c) result of BM, (d) result of FF, (e) result of CI and (f) result of SF.

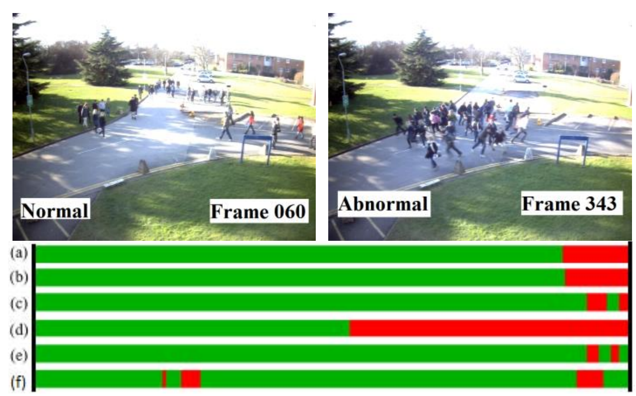


Figure 4: Crowd escape behavior detection for view 2 in the second scenario of PETS2009 dataset. (a) Ground truth, (b) result of the proposed method, (c) result of BM, (d) result of FF, (e) result of CI and (f) result of SF.

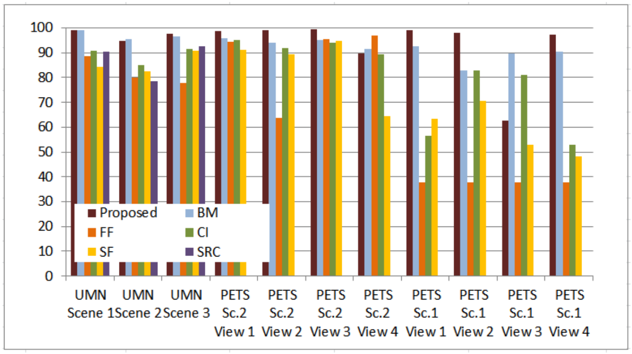


Figure 5: Accuracy comparisons of the proposed method with BM, FF, CI, SF and SRC on all scenarios for UMN and PETS2009 datasets.

4.3. The Effect of Window Size

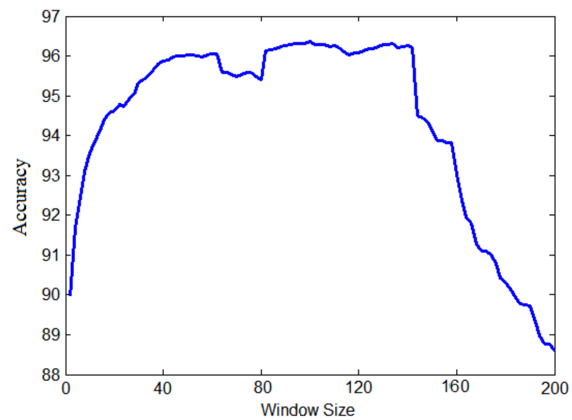
In both datasets, we present the effect of differing window size to optimize the accuracy. Figure 6 shows accuracy performances for the proposed feature with one-class SVM. In the UMN dataset, the window size ranges from 1 to 201. It is observed that the optimal window size for the proposed feature in the UMN dataset is 141 (Figure 6-a). In the second scenario of the PETS2009 dataset, the window sizes ranges from 1 to 141. In this dataset, it is observed that for the proposed feature the best accuracy is obtained when the window size is 71 (Figure 6-b).

4.4. The Effect of Angle Difference and Optical Flow Magnitude

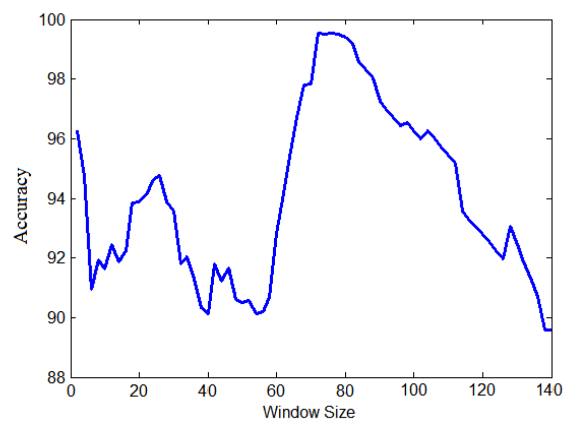
We present the influence of angle difference and optical flow magnitude, and report what the accuracy would be if only angle difference or only optical flow magnitude or combination of them (i.e. multiplication of them as given in Equation 2) was used. Table 4 shows accuracy performances on the UMN dataset. In overall, only angle difference achieves 83.73%, only optical flow magnitude achieves 89.43% and the combination achieves 96.46%. Results indicate that using the combination improves the accuracy considerably.

5. Conclusions

In this paper, we propose novel optical flow-based features for abnormal crowd behavior detection. The proposed feature is based on the angle difference computed between the optical flow vectors in the current frame and in the previous frame at each pixel location. Since there are also small noisy optical flow vectors and their angle difference would affect the observation, we multiply the angle difference with the optical flow magnitude in the current frame and select a set of maximum values to form the feature vector representing the event in the current frame. We utilize a one-class SVM to learn normal behavior, and when a test sample is significantly deviating from the normal behavior, it is detected as abnormal crowd behavior. Evaluations on UMN and PETS2009 datasets show that the proposed method is very effective and perfoms competitive results compared to the state-of-the-art methods for anormally detection.



(a)



(b)

Figure 6: The effect of window size on accuracy. (a) Accuracy performances with differing window size in the UMN dataset. (b) Accuracy performances with differing window size in the PETS2009 dataset for the second scenario.

Table 4: The effect of angle difference and optical flow magnitude on accuracy (%) in the UMN dataset

	Only Angle Difference	Only Magnitude	Combined
Scene 1	87.25	92.41	99.10
Scene 2	80.04	85.80	94.85
Scene 3	88.46	94.34	97.76
Overall Accuracy	83.73	89.43	96.46

Acknowledgements

This research is supported by Middle East Technical University – Northern Cyprus Campus, Scientific Research Project (SRP) Fund (Grant no: FEN-16-YG-10).

References

[1] J.K. Aggarwal and M.S. Ryoo: ”Human Activity Analysis: A Review”, ACM Computing Surveys, 43(3):16, 2011.

[2] T. Li, H. Chang, M. Wang, B. Ni, R. Hong and S. Yan. Crowded Scene Analysis: A Survey, in *IEEE Transactions on Circuits and Systems for Video Technology*, 25:(3):367-386, 2015.

[3] J.S. Marques, P.M. Jorge, A.J. Abrantes, and J.M. Lemos. Tracking Groups of Pedestrians in Video Sequences. Computer Vision and Pattern Recognition Workshop, 9:101, 2003.

[4] P. Tu, T. Sebastian, G. Doretto, N. Krahnstoever, J. Rittscher, and T. Yu. Unified Crowd Segmentation. European Conference on Computer Vision, 5305:691-704, 2008.

[5] G. Brostow and R. Cipolla. Unsupervised Bayesian Detection of Independent Motion in Crowds. IEEE Computer Vision and Pattern Recognition, 1:594-601, 2006.

[6] S.B. Ernesto L. Andrade and R.B. Fisher. Modelling Crowd Scenes for Event Detection. IEEE International Conference on Pattern Recognition, 1:175-178, 2006.

[7] S. Ali and M. Shah. A Lagrangian Particle Dynamics Approach for Crowd Flow Segmentation and Stability Analysis. IEEE Conference on Computer Vision and Pattern Recognition, 1-6, 2007.

[8] R. Mehran, A. Oyama, M. Shah. Abnormal Crowd Behavior Detection using Social Force Model. IEEE Conference on Computer Vision and Pattern Recognition, pages:935-942, 2009.

[9] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz. Robust Real-Time Unusual Event Detection using Multiple Fixed-Location Monitors, IEEE Transactions on Pattern Analysis and Machine Intelligence 30, 555–560.

[10] H. M. Dee and A. Caplier, Crowd behaviour analysis using histograms of motion direction, *IEEE International Conference on Image Processing*, pages: 1545-1548, 2010.

[11] L. Kratz, and K. Nishino. Anomaly Detection in Extremely Crowded Scenes using Spatio-Temporal Motion Pattern Models. IEEE Conference on Computer Vision and Pattern Recognition, pages:1446-1453, 2009.

[12] R. Mehran, A. Oyama, M. Shah. Abnormal Crowd Behavior Detection using Social Force Model. IEEE Conference on Computer Vision and Pattern Recognition, pages:935-942, 2009.

[13] S. Wu, B. E. Moore, M. Shah. Chaotic Invariants of Lagrangian Particle Trajectories for Anomaly Detection in Crowded Scenes. IEEE Conference on Computer Vision and Pattern Recognition, pages:2054-2060, 2010.

[14] L. Weixin, V. Mahadevan, and N. Vasconcelos. Anomaly Detection and Localization in Crowded Scenes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(1):18-32, 2014.

[15] D. Chen and P. Huang. “Motion-based unusual event detection in human crowds”, in Journal of Visual Communication and Image Representation, 22(2), pages: 178-186, 2011.

[16] Y. Cong, J. Yuan and J. Liu. Abnormal Event Detection in Crowded Scenes using Sparse Representation. Pattern Recognition, 46(7):1851-1864, 2013.

[17] L. Kratz, and K. Nishino. Anomaly Detection in Extremely Crowded Scenes using Spatio-Temporal Motion Pattern Models. IEEE Conference on Computer Vision and Pattern Recognition, pages:1446-1453, 2009.

[18] S. Hawkins, H. He, G. Williams, and R. Baxter. Outlier Detection Using Replicator Neural Networks. International Conference in Data Warehousing and Knowledge Discovery, 2002.

[19] Z. Fang, F. Fei, Y. Fang, C. Lee, N. Xiong, L. Shu and S. Chen. Abnormal event detection in crowded scenes based on deep learning, *Multimedia Tools and Applications*, 75(22):14617–14639, November 2016.

[20] Y. Feng, Y. Yuan, and X. Lu. Deep Representation for Abnormal Event Detection in Crowded Scenes. In *Proceedings of ACM on Multimedia Conference* (MM '16). Pages:591–595, 2016.

[21] S. Wu, H. S. Wong and Z. Yu, "A Bayesian Model for Crowd Escape Behavior Detection" in *IEEE Transactions on Circuits and Systems for Video Technology*, 24(1):85–98, January 2014.

[22] V. J. Kok, M. K. Lim, C. S. Chan. “Crowd behavior analysis: A review where physics meets biology,” in *Neurocomputing*, 177, pages:342-362, 2016.

[23] University of Minnesota, available from http://mha.cs.umn.edu/proj_events.shtml#crowd

[24] University of Reading, PETS 2009 Dataset S3 Rapid Dispersion, available from http://www.cvg.reading.ac.uk/PETS2009/a.html

[25] D. Chen and P. Huang. “Motion-based unusual event detection in human crowds”, in Journal of Visual Communication and Image Representation, 22(2), pages: 178-186, 2011.

[26] Bruce D. Lucas and Takeo Kanade. “An Iterative Image Registration Technique with an Application to Stereo Vision”, in International Joint Conference on Artificial Intelligence, pages 674-679, 1981.