IET Image Processing

Research Article

Anomaly detection and localisation in the crowd scenes using a block-based social force model

ISSN 1751-9659
Received on 23rd July 2016
Revised 2nd May 2017
Accepted on 1st October 2017
E-First on 20th November 2017
doi: 10.1049/iet-ipr.2016.0044
www.ietdl.org

Qing-Ge Ji[1], *Rui Chi*[1], *Zhe-Ming Lu*[2] ✉
[1] *School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510006, People's Republic of China*
[2] *School of Aeronautics and Astronautics, Zhejiang University, Hangzhou 310027, People's Republic of China*
✉ *E-mail: zheminglu@zju.edu.cn*

Abstract: A novel approach to detect and localise anomalous events in crowed scenes by processing surveillance videos is introduced in this study. Unusual events are those that significantly differ from current dominated behaviours. The proposed approach both detects pixel-level and block-level anomalies. In pixel level, Gaussian mixture models are used to detect abnormalities. Block-level detection segments the crowd into blocks according to pedestrian detection, and then anomalies are spotted and localised with a social force model. Experimental results using the USCD datasets Ped1 and Ped2 show that the proposed method performs favourably against state-of-the-art methods.

1 Introduction

Over the last decade, the demand for intelligent monitoring has been increasing rapidly due to public safety needs. Surveillance systems relying on special duty cannot respond to an incident immediately and cannot meet the security needs of many sectors. Abnormal behaviour detection is the main task of intelligent monitoring systems. Abnormal behaviour analysis in crowded scenes is a challenging task due to the tremendous variety of motion patterns. Labelling all the possible anomalies in supervised learning method is impossible.

Common methods for abnormal behaviour analysis are as follows: (i) for a particular normal behaviour database, abnormal behaviour is any action not in it; or (ii) given an abnormal behaviour database, abnormal behaviour is any action within it.

Most public applications rely on the second method, although both types require a good behaviour description. Various approaches treat anomaly detection in the context of computer vision. Behaviour representation, understanding and anomaly inference are the major issues [1] when it comes to anomaly detection in video.

Adam *et al.* [2] monitored integral-pixel approximations of optical flow instead of tracking objects to detect anomalies. Kim and Grauman [3] recommended a space–time Markov random field (MRF) model to detect abnormal activities in video. Boiman *et al.* [4] considered spatio-temporal patches and suspicious regions those that could not be obtained from previous frames. Mahadevan *et al.* [5] proposed a mixture of dynamic textures (MDT) model [6] to detect temporal and spatial anomalies. Roshtkhari and Levine [7] encoded spatio-temporal composition (STC) of the video volumes with a probability density function. Xu *et al.* [8] proposed a hierarchical framework to detect anomaly comprehensively considering both global and local spatio-temporal contexts with a unified abnormal energy function. Güler *et al.* [9] proposed a real-time approach using scale invariant feature transform (SIFT) features for anomaly detection without using any pre-set rules. Cheng *et al.* [1] provided a hierarchical structure for perceiving local and global anomalies via hierarchical feature representation and Gaussian process regression. Mehran *et al.* [10] introduced the social force model (SFM) to simulate the crowd dynamics of pedestrians. However, the approach in [10] simply used particles to calculate the social force instead of the concept of pedestrians, which leads to the detection results affected by outliers, such as moving arms or legs.

This paper proposes a novel approach to detect crowd anomalies using a block-based SFM (BSFM), which considers the personal desire force and interactive force in the crowd avoiding effects of single pixels. BSFM only handles anomalies related to pixels when a region model does not apply. It is organised as follows: Section 2 deals with the most unusual types of algorithms proposed so far to analyse crowd activities. The BSFM is described in Section 3. Experimental results in Section 4 compare BSFM performance with the classic SFM, MDT method and the Markov random field model with mixture of probabilistic principal component analysers (MRF-MPPCA) approach. Section 5 concludes this paper.

2 Recent anomaly detection methods for crowd scenes

2.1 Gaussian mixture model

A Gaussian mixture model (GMM) has a probabilistic nature, and it considers all the data points as coming from a mixture of a limited number of Gaussian distributions with unknown parameters. Generally, a GMM does not require knowing which subpopulation a data point belongs to, permitting the model to learn the parameters automatically. Three types of values mainly parameterise a GMM: the component weights, the component means and the variances. A GMM usually is used for background subtraction [11], and it is quite suitable to learn the 'normal' data distributions.

2.2 SIFT model for anomaly detection in crowd surveillance

Güler *et al.* [9] proposed a real-time approach to detect global anomalies, such as people running away, based on SIFT [12]. This holistic approach aims to learn normal situations and detect the abnormalities using velocity feature of the crowd estimated with SIFT feature points. Besides, this method does not require the pre-set anomaly rules and the threshold for abnormality recognition can be learned automatically.

2.3 MDTs model

Abnormality detection is usually cast as a problem of detecting outliers, where given some measurement $\boldsymbol{Y}$, then a statistical model $P_{Y}(y)$ is proposed for the distribution of $\boldsymbol{Y}$ under normal conditions. Abnormalities are defined as measurements whose probabilities are below a certain threshold according to this model. Mahadevan *et* *al.* [5] concentrated on the issue of how to devise localised video representations that permit anomaly detection in crowded scenes, thus leading to the identification of three necessary representation properties:

(1) Appearance and dynamics of crowd patterns are jointly modelled.
(2) Ability to distinguish temporal abnormalities.
(3) Capacity to detect spatial abnormalities.

Next, they came up with representations based on dynamic textures (DTs) with joint models for appearance and dynamics, which were very effective in modelling intricate dynamic scenes. Abnormalities are equated to low-probability events concerning a model of normal crowd behaviour. Then DT-based normalcy models were used over both space and time. MDTs helped modelling temporal normalcy while spatial normalcy is measured with a discriminant saliency detector based on MDTs. The MDT treats the observed video sequence of length L as a sample from one of K possible DTs. These models give a general form to some of the methods previously suggested to find abnormalities in either time or space, and can be readily incorporated into a common solution.

2.4 Generalised social force model

Helbing *et al.* [13] proposed an SFM to investigate the pedestrian behaviour in crowds where each of the N pedestrians i has mass m_i, changing his or her actual velocity v_i according to

$$F_{\mathrm{a}} = m_i \frac{\mathrm{d}v_i}{\mathrm{d}t} = F_{\mathrm{p}} + F_{\mathrm{ped}} + F_{\mathrm{w}} \tag{1}$$

as a result of the actual force F_{a}. Besides seeking to arrive at their destinations, pedestrians in crowds also try to keep distance with others and walls. So the actual force F_{a} consists of three parts: (i) personal desire force F_{p}, (ii) interaction force with pedestrians F_{ped} and (iii) interaction force with the environment F_{w}. Interaction force F_{int} is defined as follows:

$$F_{\mathrm{int}} = F_{\mathrm{ped}} + F_{\mathrm{w}}\,. \tag{2}$$

The crowd limits individual movement, so his or her actual velocity v_i will differ from his or her desired velocity v_i^0. Individuals tend to approach his or her desired velocity based on the personal desire force

$$F_{\mathrm{p}} = \frac{1}{\tau}\left(v_i^0 - v_i\right), \tag{3}$$

where τ is a relaxation parameter accounting for time adaption.

Generally speaking, the pedestrians in the crowd need to consider the effects of others' motions, indeed in the SFM the desired velocity v_i^0 can be calculated as

$$v_i^0 = (1 - p_i)v_i^p + p_i\langle v_i^c \rangle, \tag{4}$$

where p_i is a panic weight parameter, v_i^p is the ideal desired velocity for the ith pedestrian without any interference, and $\langle v_i^c \rangle$ is the average velocity of neighbouring pedestrians. The bigger panic weight parameter is, the more susceptible to his or her nearby people the pedestrian is. Overall, the generalised SFM (GSFM) can be summarised as

$$F_{\mathrm{a}} = m_i \frac{\mathrm{d}v_i}{\mathrm{d}t} = \frac{1}{\tau}\left((1 - p_i)v_i^p + p_i\langle v_i^c \rangle - v_i\right) + F_{\mathrm{int}}\,. \tag{5}$$

Hence, the interaction force F_{int} can be simply estimated as

$$F_{\mathrm{int}} = m_i \frac{\mathrm{d}v_i}{\mathrm{d}t} - \frac{1}{\tau}\left((1 - p_i)v_i^p + p_i\langle v_i^c \rangle - v_i\right). \tag{6}$$

Mehran *et al.* [10] used computer vision to detect and localise abnormal crowd behaviour using the GSFM, which describes the behaviour of the crowd as resultant from the interaction of individuals. Therefore, crowd anomaly detection is essentially an eccentric interaction of forces or states in the crowd.

This implementation avoided tracking objects to avert the typical problems arising from tracking high-density crowds like widespread clutter and dynamic occlusions. As an alternative, a holistic approach was suggested to analyse videos of crowds using the particle advection method. In this approach, a grid of particles was placed over the image and moved them along with the underlying flow field. The social force was computed between moving particles to extract interaction forces. In a crowd scene, the temporal change of interaction forces regulates the on-going behaviour of the crowd, which is captured by mapping the interaction forces to image frames. The resulting vector field is the force flow, which is used to model the normal behaviours in a bag of words approach. The regions of anomalies in the abnormal frames are localised using interaction forces.

2.5 Markov random field model with mixture of probabilistic principal component analysers

For detecting unusual activities in a video, the major difficulty is that abnormal behaviours occur unpredictably, and it is hard to distinguish a truly abnormal event from noisy normal scenes. Kim and Grauman [3] introduced a space–time MRF framework to deal with the above two primary challenges. To build an MRF graph, a video was divided into a grid of spatio-temporal local regions where each region corresponds to a single node, and links connect neighbouring nodes. Each node was associated with continual optical flow observations and learned atomic motion patterns via an MPPCAs. Based on the learned patterns, parameters for the MRF were computed. Finally, by carrying out inference on the graph, probabilistic estimates of whether each node is normal or abnormal was obtained. To efficiently adapt the model as new video data streams come in, incremental updates were devised for the MPPCA and associated MRF parameters.

2.6 Hierarchical activity pattern discovery framework with spatio-temporal context

The aforesaid methods seek to emulate activity patterns only bearing in mind the local or the global context. Such methodologies lead to some lack of global knowledge or local relationships when performing the simultaneous perception of local and global unusual motion patterns. Thus, Xu *et al.* [8] detected anomalies broadly considering both global and local spatio-temporal circumstances. A hierarchical system for learning activity patterns is recommended to accomplish this task. At global level, atomic activity patterns were discovered from low-level optical flow arrangements, and the distributions of the atomic activity patterns are modelled for higher-level activity depiction. Then, salient movement patterns were uncovered under the local context. Both layers of discovery both adopt unsupervised learning without any a priori knowledge of the anomaly. Finally, a unified abnormal energy function was designed to detect global and local pattern linked to anomalies.

3 Block-based social force model

The BSFM improves the GSFM and the model devised by Mehan *et al.* in Section 2.3, since it computes the social force based on the blocks already segmented. The BSFM assumes that for each block the ideal desired velocity v_i^p in the current frame is equal the velocity of the block in the last frame. The ideal desired velocity v_i^p can be obtained by calculating the average optical flow of the block.

In addition, the average velocity $\langle v_i^c \rangle$ can be calculated using optical flow. Besides, the block of the last frame is tracked in

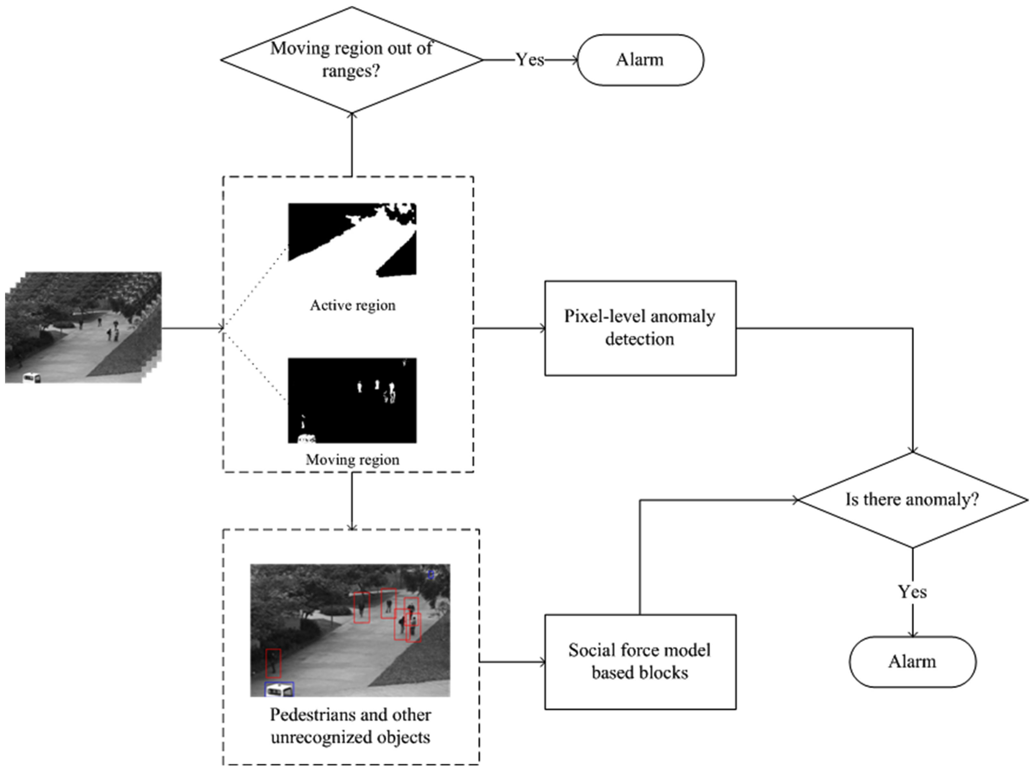

Fig. 1 *Flowchart of the proposed algorithm*

Fig. 2 *Block segmentation: red rectangles mark pedestrians, and blue rectangles mark unrecognised moving objects*

current frame, so the actual velocity v_i is equal the average optical flow of the target block in the current frame.

The tracking method proposed by this paper is based on the idea of 'tracking by detection.' These blocks are tracked by using a template-matching method combined with the SIFT feature detector. First, n candidate blocks are chosen in the following frame using a template matching method and sort them. Then, for each candidate block, we calculate the distance D_i between the target block and the candidate block and the average distance S_i of the matched SIFT points in these two blocks. It should be noted that the candidate block in which there is a matched SIFT point is necessary. If there is no qualified candidate block, the best candidate block is chosen as the target block in the current frame. Finally, the target block should satisfy

$$(1-a)|S_i| \leq |D_i| \leq (1+a)|S_i|, \tag{7}$$

where a is a given constant.

As speed and force are both directional physical variables, the interaction force components $F_{\text{int-}x}$ and $F_{\text{int-}y}$ are calculated, in horizontal direction and vertical directions, respectively, from (6). The interaction force F_{int} is given by

$$F_{\text{int}} = \sqrt{F^2_{\text{int-}x} + F^2_{\text{int-}y}}. \tag{8}$$

Generally speaking, an anomaly occurs in the place where there is the larger interaction force. The interaction force computation aids in detecting and locating anomalies. A block is labelled as an anomaly if the interaction force satisfies

$$F_{\text{int}} > T, \tag{9}$$

where T is a threshold parameter resulting from the training tests.

The BSFM algorithm has four stages, as shown in Fig. 1. First, it gets the previous active area and extracts the moving region. Next, it seeks pixel size using a GMM. Thirdly, it segments the crowds into blocks according to the presence and activity of pedestrians and other unrecognised moving objects. Finally, it analyses the interaction force in the SFM on block-based and pixel-based levels resulting in anomalies, using a simple logical AND as an appraisal measure showing that anomalies have been found.

3.1 Active region and moving region extraction

The active region is frequently a moving area taken from the crowd scenes of the training videos. Therefore, the active region is the focus theme of the scene, and it restricts the pedestrians moving ranges, which leads to a reduction of the number of necessary calculations. A GMM extracts the moving region from the test videos.

3.2 Pixel-level anomaly detection

Pixel-level abnormality detection relies on a method inspired by foreground subtraction. Throughout training phases, a GMM modelled the distribution of image intensities per pixel in completely normal videos. After training, the low probabilities following these GMMs are considered as anomalies, and the parameters of these models are not updated when performing detection.

3.3 Pedestrians and unrecognised moving objects

In this section, we try to segment the crowd into many blocks, and make these blocks interpretable and meaningful (Fig. 2). Referring

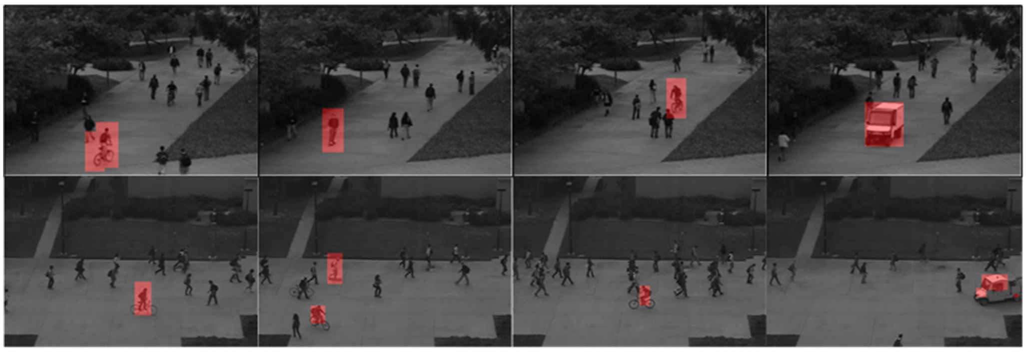

Fig. 3 *Examples of anomaly detection using the UCSD anomaly dataset*

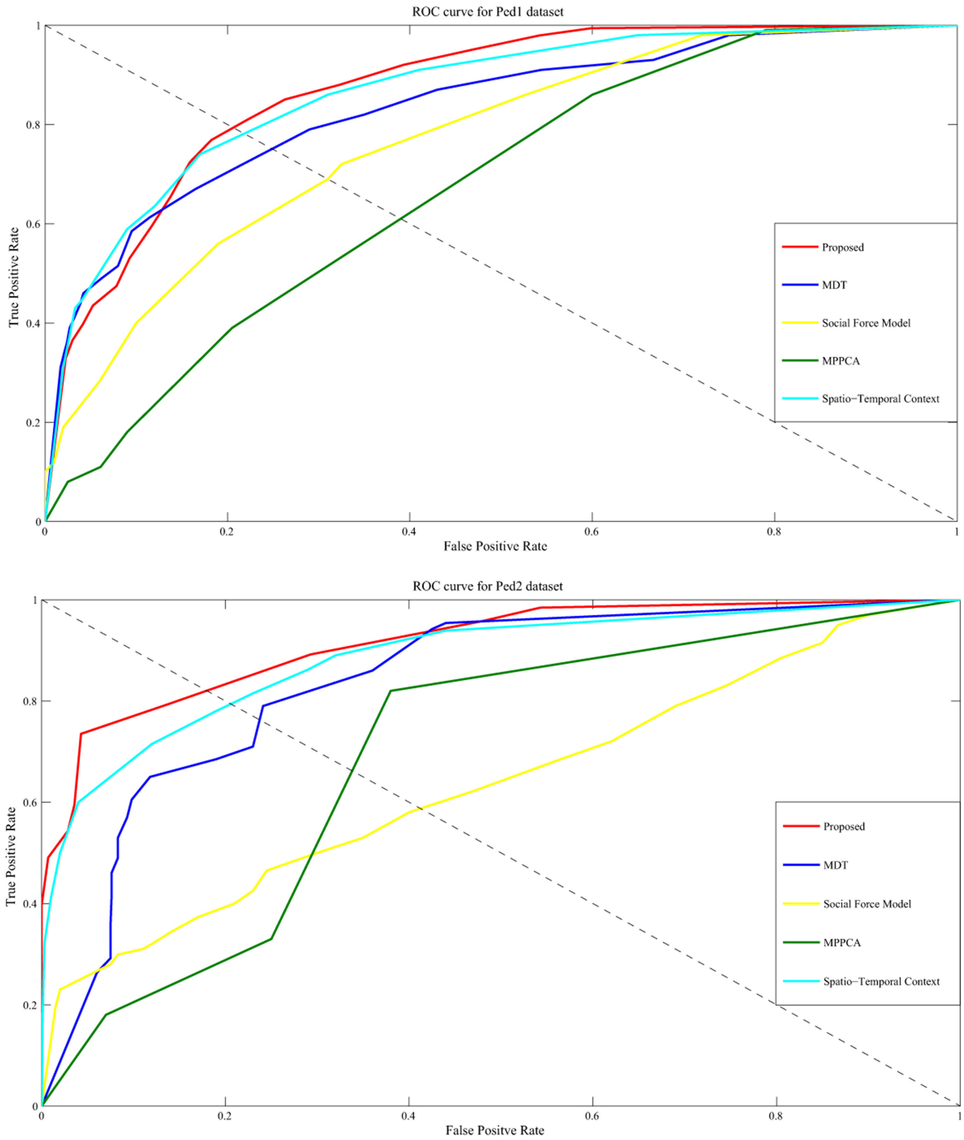

Fig. 4 *ROC curves for the UCSD anomaly dataset*

to anomaly detection in the crowd, we pay more attention to humans and other moving objects instead of the background. So the crowd is divided into two classes: pedestrians and unrecognised moving objects. The pedestrian detection method is based on the concepts of histogram of oriented gradient (HOG) feature [14] and support vector machine (SVM) classifier. The basic idea of the HOG feature extractor is to describe the local appearance of an object by its local intensity gradient distribution. Since the pedestrians have been detected in the video, the moving objects that do not belong to any pedestrians are regarded as unrecognised moving objects. Thus, the pedestrians and moving objects in the video are both well handled. Unrecognised objects can be marked using the method of flood fill.

4 Experimental results

We tested our approach on the University of California at San Diego (UCSD) anomaly dataset, which contains video clips of two scenes (Ped1 and Ped2) captured by stationary cameras with low resolution. The UCSD dataset provides training sets and testing sets for both scenes with anomalies such as people cycling, skateboarding, walking over the grass and vehicles breaking in. Some anomalies detected by the proposed approach are shown in

Table 1 Quantitative comparison of the EER using the UCSD dataset

Algorithm	Ped1, %	Ped2, %	Average, %
proposed	21	17	19
MDT	25	25	25
social force model	31	42	37
MPPCA	40	30	35
spatio-temporal context	38	42	40

Table 2 Quantitative comparison of the AUC using the UCSD dataset

Algorithm	Ped1, %	Ped2, %	Average, %
proposed	87.0	89.6	88.3
MDT	81.8	82.9	82.4
social force model	67.5	55.6	61.6
MPPCA	59.0	69.3	64.2
spatio-temporal context	85.4	88.2	86.8

Fig. 5 *Example of a false positive: the woman was crossing the walkway*

Fig. 3. The result indicates that the proposed approach can detect the common anomalies in the dataset.

We implemented the proposed method using C++ and OpenCV libraries for evaluation purposes, and the component number and the threshold possibility for the GMM in the implementation are set to 10 and 5%, respectively.

To evaluate the performance of the BSFM, we compared it to other methods including MDT-based [5], SFM-based [10], MPPCA-based [3] and STC-based [8] methods. Quantitative evaluation and comparison with other methods are presented in terms of precision-recall and ROC curves, obtained by varying the threshold parameter T in (9). The ROC curves for anomaly detection are shown in Fig. 4, and the equal error rate (EER) values are in Table 1. There are four possibilities from the prediction. If the output from the prediction is positive and the actual value is positive, then it is called true positive (TP); however, if the actual value is negative then it is called false positive (FP). Conversely, a true negative (TN) has occurred when both the prediction and actual value are negative and false negative (FN) is when the prediction outcome is negative while the actual value is positive. The true positive rate (TPR) and false positive rate (FPR) can be calculated as

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{10}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{TN} + \mathrm{FP}}. \tag{11}$$

The EER is the location on an ROC curve where the false accept rate and false reject rate (1 − TPR) are equal. In Fig. 4 and Table 1, we observed that our approach outperforms the others in the EER. We also calculate the area under the curve (AUC) as shown in Table 2. The average AUC of the proposed method is the highest one. Overall, the results show that the proposed approach is superior to the approaches for comparison in general.

Since the force is calculated in the horizontal direction and vertical direction based on the direction of speed, the proposed method is more sensitive to unusual motion patterns. On the other hand, in some situations, it may lead to false positive, such as people jaywalking, which is not regarded as an anomaly in this paper, as shown in Fig. 5. Our method may fail to distinguish normal behaviour in the situation where the definition of an anomaly is ambiguous.

5 Conclusion

This paper presents a method for anomaly detection using social force model based on blocks. The proposed approach detects anomaly both in pixel and block levels. In pixel-level anomaly detection, a GMM is used to detect pixel-level anomalies. In block level, the crowd is segmented into blocks according to pedestrian detection, then anomalies are detected and localised using a block-based social force model. Experimental results show that the proposed method outperforms state-of-the-art methods.

6 Acknowledgments

The project was supported by the Nature Science Foundation of Guangdong Province (grant no. 2015A030310172), the Science & Technology Plan Projects of Shenzhen (grant nos. JCYJ20150324140036830 and GJHZ20160226202520268).

7 References

[1] Cheng, K.W., Chen, Y.T., Fang, W.H.: 'Video anomaly detection and localization using hierarchical feature representation and Gaussian process regression'. Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition, CVPR 2015, Boston, USA, 7–12 June 2015, pp. 2909–2917

[2] Adam, A., Rivlin, E., Shimshoni, I., *et al.*: 'Robust real-time unusual event detection using multiple fixed-location monitors', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2008, **30**, (3), pp. 555–560

[3] Kim, J., Grauman, K.: 'Observe locally, infer globally: a space-time MRF for detecting abnormal activities with incremental updates'. Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition, CVPR 2009, Miami, USA, 22–24 June 2009, pp. 2921–2928

[4] Boiman, O., Irani, M.: 'Detecting irregularities in images and in video', *Int. J. Comput. Vis.*, 2007, **74**, (1), pp. 17–31

[5] Mahadevan, V., Li, W., Bhalodia, V., *et al.*: 'Anomaly detection in crowded scenes'. Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition, CVPR 2010, San Francisco, USA, 13–18 June 2010, pp. 1975–1981

[6] Chan, A.B., Vasconcelos, N.: 'Modeling, clustering, and segmenting video with mixtures of dynamic textures', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2008, **30**, (5), pp. 909–926

[7] Roshtkhari, M.J., Levine, M.D.: 'Online dominant and anomalous behavior detection in videos'. Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition, CVPR 2013, Porland, USA, 23–28 June 2013, pp. 2611–2618

[8] Xu, D., Wu, X.Y., Song, D.Z., *et al.*: 'Hierarchical activity discovery within spatio-temporal context for video anomaly detection'. Proc. IEEE Int. Conf. Image Processing, ICIP 2013, Bengaluru, India, 9–11 August 2013, pp. 3597–3601

[9] Güler, P., Temizel, A., Temizel, T.T.: 'Real-time global anomaly detection for crowd video surveillance using SIFT'. Proc. 5th Int. Conf. Imaging for Crime Detection and Prevention (ICDP 2013), London, UK, 16–17 December 2013, pp. 1–12

[10] Mehran, R., Oyama, A., Shah, M.: 'Abnormal crowd behavior detection using social force model'. Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition, CVPR 2009, Miami, USA, 22–24 June 2009, pp. 935–942

[11] Friedman, N., Russell, S.: 'Image segmentation in video sequences: a probabilistic approach'. Proc. 13th Conf. Uncertainty in Artificial Intelligence, Providence, USA, 1–3 August 1997, pp. 175–181

[12] Lowe, D.G.: 'Distinctive image features from scale-invariant keypoints', *Int. J. Comput. Vis.*, 2004, **60**, (2), pp. 91–110

[13] Helbing, D., Farkas, I., Vicsek, T.: 'Simulating dynamical features of escape panic', *Nature*, 2000, **407**, (6803), pp. 487–490

[14] Dalal, N., Triggs, B.: 'Histograms of oriented gradients for human detection'. Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition, CVPR 2005, San Diego, USA, 20–25 June 2005, vol. 1: pp. 886–893