

# Anomaly detection in crowded scenes using motion energy model

Tianyu Chen<sup>1</sup>  · Chunping Hou<sup>1</sup> · Zhipeng Wang<sup>1</sup> · Hua Chen<sup>2</sup>

Received: 27 October 2016 / Revised: 3 July 2017 / Accepted: 6 July 2017 /  
Published online: 14 July 2017  
© Springer Science+Business Media, LLC 2017

**Abstract** We present a new method for detection of abnormal behaviors in crowded scenes. Based on statistics of low-level feature—optical flow, which describes human movement efficiently, the motion energy model is proposed to represent the local motion pattern in the crowd. The model stresses the difference between normal and abnormal behaviors by considering sum of square differences (SSD) metric of motion information in the center block and its neighboring blocks. Meanwhile, data increasing rate is introduced to filter outliers to achieve boundary values between abnormal and normal motion patterns. In this model, an abnormal behavior is detected if the occurrence probability of anomaly is higher than a preset threshold, namely the motion energy value of its corresponding block is higher than that of the normal one. We evaluate the proposed method on two public available datasets, showing competitive performance with respect to state-of-the-art approaches not only in detection accuracy, but also in computational efficiency.

**Keywords** Anomaly detection · Video surveillance · Motion energy · Optical flow

## 1 Introduction

Recently, with the increasing of video surveillance system, more and more researchers in computer vision focus on digging useful information from the monitoring video data for public security. Problems such as representing human behavior in crowded scenes [12],

---

✉ Tianyu Chen  
chentianyu.7@126.com

<sup>1</sup> School of Electronics and Information Engineering, Tianjin University, Tianjin, 300072 China

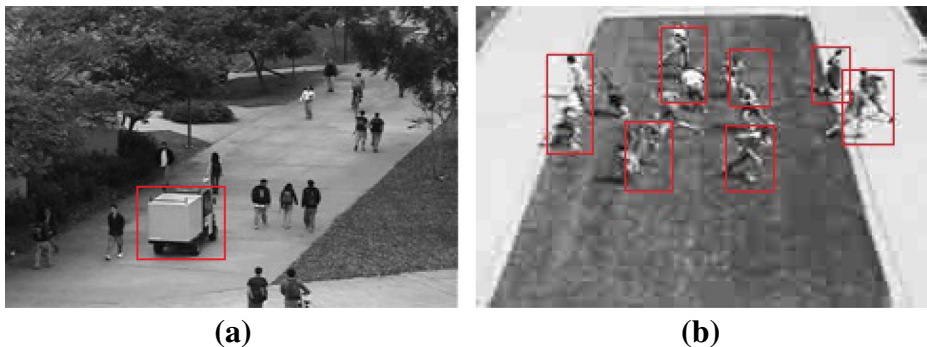
<sup>2</sup> Faculty of Information Science and Engineering, Ningbo University, Ningbo, China

recognizing different human actions [4], estimating associations of different events [11, 13] have all been subjects of research. Among them, abnormal behavior detection has got a great interest due to the importance of instant alert for anomaly in video surveillance system. Generally, it is difficult to give a standard definition of abnormal behavior due to the low occurrence frequency of it in our daily life. However, normal behaviors can be observed all the time. Therefore, the goal of anomaly detection is to distinguish uncertain abnormal behaviors from certain normal ones accurately in real-time.

Generally, based on the locality of anomaly, crowd behaviors can be classified into two categories which are local anomaly and global anomaly. Local anomaly means that an abnormal behavior exists in a local region of a frame while its surroundings are normal ones. Meanwhile, global anomaly happened in a whole frame consisting of several local anomalies [5]. Figure 1 illustrates local anomaly and global anomaly: (a) Local anomaly: A fast car appears on the sidewalk while people are walking, (b) Global anomaly: People escape suddenly in the whole surveillance scenes.

The specific task of anomaly detection always changes with various surveillance scenes since various abnormal behaviors may occur. Therefore, several anomaly detection approaches have been proposed for different surveillance tasks. Generally, existing methods in the literature can be placed into three categories: 1) analysis by statistics of relations of events or behaviors, where probability graphical model is applied to shape relation networks of different events; 2) analysis by object tracking, where trajectories of interested objects are mainly extracted; 3) analysis by extraction of statistics based low-level features such as motion and appearance information or deep learning based visual features, where various behavior patterns are modeled.

In the first category, different statistics models such as SARM [6], DDP-HMM [10] are used to summarize associations among behaviors. For example, Junseok et al. [11, 13] modeled an input video by transforming it into a graph and edit the initial graph iteratively to get reliable associations. In the second category, almost all the traditional tracking methods follow trajectories of interested objects after modeling them by gathering features such as speed and direction [2, 17, 18] while some novel methods search for spatial temporal paths for video event detection. For example, Du Tran et al. utilized max-path search to handle complex behaviors composed by moving objects [23], N. T. Nguyen et al. built HMM model to detect abnormal routes from normal ones [16]. In the third category, motion and motion-related features are extracted to model behavior patterns. For example, social force model



**Fig. 1** Examples of local anomaly and global anomaly

was proposed by Mehran et al., where crowd interaction forces are computed [7]. Dong Gyu Lee et al. built a motion influence matrix to represent crowd behaviors by using optical flow [12]. Adam et al. utilized low-level statistics collected by multiple fixed-location monitors and detected anomaly in real time [1]. However, due to limitations of motion information which do not describe appearance of objects, some special behaviors may not be detected. Therefore, novel methods using both motion and appearance features are proposed recently. For example, spatial-temporal multi-scalar visual feature was proposed for hierarchical analysis to distinguish dominant and abnormal behaviors by Roshtkhari et al. [19] and Vikas et al. analyzed crowded scenes by combining speed, size and texture of objects together [18]. Mixtures of dynamic textures [3] was proposed by Antoni et al. and it is applied in detection of dynamics of abnormal behaviors by Weixin et al. [14, 15]. Recently, some deep learning based methods are proposed by many researchers. Deep neural networks was applied to learn feature representation automatically in anomaly detection by Dan Xu et al. [27] and spatial-temporal CNN model was used to extract both motion and appearance features by Shifu Zhou et al. [29]. Jingkuan Song et al. propose a method integrating LSTM with salient-aware deep 3-D CNN features considering important subjects in video shots [26] and an encoder-decoder framework based on hierarchical LSTMs considering both visual features and language context information [22].

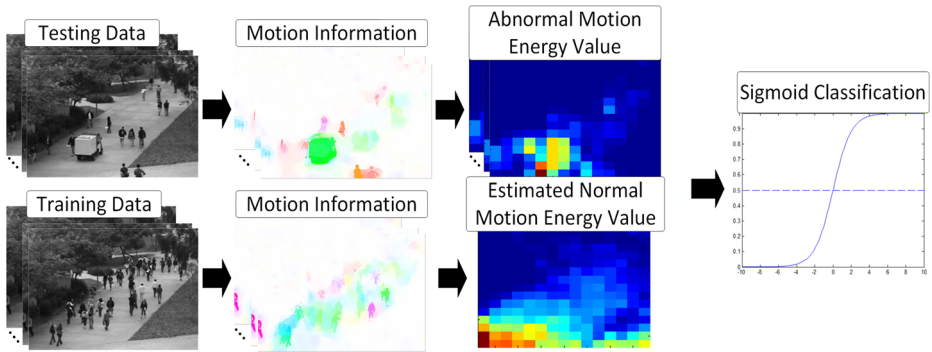
All these approaches have quite good performance on some specific scenes while their shortcomings are also obvious. Methods based on probability graphic model need huge video data to get reliable statistics. Trajectory based approaches are not suitable for crowded scenes due to occlusion and overlapping of objects. Methods combining motion and appearance information cost much time on feature extraction and feature representation, which fail to satisfy the real time requirement.

In this paper, a new anomaly detection method with low computational cost is proposed. It is suitable for crowded scenes and flexible for the requirement of the size of dataset. Unlike most methods, we do not eliminate background information before detection by foreground segmentation. Because the background information is as essential as the foreground information for the proposed crowd behavior representation—the motion energy model. Like graph models such as [11, 13, 21], the motion energy model we proposed uses neighborhood structure to propagate information which considers the motion energy of interested object and the difference between it and its neighbors synchronously. Each input frame is split into non-overlapping blocks and the motion energy value is calculated for each block which are labelled as either normal or abnormal by a binary classifier.

The paper is organized as follows: In Section 2, the proposed method is elaborated. Experiment procedure and performance comparison with other relative algorithms are given in Section 3. Section 4 concludes the paper and gives possible improvements of the method.

## 2 Proposed method

In this section, we propose a new method to represent motion characteristics of objects in crowded scenes to detect abnormal behaviors. The whole framework is illustrated in Fig. 2. Given training data and test data, the pixel-level and block-level motion information are extracted for each frame and then, the motion energy model is built and the block-level motion energy value is computed based on such motion information. The difference of motion energy values between training data and test data are input into sigmoid based classifier to generate labels of 0 or 1.



**Fig. 2** The overview of the proposed method for anomaly detection in crowded scenes

### 2.1 Motion information extraction

Most human behaviors are based on movements especially those in crowded scenes and most anomaly can be observed by abnormal movements. Therefore, the motion information is the most important characteristic in human behavior recognition. Due to the optical flow is an effective descriptor to describe motion information between two consecutive frames in a video, the crowd motion information is extracted for each frame by employing iterative coarse-to-fine Lucas-Kanade optical flow algorithm. Then, each extracted motion field is split into  $P$  by  $Q$  non-overlapping blocks of  $M \times N$  pixels. For each block, motion vectors are defined as follows:

$$B_{i,j} = \{f_{x,y}(h, v) | x = 1, 2, \dots, M, y = 1, 2, \dots, N\}, \quad i = 1, 2, \dots, P, j = 1, 2, \dots, Q \quad (1)$$

where  $f_{x,y}(h, v)$  represents the motion vector located at  $x$  and  $y$  contained in  $B_{i,j}$  which denotes the block located at  $i$  and  $j$ .  $x, y$  and  $i, j$  are index of pixel of a frame and index of block of a frame, respectively. Each motion vector includes a horizontal component  $h$  and a vertical component  $v$ . It contains both velocity information and orientation information. Due to the anomaly based on velocity is mainly considered in this work, the magnitude  $mag_{x,y}$  of each motion vector is calculated by using:

$$mag_{x,y} = \|f_{x,y}(h, v)\|_2 \quad (2)$$

In crowded scenes, the specific body movement like waving and nodding cannot be represented correctly by motion vectors, since the resolution of most surveillance videos is quite low. But the main motion trend of the crowd can be described by motion vectors reliably. Thus, we use the block-level dominant motion vector which is the vector of largest number within a block to describe the motion information of such whole block. The magnitude of dominant motion vector is extracted after quantizing each pixel-level vector into  $l$  levels. Experimentally, the proposed method performs best when  $l$  is equal to 16 by using (3):

$$mag_{x,y} = \begin{cases} 1 \cdot \max mag_{x,y}/32 & 0 \cdot \max mag_{x,y}/16 \leq mag_{x,y} < 1 \cdot \max mag_{x,y}/16 \\ 3 \cdot \max mag_{x,y}/32 & 1 \cdot \max mag_{x,y}/16 \leq mag_{x,y} < 2 \cdot \max mag_{x,y}/16 \\ \dots & \dots \\ 29 \cdot \max mag_{x,y}/32 & 14 \cdot \max mag_{x,y}/16 \leq mag_{x,y} < 15 \cdot \max mag_{x,y}/16 \\ 31 \cdot \max mag_{x,y}/32 & 15 \cdot \max mag_{x,y}/16 \leq mag_{x,y} < 16 \cdot \max mag_{x,y}/16 \end{cases} \quad (3)$$

Following above quantization rules, all vectors are quantized into 16 levels and the block-level motion information is derived from such quantized motion vectors within each block. The proposed motion energy model which is described below considers both block-level and pixel-level motion information to represent behavior of objects in crowded scenes reliably.

## 2.2 Motion energy model

In a surveillance scene, the motion based behavior pattern of an object is not only determined by itself, but also by adjacent neighbors around it, especially for objects in crowded scenes, which has been studied in [7]. For local anomaly detection, abnormal behaviors occur at a specific location in a video while their surroundings must be normal ones, and global anomaly is composed by several local anomalies. Therefore, we proposed a motion energy model to describe such characteristics for both local anomaly and global anomaly detection. The model includes a data term and a smooth term:

$$\begin{aligned} DataTerm(i, j) &= Dom \left( mag_{x,y}^{i,j} \right) \\ SmoothTerm(i, j) &= \frac{1}{8} \cdot \sum_{B_{m,n} \in S_{i,j}} \omega_{m,n} SSD(B_{i,j}, B_{m,n}) \end{aligned} \quad (4)$$

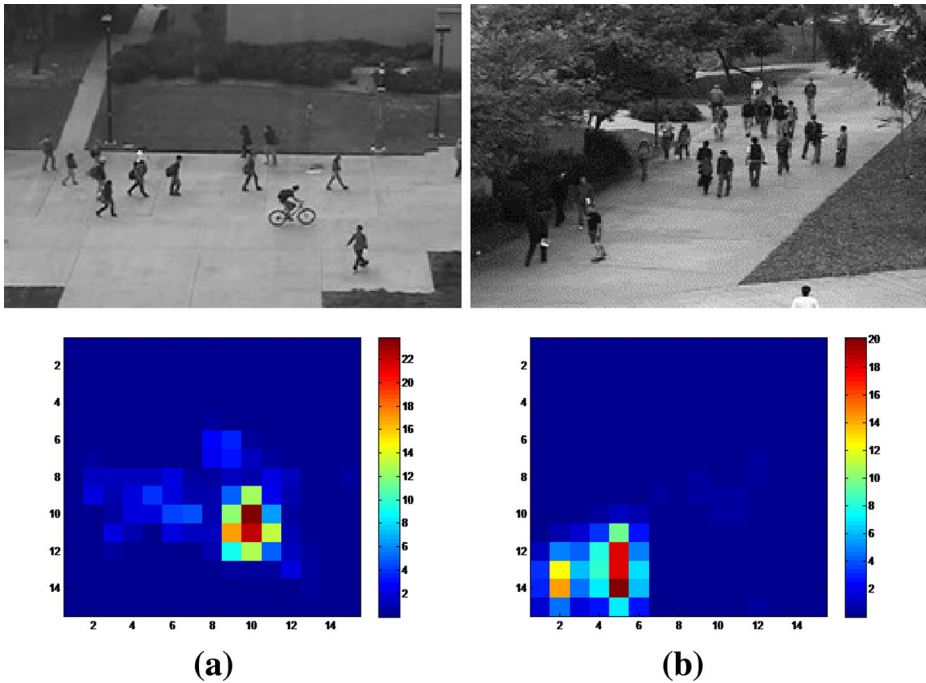
where the magnitude of dominant motion vector in block  $B_{i,j}$  is denoted as  $Dom(mag_{x,y}^{i,j})$  which is the motion vector of the largest number within a block, namely data term ( $i, j$  and  $x, y$  denote block index and pixel index respectively). In the smooth term,  $B_{m,n}$  is the block belongs to  $S_{i,j}$  which is the eight neighborhood of center block  $B_{i,j}$ . The difference of motion pattern between center block and its neighbors is measured by  $SSD$  metric shown in (5), where  $N_f$  is the number of motion vectors within block  $B_{m,n}$ . Due to closer blocks are more associated with the center block, a distance-based weight function denoted as  $\omega_{m,n}$  is employed to weigh different neighbor blocks, which is a  $3 \times 3$  Gaussian mask.

$$SSD(B_{i,j}, B_{m,n}) = \frac{1}{N_f} \sum_{1 \leq x \leq M, 1 \leq y \leq N} \left( mag_{x,y}^{i,j} - mag_{x,y}^{m,n} \right)^2, \quad m \neq i, n \neq j \quad (5)$$

Based on the rules defined above, we use (6) to represent the motion energy of block located at  $i$  and  $j$  by adding data term and smooth term together, where  $\alpha$  and  $\beta$  are parameters determined experimentally for various scenes. In this work,  $\alpha$  and  $\beta$  are both set to be 1.

$$E(i, j) = \alpha \cdot DataTerm(i, j) + \beta \cdot SmoothTerm(i, j) \quad (6)$$

Roughly, each non-overlapping block can be regarded as a small group of objects since large crowd density in crowded scenes. The data term considers the motion energy of target block itself while the smooth term considers the weighed difference of motion pattern between the target block and its associated neighbors. Two terms contribute to the motion energy value in different aspects. When local anomaly occurs, the motion energy values of corresponding blocks are higher than those of the normal ones. Figure 3 shows an example of local anomaly with calculated motion energy value. We observe that blocks around the location where anomaly occurred are in bright color (high motion energy value) compared with other normal blocks.

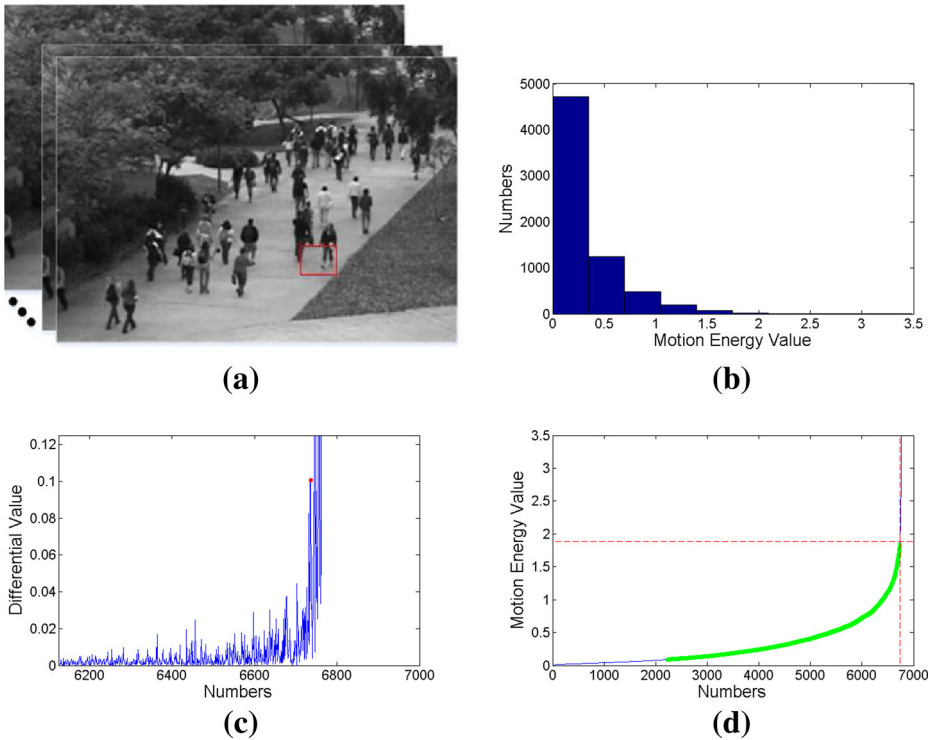


**Fig. 3** **a** and **b** are two examples of motion energy value of local anomaly: The first row are original frames and the second row are their corresponding motion energy values

### 2.3 Boundary extraction

The motion energy value has a robust characteristic that each value is derived from multiple motion vectors around it considering both block-level and pixel-level motion information. The block with higher value means it is much more different from its neighbors, which namely higher occurrence probability of anomaly in this block. Thus, how high the boundary value is must be determined to distinguish normal and abnormal behaviors. In this work, the boundary is estimated by computing motion energy value for each block of each frame in the training dataset which only contains normal behaviors. Figure 4b and d give the distribution of motion energy values of one block from low to high [28] after computing for all frames in the training dataset. Generally, most motion values are close to 0 because there is no movement in a surveillance scene in most of the time and a few values are extremely high due to aperture problem or noise in optical flow computation [1]. The values in the middle are normal ones illustrated as green in Fig. 4d.

To determine an appropriate boundary value for each block, the abnormal values which are extremely high must be eliminated. Therefore, the data increasing rate is employed in this work to filter outliers. We calculate the normalized differential values of ranked energy values for each block, which is given as Fig. 4c. If a differential value (data increasing rate) is higher than a preset threshold, which means the energy values increase beyond limitation of normal ones, the previous corresponding energy value is determined as the boundary value between abnormal behaviors and normal ones. Experimentally, the threshold of data



**Fig. 4** **b** an example of histogram of motion energy values of one block which is illustrated as a red box in **(a)**. **c** differential values of motion energy values, in which the value larger than data increasing rate(0.1) is illustrated as a red point. **d** ranked motion energy values of the red block in **(a)** of all frames and the green region represents the normal values

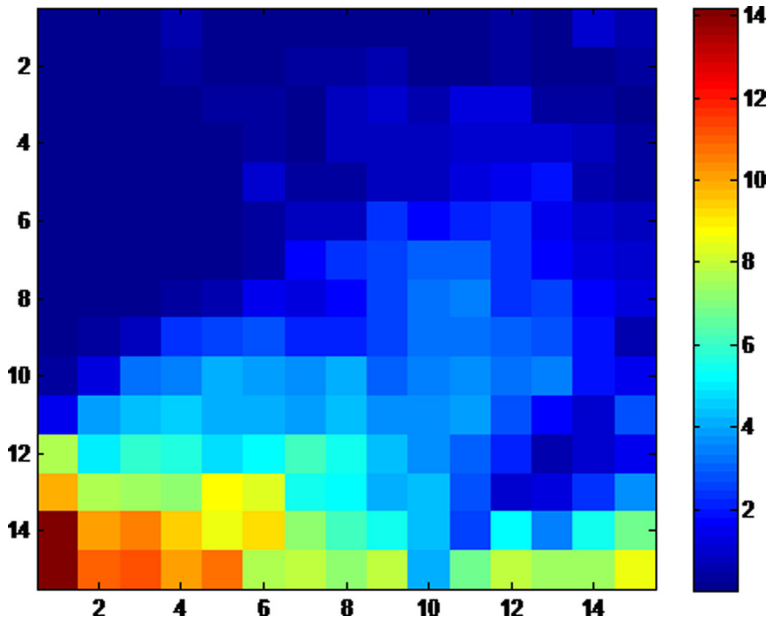
increasing rate is set to be 0.1. In Fig. 4d, the intersection of two red lines is the boundary we extract.

### 2.4 Block classification

In consideration of the camera angle, velocity of the same object will be detected differently in different locations, namely energy values of normal behaviors are different in different blocks. Generally, objects which move closer to the camera have higher energy values. Thus, after boundary extraction for each block of frames in the training dataset, different boundary values of normal behaviors in different blocks are calculated which is given by Fig. 5. The brighter block corresponds to a high motion energy value.

Similarly, the motion energy value of each block of frames in the test dataset is computed and all the blocks are classified into normal or abnormal according to corresponding energy values. In this work, each block is sequentially checked by inputting the difference between the tested motion energy value and the boundary value of normal behaviors of the same block into a sigmoid based classifier. It can map the difference to [0,1]:

$$prob = \begin{cases} \left[ 1 + \exp(-1/(Ei, j - E'(i, j))) \right]^{-1} & -1 < E(i, j) - E'(i, j) < 0 \\ \left[ 1 + \exp(-((Ei, j - E'(i, j)))) \right]^{-1} & otherwise \end{cases} \quad (7)$$



**Fig. 5** Motion energy value of normal behaviors. Brighter block at the bottom represents higher motion energy values in blocks closer to the camera

where  $E'_{i,j}$  and  $E_{i,j}$  are boundary value and tested energy value belong to the block located at  $i$  and  $j$  respectively. The more the tested energy value of a block exceeds the corresponding boundary value, the higher probability such block will get to be detected as anomaly. In our experiments, the block is classified as an abnormal one when the *prob* value of this block is higher than a preset threshold which is determined by different scenes. The preset threshold for the *prob* value is set empirically. In this work, the threshold is 0.8 for each block.

### 3 Experiments and results

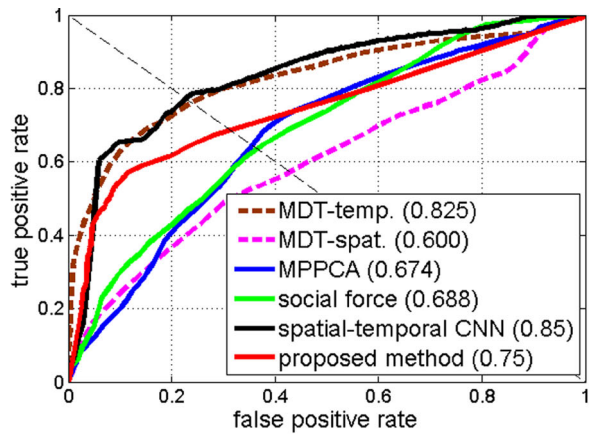
To evaluate the effectiveness of the proposed method, we performed experiments on the UCSD dataset [24] and the UMN dataset [25] which are available on the website. We use the receiver operating characteristic (ROC) curve by calculating the true positive rate (TPR) and false positive rate (FPR) at different detection thresholds to measure the performance. The Equal Error Rate (EER) for frame-level criterion and Detection Rate (DR) for pixel-level criterion of the proposed method were compared with those of other methods [8, 9, 15, 29]. In this work, the ROC curves are generated by setting several discrete threshold value from 0 to 1 with interval of 0.02.

#### 3.1 UCSD dataset

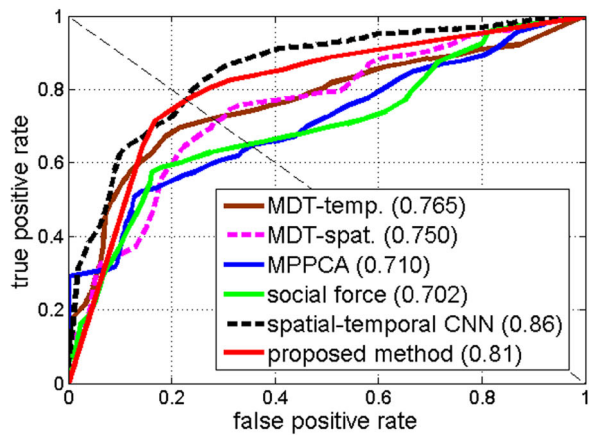
The UCSD dataset is released online by Statistical Visual Computing Lab at UCSD. It is a dataset for local anomaly detection task. Two subsets are included, named Ped1 and Ped2 which contain multiple surveillance clips of two scenes of crowd in UCSD campus. Ped1 has 34 clips for training and 36 clips for testing, while Ped2 has 16 clips for training and 12



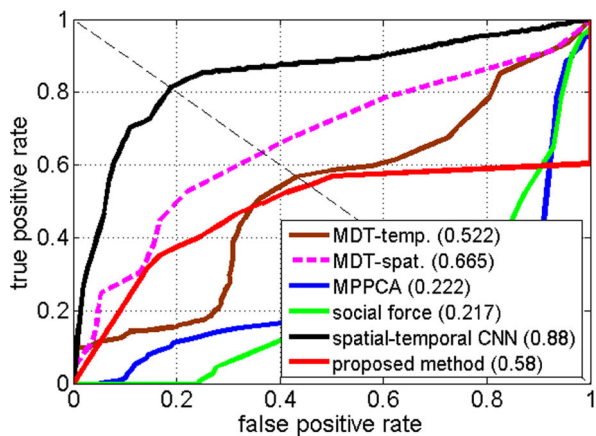
**Fig. 6** **a** and **b** The performance of proposed method for frame-level anomaly detection on Ped1 and Ped2 dataset respectively. **c** The performance of proposed method for pixel-level anomaly detection on Ped2 dataset. Numbers in brackets represent AUC values



**(a)**



**(b)**



**(c)**

**Table 1** EER for frame-level anomaly detection on ped1 and PED2

Method	Ped1(%)	Ped2(%)	Average(%)
Social force [7]	36.5	42	39.3
MPPCA [9]	39.6	31.1	35.4
MDT-spatial [15]	43.8	28.7	36.3
MDT-temporal [15]	<b>22.9</b>	27.9	25.4
Spatial-temporal CNN [29]	24.0	24.4	<b>24.2</b>
Proposed method	31.0	<b>22.0</b>	26.5

Bold numbers indicate the performance of the best method

clips for testing. The frame size of each subset is  $238 \times 158$  pixels and  $360 \times 240$  pixels respectively. The training dataset only contains pedestrians with normal behaviors and the testing dataset contains some abnormal behaviors such as riding, skating and driving.

The UCSD dataset provides two kinds of ground-truth which are frame-level ground truth for frame-level anomaly detection and pixel-level ground truth for anomaly localization within frames. For frame-level anomaly detection, if a frame contains at least one abnormal pixel, it is considered a detection [15]. For pixel-level anomaly localization, the frame is considered detected correctly if at least 40% of the truly abnormal pixels are detected [14].

For effectiveness comparison, we not only choose several other classic methods which are social force model [7] and the mixture of optical flow (denoted as MPPCA) [9], but also some state-of-the-art methods which are the spatial MDT and temporal MDT [15] and the spatial-temporal CNN. The ROC curves of frame-level anomaly detection on Ped1 and Ped2 dataset are shown in Fig. 6.

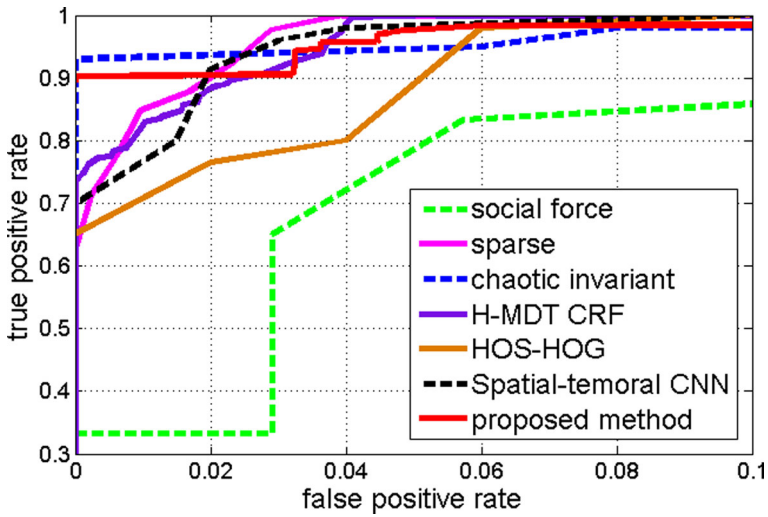
It can be seen that the proposed method is comparable with the recent brilliant method, spatial-temporal CNN, and outperforms other methods on Ped2 dataset. For Ped1 dataset, our method is at intermediate level and just performs a little worse than MDT-temporal method and spatial-temporal CNN which are the best two methods. For quantitative comparison, we generate EER of each method shown in Table 1. The average EER of the proposed method is in the leading level among competing methods.

For the pixel-level anomaly detection, DR is employed to evaluate performance of the proposed method. If the probability of anomaly of a block is higher than a preset threshold, it is labeled as 1 and localized. The ROC curves of pixel-level anomaly detection on Ped2 are shown in Fig. 6c and the DR as well as the computation efficiency are summarized in Table 2. As can be seen that the performance of proposed method is similar to temporal MDT but it is not as well as spatial MDT and spatial-temporal CNN. That is because the suitable feature for the anomaly caused by velocity, the joint motion information, is the only extracted feature in our method in consideration of the requirement of real-time detection. However, just because such acceptable sacrifice of detection accuracy, the computational efficiency of our method had been greatly improved. The method reduces the computational

**Table 2** DR and computational efficiency for pixel-level anomaly detection on PED2

Method	DR(%)	Time(Seconds per Frame)
Social force [7]	27.6	–
MPPCA [9]	22.4	–
MDT-spatial [15]	63.4	0.69
MDT-temporal [15]	56.8	0.64
Spatial-temporal CNN [29]	<b>81.9</b>	0.37
Proposed method	55	<b>0.08</b>

Bold numbers indicate the performance of the best method

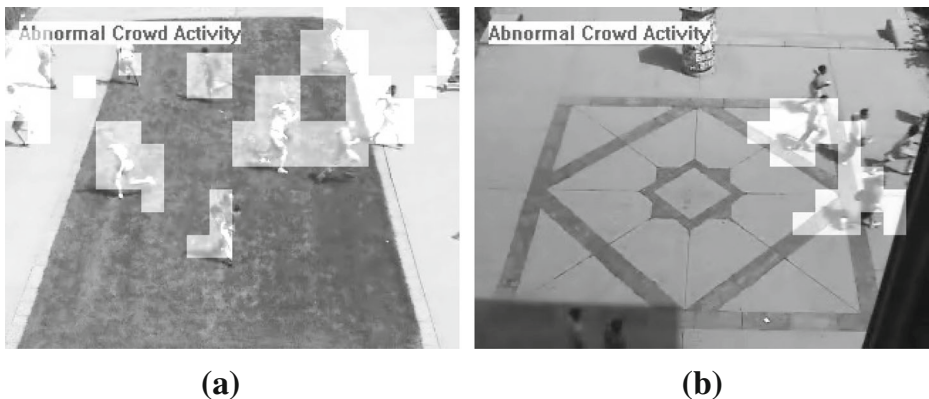


**Fig. 7** ROC curves of frame level criterion in the UMN dataset

**Table 3** AUC/EER performance for frame-level anomaly detection in UMN dataset

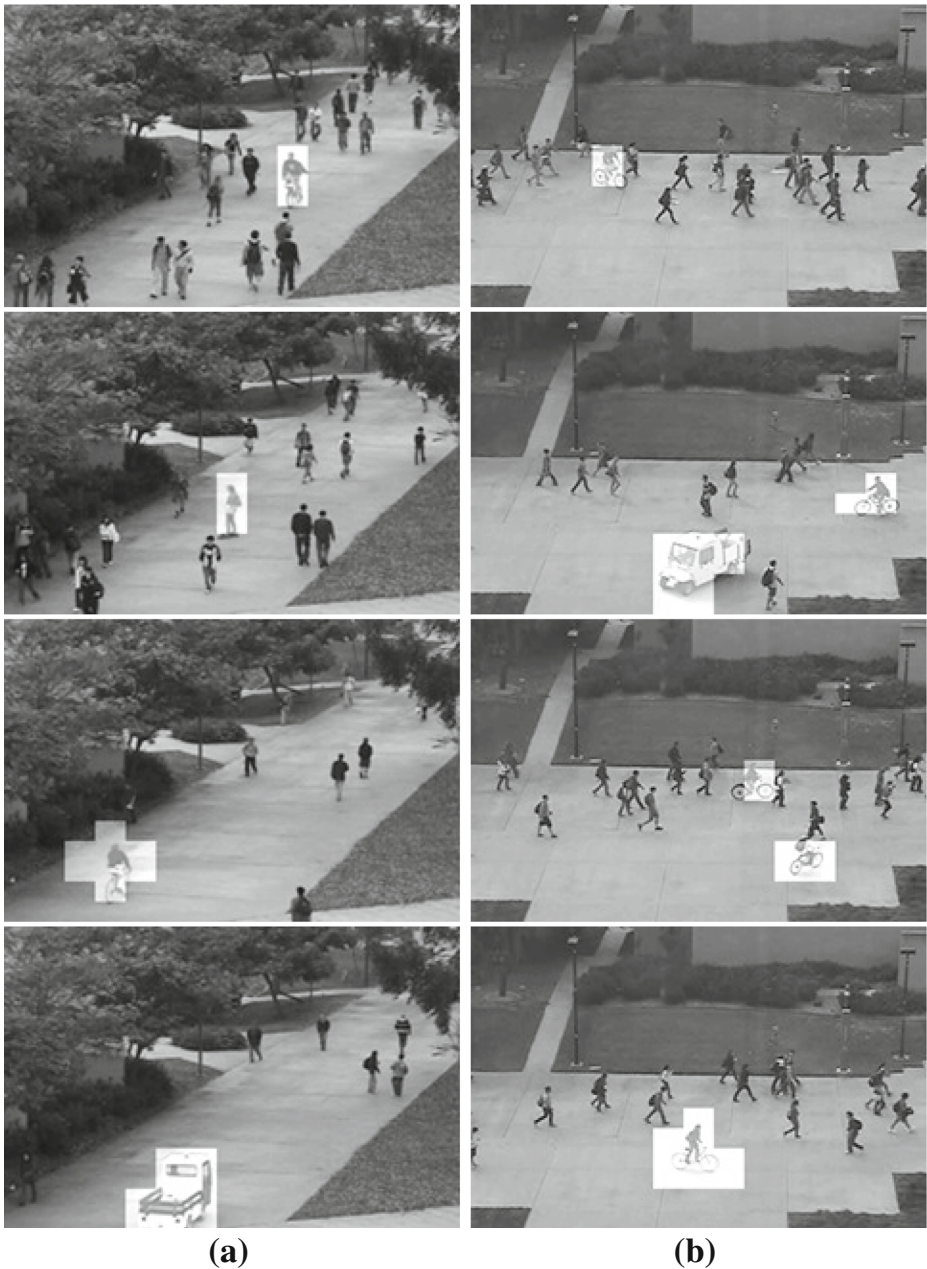
Method	AUC(%)	EER(%)
Social force [7]	94.9	12.6
Sparse [5]	<b>99.6</b>	<b>2.8</b>
HOS-HOG [8]	98.1	5.6
Chaotic invariants [20]	99.4	5.3
H-MDT CRF [14]	99.5	3.7
Spatial-temporal CNN [29]	99.6	3.3
Proposed method	98.9	4.1

Bold numbers indicate the performance of the best method



**Fig. 8** Examples of detection results in UMN dataset

complexity and improve the detection speed notably. Compared with MDT method, the proposed method computes almost ten times faster than them which are far from real-time detection with similar detection accuracy. In Fig. 9, the visualized results on Ped1 and Ped2 are shown.



**Fig. 9** a and b The visualized results of local anomaly detection on Ped1 and Ped2 respectively. The abnormal behaviors including bicycling, skating and driving are labeled in white

### 3.2 UMN dataset

The UMN dataset is a popular dataset of global anomaly detection. It contains three different escape scenes (named as lawn, indoor, and plaza, respectively) with spatial resolution of  $240 \times 320$ , in which people escape suddenly in different directions from normal walking. We extracted 4890 normal frames totally in three scenes for training and 2069 frames with normal and abnormal behaviors for testing. As the UMN dataset only provides global ground-truth, we follow the standard Area Under Curve (AUC) criterion, which is the area within the ROC curve and the axis. The comparison with other state-of-the-art methods and some classic methods is shown in Fig. 7 and Table 3. Our method shows comparable performance in detecting escaping pedestrians with high speed against others. The examples of detection results are shown in Figs. 8 and 9.

## 4 Conclusion

In this paper, we propose a motion energy model to represent the motion based behavior pattern in crowded scenes for anomaly detection. Our method has two advantages: 1) taking center object and its neighbors both into consideration; 2) no need of foreground segmentation or other human detection method. The framework is just simple to generate block-level local anomaly probabilities which are classified by a single threshold. Unlike other low-level characteristics based methods, the proposed method not only performs well relatively in detection accuracy, but also much more efficiently in computation. This is of great importance for video surveillance system due to requirement of real time monitoring. The future efforts will focus on auto-adaptive block size in order to apply our method to various scenes.

**Acknowledgements** This work was supported by the National Natural Science Foundation of China under Grant 61471262, the International (Regional) Cooperation and Exchange under Grant 61520106002, and the Doctoral Fund of Ministry of Education of China under Grant 20130032110010.

## References

1. Adam A, Rivlin E, Shimshoni I, Reinitz D (2007) Robust real-time unusual event detection using multiple fixed-location monitors. *IEEE Trans Pattern Anal Mach Intell* 30(3):555–560
2. Basharat A, Gritai A, Shah M (2008) Learning object motion patterns for anomaly detection and improved object detection, pp 1–8
3. Chan AB, Vasconcelos N (2005) Mixtures of dynamic textures. In: Tenth IEEE International Conference on Computer Vision, pp 641–647
4. Chaudhry R, Ravichandran A, Hager G, Vidal R (2009) Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions. 94(10):1932–1939
5. Cong Y, Yuan J, Liu J (2011) Sparse reconstruction cost for abnormal event detection. 32(14):3449–3456
6. Hamalainen W, Nykanen M (2008) Efficient discovery of statistically significant association rules, pp 203–212
7. Helbing D, Molnr P (1995) Social force model for pedestrian dynamics. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Top* 51(5):4282–4286
8. Kaltsa V, Briassouli A, Kompatsiaris I, Hadjileontiadias LJ, Strintzis M (2015) Swarm intelligence for detecting interesting events in crowded environments. *IEEE Trans Image Process A Publ IEEE Signal Process Soc* 24(7):2153
9. Kim J, Grauman K (2009) Observe locally, infer globally: A space-time mrf for detecting abnormal activities with incremental updates. In: *IEEE Conference on computer vision and pattern recognition*, pp 2921–2928

10. Kuettel D, Breitenstein MD, Van Gool L, Ferrari V (2010) What's going on? Discovering spatio-temporal dependencies in dynamic scenes. *10(9)*:1951–1958
11. Kwon J, Lee KM (2012) A unified framework for event summarization and rare event detection from multiple views. *37(9)*:1266–1273
12. Lee DG, Suk HI, Lee SW (2013) Crowd behavior representation using motion influence matrix for anomaly detection. In: *Iaprr Asian conference on pattern recognition*, pp 110–114
13. Lee Kyoungmu, Kwon Junseok (2015) A unified framework for event summarization and rare event detection. In: *IEEE Conference on computer vision and pattern recognition*, pp 1737–50
14. Li W, Mahadevan V, Vasconcelos N (2013) Anomaly detection and localization in crowded scenes. *IEEE Trans Pattern Anal Mach Intell* *36(1)*:18–32
15. Mahadevan Vijay, Li Weixin, Bhalodia Viral, Vasconcelos Nuno (2010) Anomaly detection in crowded scenes. In: *IEEE Conference on computer vision and pattern recognition*, pp 1975–1981
16. Nguyen NT, Phung DQ, Venkatesh S, Bui H (2005) Learning and detecting activities from movement trajectories using the hierarchical hidden markov model. In: *IEEE Computer society conference on computer vision and pattern recognition*, vol 2, pp 955–960
17. Piciarelli Claudio, Micheloni Christian, Foresti GianLuca (2008) Trajectory-based anomalous event detection. *IEEE Trans Circ Syst Video Technol* *18(11)*:1544–1554
18. Reddy V, Sanderson C, Lovell BC (2011) Improved anomaly detection in crowded scenes via cell-based analysis of foreground speed, size and texture, pp 55–61
19. Roshtkhari M, Levine M (2013) Online dominant and anomalous behavior detection in videos. In: *Computer Vision and Pattern Recognition*, pp 2611–2618
20. Shandong W, Moore BE, Shah M (2010) Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes. In: *Computer Vision and Pattern Recognition*, pp 2054–2060
21. Song J, Gao L, Nie F, Shen H, Yan Y, Sebe N (2016) Optimized graph learning with partial tags and multiple features for image and video annotation. *IEEE Trans Image Process A Publ IEEE Signal Process Soc* *25(11)*:4999–5011
22. Song J, Guo Z, Gao L, Liu W, Zhang D, Shen HT (2017) Hierarchical lstm with adjusted temporal attention for video captioning
23. Tran D, Yuan J, Forsyth D (2014) Video event detection: From subvolume localization to spatiotemporal path search. *IEEE Trans Pattern Anal Mach Intell* *36(2)*:404–416
24. Ucsd anomaly dataset. <http://www.svcl.ucsd.edu/projects/anomaly>
25. Umn anomaly dataset. <http://mha.cs.umn.edu/Movies/Crowd-Activity-All.avi>
26. Wang X, Gao L, Song J, Shen H (2017) Beyond frame-level cnn: Saliency-aware 3-d cnn with lstm for video action recognition. *IEEE Signal Process Lett* *24(4)*:510–514
27. Xu D, Ricci E, Yan Y, Song J, Sebe N (2015) Learning deep representations of appearance and motion for anomalous event detection
28. Zhang Y, Huchuan L, Zhang L, Xiang R (2016) Combining motion and appearance cues for anomaly detection. *Pattern Recogn* *51(C)*:443–452
29. Zhou S, Shen W, Zeng D, Fang M, Wei Y, Zhang Z (2016) Spatial-temporal convolutional neural networks for anomaly detection and localization in crowded scenes. *Signal Process Image Commun* *47*:358–368



**Tianyu Chen** is currently pursuing the M.D. degree in Electronic Information Engineering Institute, Tianjin University, Tianjin, China. His research interests including video analysis and event detection.



**Chunping Hou** received the M.Eng. and Ph.D. degrees, both in electronic engineering, from Tianjin University, Tianjin, China, in 1986 and 1998, respectively. Since 1986, she has been with the faculty of the School of Electronic Information Engineering, Tianjin University, where she is currently a Full Professor and the Director of the Broadband Wireless Communications and 3D Imaging Institute. Her current research interests include wireless communication, 3D image processing, and the design and applications of communication systems.



**Zhipeng Wang** is currently pursuing the M.D. degree in Electronic Information Engineering Institute, Tianjin University, Tianjin, China. His research interests including video analysis and action detection.



**Hua Chen** received the M.Eng. degree and Ph.D. degree in Information and Communication Engineering from Tianjin University, Tianjin, China, in 2013 and 2017, respectively. From Aug. 2015 to Aug. 2016, he was a one-year visiting scholar at the Department of ECE, Concordia University. He is now with the Faculty of Information Science and Engineering, Ningbo University, Ningbo, Zhejiang Province, 315211, China. His research interests include array signal processing, MIMO radar and passive radar.