CrossMark

Local abnormal behavior detection based on optical flow and spatio-temporal gradient

Songhao Zhu[1] · **Juanjuan Hu**[1] · **Zhe Shi**[1]

Abstract To improve the accuracy of the detection of local abnormal behavior, a novel method is here proposed. The main idea of the proposed method is described as follows: firstly, a video sequence is divided into spatio-temporal blobs; then, a statistical method based on the semi-parametric model is adopted to detect these blobs where abnormal behaviors most likely to appear; finally, maximum optical flow energy and local nearest descriptor are utilized to determinate whether these suspicious blobs really contain abnormal behaviors. The experimental results conducted on several benchmarks ademonstrate the effectiveness of the proposed method.

Keywords Abnormal behavior · Semi-parametric model · Optical flow energy · Local nearest descriptor

1 Introduction

Abnormal behavior detection is one of the key issues in the field of intelligent video surveillance, where the main tasks contain the following two aspects: one is to detect various abnormal behaviors automatically from surveillance videos, and the other is to remind security persons to deal with these unusual events timely.

Nowadays, various approaches of abnormal behavior detection can be divided into the following two categories: one category is object tracking based method [3, 5, 8, 9, 12, 17, 22, 24, 25] and the other category is group features based method [2, 5, 7, 10, 14–16, 23, 26]. The first method tracks each moving object individually, and then the obtained motion information of each moving object is utilized to complete the detection of abnormal behavior. However, the occlusion between objects in complex scenes seriously affects the performance of the first method. The second method directly extracts the motion information from the whole video sequence. In [5, 14], the information of both normal behaviors and abnormal behaviors are utilized to train a support vector machine to complete the anomaly detection. In [2], only the information of normal

✉ Songhao Zhu
  zhush@njupt.edu.cn

[1] School of Automatic, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

examples is utilized to detect abnormal examples. In [10], the information of normal examples is utilized to construct the Probabilistic Principal component analysis model, and those behaviors which could not be represented by the proposed model are considered as abnormal examples. In [7, 15], the Latent Dirichlet Allocation method is utilized to detect abnormal behaviors. To detect anomaly behaviors in video sequences, existing methods usually scan every video pixel or every video region carefully, which increases the computation cost and reduces the detection accuracy.

To improve the detection accuracy of local abnormal behaviors, a novel method is here proposed. Firstly, bag of words [21] is adopted to describe the optical flow information of each blob, and the semi-parametric based model [4] is adopted to detect suspicious local abnormal blobs; then, suspicious abnormal blobs are divided into rectangular cells equally, and the maximum optical flow energy method [27] is adopted to detect undoubted abnormal cells; finally, the local nearest descriptor and Mixed Naïve Bayes model [19] are adopted to determinate anomaly behaviors.

The rest of the paper is organized as follows. Semi-parametric model based suspicious abnormal blob detection is described in Section 2. The detection of anomaly behavior is described in Section 3. In Section 4, we first introduce the experimental datasets and evaluation methods, and then report the experimental result. Section 5 concludes the paper.

2 Abnormal blob detection

Each video image frame is first divided into blobs, and the process of considering a blob as a suspicious abnormal blob is described as follows.

First, x is utilized to describe a group of features of a blob S; then, the blob S can be considered as a suspicious abnormal blob when the following inequality is true:

$$\lambda(S) = \frac{\Pr\left(x \middle| H_1(S)\right)}{\Pr\left(x \middle| H_0(S)\right)} > \rho \tag{1}$$

where $\Pr(x|H_0(S))$ is the likelihood probability of x when S is a normal blob, and $\Pr(x|H_1(S))$ is the likelihood probability of x when S is an abnormal blob.

An appropriate likelihood probability model of $\Pr(x|H_i)$ needs to be built to compute the likelihood ratio $\lambda(S)$. Unlike existing methods computing $\lambda(S)$ by a probability density model with specific parameters, a semi-parametric probability density model is here adopted to compute $\lambda(S)$.

2.1 Semi-parametric probability density model

The modeling process of the semi-parametric probability density function is described as below.

- First, x_1 and x_2 are feature set inside and outside the blob S, and $f(x)$ and $g(x)$ are their corresponding probability density functions:

$$\begin{cases} x_1 = \left(x_{11}, x_{12}, \ldots\ldots, x_{1n_1}\right)^T \widetilde{} f(x) \\ x_2 = \left(x_{21}, x_{22}, \ldots\ldots, x_{2n_2}\right)^T \widetilde{} g(x) \\ \dfrac{f(x)}{g(x)} = \exp\left(\alpha + \beta^T h(x)\right) \end{cases} \tag{2}$$

where n_1 and n_2 are the size of blob inside and outside S, and $h(x)$ is a pre-defined function.

- Then, the likelihood probability of semi-parametric for S can be formulated as:

$$L(\alpha, \beta, G) = \prod_{i=1}^{n} p_i \prod_{j=1}^{n_1} \exp\left(\alpha + \beta^T h(x_{1j})\right) \tag{3}$$

where $n=n_1+n_2$, $p_i=dG(t_i)=\Pr(X=t_i)$, $t=(t_1,t_2, \ldots,t_n)^T=(x_{11}, x_{12},\ldots, x_{1n1}, x_{21},\ldots, x_{2n2})^T$, and $G(x)$ is the cumulative distribution function of $g(x)$.

- Finally, the likelihood probability of $\lambda(S)$ in Eq. (1) can be computed with $f(x)=g(x)$ or $f(x)\neq g(x)$ respectively.

- When $\alpha=0$ and $\beta=0$, $f(x)=g(x)$, which means that the feature distribution inside and outside S are same and $\lambda(S)$ is 1.
- When $\alpha\neq0$ and/or $\beta\neq0$, L in Eq. (3) can be maximized with the constraints of $\sum p_i=1$ and $\sum p_i[\omega(t_i)-1]=0$. The detail of the parameter estimation is described as follows. α and β are first utilized to represent p_i by Lagrangian relaxation; then, p_i is substituted into the log-likelihood $l=\ln L(\alpha, \beta, G)$; next, the maximum value of L can be obtained under $p_i= n_2^{-1}[1+\rho\omega(t_i)]^{-1}$ $(\rho=n_1/n_2)$ and $\partial l/\partial \alpha=0$, and Eq. (4) with only α and β can be obtained by ignoring constants; finally, α and β can be estimated by using the method of Newton maximization.

$$l(\alpha, \beta) = -\sum_{i=1}^{n} \log\left[1 + \rho exp\left(\alpha + \beta^T h(t_i)\right)\right] + \sum_{j=1}^{n_1}\left[\alpha + \beta^T h(t_i)\right] \tag{4}$$

- Set $\alpha^{\wedge}$ and $\beta^{\wedge}$ be the maximum likelihood estimation of α and β, then $l\left(\alpha^{\wedge}, \beta^{\wedge}\right)$ is the log-likelihood of $f(x)\neq g(x)$:

$$\begin{aligned}\lambda(S) &\equiv -2\left[l(0,0)-l\left(\hat{\alpha},\hat{\beta}\right)\right]-l(\alpha, \beta) \\ &= -2\sum_{i=1}^{n} \log\left[1 + \rho exp\left(\alpha + \hat{\beta}^T h(t_i)\right)\right] + 2\sum_{j=1}^{n_1}\left[\alpha + \hat{\beta}^T h(t_i)\right] + 2n\log(1 + \rho)\end{aligned} \tag{5}$$

The method of semi-parametric probability density model has the following two merits. Firstly, the feature set t around each blob is here utilized to estimate α, β and their corresponding distribution simultaneously, since it is difficult to obtain the distribution model of α and β separately when the blob is too small; secondly, there are no specific parametric probabilistic models needed to be assigned, since different forms of $h(x)$ can be selected for different feature distributions.

2.2 Feature description

The process of feature exaction for abnormal behavior detection is described as follows.

- First, each image is divided into blobs equally, and the optical flow in each blob are extracted respectively.
- Next, all optical flow features are clustered into several categories and every category is represented as a visual word.
- Finally, a histogram of visual words is utilized to describe the behavior feature of the whole video sequence, which can compress image information and keep local features of interesting images simultaneously.

The histogram of visual word simplifies the computation of log-likelihood $l(\alpha, \beta)$. Set $b=(b_1, b_2, \ldots, b_K)^T$ be the visual word histogram in an image frame, where $b_k(k=1, 2, \ldots, K)$ is the number of blobs belonging to the kth visual word. Set $b^w=(b_1{}^w, b_2{}^w, \ldots, b_k{}^w)^T$ be the histogram within the whole video sequence, then Eq. (4) can be rewritten as below:

$$l(\alpha, \beta) = -\sum_{k=1}^{K} b_k^{\omega} \log\big[1 + \rho exp\big(\alpha + \beta^T h(k)\big)\big] + \sum_{k=1}^{K} b_k\big[\alpha + \beta^T h(k)\big] \qquad (6)$$

Similarly, Eq. (6) can be simplified as below:

$$\lambda(S) = -2\sum_{k=1}^{K} b_k^{\omega} \log\left[1 + \rho exp\left(\hat{\alpha} + \hat{\beta}^T h(k)\right)\right] + 2\sum_{k=1}^{K} b_k\left[\hat{\alpha} + \hat{\beta}^T h(k)\right] + K\log(1 + \rho) \qquad (7)$$

2.3 Abnormal blob detection

Based on the definition in Eq. (1), it can be seen that the value of $\lambda(S)$ of abnormal behavior regions is larger than that of normal behavior regions. Therefore, if the value of $\lambda(S)$ is larger than the threshold ρ, the blob S can be considered a suspicious abnormal blob.

However, some normal blobs with interactive motion or complex motion may also have large $\lambda(S)$. Therefore, suspicious abnormal behavior blobs need to be further confirmed to reduce false alarms.

3 Abnormal behavior detection

3.1 Basic idea

Compared with images blobs of normal behavior, images blobs of abnormal behavior often have obvious characteristics of higher motion velocity and more disordered motion direction. Therefore, the optical flow energy of an abnormal behavior blob is larger than that of a normal behavior blob. Based on the discussed above, the basic idea of local abnormal behavior detection in this paper is described as follows: the mixed naive Bayes model is first utilized to train nearest neighbor descriptor of normal cells, and then the trained nearest neighbor

descriptor is utilized to determinate whether each test cell is an abnormal cell or not. The detail of abnormal behavior detection is shown in Algorithm 1.

Algorithm 1: Abnormal behavior detection process

• Condition: known abnormal blobs
• Dividing each abnormal blob into cells with same size, and computing the optical flow energy in each cell.
• Searching abnormal cells with the largest optical flow energy according to the mixed naive Bayes model. If one cell is abnormal, then execute the next step; otherwise, the current abnormal blob is considered as a false abnormal blob, and the search process terminates.
• Continuing to search abnormal cells within the 4-neighbors of the current abnormal cell according to the mixed naive Bayes model.
• Marking all abnormal cells with red

3.2 Nearest neighbor descriptor

The construction process of the nearest neighbor descriptor is described as below.

• Firstly, each image is divided into cells with the same size of $h \times w$.
• Then, the spatio-temporal gradient magnitude of each pixel (i, j) v_{ij}, the variance M_2, the skewness M_3, and the kurtosis M_4 of each cell are computed respectively:

$$\begin{cases} M_r = [m_{i,j}], & i = 1, 2, \ldots, h, \ \ j = 1, 2, \ldots, w \\ m_{i,j} = \dfrac{1}{h*w} \sum\nolimits_{i,j} (\nu_{ij})^r, r = 2, 3, 4 \end{cases} \tag{8}$$

• Next, a vector $M \in \mathrm{R}^{3 \times h \times w}$ is constructed as below:

$$M = [m_2 \ \ m_3 \ \ m_4] \tag{9}$$

• Finally, the distance between cell S and S' is computed based on the same formula as in the reference [28]:

$$d(V_s, V_{s'}) = \sum_v 2^{-2v} WAV_v(|M_s| - |M_{s'}|) \tag{10}$$

and K spatio-temporal nearest neighbors of one cell in video sequence are formulated as:

$$X^{sd} = [d_1, d_2, \ldots, d_k, \ldots, d_K]^T \tag{11}$$

where d_k is the distance between one cell and its kth nearest neighbor.

3.3 Anomaly detection using naive Bayes model

As shown in Fig. 1, the basic idea of the graphical model for mixed naive Bayes model is described as below.

- Select a mixed-membership vector $\pi \sim$ Dirichlet(ς).
- For each feature x_j of X:

- Select a component $z_j = c \sim$ dicrete(π).
- Select a feature value $x_j \sim p_{\Psi_j}(x_j|\theta_{jc})$, where Ψ_j and θ_{jc} jointly determine whether the feature j and the component c can construct an exponential distribution function.

    The naive Bayes model can be obtained by training video sequences with normal behaviors, and the detail training process is described as follows.

- Given the model parameter φ and a set of Gaussian distributions $\Omega = (\mu_{jc}, \sigma_{jc}, [j]_1^{\,d}, [c]_1^{\,k})$, the probability density function of X within the mixed naive Bayes model can be formulated as below:

$$p(X|\varphi, \Omega) = \int_\pi p(\pi|\alpha) \times \left(\prod_{\substack{j\,=\,1 \\ \exists x_j}}^{d} \sum_{c=1}^{k} p(z_j = c|\pi) \frac{1}{\sqrt{2\pi\sigma_{jc}^2}} \exp\left(-\frac{\left(x_j - \mu_{jc}\right)^2}{2\sigma_{jc}^2} \right) \right) d\pi \quad (12)$$

where μ_{jc} and σ^2_{jc} are the mean and variance of the cth component within the jth Gaussian distribution.

- Given training sets $X = [X_1, X_2, \ldots, X_L]$, the optimal parameters of φ^* and Ω^* can be obtained by maximizing the likelihood of the whole dataset $p(X|\varphi, \Omega)$:

$$\left(\varphi^*, \Omega^*\right) = \underset{(\varphi, \Omega)}{\operatorname{argmax}}\, p(X|\varphi, \Omega) \quad (13)$$

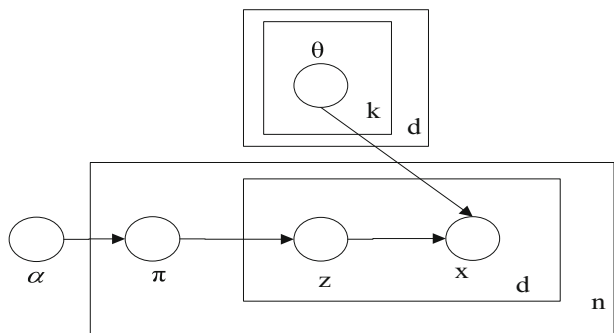**Fig. 1** Graphical model for the mixed naive Bayes model

Fig. 2 Examples from the Ped1 dataset. **a** An example of normal walking behavior. **b** An example of abnormal behavior of biking. **c** An example of abnormal behavior of driving

- In the phase of learning optimal parameters φ^* and Ω^*, a variational expectation maximization algorithm discussed in reference [18] is here utilized to achieve the expression of the mixed naive Bayes model quickly.

During the test phase, the learned mixed naive Bayes model is utilized to compute the log-likelihood $l = \log p(X|\alpha, \Omega)$ of nearest neighbor descriptors, and the behavior in one cell cab be considered as anomaly if the following inequality is true:

$$|l| < T \qquad (14)$$

where T is an appropriate threshold.

4 Experimental results

Experiments are conducted on two benchmarks to demonstrate the capabilities of the proposed method for abnormal activity swtection. In the following, the benchmarks are first described, and then the evaluation criterias are introduced. Finally, experimental results and extensive evaluation are presented.
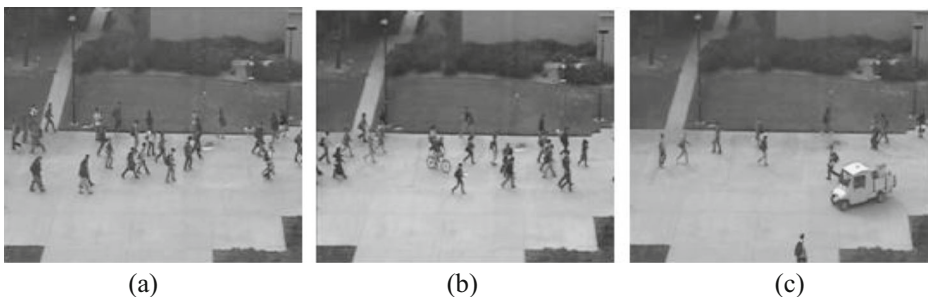

Fig. 3 Examples from the Ped1 dataset. **a** An example of normal walking behavior. **b** An example of abnormal behavior of biking. **c** An example of abnormal behavior of driving

Fig. 4 Examples from the subway dataset. **a** An example of normal behavior from the entrance video sequence. **b**–**c** Examples of abnormal behavior where a person is exiting through the entrance gate. **d** An example of abnormal behavior where two persons are exiting through the entrance gate. **e** An example of normal behavior from the exit video sequence. **f**–**h** Examples of abnormal behavior where a person is entering through the exit gate

4.1 Dataset

UCSD The UCSD dataset includes ped1 and ped2 subsets for detecting local abnormal behavior[6, 13]. The ped1 video subset consists of 20 training video sequences, 14 validation video sequences and 36 testing video sequences, where each video sequence includes 200 frames in total. The ped2 subset consists of 10 training video sequences, 6 validation video sequences and 12 testing video sequences, where the frame number of each video sequence is 120, 150 or 180. The normal behaviors on the UCSD datasets are defined as walking with normal speed. The local abnormal behaviors mainly include irregular moving, such as skating, biking, and driving. Figures 2 and 3 show several frames of normal and abnormal behaviors from the ped1 dataset and ped2 dataset respectively.
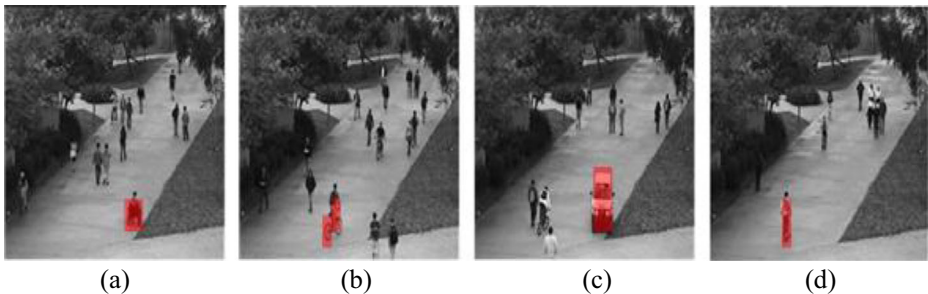

Fig. 5 Examples of anomaly behavior detection from the Ped1 dataset. **a** A pedestrian drives a wheelchair. **b** A pedestrian rides a bicycle. **c** A pedestrian drives a car. **d** A pedestrian rides a skateboard
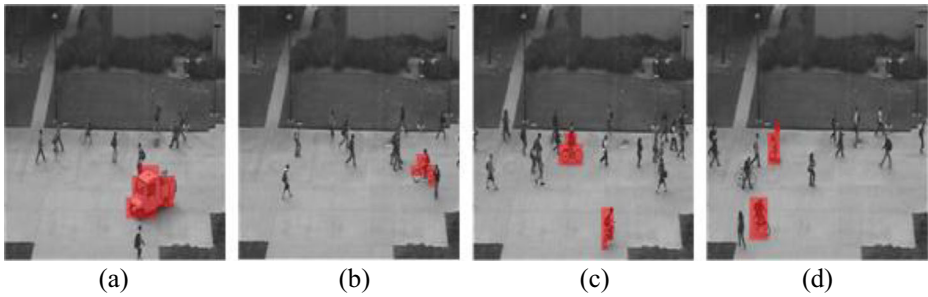
Fig. 6 Examples of anomaly behavior detection from the Ped2 dataset. **a** A pedestrian drives a car. **b** A pedestrian drives a bicycle. **c** Two pedestrians drive bicycles. **d** A pedestrian drives a bicycle and A pedestrian rides a skateboard

SubWay The subway dataset contains two video sequences recorded by a camera at the entrance gate and a camera at the exit gate respectively [1, 20]. The first video sequence, the entrance gate video sequence is 96 min long and contains normal behaviors including going down through the turnstiles and entering the platform. There are also 66 abnormal behaviors, including walking in the wrong direction, irregular interactions between people, sudden stopping, running fast. The second one, the exit gate surveillance video, is 43 min long and contains 19 anomalous events, mainly walking in the wrong direction and loitering near the exit gate. Neither the surveillance videos nor groups of frames within them are labeled as training or testing videos. Figure 4 shows several examples from the subwy dataset.
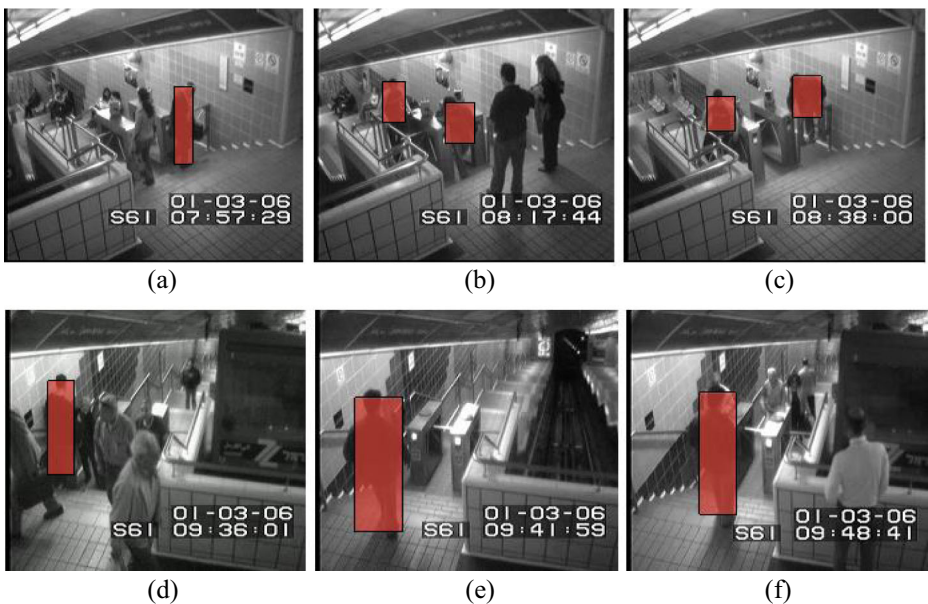

Fig. 7 Examples of abnormal behavior detection from the subway dataset. **a–b** Examples of abnormal behavior where a person is exiting through the entrance gate. **c** Examples of abnormal behavior where two persons are exiting through the entrance gate. **d–f** Examples of abnormal behavior where a person is entering through the exit gate
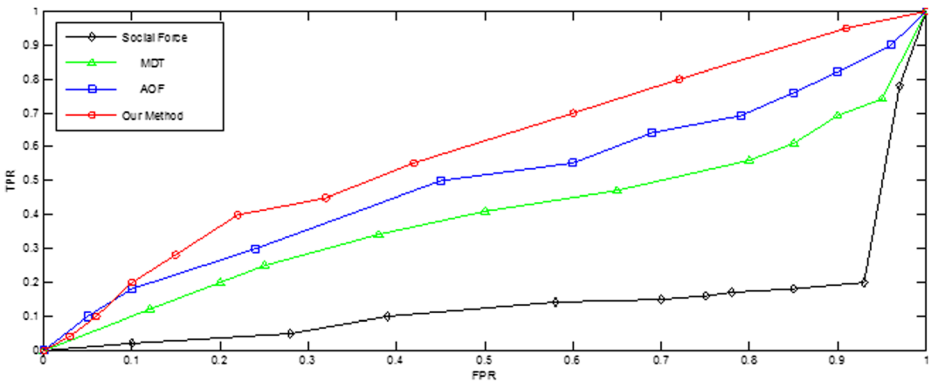
Fig. 8 Pixel-level receiver operating characteristic of the Ped1 dataset

4.2 Evaluation criteria

To better evaluate the performance of the proposed approach, a pixel-level based criterion is here adopted to compare the local abnormal behavior results with the ground-truth anomaly behaviors. Specifically, if at least 40 % of anomalous pixels within a detected anomaly image frame are real anomaly pixels, then the corresponding image frame is considered as an abnormal frame.

The pixel-level based receiver operating characteristic (shorted as ROC) curve is here adopted as the performance evaluation criterion, which is the integrated index of true positive rate (shorted as TPR) and false positive rate (shorted as FPR). Specifically, the detection accuracy is high when the area under the ROC curve is large. Furthermore, the more the ROC curve closes to the upper left, the larger the value of TPR and FPR are.

4.3 Prameter setting

There are several parameters needed to be set: K, υ, h, w, ρ and T. The algorithm is not very sensitive to the number of neighbors. In the current implementation, K is set as 9. The wavelet transform scale
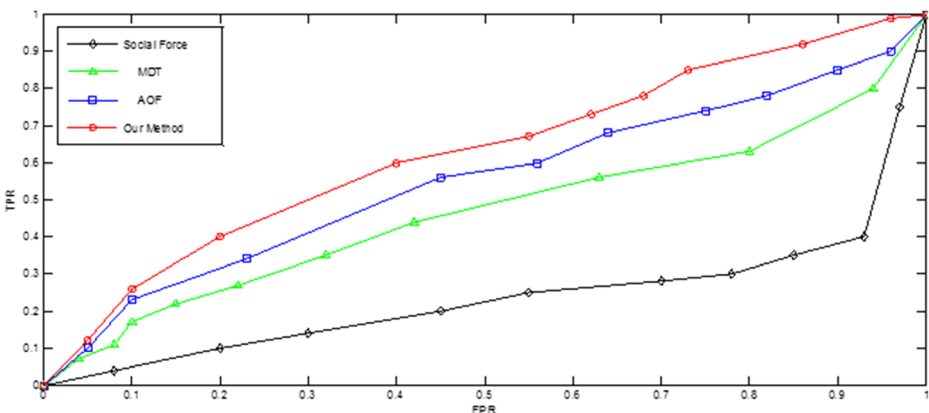

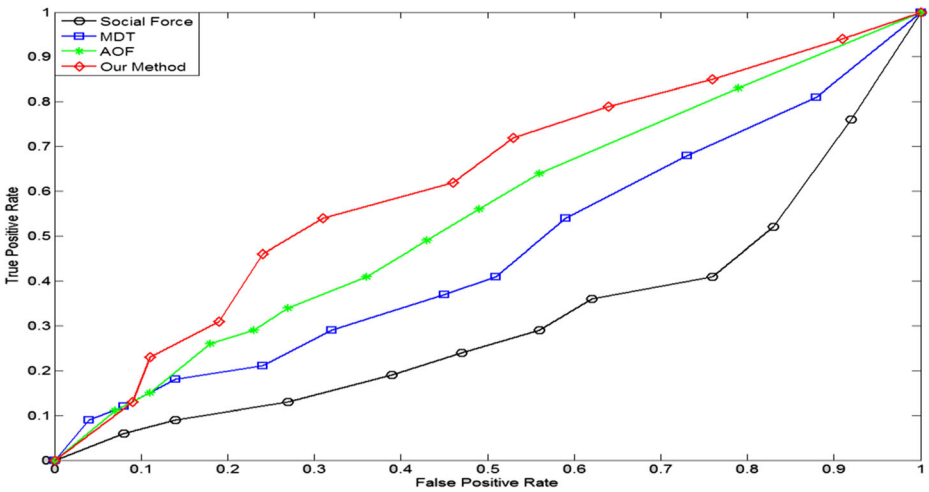**Fig. 9** Pixel-level receiver operating characteristic of the Ped2 dataset

Fig. 10 Pixel-level receiver operating characteristic of the subway dataset

value υ is set as 8, since there is no improvement in performance can be obtained with more scale values. After cross-validation, the size of each blob $h\times w$ is fixed at 60×40, and the threshold in the mixed naive Bayes model T is set as 1.25. Furthermore, the threshold in Eq. (1), ρ is set as 1.5.

4.4 Abnormal detection results on UCSD

In our work, the detection results of local abnormal behaviors are marked with red rectangular boxes. Some examples of anomaly behavior detection on USCD and subway datasets are given in Figs. 5, 6 and 7 respectively.

It can be seen from the results, the proposed method can detect different kinds of local anomalies, such as a pedestrian drives a wheelchair in Fig. 5a, a pedestrian rides a bicycle in Fig. 5b, a pedestrian drives a car in Fig. 5c, a pedestrian rides a skateboard in Fig. 5d, a pedestrian drives a car in Fig. 6a, a pedestrian drives a bicycle in Fig. 6b, two pedestrians drive bicycles in Fig. 6c, a pedestrian drives a bicycle and a pedestrian rides a skateboard respectively in Fig. 6d, a person is exiting through the entrance gate in Fig. 7a, two persons are exiting through the entrance gate in Fig. 7b and c, a person is entering through the exit gate in Fig. 7d–f. It can be seen from the experimental results in Figs. 5, 6 and 7 that the proposed approach also has the omission phenomenon of slow abnormal behaviors, such as a pedestrian

Table 1 The area under receiver operating characteristic of four methods

Method	Ped1	Ped2	Subway	Average
Social Force	17.9 %	26.3 %	37.5 %	27.23 %
MDT	44.1 %	50.7 %	53.8 %	49.53 %
AOF	51.8 %	60.1 %	63.4 %	58.43 %
Our Method	58.6 %	68.3 %	76.2 %	67.70 %

Table 2 The required computational time of four methods

Method	Ped1	Ped2	Subway
Social Force	32	46	69
MDT	24	33	48
AOF	12	19	27
Our Method	0.37	0.41	0.47

cycling slowly in Fig. 6d, which is caused that slow abnormal behaviors cannot be detected by the information of optical flow.

The receiver operating characteristic curves of the proposed approach and other three approaches on Ped1 and Ped2 datasets are given in Figs. 8 and 9 respectively, where the compared approaches contain social force based method [16], mixed dynamic texture based method [13] (shorted as MDT), adaptive optical flow filtering based method [11] (shorted as AOF). It can also be seen from the experimental results in Figs. 8, 9 and 10, the proposed approach is superior to other three approaches.

The area under receiver operating characteristic (shorted as AUC) of four approaches is shown in Table 1. The results illustrate the same conclusion that the proposed approach works better than other three approaches.

The processing time per frame in seconds of tested methods using different datasets are liiustrated in Table 2, where the approximate computational time is obtained on a PC with an Intel E6700 CPU and 2GB of RAM. It can be seen from Table 2 that the proposed method is considerably faster andit requires much less memory to store the learnt data.

5 Conclusion

This paper proposes a novel approach for local abnormal behavior detection. The proposed approach needs no precise object detection or precise target tracking, and it also has some degree of robust. Semi-parameter based statistical model and largest optical flow energy model are utilized to reduce the search range of abnormal behaviors and improve the search efficiency. The experimental results on UCSD and subway public datasets demonstrate the effectiveness and superiority of the proposed approach.

There are some future works along this direction. One direction is that image segmentation technology can be utilized to locate abnormal tragets to improve abnormal behavior detection accuracy. Another direction is that a more comprehensive database needs to be established, such as fighting, escaping.

References

1. Adam A, Rivlin E, Shimshoni I et al (2008) Robust real-time unusual event detection using multiple fixed-location monitors. IEEE Trans Pattern Anal Mach Intell 30(3):555–560
2. Breitenstein M, Grabner H, Gool L (2009) Hunting nessie-real-time abnormality detection from webcams. IEEE Conf Vis Surveill 2009:76–83
3. Chen X, Yuan J, Nie L, et al. (2010) TRECVID 2010 Known-item Search by NUS. IEEE Conference on Text and Video Retrieval 2010:1–12
4. Cheng K, Chu C (2004) Semiparametric density estimation under a two-sample density ratio model. Bernoulli 10(4):583–604
5. Cui X, Liu Q, Gao M (2011) Abnormal detection using interaction energy potentials. IEEE Conf Comput Vis Pattern Recognit 2011:3161–3167
6. Dee H and Hogg D (2004) Detecting inexplicable behaviour. British machine vision conference, 477–486
7. Hendel A, Weinshall D, Peleg S (2010) Identifying surprising events in videos using bayesian topic models. Asian Conf Comput Vis 6494:448–459
8. Hong R, Li G, Nie L et al (2010) Exploring large scale data for multimedia QA: an initial study. ACM Conf Content-Based Image Video Retr 2010:74–81
9. Hou A, Guo J, Wang C (2013) Abnormal behavior recognition based on trajectory feature and regional optical flow. ICIG 2013:643–649
10. Kim J and Grauman K (2009) Observe locally, infer globally: a space-time MRF for detecting abnormal activities with incremental updates. IEEE Conf Comput Vis Pattern Recognit 2921–2928. doi:10.1109/CVPR.2009.5206569
11. Lee D, Suk H, Lee S (2014) Modeling crowd motions for abnormal activity detection. IEEE Conf Adv Video Signal Based Surveill 2014:325–330
12. Li C, Han Z, Ye Q (2013) Visual abnormal behavior detection based on trajectory sparse reconstruction analysis. Neurocomputing 119(1):94–100
13. Mahadevan V, Weixin L, Bhalodia V et al (2010) Anomaly detection in crowded scenes. IEEE Conf Comput Vis Pattern Recognit 2010:1975–1981
14. Mehran R, Moore B, Shah M (2010) A streakline representation of flow in crowded scenes. Eur Conf Comput Vis 6313:439–452
15. Mehran R, Oyama A, Shah M (2009) Abnormal crowd behavior detection using social force model. IEEE Conf Comput Vis Pattern Recognit 2009:935–942
16. Nie L, Wang M, Gao Y et al (2013) Beyond text qa: multimedia answer generation by harvesting web information. IEEE Trans Multimed 15(2):426–441
17. Nie L, Wang M, Zha Z et al (2011) Multimedia answering: enriching text QA with media information. ACM SIGIR Conf 2011:695–704
18. Palmer J, Kreutz K, Rao B (2005) Variational EM algorithms for non-Gaussian latent variable models. IEEE Conf Adv Neural Inf Process Syst 2005:1059–1066
19. Shan H, Banerjee A (2011) Mixed-membership naive Bayes models. Data Min Knowl Disc 23(1):1–62
20. Shet V, Harwood D, Davis L (2006) Multivalued default logic for identity maintenance in visual surveillance. Eur Conf Comput Vis 3954:119–132
21. Wallach H (2006) Topic modeling: beyond bag-of-words. ACM Conf Mach Learn 2006:977–984
22. Yan Y, Ricci E, Liu G et al (2015) Egocentric daily activity recognition via multitask clustering. IEEE Trans Image Process 24(10):2984–2995
23. Yan Y, Ricci E, Subramanian R et al (2013) No matter where you are: flexible graph-guided multi-task learning for multi-view head pose classification under target motion. IEEE Int Conf Comput Vis 2013:1174–1183
24. Yan Y, Ricci E, Subramanian R et al (2014) Multi-task linear discriminant analysis for multi-view action recognition. IEEE Trans Image Process 23(12):5599–5611
25. Yan Y, Ricci E, Subramanian R, et al. (2015) Multi-task learning framework for head pose estimation under target motion. IEEE Trans Pattern Anal Mach Intell. doi:10.1109/TPAMI.2015.2477843
26. Yan Y, Yang Y, Meng D et al (2015) Event oriented dictionary learning for complex event detection. IEEE Trans Image Process 24(6):1867–1878
27. Zach C, Pock T, Bischof H (2007) A duality based approach for realtime TV-L 1 optical flow. Pattern Recogn 4713:214–223
28. Zhu X, Liu J, Wang J (2014) Sparse representation for robust abnormality detection in crowded scenes. Pattern Recogn 47(5):1791–1799

Songhao Zhu received his PhD degree in image processing and pattern recognition from Shanghai Jiao Tong University. In September 2009, he joined Nanjing University of Posts and Telecommunications. His research interests include image processing, pattern recognition, multimedia communication and computer vision.

Juanjuan Hu is currently a Master candidate at Institute of Image Processing and Pattern Recognition, Nanjing University of Posts and Telecommunications. Her research interests include image processing, pattern recognition, multimedia communication and computer vision.

Zhe Shi is currently a Master candidate at Institute of Image Processing and Pattern Recognition, Nanjing University of Posts and Telecommunications. His research interests include image processing, pattern recognition, multimedia communication and computer vision.