SCIENCE CHINA
Information Sciences

• MOOP •

May 2020, Vol. 63 154101:1–154101:3
https://doi.org/10.1007/s11432-018-9792-8

Anomaly detection by exploiting the tracking trajectory in surveillance videos

Zixuan XUE[1] & Wei WU[2*]

[1]*Faculty of Information Technology, Beijing University of Technology, Beijing* 100124*, China;*
[2]*State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing* 100191*, China*

Received 18 September 2018/Revised 8 December 2018/Accepted 15 February 2019/Published online 25 February 2020

Citation Xue Z X, Wu W. Anomaly detection by exploiting the tracking trajectory in surveillance videos. Sci China Inf Sci, 2020, 63(5): 154101, https://doi.org/10.1007/s11432-018-9792-8

Surveillance systems have become increasingly ubiquitous, which has led to a requirement to detect anomalies for efficiently preventing terrorism and reducing crimes. The increasing number of surveillance networks has imposed major technical challenges on intelligent anomaly detection because of the inconsistent appearance of pedestrians owing to posture deformation and clutter in video streams. Furthermore, partial occlusions significantly increase the difficulty of anomaly detection in case of a single camera view.

Researchers have extensively worked to develop anomaly detection in surveillance videos. The method of clustering [1] is used quite commonly for performing anomaly detection, where the spatially isolated search clusters are employed to detect anomalies. Kaltsa et al. [2] adopted histograms of oriented swarms to analyze the dynamics of anomalies in crowds; however, they neglected the fact that the discrimination of hand-crafted appearance cues will be weakened when the moving orientations change among various surveillance scenarios. Apart from the appearance cues, considerable efforts have been invested toward the development of the social force model [3] and sparse-coding [4] for anomaly detection. Although these methods have exhibited promising results in highly structured monitoring environments, their lack of scalability and time complexity considerably limit their capabilities. Tracking trajectory [5] has been applied to detect anomalies. Multi-object tracking provides the strongest clues for detecting abnormal events because it can directly obtain object-level semantic interpretations of anomalies. However, partial occlusions and posture deformation restrict the ability of multi-object tracking while capturing the feature representations of moving objects, especially in cluttered environments, which has not been explored yet. In case of unsolved problems, we propose an effective method based on trajectory tracking for performing intelligent anomaly detection. Specifically, we derive spatial and temporal feature representations for trajectory association between consecutive frames by distilling the motion information about moving objects. Based on the results of trajectory tracking, algorithms are designed for each type of anomaly. Furthermore, we improve the frame differential using the double fusion method during the foreground extraction process, which is especially crucial for detecting the abandoned objects.

Feature representations of spatial and temporal. We employ the faster R-CNN and deep residual network for performing object detection. Further, we adopt the GoogLeNet Inception V3 network to enhance the discrimination of the appearance features, which explicitly leverages the appearance cues to improve the trajectory associations. Because frequent deformations and occlusions are the main problems that lead to the inconsistent appearances of pedestrians, we use spatial and temporal feature representations with discriminative mixed features to overcome these problems. The trajectory association of the pedestrians is formu-

[*] Corresponding author (email: wuwei@buaa.edu.cn)

lated as

$$A(q^m_{\varphi_i}, q^n_{\varphi_{i-1}}) = \omega_1 E_\theta(q^m_{\varphi_i}, p^n_{\varphi_{i-1}}) + \omega_2 E_l(q^m_{\varphi_i}, q^n_{\varphi_{i-1}}), \tag{1}$$

where $A(q^m_{\varphi_i}, q^n_{\varphi_{i-1}})$ can be used to calculate the similarity of the pedestrians between the i-th and $(i-1)$-th frames. $q^m_{\varphi_i}$ denotes the feature representation of the m-th pedestrian in the i-th frame. $p^n_{\varphi_{i-1}}$ demonstrates the mixed feature representation of the n-th pedestrian in the $(i-1)$-th frame. Further, we use the Euclidean distance to measure the similarity between consecutive frames, which is a major evidence of the trajectory associations.

Based on the comparative analysis of multiple experiments, the trajectory association performance is observed to be superior when ω_1 and ω_2 are set in

$$[\omega_1, \omega_2] = \begin{cases} [0.2, 0.8], & \text{if } K_L(x) < \xi_L \\ & \quad \cup K_R(x) > \xi_R, \\ [0.3, 0.7], & \text{if } K_D(x) < \xi_D \\ & \quad \cup K_U(x) > \xi_U, \\ [0.75, 0.25], & \text{otherwise,} \end{cases} \tag{2}$$

where $K_L(x)$, $K_R(x)$, $K_D(x)$, and $K_U(x)$ refer to the left, right, down, and up boundaries for the location of x, respectively. The $K(x) = \ldots, K_{t-1}(x), K_t(x), K_{t+1}(x), \ldots$ sequences are generated according to the coordinates extracted from deep learning-based object detection.

Meanwhile, the mixed feature is obtained from the combination of history state and appearance feature in the i-th frame, and it improves the discriminative feature representations in the face of partial occlusions and posture deformations; it is expressed as

$$p_{\varphi_i} = \lambda_1^{\tau-1}\phi_1 + \sum_{i=1}^{\tau-1} \lambda_2 \lambda_1^{\tau-1-i}\phi_{i+1}, \tag{3}$$

where ϕ_i indicates the appearance feature vector. We consider the pedestrians in the first frame as the initial state of the pedestrian library. The pedestrian library is updated after each frame using the mixed appearance feature between the $(i-1)$-th frame and the historic frame states in which the number of frames in the video sequence is τ. λ_1 and λ_2 are the elastic thresholds of the appearance feature weight, and they are empirically set to 0.65 and 0.35, respectively.

Anomaly detection with algorithms.

- Detection of abandoned objects. We divide clips into segments and consider the first frame of each segment as the benchmark frame. Further, we use the frame differential to obtain foreground sequences and complete the first fusion. We eliminate the single pixels using the median filter algorithm and complete the second fusion to simultaneously obtain the foreground. The double fusion method addresses the inaccurate detection of frame differential, because of the low velocity of the moving objects. The process of our double fusion method is depicted in Figure 1(d), and it is expressed by

$$\eta_{c,r} = \begin{cases} \displaystyle\sum_{a=1}^{C}\sum_{b=1}^{R} O^{\varsigma_i}_{a,b}, & \text{if } |\eta^{\varsigma_i}_{c,r} - \eta^{\varsigma_{i-1}}_{c,r}| \geqslant \delta, \\ 0, & \text{otherwise,} \end{cases} \tag{4}$$

$$\tilde{\eta}_{c,r} = \begin{cases} \eta_{c,r}, & \text{if } \eta_{c,r} \in \text{moving objects}, \\ 0, & \text{otherwise,} \end{cases} \tag{5}$$

where $O^{\varsigma_i}_{a,b}$ represents the pixel value of the patches, and each frame is divided into R rows and C columns. δ is predefined to measure the change between the i-th and $(i-1)$-th frames and is set to 3. $\eta_{c,r}$ refers to the matrix of the patch values. $\tilde{\eta}_{c,r}$ denotes the foreground matrix; we extract the moving objects by updating each element of the patches.

- Detection of abnormal behavior in crowds. We consider the intersection of the bounding boxes as an important factor for performing crowd anomaly detection. Based on the tracking trajectory results and the Union-Find algorithm, a crowd anomaly is assumed to occur if more than three bounding boxes intersect or if those intersecting bounding boxes separate.

- Detection of the fast-moving objects. Once a pedestrian exceeds the finite fluctuation around the average velocity of the crowd, we consider this behavior to be running. Further, the deflection angle between the pedestrians in consecutive frames can be given by

$$\rho = \arccos\Big\{(v^m_i \times v^m_{i+1} + v^n_i \times v^n_{i+1}) \Big/ \sqrt{\big((v^m_i)^2 + (v^n_i)^2\big) \times \big((v^m_{i+1})^2 + (v^n_{i+1})^2\big)}\Big\}, \tag{6}$$

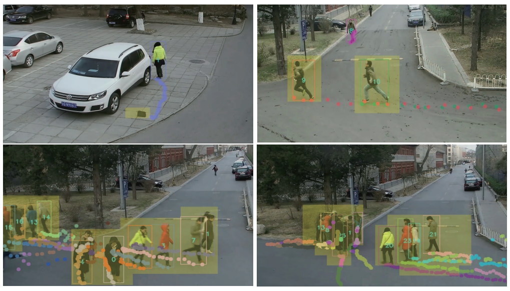
where ρ denotes the deflection angle. v^m_i and v^n_i denote the velocities of the m-th and n-th pedestrians, respectively, in the i-th frame. The maximum deflection angle, maximum difference in velocity, and minimum anomaly velocity are set to 7.5, 0.3, and 2.2, respectively.

Experiments and comparisons. We performed both qualitative and quantitative experiments by applying our method to the UCSD Ped1, UMN, and PKU-SVD-B datasets to evaluate its performance, as depicted in Figures 1(a), (b), (c). The

	F_{Clip}	F_{Frame}	F_{Clip}	F_{Frame}
SF [3]	61.2	59.5	64.4	61.9
Sparse [4]	76.1	74.5	83.3	82.5
Ours	**88.7**	**85.6**	**92.7**	**90.6**

	F_{Clip}	F_{Frame}	Score
Abandoned object	92.84	91.44	92.36
Aggregated	90.16	88.43	89.57
Dispersed	89.72	86.49	88.16
Fast-moving	**94.27**	**92.16**	**93.55**

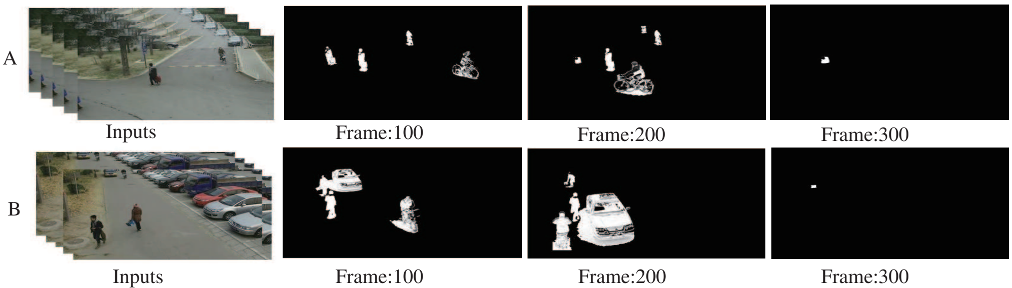
(a)



(b)

Tracker	MOTA	MOTP	IDF1	MT (%)	ML (%)	FP	FN	IDsw	Frag	FPS
TRID [6]	55.7	76.5	**61.0**	**40.6**	**25.8**	**6273**	**20611**	**351**	667	3.9
JointMC [6]	35.6	71.9	45.1	23.2	39.3	10580	28508	457	969	0.6
Ours	**56.2**	**78.1**	57.4	32.1	37.6	8829	21754	443	**647**	**5.2**

(c)



(d)

Figure 1 (Color online) (a) Results of the application of proposed method to the UCSD Ped1, UMN, and PKU-SVD-B datasets; (b) anomaly detection using the PKU-SVD-B dataset; (c) comparative results of the tracking trajectory with respect to the 2D MOT 2015 benchmark; (d) the process of double fusion method. A and B represent different video sequences.

results demonstrate that our method significantly outperforms the existing methods by achieving high accuracy on both the frame level and the clip level. The spatial and temporal feature representations can effectively alleviate the mismatching of trajectories because of posture deformation and partial occlusions, which ensures the consistency of appearance. Furthermore, the improvement in time efficiency is of considerable significance for promoting the online trajectory association and for facilitating anomaly detection in large-scale surveillance environments.

Conclusion and discussion. We proposed an anomaly detection method by exploiting the tracking trajectory to alleviate the problems that are caused by partial occlusions, posture deformation, and clutter in views. We utilized the spatial and temporal feature representations generated by trajectory tracking for associating the trajectories between consecutive frames. Based on this, algorithms have been designed for specific anomalies. More importantly, in combination, the aforementioned steps achieved superior performance in terms of both accuracy and efficiency. Further, we denoted the challenges associated with intelligent anomaly detection, and we will continue to work on real-time detection in the future.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61872024, 61472020).

Supporting information Videos and other supplemental documents. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Zhou S F, Shen W, Zeng D, et al. Spatial-temporal convolutional neural networks for anomaly detection and localization in crowded scenes. Signal Process Image Commun, 2016, 47: 358–368

2 Kaltsa V, Briassouli A, Kompatsiaris I, et al. Swarm intelligence for detecting interesting events in crowded environments. IEEE Trans Image Process, 2015, 24: 2153–2166

3 Mehran R, Oyama A, Shah M. Abnormal crowd behavior detection using social force model. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Miami, 2009. 935–942

4 Adler A, Elad M, Hel-Or Y, et al. Sparse coding with anomaly detection. J Sign Process Syst, 2015, 79: 179–188

5 Bae S, Yoon K. Robust online multi-object tracking based on tracklet confidence and online discriminative appearance learning. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Boston, 2014. 1218–1225

6 Multiple object tracking benchmark. https://motchallenge.net/results/2D_MOT_2015/