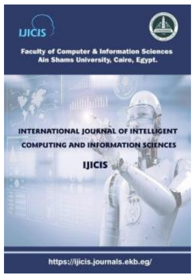
IJICIS, Vol.21, No.3, 84-94
DOI: 10.21608/ijicis.2021.84588.1112

International Journal of Intelligent Computing and Information Sciences

https://ijicis.journals.ekb.eg/



Comparative Study for Anomaly Detection in Crowded Scenes

Mohamed Ali*
Computer Science department, Computer and Information Science, Ain Shams University, Cairo, Egypt
mohamed.abdelghafour@cis.asu.edu.eg

Maryam Al-Berry
Scientific Computing department Computer and Information Science, Ain Shams University, Cairo, Egypt
maryam_nabil@cis.asu.edu.eg

Zaki Taha
Computer Sciense department, Computer and Information Science, Ain Shams University, Cairo, Egypt
ztfayed@hotmail.com

Received 2021-7-7; Revised 2021-9-8; Accepted 2021-9-14

Abstract: *Nowadays, video analysis is an important research area especially from a security point of view. The discovery of unusual activities is important because it is a difficult task for humans especially with increasing number of surveillance cameras in all crowded places. That is because it requires a lot of human effort, and these activities happen rarely. Also the definition of anomaly events is different based on the location of the event. For example running in the park is a normal event but running in a restaurant is an abnormal event. The event is the same but the place was the factor of making it normal or not. The main objective of this paper is to compile what has been achieved in the field of anomaly detection and compare them, and to look at the different datasets used in the recent period. We will show how to detect and identify anomalies in videos, approaches for video anomaly detection and also what are the latest learning frameworks.*

Keywords: *Abnormal event detection, Video surveillance, Unsupervised learning.*

1. Introduction

Now, Closed-circuit televisions (CCTVs) are everywhere, especially crowded places. So, it is very important to use machine learning to help detect anomalies in those places. Without the help of artificial

* Corresponding author: Mohamed Ali
Computer Science department, Computer and Information Science, Ain Shams University, Cairo, Egypt
E-mail address: mohamed.abdelghafour@cis.asu.edu.eg

intelligence, this task will require a lot of human efforts to monitor a lot of screens and discover these cases [1], knowing that they are rare [2].

There is no fixed definition of anomalous events because the event varies according to the location of the event [3][4]. For example, running in parks is normal, but running in a restaurant is strange. The event, which is traction, changes its evaluation according to the place in which it occurred. Those differences add more challenges to machine learning to discover these events in applications in our daily life. There are also some other challenges like environmental conditions (presence of shadow, different lighting, appearance of obstructions, background problems, etc.), crowd density, the complex nature of human behavior. These things add further challenge to machine learning to discover anomalies.

Some modern works deal with the problem of discovering anomalies as a binary classification problem (normal and anomalies) [5], and some other works focus on classifying each event separately, as each moving object is tracked and behavior is studied to determine if it is normal or not.

Fixed surveillance cameras can be used, such as surveillance cameras in the streets or in front of stores, or mobile surveillance cameras such as front car cameras or police cameras. However, there is little research to detect video anomalies using mobile cameras, so this research will focus on machine learning methods to detect anomalies using fixed cameras.

The rest of the paper is organized as follows: In Section 2, the methods of classification. The recent training and learning frameworks are discussed in Section 3. Section 4 shows approaches for video anomaly detection. Section 5 lists the available data sets and a comparison between them. The final section gives concluding remarks on some future research areas.

2. Classification of video anomalies

There are two factors to detect and identify anomalies in the video 1) complexity of the environment, 2) the type of anomaly.

Moreover, the complexity of the environment depends on the density of moving targets in the scene. Based on the density of moving objects, the environment can be classified into three types which are 1) a slightly crowded environment (10 sqft/person), 2) a moderately crowded environment (4.5 sqft/person), and 3) a crowded environment (2.5 sqft/person) [6].

The types of anomalies can be classified as local anomalies and global anomalies. A local anomaly is a significant deviation from neighboring activities such as a car moving in the opposite direction [7][8][9][10][11]. You will find that the behavior of the individual is different from his neighbors [12] like someone is walking on grass and other people are walking on the road. On the contrary, the global anomaly is the activities that interact with each other strangely, regardless of individual behavior [10] like everyone are running suddenly on the same time.

3. Training and learning frameworks

Learning methods used in deep artificial neural networks to detect anomalies in crowded scene videos are classified into four categories: they are supervised, unsupervised, semi-supervised and active learning [11][13]. Human intervention and prior knowledge of labels during training are essential pillars of this classification.

3.1. Supervised video anomaly detection

The main pillar in the use of supervised learning is the use of binary classification using labels associated with anomalous or normal events. Therefore, each label must be clearly identified whether it is anomalous or not. Also, balance must be taken into account in the labels, as well as providing a lot of training videos. Despite this, there are some problems, such as a lack of frames that contain anomalies, or there are problems in determining whether that frame will be anomalous or not [11]. Therefore, the prevailing trend is to use the other two methods, which are unsupervised and semi-supervised training.

3.2. Unsupervised video anomaly detection

Unsupervised learning uses co-occurrence statistical concepts from the unlabeled video data [13]. These methods are used in the auto-labeling of the unlabeled video data [11]. Learning methods without supervision require a lot of video dataset and massive computational resources. When they are available, methods of learning without supervision surely excel over methods of supervised learning [11].

3.3. Semi-supervised video anomaly detection

The main pillar in using semi-supervised learning is the use of weakly-labeled normal instances of video. This type is widely used because it uses the advantages of both supervised and unsupervised techniques. In the last period, models based on deep-autoencoder are trained with training data that includes only normal events so as to produce minimal reconstruction errors for normal activities [11].

3.4. Active learning-based video anomaly detection

The methods within this category do not require training data either in the labeled or unlabeled form. Instead, these models are based on knowledge of the external domain [2]. Learning in such models is immediately achieved based on parameter adjustments / adaptations during inferences on the test video.

A summary of the previous four different types of methods is described in table 1. It shows some works for each category.

Table 1: Categorization of learning methods to detect anomalies in crowded scene.

Category	Reference	Feature	Dataset	AUC frame level
Supervised	Mahadevan et al. [7]	Dynamic textures	UCSD	
	Cong et al. [14]	Spatio-temporal	UCSD	47.1%, 86.8%
	Biswas and Gupta [15]	Motion	UCSD, UMN	88.71%, 96.70% 98.34%
	Luo et al. [16]	Convolutional neural network	UCSD ped2, Avenue	92.21%, 81.71%
	Ravanbakhsh et al. [17]	Motion	UCSD, UMN	97.4%, 93.5%, 99%
	Luo et al. [18]	Convolutional neural network	UCSD, Avenue	75.5%, 88.1%, 77%
	Sabokrou et al. [19]	Convolutional neural network	UCSD Ped2	N/A
	Ullah et al. [20]	Spatio-temporal	UCSD, UMN	N/A
	Ionescu et al. [21]	Appearance and	UCSD Ped2, Avenue	97.8%, 90.4%
	Khan et al. [22]	Histogram of	UCSD, UMN	81.08%, 93.75%, 98.18%
	Zhou et al. [23]	Fused Motion	UCSD, UMN, Avenue	83.5%, 94.9%, 86.1%, 86.1%
	Singh et al. [24]	Convolutional neural network	UCSD, Avenue	94.6%, 95.9%, 89.3%
Unsupervised	Xu et al. [25]	Motion	UCSD	92.1%, 90.8%
	Sabokrou et al. [26]	Spatio-temporal	UCSD ped2, UMN	N/A, 99.6%
	Li et al. [27]	Spatio-temporal	UCSD, UMN	87.2%, 89.1%, 93%
	Ionescu et al. [28]	Spatio-temporal	UCSD, Avenue	68.4%, 82.2%, 80.6%
	Hu et al. [29]	Motion	UCSD, UMN, Avenue	62.62%, 74.04%, 99.3%, 87.19%
	Bansod et al. [30]	Histogram of magnitude and momentum	UCSD, UMN	N/A
Semi-supervised	Bertini et al. [31]	Spatio-temporal	UCSD	N/A
	Hasan et al. [32]	Convolutional neural network and motion	UCSD, Avenue	92.7%, 90.8%, N/A
	Liu et al. [33]	Intensity and motion	UCSD, Avenue	83.1%, 95.4%, 85.1%
Active learning-based	Roshtkhari et al. [34]	Spatio-temporal	UCSD	N/A
	Yuan et al. [35]	Structural Motion	UCSD, UMN	91%, 92.5%, 99.67%
	Nawaratne et al. [36]	Convolutional neural network	UCSD, Avenue	75.2%, 91.1%, 76.8%

4. Approaches for video anomaly detection

There are two methods of detecting anomalies in videos: accuracy-oriented approaches and processing-time-oriented approaches [37].

4.1. Accuracy-oriented approaches

Accuracy-oriented approaches detect and localize anomalies with high accuracy and low false alarms. Complex models are trained with the largest number of features to achieve the high precision required at the expense of high processing time [37]. It wants to make video anomaly detection methods suitable for offline applications using all available training video datasets [11]. Some of the important research works are based on temporal regularity model [32], generative models [38], spatio-temporal / predictive models [23], hybrid models [39], and models using detection of the STIPs [40].

4.2. Processing-time-oriented approaches

Processing-time-oriented approaches detect and localize anomalies with minimum frame processing time and a competitive level of accuracy. It wants to make video anomaly detection methods suitable for online applications using high computational speed and less computation space [37]. Few important research works are based on generative models [41], temporal regularity model [42], spatio-temporal / predictive models [43], hybrid models, cell-based models [37].

5. Datasets

The available data sets are rather scarce, due to the fact that the research topic is rather new and also due to the scarcity of anomalies in the crowded scenes. We will talk about the most common datasets, and they are listed in the table 2. It shows a brief for available dataset.

Table 2: Comparison between the public datasets contain anomaly events in crowded scenes

Dataset	Description	Number of Images or Videos
University of California, San Diego (UCSD) [44]	**Peds1:** Clips of groups of people walking towards and away from the camera, and some amount of perspective distortion. **Peds2:** Scenes with pedestrian movement parallel to the camera plane.	**Peds1:** Contains 34 training video samples and 36 testing video samples. **Peds2:** Contains 16 training video samples and 12 testing video samples.
University of Minnesota (UMN) [45]	Three scenes for crowds running.	Crowd Escape Panic, 11 Videos, 3 Scenes.
Avenue [46]	The videos are captured in CUHK campus avenue with 30652 (15328 training, 15324 testing) frames in total.	Contains 16 training and 21 testing video clip.
PETS 2009 [47]	The datasets are multisensor sequences containing different crowd activities. S1: Concerns person count and density estimation. S2: Addresses people tracking. S3: Involves flow analysis and event recognition.	Contains 11 videos.
Anomalous Behavior/ York [48]	Collection of 6 datasets with names: Traffic-Train Belleview Boat-Sea Boat-River Caouflage Airport-WrongDir	Traffic-Train: 19218 frames. Belleview: 2918 frames. Boat-Sea: 450 frames. Boat-River: 250 frames. Caouflage: 1629 frames. Airport-WrongDir: 2200 frames.
Queen Mary University of London (QMUL) Junction [49][50][51][52]	A busy traffic dataset for 1 hour	Contains 90000 frames
MIT Traffic [53]	It includes a traffic video sequence of 90 minutes long. It is recorded by a stationary camera.	Contains 20 clips.
Violent flows [54]	A database of real-world, video footage of crowd violence, along with standard benchmark protocols designed to test both violent/non-violent classification and violence outbreak detections. The shortest clip duration is 1.04 seconds, the longest clip is 6.52 seconds, and the average length of a video clip is 3.60 seconds.	Contains 246 videos.
BOSS [55]	It is a fairly realistic dataset for pose, action and interaction recognition. It consists of video/audio recordings of acted actions inside a moving train. It used multiple calibrated cameras.	Contains 130 videos for 19 events.
Crowd [56]	There are a total of 470K human instances from train and validation subsets and 23 persons per image, with various kinds of occlusions in the dataset.	Contains 15000, 4370 and 5000 images for training, validation, and testing, respectively.
BEHAVE [57]	It comprises of two views of various scenarios of people acting out various interactions.	Contains 4 clips.

6. Conclusion

The topic of anomaly detection in crowded scenes is important and interesting. In this article, we discussed how to classify videos as anomaly and mentioned each classification of them. We also discussed modern training and learning frameworks and dealt with each type of them. We covered the latest approaches to discovering anomalies in video, and what is the differences among them. We also collected most of the databases used in this field of research, some of them are used a lot and some of them are of little use. Finally, there are still some challenges in this field of research, and researchers are still providing new solutions.

Video anomaly detection can be applied in many fields of video surveillance applications such as traffic violations, abnormal crowd behavior and criminal activities. It will save a lot of time and efforts. Also it will make the anomaly detection more accurate.

For future work we recommend to use datasets UCSD, Avenue because they have different anomaly events and also they are in park and in campus so these events are mostly like other events in crowded scenes. Also we suggest a model in which Convolutional Autoencoder learns the spatial features to detect anomaly events from video sequences. We will extract local features from the videos, which are then used to train a normative model.

References

1. Haering, Niels, Péter L. Venetianer, and Alan Lipton. "The evolution of video surveillance: an overview." Machine Vision and Applications 19.5 (2008): 279-290.
2. Sodemann, Angela A., Matthew P. Ross, and Brett J. Borghetti. "A review of anomaly detection in automated surveillance." IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 42.6 (2012): 1257-1272.
3. Zhou, Shifu, et al. "Spatial–temporal convolutional neural networks for anomaly detection and localization in crowded scenes." Signal Processing: Image Communication 47 (2016): 358-368.
4. Zhou, Shifu, et al. "Unusual event detection in crowded scenes by trajectory analysis." 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2015.
5. Taha, Ayman, and Ali S. Hadi. "Anomaly detection methods for categorical data: A review." ACM Computing Surveys (CSUR) 52.2 (2019): 1-35.
6. Jacobs, Herbert. "To count a crowd." Columbia Journalism Review 6.1 (1967): 37.
7. Mahadevan, Vijay, et al. "Anomaly detection in crowded scenes." 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2010.
8. Saligrama, Venkatesh, and Zhu Chen. "Video anomaly detection based on local statistical aggregates." 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012.
9. Kaltsa, Vagia, et al. "Swarm intelligence for detecting interesting events in crowded environments." IEEE transactions on image processing 24.7 (2015): 2153-2166.
10. Pawar, Karishma, and Vahida Attar. "Deep learning approaches for video-based anomalous activity detection." World Wide Web 22.2 (2019): 571-601.
11. Chalapathy, Raghavendra, and Sanjay Chawla. "Deep learning for anomaly detection: A survey." arXiv preprint arXiv:1901.03407 (2019).
12. Hu, Xing, et al. "Video anomaly detection using deep incremental slow feature analysis network." IET Computer Vision 10.4 (2016): 258-267.
13. Popoola, Oluwatoyin P., and Kejun Wang. "Video-based abnormal human behavior recognition—A review." IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 42.6 (2012): 865-878.
14. Cong, Yang, Junsong Yuan, and Yandong Tang. "Video anomaly search in crowded scenes via spatio-temporal motion context." IEEE transactions on information forensics and security 8.10 (2013): 1590-1599.
15. Biswas, Soma, and Vikas Gupta. "Abnormality detection in crowd videos by tracking sparse components." Machine Vision and Applications 28.1-2 (2017): 35-48.
16. Luo, Weixin, Wen Liu, and Shenghua Gao. "A revisit of sparse coding based anomaly detection in stacked rnn framework." Proceedings of the IEEE International Conference on Computer Vision. 2017.
17. Ravanbakhsh, Mahdyar, et al. "Abnormal event detection in videos using generative adversarial nets." 2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2017.
18. Luo, Weixin, Wen Liu, and Shenghua Gao. "Remembering history with convolutional lstm for anomaly detection." 2017 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2017.

19. Sabokrou, Mohammad, et al. "Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes." Computer Vision and Image Understanding 172 (2018): 88-97.
20. Ullah, Habib, et al. "Anomalous entities detection and localization in pedestrian flows." Neurocomputing 290 (2018): 74-86.
21. Ionescu, Radu Tudor, et al. "Object-centric auto-encoders and dummy anomalies for abnormal event detection in video." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.
22. Khan, Muhammad Umar Karim, Hyun-Sang Park, and Chong-Min Kyung. "Rejecting motion outliers for efficient crowd anomaly detection." IEEE Transactions on Information Forensics and Security 14.2 (2018): 541-556.
23. Zhou, Joey Tianyi, et al. "AnomalyNet: An anomaly detection network for video surveillance." IEEE Transactions on Information Forensics and Security 14.10 (2019): 2537-2550.
24. Singh, Kuldeep, et al. "Crowd anomaly detection using aggregation of ensembles of fine-tuned convnets." Neurocomputing 371 (2020): 188-198.
25. Xu, Dan, et al. "Learning deep representations of appearance and motion for anomalous event detection." arXiv preprint arXiv:1510.01553 (2015).
26. Sabokrou, Mohammad, et al. "Real-time anomaly detection and localization in crowded scenes." Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2015.
27. Li, Nannan, et al. "Spatio-temporal context analysis within video volumes for anomalous-event detection and localization." Neurocomputing 155 (2015): 309-319.
28. Tudor Ionescu, Radu, et al. "Unmasking the abnormal events in video." Proceedings of the IEEE International Conference on Computer Vision. 2017.
29. Hu, Xing, et al. "Abnormal event detection in crowded scenes using histogram of oriented contextual gradient descriptor." EURASIP Journal on Advances in Signal Processing 2018.1 (2018): 1-15.
30. Bansod, Suprit D., and Abhijeet V. Nandedkar. "Crowd anomaly detection and localization using histogram of magnitude and momentum." The Visual Computer 36.3 (2020): 609-620.
31. Bertini, Marco, Alberto Del Bimbo, and Lorenzo Seidenari. "Multi-scale and real-time non-parametric approach for anomaly detection and localization." Computer Vision and Image Understanding 116.3 (2012): 320-329.
32. Hasan, Mahmudul, et al. "Learning temporal regularity in video sequences." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
33. Liu, Wen, et al. "Future frame prediction for anomaly detection–a new baseline." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.
34. Javan Roshtkhari, Mehrsan, and Martin D. Levine. "Online dominant and anomalous behavior detection in videos." Proceedings of the IEEE conference on computer vision and pattern recognition. 2013.
35. Yuan, Yuan, Jianwu Fang, and Qi Wang. "Online anomaly detection in crowd scenes via structure analysis." IEEE transactions on cybernetics 45.3 (2014): 548-561.
36. Nawaratne, Rashmika, et al. "Spatiotemporal anomaly detection using deep learning for real-time video surveillance." IEEE Transactions on Industrial Informatics 16.1 (2019): 393-402.

37. Leyva, Roberto, Victor Sanchez, and Chang-Tsun Li. "Video anomaly detection with compact feature sets for online performance." IEEE Transactions on Image Processing 26.7 (2017): 3463-3478.
38. Wang, Tian, et al. "Generative neural networks for anomaly detection in crowded scenes." IEEE Transactions on Information Forensics and Security 14.5 (2018): 1390-1399.
39. Zhao, Yiru, et al. "Spatio-temporal autoencoder for video anomaly detection." Proceedings of the 25th ACM international conference on Multimedia. 2017.
40. Cheng, Kai-Wen, Yie-Tarng Chen, and Wen-Hsien Fang. "Video anomaly detection and localization using hierarchical feature representation and Gaussian process regression." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.
41. Wu, Peng, Jing Liu, and Fang Shen. "A deep one-class neural network for anomalous event detection in complex scenes." IEEE transactions on neural networks and learning systems 31.7 (2019): 2609-2622.
42. Chong, Yong Shean, and Yong Haur Tay. "Abnormal event detection in videos using spatiotemporal autoencoder." International symposium on neural networks. Springer, Cham, 2017.
43. Sultani, Waqas, Chen Chen, and Mubarak Shah. "Real-world anomaly detection in surveillance videos." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
44. Statistical Visual Computing Lab. UCSD Anomaly Data Set. 2014. Available online: http://www.svcl.ucsd.edu/projects/anomaly.
45. University of Minnesota, Department of Computer Science and Engineering, available from http://mha.cs.umn.edu/proj_events.shtml#crowd
46. Lu, Cewu, Jianping Shi, and Jiaya Jia. "Abnormal event detection at 150 fps in matlab." Proceedings of the IEEE international conference on computer vision. 2013.
47. Ferryman, J. A. M. E. S., J. Crowley, and A. Shahrokni. "Pets 2009 benchmark data." Computer Vision and Pattern Recognition, Miami, Florida: IEEE Conference on. 2009.
48. Zaharescu, Andrei, and Richard Wildes. "Anomalous behaviour detection using spatiotemporal oriented energies, subset inclusion histogram comparison and event-driven processing." European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2010.
49. Loy, Chen Change, Tao Xiang, and Shaogang Gong. "Modelling Multi-object Activity by Gaussian Processes." BMVC. 2009.
50. Loy, Chen Change, Tao Xiang, and Shaogang Gong. "Stream-based active unusual event detection." Asian Conference on Computer Vision. Springer, Berlin, Heidelberg, 2010.
51. Loy, Chen Change, Tao Xiang, and Shaogang Gong. "Detecting and discriminating behavioural anomalies." Pattern Recognition 44.1 (2011): 117-132.
52. Loy, Chen Change, et al. "Stream-based joint exploration-exploitation active learning." 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012.
53. Wang, Xiaogang, Xiaoxu Ma, and W. Eric L. Grimson. "Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models." IEEE Transactions on pattern analysis and machine intelligence 31.3 (2008): 539-555.
54. Hassner, Tal, Yossi Itcher, and Orit Kliper-Gross. "Violent flows: Real-time detection of violent crowd behavior." 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. IEEE, 2012.

55. Velastin, Sergio A., and Diego A. Gómez-Lira. "People detection and pose classification inside a moving train using computer vision." International Visual Informatics Conference. Springer, Cham, 2017.
56. Shao, Shuai, et al. "Crowdhuman: A benchmark for detecting human in a crowd." arXiv preprint arXiv:1805.00123 (2018).
57. Blunsden, Scott, and R. B. Fisher. "The BEHAVE video dataset: ground truthed video for multi-person behavior classification." Annals of the BMVA 4.1-12 (2010): 4.