International Journal of Research in Engineering, Science and Management
Volume-2, Issue-2, February-2019
www.ijresm.com | ISSN (Online): 2581-5792

A Survey on Real World Anomaly Detection in Live Video Surveillance Techniques

Vina Lomte[1], Satish Singh[2], Siddharth Patil[3], Siddheshwar Patil[4], Durgesh Pahurkar[5]
[1]*Professor & HoD, Dept. of Computer Engineering, RMDSSOE, Savitribai Phule Pune University, Pune, India*
[2,3,4,5]*Student, Dept. of Computer Engineering, RMDSSOE, Savitribai Phule Pune University, Pune, India*

***Abstract*: The creation of various technologies main objective is to improve our society and to maintain peace in our society. In this paper we try to focus on one of the problem that our society faces, that is various crimes and anomalies that lead to creation of tension in the society. In this paper we have done a survey of various available techniques for anomaly detection in live video surveillance and give a comparative analysis between these available techniques. Real-time anomaly detection will decrease human efforts that goes into watching surveillance videos and will be useful in crime detection.**

***Keywords*: Crimes, live video surveillance, anomaly detection.**

1. Introduction

Nowadays due to the presence of surveillance cameras and video cameras we can see how the crime took place. But it does no good to prevent it, so we need some methodology for automatically detecting and classifying whether the abnormalities or anomalies will take place or not. Also notifying and alarming which will result will not only in reduction of crimes but also prevention of it. This problem of real time anomaly detection falls under Emergency management and it is very important to reduce the impact of emergencies. Hence a creation of a model or a system is essential that can solve this problem and can be used in various places like on the road, national parks, inside banks, shops, airports, metros, streets and swimming pools. As the cameras are already present in these places these will work efficiently. This chapter presents a review and systematic comparison of the state of the art on crowd video analysis. The rationale of our review is justified by a recent increase in intelligent video surveillance algorithms capable of analyzing automatically visual streams of very crowded and cluttered scenes, such as those of airport concourses, railway stations, shopping malls and the like. Since the safety and security of potentially very crowded public spaces have become a priority, computer vision researchers have focused their research on intelligent solutions. The aim of this chapter is to propose a critical review of existing literature pertaining to the automatic analysis of complex and crowded scenes. We discuss the merits and weaknesses of various approaches for each topic and provide a recommendation on how existing methods can be improved.

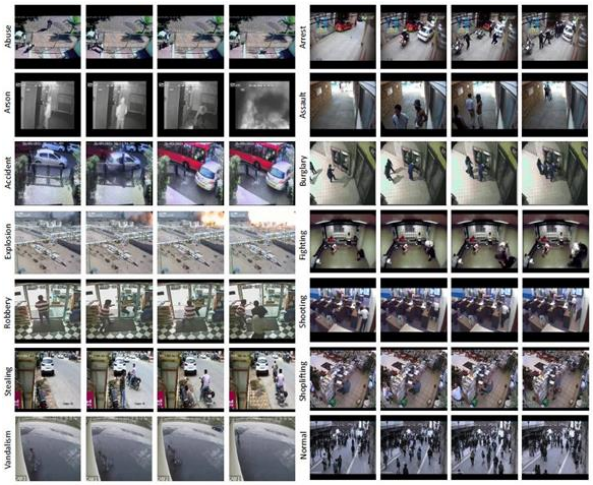

Fig. 1. Types of anomalies

2. Approaches for anomaly detection

Anomaly detection can be approached in many ways depending on the nature of data and circumstances. Following is a classification of some of those techniques.

A. Static rules approach

The most simple, and maybe the best approach to start with, is using static rules. The Idea is to identify a list of known anomalies and then write rules to detect those anomalies. Rules identification is done by a domain expert, by using pattern mining techniques, or a by combination of both.

Static rules are used with the hypothesis that anomalies follow the 80/20 rule where most anomalous occurrences belong to few anomaly types. If the hypothesis is true, then we can detect most anomalies by finding few rules that describe those anomalies.

Implementing those rules can be done using one of three following methods.

If they are simple and no inference is needed, you can code them using your favorite programming language If decisions need inference, then you can use a rule-based or expert system (e.g. Drools) If decisions have temporal conditions, you can use a Complex Event Processing System (e.g. WSO2 CEP, Esper) Although simple, static rules-based systems tend to be brittle and complex. Furthermore, identifying those rules is often a

complex and subjective task. Therefore, statistical or machine learning based approach, which automatically learns the general rules, are preferred to static rules.

B. When we have training data

Anomalies are rare under most conditions. Hence, even when training data is available, often there will be few dozen anomalies exists among millions of regular data points. The standard classification methods such as SVM or Random Forest will classify almost all data as normal because doing that will provide a very high accuracy score (e.g. accuracy is 99.9 if anomalies are one in thousand).

Generally, the class imbalance is solved using an ensemble built by resampling data many times. The idea is to first create new datasets by taking all anomalous data points and adding a subset of normal data points (e.g. as 4 times as anomalous data points). Then a classifier is built for each data set using SVM or Random Forest, and those classifiers are combined using ensemble learning. This approach has worked well and produced very good results.

If the data points are auto correlated with each other, then simple classifiers would not work well. We handle those use cases using time series classification techniques or Recurrent Neural networks.

C. When there is no training data

If you do not have training data, still it is possible to do anomaly detection using unsupervised learning and semi-supervised learning. However, after building the model, you will have no idea how well it is doing as you have nothing to test it against. Hence, the results of those methods need to be tested in the field before placing them in the critical path.

3. Literature survey

[1] Proposed method presents an efficient method for detecting anomalies in videos. Recent applications of convolutional neural networks have shown promises of convolutional layers for object detection and recognition, especially in images. However, convolutional neural networks are supervised and require labels as learning signals. They propose a spatiotemporal architecture for anomaly detection in videos including crowded scenes. Their architecture includes two main components, one for spatial feature representation, and one for learning the temporal evolution of the spatial features.

Experimental results on Avenue, Subway and UCSD benchmarks confirm that the detection accuracy of their method is comparable to state-of-the-art methods at a considerable speed of up to 140 fps.

[2] Proposed method learns anomalies by exploiting both normal and anomalous videos. To avoid annotating the anomalous segments or clips in training videos, which is very time consuming, they propose to learn anomaly through the deep multiple instance ranking framework by leveraging weakly labeled training videos, i.e. the training labels (anomalous or normal) are at video level instead of clip-level. Their method considers normal and anomalous videos as bags and video segments as instances in multiple instance learning (MIL), and automatically learn a deep anomaly ranking model that predicts high anomaly scores for anomalous video segments. They introduce sparsity and temporal smoothness constraints in the ranking loss function to better localize anomaly during training. They also introduce a new large-scale first of its kind dataset of 128 hours of videos. It consists of 1900 long and untrimmed real-world surveillance videos, with 13 realistic anomalies such as fighting, road accident, burglary, robbery, etc. as well as normal activities. This dataset can be used for two tasks. First, general anomaly detection considering all anomalies in one group and all normal activities in another group. Second, for recognizing each of 13 anomalous activities. Their experimental results show that the MIL method for anomaly detection achieves significant improvement on anomaly detection performance as compared to the state-of-the-art approaches. They provide the results of several recent deep learning baselines on anomalous activity recognition. The low recognition performance of these baselines reveals that their dataset is very challenging and opens more opportunities for future work.

[3] Nowadays, abnormal activity detection has been most important part in public safety. Most existing studies do not account for the processing time and the continuity of abnormal behavior characteristics. In this paper, they have proposed a new motion feature called as Sensitive Movement Point (SMP) and new model called as Gaussian Mixture Model (GMM). The Gaussian Mixture Model (GMM) is used for modeling the abnormal crowd behavior with full consideration of the characteristics of crowd abnormal behavior. Firstly, they analyze video by Gaussian Mixture Model to extract sensitive movement point by setting up a certain threshold value of GMM. After that, Sensitive Movement Points are analyzed through the spatial and temporal model. The algorithm can be implemented with automatic adapt to environmental change and online learning, without tracking individuals of crowd and large scale training in detection process. Experiments involving the UMN datasets and the videos taken by them show that the proposed algorithm can real-time effectively identify various types of anomalies and that the recognition results and processing time are better than existing algorithms.

[4] Generally, there are two different methods for analyzing crowd behavior: the method which is based on tracking moving objects in which every person or moving object is traced separately in the scene, and holistic method which investigates the crowd as a whole. The proposed method identifies the abnormal behaviors in public places through applying holistic methods (only pedestrians are presents in these places). Crowd behavior is modeled as a collection of basis; identifying and locating the abnormal behaviors are done by Sparse Coding. General training features are saved in a two- part dictionary, and test movements are analyzed through rebuilding the extracted features from the test video based on the available dictionary which is formed in an unsupervised way using sparse

combinations. High errors on this stage show the lack of suitable rebuilding of the test video based on available behaviors in the dictionary, so the algorithm detects and locates abnormal behaviors. Proposed algorithm is performed on UCSD datasets, ROC curve is calculated and EER values are 0.29 and 0.35 respectively. The results show the ability of the proposed algorithm for real time detection of abnormal behaviors.

[5] proposed a technique for real-time anomaly detection and localization in crowded scenes. Each video is denied as a set of non-overlapping cubic patches, and is described using two local and global descriptors. These descriptors catch the video properties from various perspectives. The local and global features depend on structure similarity between adjacent patches and the features learned in an unsupervised way, using a sparse auto encoder. Their system can detect and localize anomalies as soon as they happen in a video.

[6] developed a method incorporating the most acclaimed histogram of Oriented Gradient and the saliency prediction model Deep Multi-Level Network to detect humans in video sequences. Further they implemented the K-means algorithm to cluster the HOG feature vectors of the positively detected windows and determined the path followed by a person in the video. Their model achieved a detection precision of 83.11% and recall of 41.27%.

[7] proposed a data-driven method based on broken windows theory and spatial analysis crime data using machine mining algorithm. Further they improve the model performance by accumulating results. They visualize potential crime hotspots on a map, and observe whether the model can identify true hotspots.

[8] Tackled the problem of large scale visual place recognition, where the task is to quickly and accurately recognize the location of a given query photograph. They present the following three principal contributions. First, they developed a convolutional neural network (CNN) architecture that is train-able in an end-to-end manner directly for the place recognition task. The main component of this architecture, NetVLAD, is a new generalized VLAD layer, inspired by the Vector of Locally Aggregated Descriptors image representation commonly used in image retrieval. The layer is readily pluggable into any CNN architecture and amenable to training via backpropagation. Second, they developed a training procedure, based on a new weakly supervised ranking loss, to learn parameters of the architecture in an end-to-end manner from images depicting the same places over time downloaded from

Google Street View Time Machine. Finally, they showed that the proposed architecture significantly outperforms non-learnt image representations and off-the-shelf CNN descriptors on two challenging place recognition benchmarks, and improves over current state-of-the-art compact image representation.

[9] Perceiving meaningful activities in a long video sequence is a challenging problem due to ambiguous definition of meaningfulness as well as clutters in the scene. Proposed approach addresses this problem by learning a generative model for regular motion patterns (termed as regularity) using multiple sources with very limited supervision. They propose two methods that are built upon the auto encoders for their ability to work with little to no supervision. They first leverage the conventional handcrafted spatio-temporal local features and learn a fully connected auto encoder on them. Second, they build a fully convolutional feed-forward auto encoder to learn both the local features and the classifiers as an end-to-end learning framework. Their model can capture the regularities from multiple datasets. They evaluate their methods in both qualitative and quantitative ways - showing the learned regularity of videos in various aspects and demonstrating competitive performance on anomaly detection datasets as an application.

[10] Nowadays, with so many surveillance cameras having been installed, the market demand for intelligent violence detection is continuously growing, while it is still a challenging topic in research area. Therefore, proposed method attempts to make some improvements of existing violence detectors. The primary contributions of proposed system are two-fold. Firstly, a novel feature extraction method named Oriented Violent

Table 1
Comparison

Author	Year	Approach	Description
Zhaohui Luo,Weisheng He,Minghui Liwang,Lianfen Huang and Yifeng Zhao	2017	Gaussian Mixture Model	Spatial and Temporal model is used for learning
S.Maryam Masoudirad and Jawad Hadadnia	2017	Sparse Dictionary	Two-part Sparse Dictionary is used for learning
Mohammad Sabokrou, Mahmood Fathy, Mojtaba Hoseini, Reinhard Klette	2015	Localization	Objects are localized to detect anomaly
Waqas Sultani, Chen Chen, Mubarak Shah	2018	Multiple instance learning(MIL)	Model is trained on both normal and abnormal videos
Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.k., Woo, W.c.	2015	Convolutional Long-Short Term Memory(LSTM)	Spatio-temporal patterns are recognized using convolutional LSTM
Patraucean, V., Handa, A., Cipolla, R	2015	Autoencoders	Spatio-temporal autoencoders used for learning
Chong Y.S., Tay Y.H.	2017	Convolutional Long-Short Term Memory(LSTM)cand autoencoders	Spatio-temporal patterns are recognized using convolutional LSTM
Gao, Yuan et al.	2016	AdaBoost and Linear SVM	takes full advantage of the motion magnitude change information in statistical motion orientations,combination and multi-classi er combination strategies are adopted

Flows (OViF), which takes full advantage of the motion magnitude change information in statistical motion orientations, is proposed for practical violence detection in videos. The comparison of OViF and baseline approaches on two public databases demonstrates the efficiency of the proposed method. Secondly, feature combination and multi-classifier combination strategies are adopted and excellent results are obtained. Experimental results show that using combined features with AdaBoost+Linear-SVM achieves improved performance over the state-of-the-art on the Violent-Flows benchmark.

[11] This paper presents a hierarchical framework for detecting local and global anomalies via hierarchical feature representation and Gaussian process regression (GPR) which is fully non-parametric and robust to the noisy training data, and supports sparse features. They have focused on global anomalies that involve multiple normal events interacting in an unusual manner, such as car accidents. To simultaneously detect local and global anomalies, they cast the extraction of normal interactions from the training videos as a problem of finding the frequent geometric relations of the nearby sparse spatio-temporal interest points (STIPs). A codebook of interaction templates is then constructed and modeled using the GPR, based on which a novel inference method for computing the likelihood of an observed interaction is also developed.

4. Conclusion

In this paper, before coming up with new solutions, it is necessary to survey state of the art techniques for learning purposes for video anomaly detection. We covered multiple aspects of research problem of anomaly detection, stated various anomaly types. We surveyed various available techniques, compared various anomaly detection systems, techniques and approaches and gave comparative analysis on them.

References

[1] Chong Y.S., Tay Y.H. (2017) Abnormal Event Detection in Videos Using Spatiotemporal Autoencoder. In: Cong F., Leung A., Wei Q. (eds) Advances in Neural Networks - Lecture Notes in Computer Science, vol 10262. Springer, Cham

[2] Sultani, Waqas et al. Real-World Anomaly Detection in Surveillance Videos. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (2018): 6479-6488.

[3] Luo, Zhaohui et al. Real-time detection algorithm of abnormal behavior in crowds based on Gaussian mixture model. 2017 12th International Conference on Computer Science and Education (ICCSE) (2017): 183-187.

[4] Masoudirad, S. Maryam and Jawad Hadadnia. Anomaly detection in video using two-part sparse dictionary in 170 FPS. 2017 3rd Interna-tional Conference on Pattern Recognition and Image Analysis (IPRIA) (2017): 133-139.

[5] Sabokrou, Mohammad Fathy, Mahmood Mojtaba, H Klette, Reinhard. (2015). Real-Time Anomaly Detection and Localization in Crowded Scenes.

[6] Gajjar, Vandit et al. Human Detection and Tracking for Video Surveillance: A Cognitive Science Approach. 2017 IEEE International Conference on Computer Vision Workshops (ICCVW) (2017): 2805-2809.

[7] Lin, Ying-Lung et al. Using Machine Learning to Assist Crime Prevention. 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI) (2017): 1029-1030.

[8] Arandjelovi, Relja Gronat, Petr Torii, Akihiko Pajdla, Tomas Sivic, Josef. (2015). NetVLAD: CNN architecture for weakly supervised place recognition.

[9] Hasan, M., Choi, J., Neumann, J., Roy-Chowdhury, A.K., Davis, L.S.: Learning temporal regularity in video sequences. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 733-742, June 2016.

[10] Gao, Yuan et al. Violence detection using Oriented VIolent Flows. Image Vision Comput. 48-49 (2016): 37-41.

[11] Cheng, Kai-Wen et.al. Gaussian Process Regression-Based Video Anomaly Detection and Localization with Hierarchical Feature Representation. IEEE Transactions on Image Processing 24 (2015): 5288-5301.

[12] A. Sodemann, M. P. Ross, and B. J. Borghetti, A review of anomaly detection in automated surveillance, Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, vol. 42, no. 6, pp. 1257-1272, 2012.

[13] Kratz, Louis and Ko Nishino. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. 2009 IEEE Conference on Computer Vision and Pattern Recognition (2009): 1446-1453.

[14] S. Yi, H. Li, and X. Wang, Understanding pedestrian behaviors from stationary crowd groups, 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 34883496, June 2015.

[15] R. A. A. Rupasinghe, S. G. M. P. Senanayake, D. A. Padmasiri, M. P. B. Ekanayake, G. M. R. I. Godaliyadda, and J. V. Wijayakulasooriya, Modes of clustering for motion pattern analysis in video surveillance, in 2016 IEEE International Conference on Information and Automation for Sustainability (ICIAfS), Dec 2016, pp. 16.

[16] S. Andrews, I. Tsochantaridis, and T. Hofmann. Support vector machines for multiple-instance learning. In NIPS, pages 577584, Cambridge, MA, USA, 2002. MIT Press.

[17] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic. NetVLAD: CNN architecture for weakly supervised place recognition. In CVPR, 2016.

[18] A. Gordo, J. Almazan, J. Revaud, and D. Larlus. Deep image retrieval: Learning global representations for image search. In ECCV, 2016.

[19] M. J. Roshtkhari, and M. D. Levine, An on-line, real-time learning method for detecting anomalies in videos using spatiotemporal compositions, Computer Vision and Image Understanding, vol. 117, no. 10, pp. 1436-1452, 2013.

[20] W. Hu, T. Tan, L. Wang, and S. Maybank, A survey on visual surveillance of object motion and behaviors, Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, vol. 34, no. 3, pp. 334-352, 2004.

[21] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma, Toward a practical face recognition system: Robust alignment and illumination by sparse representation, Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 34, no. 2, pp. 372-386, 2012.

[22] Mahadevan, Vijay et al. Anomaly detection in crowded scenes. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2010): 1975-1981.

[23] Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.k., Woo, W.c.: Convolutional lstm network: A machine learning approach for precipitation nowcasting. In: Proceedings of the 28th International Conference on Neural Information Processing Systems. pp. 802-810. NIPS'15, MIT Press, Cambridge, MA, USA (2015).

[24] Patraucean, V., Handa, A., Cipolla, R.: Spatio-temporal video autoencoder with differentiable memory. International Conference On Learning Representations (2015), 1-10, 2016.

[25] Cong, Y., Yuan, J., Liu, J.: Sparse reconstruction cost for abnormal event detection. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 3449-3456, 2011.

[26] https://iwringer.wordpress.com/2015/11/17/anomaly-detection-concepts-and-techniques/