Detecting Abnormal Behavior in Examination Surveillance with 3D Convolutional Neural Networks

Manish Ravindra Borse[1], Vijay Rathi[2], Prasad Gajanan Pawar[3], Piyush Dilip Mirase[4], Yogesh Yashwant Paste[5]

[1,3,4,5]*Student, Department of Computer Engineering, Parvatibai Genba Moze College of Engg., Pune, India*
[2]*Professor, Department of Computer Engineering, Parvatibai Genba Moze College of Engg., Pune, India*

***Abstract*: Video analytics is the method of processing a video, gathering data and analysing the data for getting domain specific information. In the current trend, besides analyzing any video for information retrieval, analyzing live surveillance videos for detecting activities that take place in its coverage area has become more important. Such systems will be implemented real time. Automated face recognition from surveillance videos becomes easier while using a training model such as Artificial Neural Network. Hand detection is assisted by skin color estimation. This research work aims to detect suspicious activities such as object exchange, entry of a new person, peeping into other's answer sheet and person exchange from the video captured by a surveillance camera during examinations. Nowadays, people pay more attention to fairness of examination, so it is meaningful to detect abnormal behavior to ensure the order of examination. Most current methods propose models for particular cheating behavior. In this system, we extract the optical flow of video data and propose a 3D convolution neural networks model to deal with the problem. This requires the process of face recognition, hand recognition and detecting the contact between the face and hands of the same person and that among different persons. Automation of 'suspicious activity detection' will help decrease error rate due to manual monitoring.**

***Keywords*: 3D convolution networks, Optical flow, Abnormal behavior, Examination surveillance video.**

1. Introduction

The human face and human behavioral pattern play a significant role in person identification. Perceptible information is a key source for such identifications. Surveillance videos provide such visual information which can be viewed as live videos, or it can be played back for future references. The recent trend of 'automation' has its impact uniform in the field of video analytics. Video analytics is used for different applications such as motion detection, human activity prediction, person identification, abnormal activity recognition, people counting at congest places, etc. In this domain, the two factors which are used for person identification are technically termed as face recognition and gait recognition respectively. In number of techniques these two techniques, face recognition is more versatile for automated person identification through surveillance videos. Face recognition can be used to predict the orientation of a person's head, which in turn will help to predict a person's behavior. Motion recognition with face recognition is very helpful in many applications such as validation of a person, identification of a person and detecting presence or absence of a person at particular place and time. In addition, human interfaces such as fine contact among two entities, hand gesture recognition, head motion detection and estimate are used to devise a system that can identify and recognize doubtful behavior among pupil in an examination hall successfully. This approach provides a methodology for doubtful human activity detection through face recognition. Video processing is used in two main domains such as safety and research. Such a technology uses intelligent algorithms to monitor live videos. Computational complexities and time complexities are some of the key factors while designing a real-time system. The system which uses an algorithm with a relatively lower time complexity, using some amount of hardware resources and which produces good results will be more useful for time-critical applications like bank robbery detection, patient monitoring system, detecting and reporting suspicious activities at the railway station, exam holes etc.

2. Literature Review

C. Lu, J. Shi, and J. Jia, "Abnormal event detection at 150 fps in matlab", in Proceedings of the IEEE international conference on computer vision, 2013.

In this paper propose an efficient sparse combination learning frame- work. It achieves decent performance in the detection phase without compromising result quality. The short running time is guaranteed because the new method effectively turns the original complicated problem to one in which only a few costless small-scale least square optimization steps are involved. Our method reaches high detection rates on benchmark datasets at a speed of 140∼150 frames per second on average when computing on an ordinary desktop PC using MATLAB.

M. Sabokrou, M. Fathy, M. Hoseini, and R. Klette, "Real-time anomaly detection and localization in crowded scenes," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2015.

In this paper distinguish normal activities and anomalies in videos. The local and global features are based on structure similarity between adjacent patches and the features learned in an unsupervised way, using a sparse auto encoder. Experimental results show that our algorithm is comparable to a state-of-the-art procedure on UCSD ped2 and UMN benchmarks, but even more time-efficient. The experiments confirm that our system can reliably detect and localize anomalies as soon as they happen in a video.

M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis, "Learning temporal regularity in video sequences," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

In this project propose two methods that are built upon the auto encoders for their ability to work with little to no supervision. first leverage the conventional handcrafted spatio-temporal local features and learn a fully connected auto encoder on them. Second, we build a fully convolutional feed-forward auto encoder to learn both the local features and the classifiers as an end-to-end learning framework. Our model can capture the regularities from multiple datasets. We evaluate our 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.

B. Zhao, L. Fei-Fei, and E. P. Xing, "Online detection of unusual events in videos via dynamic sparse coding," in Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011.

In this paper propose a fully unsupervised dynamic sparse coding approach for detecting unusual events in videos based on online sparse reconstructibility of query signals from an atomically learned event dictionary, which forms a sparse coding bases. Based on an intuition that usual events in a video are more likely to be constructible from an event dictionary, whereas unusual events are not, our algorithm employs a principled convex optimization formulation that allows both a sparse reconstruction code, and an online dictionary to be jointly inferred and updated. Our algorithm is completely unsupervised, making no prior assumptions of what unusual events may look like and the settings of the cameras. The fact that the bases dictionary is updated in an online fashion as the algorithm observes more data, avoids any issues with concept drift. Experimental results on hours of real world surveillance video and several YouTube videos show that the proposed algorithm could reliably locate the unusual events in the video sequence, outperforming the current state-of the art methods.

W. Luo, W. Liu, and S. Gao, "A revisit of sparse coding based anomaly detection in stacked RNN framework," in The IEEE International Conference on Computer Vision (ICCV), Oct. 2017.

In this paper propose a TSC, which can be mapped to a sRNN which facilitates the parameter optimization and accelerates the anomaly prediction. ii) they build a very large dataset which is even larger than the summation of all existing dataset for anomaly detection in terms of both the volume of data and the diversity of scenes. Extensive experiments on both a toy dataset and real datasets demonstrate that our TSC based and sRNN based method consistently outperform existing methods, which validates the effectiveness of our method.

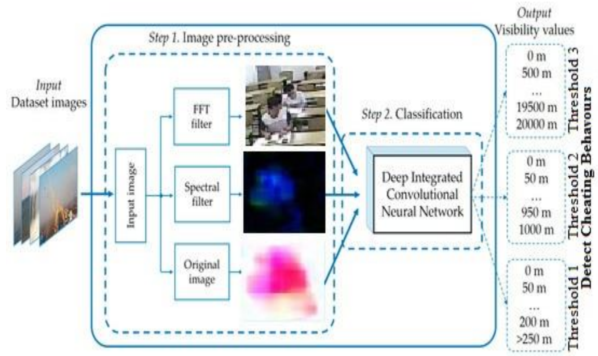

Fig. 1. System architecture

3. Mathematical model

Let S be the Whole system which consists:
S= {IP, V, PRO,OP}.
Where,
IP is the input of the system.
Pro is the procedure applied to the system to process the given input.
OP is the output of the system.

A. Surveillance video

V= Is the Input Video
Input: {Video dataset}
Output: {Video convert into module}

I. Shows the architecture of our 3D CNN model.
The model has convert convolution layers,
1. Pooling layers,
2. Fully-connected layers
3. Softmax layer.
Apply filters for 2 convolution layers are 64 and 128.
Apply algorithm: CNN model is a binary classification

B. Video strumming

Input: {video}
Output: {Abnormal Activity Identification}
Significant to detect abnormal behavior in examination surveillance video. If identify the abnormal behaviour it will be marked in corresponding testing clip by red box.

C. Save video

4. Conclusion

We propose a unified deep learning based framework for abnormal event detection from exam hall. The proposed system consists of three blocks which are designed to achieve three keys of anomaly detection in neural networks. In short, the motion fusion block is designed to keep the temporal and spatial connection between the motion and appearance cues. The feature transfer block is used to extract discriminative features by exploiting the transferability of the neural network from different tasks/domains. The coding block is a novel LSTM to achieve fast sparse coding, which could enjoy fast inference and end-to-end learning. Extensive experiments show the promising performance of our method in image reconstruction and abnormal events detection in surveillance.

Acknowledgment

I would prefer to give thanks the researchers likewise publishers for creating their resources available. I'm conjointly grateful to guide, reviewer for their valuable suggestions and also thank the college authorities for providing the required infrastructure and support.

References

[1] Y. Cong, J. Yuan and J. Liu, "Sparse reconstruction cost for abnormal event detection," IEEE Conference on Computer Vision and Pattern Recognition, pp. 3449-3456, 2011.
[2] L. Kratz and K. Nishino, "Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models," IEEE Conference on Computer Vision and Pattern Recognition, pp. 1446-1453, 2009.
[3] R. Mehran, A. Oyama and M. Shah, "Abnormal crowd behavior detection using social force model," IEEE Conference on Computer Vision and Pattern Recognition, pp. 935-942, 2009.
[4] Y. B. Wang, "An intelligent algorithm for detecting cheating behavior at the exam site," Chang Sha: National University of Defense Technology, 2011.
[5] J. Li, "Detection and analysis of abnormal behavior in electronic invigilating," Tai Yuan: Taiyuan University of Technology, 2013.
[6] D. Gowsikhaa, Manjunath and S. Abirami, "Suspicious human activity detection from surveillance videos," International Journal on Internet and Distributed Computing Systems, vol. 2, no. 2, 2012.
[7] S. Xiong, "Research of SVM-based abnormal behavior detection in electronic invigilated examination room," Journal of Hubei University of Education, vol. 30, no. 8, pp. 42-46, 2013.
[8] S. W. Ji, W. Xu, M. Yang and K. Yu, "3D convolutional neural networks for human action recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 1, pp. 221-231, 2013.
[9] W. Z. Nie, W. Li, A. A. Liu, T. Hao and Y. T. Su, "3D convolutional networks-based mitotic event detection in time-lapse phase contrast microscopy image sequences of stem cell populations," IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 55-62, 2016.
[10] T. Du, L. Bourdev, R. Fergus, L. Torresani and M. Paluri, "Learning spatiotemporal features with 3D convolutional networks," IEEE International Conference on Computer Vision and Pattern Recognition, 2015.
[11] G. Farneback, "Two-frame motion estimation based on polynomial expansion," Scandinavian Conference on Image Analysis, pp. 363-370, 2003.
[12] J. Y. Ng, M. Hausknecht, S. Vijayanarasimhan, et al, "Beyond short snippets: Deep networks for video classification," IEEE Conference on Computer Vision and Pattern Recognition, pp. 1613-1631, 2015.
[13] S. Baker, D. Scharstein, J. P. Lewis, et al, "A database and evaluation methodology for optical flow," International Journal of Computer Vision, vol. 92, no. 1, pp. 1-31, 2007.