International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 8, Issue 9, September 2019)

Video Anomaly Detection using Ensemble Learning.

Prof.Vina Lomte (H.O.D)[1], Durgesh Pahurkar[2], Siddheshwar Patil[3], Siddharth Patil[4], Satish Singh[5]

[1,2,3,4,5]*Department of Computer Engineering RMD Sinhgad School of Engineering (Savitribai Phule Pune University) Pune, India*

comphod.rmdssoe@sinhgad.edu[1],durgesh.ps1997@gmail.com[2],siddheshpatil223@gmail.com[3], patilsiddharth17@gmail.com[4],sattish1996@gmail.com[5]

***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. Anomaly detection is a method of identifying an abnormal activity through the live surveillance video. Our proposed system uses Sparse dictionary and auto- encoders for detecting anomaly activities. The model uses Bayes classifier to detect the type of abnormal activity occurred in live surveillance. Ensemble learning is used to enhance the system by combining decisions of Sparse Dictionary and auto-encoders (with convolutional LSTM).**

***Index Terms*—Ensemble learning, Sparse Dictionary, auto-encoders, anomaly detection.**

I. 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 abnor- malities 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.

Abnormal behavior contains the issues like overcrowding of people, noise, happening of none structured events, etc. The hardest task while real time processing is tracking a dynamic environment that has multiple moving entities. The first model used is sparse dictionary learning for detecting outlier frame by frame.

Sparse dictionary learning has been successfully implied in image, video processing tasks and the other model used is Convolutional long short term memory(ConvLSTM) which is type of artificial neural network that uses spatial and temporal data for outlier detection. The result of both these model is given to ensemble leaning which is used for multiple learning algorithm to obtain better predictive performance. In this process the multiple models used are strategically generated and combined to solve a problem. The result of both these model is combined and given to the Bayesian model. Bayesian network classifier works on the classification and the classified anomaly is given as the result. In order to perform this as project, two methods are combined and classification is done by the classifier.

II. Related Work

Video anomaly detection is a field which is getting lot of attention these days due to increasing crime rates and avail- ability of video surveillance for most parts of the city. Many researchers have come up with working models which despite of having some advantages are not suitable for real world application due to their high learning complexity and unac- ceptable error rates.

[1] Proposed method presents a spatiotemporal architecture where convolutional neural networks are used for learning spa- tial features and dynamical changes in those spatial features. They have achieved 140 fps on USCD, Subway and Avenue datasets.

[2] Have Proposed a method that uses both normal and abnormal video for training using Multiple Instance Learn- ing(MIL). Training generates an appropriate anomaly score which helps them to classify video bags(segments). The authors have used their own dataset which has collection of both normal and abnormal events and is 1900 hours long. They are the only ones who have further classified abnormal events into 13 classes.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 8, Issue 9, September 2019)

[3] In this paper, authors have considered new feature for detecting motion called Sensitive Movement Point(SMP). Video is analyzed using Gaussian Mixture Model(GMM) to find SMP and then SMP is further analyzed on models based on temporal and spatial characteristics. It is also suitable for online learning. They used UMN Dataset for experiments.

[4] Suggested that we can track every person or treat whole crowd as a singular entity for anomaly detection. The proposed method treats whole crowd as a singular entity; identifies and locates abnormal events using Sparse Coding. A Two-part Sparse Dictionary is trained using only normal videos, for finding abnormal videos test videos are reconstructed using learnt dictionary. UCSD dataset is used and EER values are found to be 0.35 and ROC curve gives value of 0.29.

[5] proposed a technique for localization of anomaly with anomaly detection. Videos are considered as set of non-overlapping cubic patches and are given two descriptors, global and local which are used to catch video properties. Captured properties are used to identify videos containing abnormal patches. Their system can detect and localize anomalies as soon as they happen in a video.

III. Our Contribution

Our contribution to this field are as follows:

1) We propose an ensemble learning approach to anomaly detection problems which combines two or more classi- fication models for combining their advantages and thus giving better results than all other previous methods.
2) We try to find the type of anomaly occurred based on reconstruction values above threshold,this will help in prioritizing and planning actions against various type of crime or abnormal situations.

IV. Proposed System

A. Problem Statement

To process real time video and generate alerts(notifications) when some predefined type of anomaly is generated.

B. System Architecture

In our proposed system, we provide input video frame to both of the models and they process data on their own giving probabilities of event being normal or abnormal.

Sparse learning encodes the input image frame into less complex image frame and stores the learned matrix as dictionary instance while training. We build a dictionary of only normal videos while training. We are using auto-encoder for reduc- ing dimensions of the input video frame while processing to require less computational resources. While training we generate values for reconstruction threshold of Convolutional LSTM (Long Short term memory) for normal and abnormal events. We build up a temporal library (history library) to train on both normal and abnormal events. The output of both our individual models help us in generating a more general sense of abnormal and normal events using an ensemble learning technique. The output value if above threshold decided for normal events is passed onto Bayesian classifier for getting a class of anomaly occurring. Anomaly classes are generated from output combinations of both models above threshold value.

During testing and operation phase we provide an unknown video sequence to the model and it alarms the system user of any anomaly occurring.

Figure 1 shows the architecture of this system.

C. Mathematical Model

1) *Convolutional LSTM:* [1] We are using ConvLSTM architecture that is well proven for anomaly detection, where all the inputs x1, . . ., xt, cell outputs c1, . . ., ct, hidden states h1, . . ., ht, and gates it, ft, ot of the ConvLSTM are 3D tensors whose last two dimensions are spatial dimensions.

The formulation of the ConvLSTM unit can be summarized with (1) through (5).

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci} \sim c_{t1} + b_i) \quad (1)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf} \sim c_{t1} + b_f) \quad (2)$$

$$c_t = f_t \sim c_{t-1} + i_t \sim tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (3)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co} \sim c_t + b_o) \quad (4)$$

$$h_t = o_t \sim tanh(c_t) \quad (5)$$

Model is given images as input and it modifies weights as convolutional filters (the symbol ~ denotes a convolution operation).

Sparse Dictionary Learning: [4] We assume our data X satisfies

$$X \approx \sum_{i=1} a_i D_i = aD \quad (6)$$

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 8, Issue 9, September 2019)

We learn a dictionary X, using our input video frames using their sparse representation.

$$Xj; j \in (1, m)$$

Learn dictionary D and sparse code a.

Once we learn dictionary for normal data we try to recon- struct the test video using the dictionary. If the reconstruction error exceeds the set threshold, then the frame is flagged as abnormal i.e. containing anomaly.

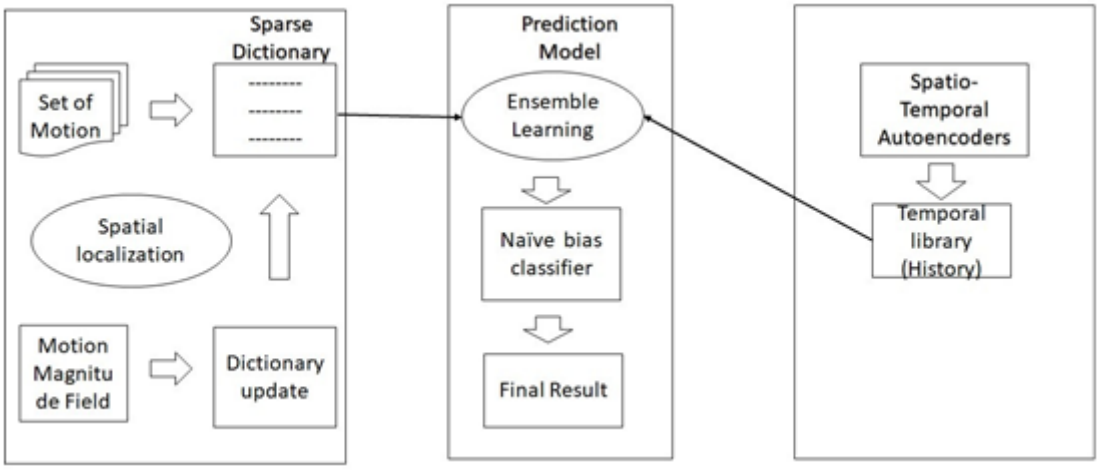

Fig. 1. System architecture

D. Algorithm

input: Test Frames $f_t \; t\{.\}$ threshold 'N' & 'A'.

output: Final frame with anomaly Detected(Y)

1) Sparse Learning

```
for t←−1 do
    for MMF ∈ SOM do
        MMF←− sparseDictionaryLearning(0,1)
        MOS←−reconstruct(MMF)
        if MMF and MOS = SDL(1)
            for MMF ∈ SDL do
            Y(A)    ←− 1
            end for
        else
            for MMF ∈ SDL do
                Y(A) ←− 0 ... (N −→Normal)
            end for
        end if
    end for
end for
```

2) *Spatio*-Temporal Auto Encoders

1 Encode using spatial encoder
2 Encode using ConvLSTM(Temporal encoder)
3 Decode using COnvLSTM(Temporal Decoder)
4 Decode using spatial decoder
5 generate output state(abnormal, normal)

3) feed output of 1 and 2 in Ensemble learning
4) feed output of ensemble learning to bayesian classifier
5) output type of anomaly with Notification

V. Results

As our approach uses two models simultaneously it is resource intensive but what is good about it is that it gives more accuracy by using them both. Further this method was only trained using low computational resources (Intel i5 6th Gen processor and 8 GB RAM), If this model is trained on a better system it will give more accuracy (as is the case with any machine learning model up to some extent). We currently do not know to what extent the accuracy can be extended, but the results will definitely be better than the current one.

As we can see from the below graph that as we keep increasing the number of epochs(iteration) for training our ConvLSTM model, we get the better accuracy. The more the number of epochs, the better the accuracy we get. And the accuracy is found to be near around 73%.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 8, Issue 9, September 2019)

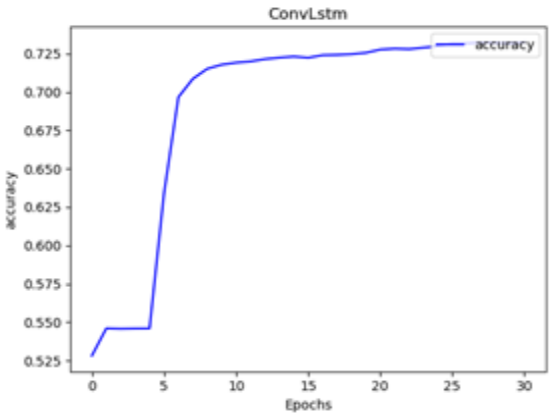

Fig. 2. Accuracy

As we can see from the below graph that as we keep increasing the number of epochs (iteration) for training our ConvLSTM model, the loss of some important features gets decreased.

VI. Conclusion And Future Applications

Our method being a combination of speed (Sparse learning) and accuracy (ConvLSTM) is expected to perform better at anomaly detection and will be more suitable for real-time application such as live surveillance and crime detection.

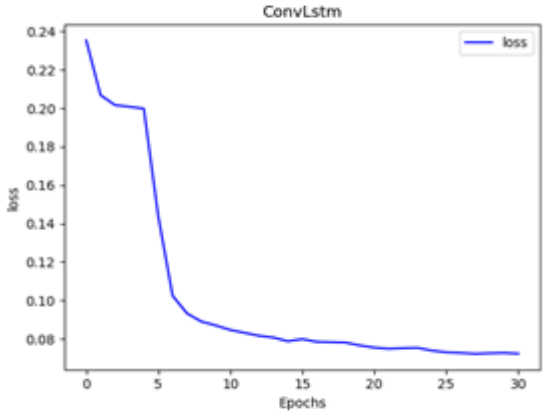

Fig. 3. Loss

This system can be further integrated with audio sensor to increase accuracy or confirm results of video processing sys- tem. We can further collaborate with government and integrate this system with public surveillance to detect public crimes and get faster crime response than what it is currently. This system can also help private organizations and stores to monitor large amount of employees/people.

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 - ISNN 2017. ISNN 2017. 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 International 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. 10.1109/CVPRW.2015.7301284.

[6] Gajjar, Vandit et al. Human Detection and Tracking for Video Surveil- lance: A Cognitive Science Approach. 2017 IEEE International Confer- ence 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 Confer- ence 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 Confer- ence 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] https://iwringer.wordpress.com/2015/11/17/anomaly-detection-concepts- and-techniques/