International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249 – 8958 (Online), Volume-9 Issue-4, April 2020

A Novel Framework for Anomaly Detection in Video Surveillance using Convolutional LSTM

Lovleen siddhu, Ranganathan Sridhar

Abstract*: Today, due to public safety requirements, surveillance systems have gained increased attention. Video data processing technologies such as the identification of activity [1], object tracking [2], crowd counting [3], and the detection of anomalies [4] have therefore been rapidly developing. In this study, we establish an unattended method for the detection of anomaly events in videos based on a ConvLSTM encoder-decoder to learn about the evolution of spatial characteristics. Our model only covers typical video events during preparation, whereas in testing the videos are both usual and abnormal. Experiments on the UCSD datasets confirm the validity of the suggested approach to abnormal event detection.

***Keywords*: Anomaly event detection, Autoencoder, LSTM, UCSD dataset, Convolutional neural networks.**

I. INTRODUCTION

The need for automatic surveillance systems is now increasing because of increased safety concerns. Advances in technology and lower costs culminated in the rapid deployment in both public and private buildings of surveillance cameras. Traditionally, human operators are responsible for monitoring video feeds from multiple cameras that need visual inspection simultaneously. Nonetheless, after long periods of gazing on display screens, even committed staff has been affected by diminished visual attention. [1].

The classical research topic of computer vision or related fields in so many activities, from domestic security, medical diagnostics or military and industrial operations, to picture sequence abnormality.[2], [3]. The growing number of applications and the need for accurate results increase demand for alternative solutions that are less dependent on people. Besides being a commonly known problem, the automatic detection of anomalies is still a challenging and challenging issue because of several complex problems, such as camera position, lighting, shadows, occlusions, weather conditions, camera jitter, etc. [4].

Revised Manuscript Received on March 16, 2020.
* Correspondence Author
Lovleen Siddhu*, computer science department, Vellore Institute of Technology (VIT) University, Raipur, Chhattisgarh. Email: lovleensiddhu@gmail.com
Ranganathan Sridhar, school of computer science and engineering (SCOPE), Vellore Institute of Technology (VIT) University, Chennai, India. Email: Sridhar@vit.ac.in

Additional cameras should be used to fix several occlusions in several monitoring activities. Some activities are usually multiple perspectives for proper inspection, particularly in disastrous environments like industrial plants and offshore petroleum platforms [5]. In dangerous conditions and areas where access is difficult, such a need is even greater [6], [7], [8]. Collective no. of cameras often increases the amount that can be monitored in large installations.

The use of a single movable camera is an interesting approach to dealing with this issue. In action, on a moving platform (e.g. car, robot, or drone), a traditional Camera is mountable that carries a camera along a predefined trajectory to the desired position. This method will reduce dramatically the number of cameras needed but enables monitoring of the range of different viewing points at the same time as automating repeat inspections. Together with extensive use of portable cameras, these factors stimulate interest in monitoring & background / earth separation problems through movable cameras [6], [9], [10], [11], [12].

There are primarily two issues with anomaly event identification. Firstly, it is rare (i.e. with low probability) for an irregular occurrence to occur. Furthermore, it is not very easy to describe abnormality. For example, someone who runs in a bank can be perceived to be an anomalous occurrence, but he or she runs in the Park like a normal event. However, video data modeling is a challenge because of its high-dimensionality, noise, large events or interaction properties.

The literature has extensively studied understanding of human activity in this sense. In this field, most methods specifically model such events a priori and therefore their implementation is restricted, typically in controlled scenarios, to the detection of such events. For instance, abandoned object detection or violation of prohibited areas is an example of event detection. More specifically, the emphasis on anomaly detection has been increased despite specific modeling. a fact which video surveillance incidents are difficult to guess yet pose a small incident rate lies behind the rise of the interest.2[13], In noisy situations in particular. For these purposes, it became a research topic3[14] to identify these events. Intuitively, in a video sequence, we can describe an anomaly as an action which separates time and space from its context 4[15]. This allows an anomaly detection statistical method by treating anomalies as low probability occurrences about a normal behavior statistical model. A key element, in this case, is the identification of irregularities involving complex Spatio-temporal interaction between artifacts in the scene.

Anomalies in the points and conceptual anomalies can be listed as current approaches [14]. The former uses a certain position (size, speed) with the salience of an attribute.

Later, information from a temporal (trajectory) or spatial (nearly observed) background is included by use of features such as pixel shift rate 5[16], foreground pixel filler ratio 6[17], optical flow (optical flow) object dimensions and motion vectors 7[18], and foreground data 8[19].

This paper deals with the question of detecting changes in video sequences taken from camera configurations of this kind. The approach suggested breaks down potentially anomalous video targets into a small combination of the frames in an anomalous video without reference plus a small residual, referring to potential interest anomalies. The basic assumption that the algorithm will be successful is that the position and trajectory of the camera during objective and reference video processing are identical to the data in the frames of the target film. Under such conditions, a linear convex optimization can be used to decompose target frames as a sparse combination of reference frames. However, the camera shakes as it moves in real-world scenarios, and thus its location and trajectory can change about the recording of video reference. One way to solve this issue is an additional non-linear domain shift definition that involves an iterative linearization approach. This transition of the domain allows the system to better match reference as well as a target video frame, creating less false detection for the algorithm

II. LITERATURE REVIEW

The potential applications in different fields [20], like vision detection and control, human-computer interfaces, content-based video study or behavioral biometrics, have brought significant attention to identification and comprehension of human activity in Video. In the last few years, anomalous events have been observed in the change in paradigm. Here we discuss concepts and methods for identification of phenomena in a video for short.

Notwithstanding the variety of approaches and implementations, the concept of anomalies is universal. Anomalies were identified in literature as' unknowns'[21]," anomalous occurrences'[22],' anomaly'[15],' suspects of action' or' anomalies'[24]. Intuitively, an anomaly is a trend in each sense that is not consistent with normal behavior [4]. Pattern recognition methods, in video analysis, are often the preferred way to identify or recognize events by modeling specific training data events. Detection is carried out through the interaction between new patterns and those already educated. In comparison, a model of usual course is perceived or anomalies are detected as nonconformities from this pattern in anomalous event detection.

We may differentiate between point anomalies and contextual abnormalities depending on the contextual extent at that anomalies are taken into account[15]. The first suggests that at a certain spot, the value of the extracted features differs significantly from the normal value. Sometimes, they mean a particular location (e.g. height, speed) for an attribute. It includes information from the temporal (trajectory) or spatial (near observations) background. Anomalies that understand the time background, also known as sequential anomalies, analyze inconsistencies in the time series of a particular feature removed. Most works in this area detect anomalous trajectories, as inferred by object tracking algorithms, shown by moving objects. Nonetheless, due to the current difficulties in object tracking, they are not good at crowded environments, so that they are not suitable for unlimited scenarios.

Approaches dealing with features derived directly on the pixel level are more suited for use unconditionally in the infinite number of objects than with object abstractions. Such techniques, however, are largely restricted to the detection of salience and are not therefore appropriate for detecting contextual anomalies. Functions collected include frequency of pixel change [16], front-end pixel filling ratio[17], object sizes and vectors of motion (optical flow).

In [19], a descriptor for object sized is proposed to detect anomalous motion effectively, regardless of the form of item. Such methods also divide the image into blocks at the extraction level of a feature [18] [24] or calculate features in each pixel from a fifth. The detected objects may be constrained by a set block or neighborhood size since it is responsible for both video and normal object size. Such methods cannot, therefore, be tailored to situations that shift objects of various sizes. In [24], the authors use blocks for various video resolutions in different blocks.

III. ANOMALY DETECTION IN VIDEOS

Imagine we've got thousands of security cameras running all the time, some in remote places or in streets where anything dangerous is unlikely to take place, others in busy streets or city squares. A wide range of abnormal events that take place even at one location; from place to place and from time to time the concept of the abnormal event varies.

It is highly desirable to use automated systems to detect unusual events in this scenario and to improve safety and broader supervision. In general, it is a difficult task to identify suspicious events in videos This also has wide applications across the vertical industry and, most recently, has become one of the main tasks of video analysis and is currently attracting great interest from researchers. A rapid and accurate approach to anomaly detection in real-world applications is very much needed.

IV. PROPOSED METHODOLOGY

Why don't we use a Supervised Learning Technique for Anomaly Detection?

If we crave to consider the problem as a matter of Binary classification is required, and in this case, it is difficult to collect labeled data for the following reasons:

- Anomalous occurrences are difficult to achieve because of their rareness.
- There is a wide range of anomalous incidents, which requires a lot of workplace manual identification and marking.

The above factors have promoted the need to use unattended or semi-monitored means such as dictionary learning, space-time, and self-encoding. And self-encoders. nothing like regulated methods,

unregulated video footage containing little or no suspicious incidents that are readily available for real-time applications need only these methods.

A. Autoencoders

Autoencoders Neural networks are trained to reconstruct the data. There are two autoencoders parts:

- The encoder: Capable of learning efficient Input data representations (x) is called f(x) encoding. The bottleneck which contains the entry is the final layer of the encoder representation f(x).
- The decoder: reconstructs the data r = g(f(x)) by means of the bottleneck encoding.

The reconstruction will be used for training the proposed model used in this research.

B. Training the Model

The training set consists of regular sequences of video frames and the model is trained in the reconstruction of these sequences. Let us, therefore, get the data ready to feed our model in these three phases.:

- Divide the training video frames into temporal sequences, each of size 10 using the sliding window technique.
- To ensure that each frame has the same resolution, resize each frame to 256 x 256.
- Scale pixels values between 0 & 1 by dividing each pixel by 256.

C. Proposed Model

In this research, we have used Neural Network-based Convolutional LSTM.

A special kind of RNN that can learn long-term dependencies–generally known only as "LSTM" –is a short-term memory network. We were developed by Hoch Reiter & Schmid Huber team (1997) and have been improved and popularized by a significant number of employees in subsequent work1.

LSTMs have been planned specifically to prevent the problem of long-term dependency. The long-term reminder of knowledge is your default actions, not what you are trying to learn!

All recurrent neural networks are formed by a chain of neural network repeat modules. This module has a simple system for regular RNNs.

LSTM as a unique RNN structure is stable and secure for the conservation of long-range dependencies for general sequence models.

Here we use Convolutional LSTM layers rather than fully connected LSTM layers because FC-LSTM layers do not keep the spatial data very well Due to the use of both input-to-state & state-to-state links where no spatial data is encoded.

CNN LSTM architecture uses CNN layers for extracting input data in conjunction with LSTMs to assist the sequence prediction. CNN LSTM architecture.

CNN LSTMs have been developed to help predict visual time series problems and to apply textual descriptions from image sequences (e.g. videos). The problems of in particular:

- *Activity Recognition:* Generate a textual operation summary shown in the picture series.
- *Image Description:* Generate a single picture text summary.
- *Video Description:* Generate a textual image series summary.

"[CNN LSTMs are] a highly space-based and time-based models class with the flexibility to perform several sequential inputs to output visibility tasks."

D. Testing the Model

Each video will be checked separately. 34 research videos are available from the UCSD dataset. Each testing video has 200 frames. We use the window slider approach to obtain all 10 frame sequences consecutively. In other words, regularity of the Sr(t) series starting from frame(t) & ending at frame(t) is determined for every t between 0 & 190 (t+9).

We measure reconstruction error of a value of a pixel I in video's frame with the L2-standard (x, y):

$$e(x,y,t) = ||I(x,y,t) - f_w(x,y,t))||$$

Wherever f_w is an autoencoder trained developed by the LSTM. We measure than the frame t reconstruction error by summarizing all the errors in pixels:

$$e(t) = \Sigma_{(x,y)}\, e(x,y,t))$$

The cost of rebuilding a ten-frame sequence which begins at t can be calculated as follows:

$$Sequence\_reconstrution\_cost(t) = \Sigma \sum_{t'=t}^{t+10} e(t')$$

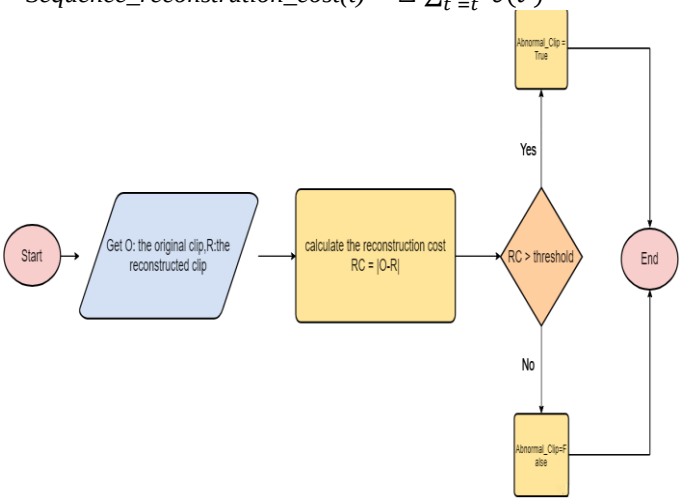

Fig. 1. Flow Diagram of the research Methodology.

We then quantify abnormality by scaling Sa(t) between 0 & 1.

$$S_at = \frac{Sequence\_reconstrution\_cost(t) - Sequence\_reconstrution\_cost(t)_{min}}{Sequence\_reconstrution\_cost(t)_{min}}$$

We can derive regularity score Sr(t) by subtracting abnormality scores from **1**.

$$S_r(t) = 1 - s_a(t)$$

After we compute the regularity score Sr(t) for each **t** in range [0,190], we draw Sr(t).

V. RESULTS AND DISCUSSIONS

Let us understand with an example, at the beginning of any video, there is a walkway bicycle this explains the low regularity rating. After the left bike, the score increases regularly. This means that another wheel in frame 60 again reduces the regularity rate and increases right after it is left.

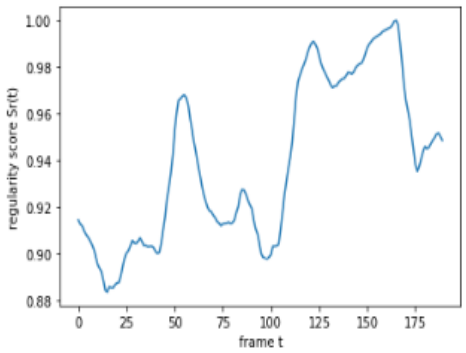

Figure 2 shows the regularity score of the given dataset

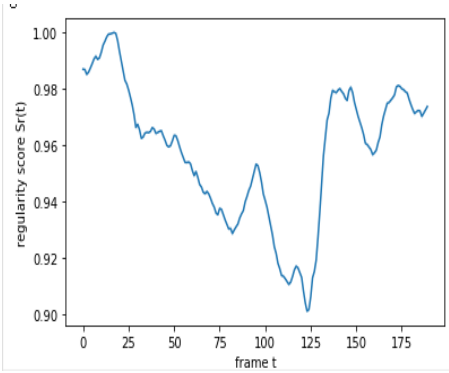

Figure 3 Another regularity score of the same dataset

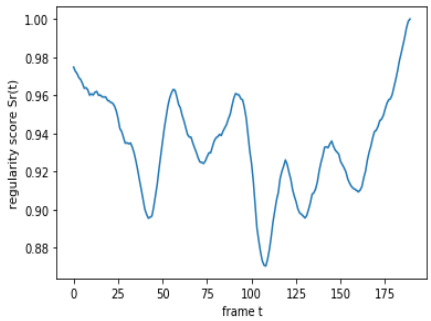

Figure 3 Another regularity score of the same dataset

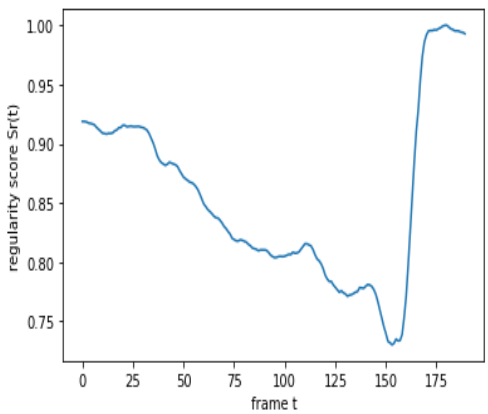

Figure 3 Regularity score of another frame

The variation in the regularity score indicates that in all the above figures regularity score increases when there is an anomaly detected on the road. The graph raises in that condition and decreases when the road is in a safe zone.

VI. DATA SET USED

A stationary camera with an elevation overlooking the footbridge was connected to the UCSD Anomaly Detection Dataset. The density of the crowds in the paths ranged from scarce to packed. The video only shows pedestrians in the normal setting. Anomalous events are attributable:

- The flow of non-pedestrian beings through the paths.
- Anomalous movement trends Football.

Small cars, skaters, Bikers, & people who walk through a footbridge or in surrounding grass often produce anomalies. There have also been several cases of wheelchair users. All anomalies occur naturally, i.e. for assembly of the data set they have not been produced. The data were divided into 2 sub-sets, which suit each scene. The film from each scene was divided into different videos of approximately 200 images.

A. Peds1

Clips of people groups moving to and from the shot and a degree of distortion of perspective. This consists of 34 video training and 36 video test samples.

B. Peds2

Football scenes parallel to the camera plane. Includes 16 video samples & 12 video checks.

That clip covers binary flag per frame for each annotation which indicates whether there is an anomaly in that frame. Furthermore, a subset of ten pixel-level binary masks for Peds1 and 12-pixel clips for Peds2 is given which determines anomaly-related areas. This is designed to enable the evaluation of the performance of algorithms to detect anomalies.

VII. CONCLUSION

Incident identification anomaly is a vital function for the automated tracking of incidents such as assaults, traffic accidents or illegally happening in videos. Current monitoring systems, by comparison, involve manually detecting irregularities, which is a very detailed process that often needs more effort than is generally possible. In the paper, we suggested an unsupervised approach to the identification of anomalies based on Convolutional Network encoder decoders for the extraction of spatial features and a Convolutional LSTM encoder-decoder to learn the temporal evolution of spatial characteristics. The suggested method for the identification of anomalies was evaluated using UCSD dataset experiments. We can combine multiple data sets in the future and test whether the model still works well. We can also find a way to improve the anomaly detection process, such as using fewer sequences during the testing phase. New methods such as extraction & reliability can also be based on future research.

Retrieval Number: D6476049420/2020©BEIESP
DOI: 10.35940/ijeat.D6476.049420
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

REFERENCES

1. C. Regazzoni, A. Cavallaro, Y. Wu, J. Konrad, and A. Hampapur, "Video Analytics for Surveillance: Theory and Practice," IEEE Signal Process. Mag., 27(5):16–17, May 2010.
2. R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: A systematic survey," IEEE Transactions on Image Processing, vol. 14, no. 3, pp. 294–307, Mar. 2005.
3. V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," ACM Computer Survey vol. 41, no. 3, art. 15, July 2009.
4. N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "ChangeDetection.NET: A new change detection benchmark dataset," Proc. IEEE Workshop on Change Detection, Providence, USA, June 2012.
5. Y. Tomioka, A. Takara, and H. Kitazawa, "Generation of an optimum patrol course for a mobile surveillance camera," IEEE Transactions on Circuits and Systems for Video Technology, vol. 2, no. 22, pp. 216–224, Feb. 2012.
6. G. Carvalho, L. A. Thomaz, A. F. da Silva, E. A. B. da Silva, and S. L. Netto, "Anomaly detection with a moving camera using multiscale video analysis," Multidimensional Systems and Signal Processing, pp.1– 32, Feb. 2018.
7. E. Jardim, X. Bian, E. A. B. da Silva, S. L. Netto, and H. Krim, "On the detection of abandoned objects with a moving camera using robust subspace recovery and sparse representation," Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Brisbane, Australia, pp. 1295–1299, Apr. 2015.
8. L. A. Thomaz, A. F. da Silva, E. A. B. da Silva, S. L. Netto, X. Bian, and H. Krim, "Abandoned object detection using operator-space pursuit," Proc. IEEE International Conference on Image Processing, Quebec City, Canada, vol. 2, pp. 1980–1984, Sept. 2015.
9. Y. Sheikh, O. Javed, and T. Kanade, "Background subtraction for freely moving cameras," Proc. IEEE International Conference on Computer Vision, pp. 1219–1225, 2009.
10. O. M. Sincan, V. B. Ajabshir, H. Y. Keles, and S. Tosun, "Moving object detection by a mounted moving camera," Proc. IEEE International Conference on Computer as a Tool, pp. 1–6, Sept. 2015.
11. H. Kong, J.-Y. Audibert, and J. Ponce, "Detecting abandoned objects with a moving camera," IEEE Transactions on Image Processing, vol. 19, no. 8, pp. 2201–2210, Aug. 2010.
12. X. Cui, J. Huang, S. Zhang, and D. N. Metaxas, "Background subtraction using low rank and group sparsity constraints," Proc. European Conference on Computer Vision, Part I, LNCS 7572, pp. 612–625, Oct. 2012.
13. W. Li, V. Mahadevan, N. Vasconcelos, "Anomaly Detection and Localization in Crowded Scenes", IEEE Trans. Pattern Anal. Mach. Intell., 36(1):18–32, Jan 2014.
14. P. Popoola, K. Wang, "Video-Based Abnormal Human Behavior Recognition—A Review", IEEE Trans. Syst. Man. Cybern. Part C, 42 (6):865–878, June 2012.
15. V. Chandola, A. Banerjee, V. Kumar, "Anomaly detection: A survey", ACM Comput. Surv., 41(3): 1–58, Mar. 2009.
16. P. Cui, L.-F. Sun, Z.-Q. Liu, and S.-Q. Yang, "A Sequential Monte Carlo Approach to Anomaly Detection in Tracking Visual Events," in.
17. T. Xiang, S. Gong, "Video Behavior Profiling for Anomaly Detection", IEEE Trans. Pattern Anal. Mach. Intell., 30 (5): 893–908, May 2008.
18. V. Reddy, C. Sanderson, B. Lovell, "Improved anomaly detection in crowded scenes via cell-based analysis of foreground speed, size and texture", IEEE CVPRW, pp. 55–61, June 2011.
19. P. Jodoin, Saligrama V., Konrad J., "Behavior Subtraction", IEEE Trans. I'm. Proc., 21 (9): 4244–4255, Sept. 2012.
20. A. Agrawal & S. Vishwakarma, "A survey on activity recognition and behavior understanding in video surveillance," The Visual Computer, 29(10):983–1009, Oct. 2013.
21. Xu D., Wu X., Song D., Li N., Chen Y-L., "Hierarchical Activity Discovery within Spatio-Temporal Context for Video Anomaly Detection", IEEE Int. Conf. Image Processing, pp. 3597-3601, Sep 2013.
22. Zhao B., Fei-Fei L., and Xing E. P., "Online detection of unusual events in videos via dynamic sparse coding," in IEEE Conf. Computer Vision and Pattern Recognition, pp. 3313–3320, 2011.
23. Jiang F., Yuan J., Tsaftaris S. A., &. Katsaggelos A. K, "Anomalous video event detection using spatiotemporal context," Computer Vision and Image Understanding, 115(3): 323–333, Mar. 2011.
24. Boiman O. & Irani M., "Detecting Irregularities in Images and Video," Int J Comput. Vision, 74(1):17–31, Jan. 2007.

AUTHORS PROFILE

Lovleen siddhu, received her bachelor's degree from Pandit Ravishankar University, Raipur, Chattisgarh and master's degree from the Vellore Institute of Technology (VIT), Chennai. She has a strong interest in the machine and deep learning.

Ranganathan Sridhar, obtained his Bachelor's degree in Electrical engineering from NIT [Formerly REC] Trichy in 1984. He has completed his M.Tech in Computer Science from the Indian Institute of Technology, Madras in 1994. He has completed his M.Phil in Computer science from Alagappa University. He has a total of 19 years of Industrial experience and 12 years of academic experience. He has published 4 papers in various national and international peer-reviewed journals and conferences. He is currently Associate Professor in SCOPE, VIT Chennai. His teaching and research expertise covers a wide range of subject areas including Digital Signal Processing, Electronics, Image processing, Knowledge mining, Web mining and Natural Language processing.