

Advanced Video Anomaly Detection Using Artificial Intelligence Techniques: A Review

Ritu Bai Sharma and Kavita

M.Tech Scholar, Department of Computer Science Engineering
Assistant Professor, Department of Computer Science Engineering
Mata Raj Kaur Institute of Engineering & Tech (Saharanwas), Rewari
ritusharma2314@gmail.com

Abstract: *As surveillance becomes ubiquitous, the amount of data to be processed grows along with the demand for manpower to interpret the data. A key goal of surveillance is to detect behaviors that can be considered anomalous. As a result, an extensive body of research in automated surveillance has been developed, often with the goal of automatic detection of anomalies. Research into anomaly detection in automated surveillance covers a wide range of domains, employing a vast array of techniques. This review presents an overview of recent research approaches on the topic of anomaly detection in automated surveillance. The reviewed studies are analyzed across five aspects: surveillance target, anomaly definitions and assumptions, types of sensors used and the feature extraction processes, learning methods, and modeling algorithms*

Keywords: Anomaly Detection, Motion Pixel, Unsupervised Approach

I. INTRODUCTION

Anomaly detection in video surveillance (VS) is a tough task to deal with many problems, e.g. noise, illumination change and interaction between different event and multi-dimensional video sequences are captured often under illumination and lighting condition. Multiple cameras may have like positions, orientations, and zooming factors. From fundamental point of view, techniques in video investigation are inspired by need to expand machine learning algorithms that can emulate abilities of human visual frameworks. We think about moving items in video as some compact areas with various obvious movements from background. In particular, if area is moving in specific frame, we think of it as moving object all through video. For instance in video, kid walks for some time and stops to swing his hands. It is important to regard his entire body as moving item rather than his hands just in late frames. Based on moving object definition, we think about movement as significant prompt to recognize it. If pixel has noteworthy diverse clear movement from background, it generally likely goes with moving object. In like manner, movement signs are characterized as contrast between pixel movement and background movement.

From point of view of threat detection in real time, it is outstanding that visual consideration dips under acceptance level, notwithstanding when prepared individual and assignment of visual observing. Video investigation innovations can be connected to create keen surveillance frameworks that can be help to real time threat identification. Specifically, multi scale tracking advances are subsequent stage in applying programmed video examination to observation systems. It is the documentation of rare explanations which raise doubts by contrasting altogether from most of the data. Naturally anomalous things will mean to some sort of issue such as bank fraud, therapeutic issues or mistakes in content. Anomalies are additionally alluded to as anomalies, anomalies, noise, deviations and exceptions. In specific, manhandle and arrange interruption identification, however unforeseen rushes in action. Numerous anomaly identification methods will bomb on such information, except if it is collected properly. Group investigation calculation have option to classify small scale bunches shaped. [4] Unsupervised detection procedures recognize discrepancies in unlabeled test informational collection under suspicion that most of examples in informational collection are ordinary by searching for occasions that appear to fit least to rest of informational index. Regulated anomaly detection methods



require informational collection that has been marked as “would be expected” and “irregular” and include preparing classifier (the key contrast to numerous other factual arrangement issues is in born unequal nature of anomaly discovery). Human mind is most smart framework, which use subjective thinking and procedures visual information for anomalous state semantic understanding and increasing specific situational mindfulness. In course of most recent couple of years, PC vision analysts have endeavored comparative capacities to video examination frameworks. With accessibility of less expensive visual sensors, requirement for seeing huge amount of video information is likewise expanded. Noteworthy use of video examination is clever traffic reconnaissance framework, which is utilized for scientific investigation or anomalous occasion location/expectation.

II. BACKGROUND

Hamid Izadina., et al., (2013) designed a multi-modal analysis for recognition & segmentation of actuating objects. Quick Shift algorithm grouped over time through K-means to achieve the spacio temporal video division or segmentation. This video is presented with the gesture or motion characteristics projected above item by item sections. The MFCC or constant of audio type signal and 1st order of derivatives are applied to classify the audio.

Daniela Giordano.,etal.,(2015) for segmenting or sectioning objects in videos for difficult scenes with many targets. The method failed to create particular on videos and depending on the objects that recognized through humans consistent with the Gestalt laws. Coarse foreground segmentation is created with the predictions about motion regions through examining the super pixel segmentation variations in successive frames.

A real-time H.264/AVC compressed or contracted the domain actuating object segmentation or division and tracking or crossing algorithm is planned by Mansi Patel., et al., (2014) for surveillance videos.

The algorithm has moving object recognition and bounding box matching. In addition, the spatiotemporal combination, split or burst concluding and trajectory smoothening with modernization in the information are carried out by the forecasting modes into the motion detection as well as trajectory construction. A new method of candidate smoke region segmentation is presented by Yaqin Zhao., (2015) depending on rough set theory. Kalman filtering technique is utilized to update or modify the video environment to remove the intervention of static smoke-color objects. RGB color spaces parts are divided through showing the upper estimation, lower estimation and irregularity of smoke-color distribution or dispersion.

A method is designed by FuyuanXu., et al., (2014) for motion segmentation by moving or actuating camera. The method categorizes all image or picture pixels in a particular image sequence as background with three-view constraint termed as parallax-based multi planer constraint. The three-view drawback is based on concern structure of two points in three views as per discussed above. The parallax based multi planar constraint addresses the former geometry drawbacks and failed to need the reference plane across many views. A local adaptive thresholding approach planned by Y.P.Guan., (2012) removes the binary motion or apparent movement masks depending on to the performances of wavelet translation. A highlight classification achieves optimal segmentation through combining the filtering process, connectivity process and spatial-temporal correlation.

A robust algorithm is designed by Zhengyang Wu., et.al., (2015) to create video segment schemes. The designed methods failed to take particular motion models and less ability to create across videos. Least squares tracking or chasing framework is planned where image segment methods are created and covered or tracked by learned appearing models. The modernism lies in the utilization of two moves, the blend move and freely addition initiates the segments from any of the frame and tracks them through the complete blockage lacking extra computation.

A moving object segmentation algorithm is designed by Yanli Wan., et al., (2014) for freely actuating cameras. It is essential for out of door surveillance system, the car or vechicle build-in surveillance system model and the robot navigation system. A two-leveled based affine translation model optimization or improvement method is designed for camera compensation or recompense idea. The outermost layer iteration or looping removes the non- background characteristics points as per considerable and the inner layer iteration or looping account refined affine model carried from the RANSAC method or technique. The characteristics points are categories into the highlight and background



coherent with the detected or observed motion or actuating data. A geodesic based graph cut algorithm extracts the moving foreground depending on classified features.

A moving object segmentation algorithm is designed by Yanli Wan., et al., (2014) for moving cameras in out of door surveillance system, car or vehicle build in surveillance model and the robot navigation or piloting model. A two-layer or leveled based affine transformation system optimization or navigating technique is planned for camera compensation or recompense idea or approximation where the outermost layer or leveled looping is applied to dispatch the non-background characteristics points. The inner layer iteration or looping is applied to calculate refined affine model or system depending on the RANSAC technique. The characteristics points are categorized into foreground or highlight and background with described motion or actuating information or data. A geodesic founded graph cut algorithm is used to dispatch moving or actuating highlight depending on classified characteristics.

Video object segmentation is designed by Chenguang Zhang and HaizhouAi (2011) for hierarchical localized categorization of regions and video Object division is to split the preferred object from video sequence, the main limitation like as shape deformation or distortion, appeasable difference and background clutter. To solve the above problems the above listed technique is designed. The complete technique is to apply the Gaussian Mixture Color Models technique. Ultimate and quality points, after combining or joining the local as well as global outputs in-to a prediction based mask, So, the graph cut method gets the finalized segmentation consequence. A lot of challenging video sequences is applied to represent the effectuality and flexibleness of the HLCR method.

A stereo video object segmentation or division algorithm by Lingyun Wang et al. (2015) tells the motion detection and disparity. In the stereo system video, the knowledge or information of the stereo video object segmentation or division has bit by bit improved to a research hotpot. The motion or actuating detection and dispassion applied for handling the motion or actuating information, dispassion and edging characteristics to dispatch actuating objects from video sequences. Basically, the eight neighbor motion or actuating detections are used to get the alter detection mask of moving or actuating objects. After that, the stereo matching is acquired to obtain the dispassion map which is basically used to changed the alter detection mask. At last, the edging information or data and mathematic geomorphology post-processing are established to obtain the effective video object. In accession to explain the algorithm , affection or effectively applied for extracting the moving or actuating objects from stereo video sequences whose background has other moving or actuating objects.

III. PROBLEM FORMULATION

It is important to point out that many papers on video anomaly detection have addressed different formulations of the problem than the single-scene formulation on which this survey focuses. We have already discussed the multi-view formulation ([7], [8], [9], [10], [11]) above in some detail. Another alternate formulation for video anomaly detection that has been used in a number of papers ([12], [13], [14], [15], [16]) is training-free video anomaly detection. In this formulation, no normal training video is provided and the task is to either detect changes in the testing video or else to detect the most unusual segments of the testing video as proxies for anomalousness. Detecting the most unusual segments of testing video is analogous to discord detection in time series analysis [17]. While these formulations of the problem are also useful, they are significantly different from single-scene video anomaly detection and require different datasets and ground truth annotations. Many existing research papers do not clearly distinguish which problem formulation they are using. This leads to ambiguities and confusion about what datasets should be tested on and which methods should be compared against. It also leads to differences in understanding the performance of different methods. We think it is important to make clear the problem formulation being used in any paper on video anomaly detection. In this survey, we have chosen to focus on the single-scene video anomaly detection formulation because it encompasses a very common scenario and has many practical applications, such as in surveillance, security, factory automation (monitoring the activity of workshop floors), video search and video summarization.



IV. PERFORMANCE PARAMETERS

True Positives (TP) - These are the correctly predicted positive values which means that the value of actual class is yes and the value of predicted class is also yes. E.g. if actual class value indicates that this passenger survived and predicted class tells you the same thing.

True Negatives (TN) - These are the correctly predicted negative values which means that the value of actual class is no and value of predicted class is also no. E.g. if actual class says this passenger did not survive and predicted class tells you the same thing.

False positives and false negatives, these values occur when your actual class contradicts with the predicted class.

False Positives (FP) – When actual class is no and predicted class is yes. E.g. if actual class says this passenger did not survive but predicted class tells you that this passenger will survive.

False Negatives (FN) – When actual class is yes but predicted class in no. E.g. if actual class value indicates that this passenger survived and predicted class tells you that passenger will die. Once you understand these four parameters then we can calculate Accuracy, Precision, Recall and F1 score.

Accuracy - Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations. One may think that, if we have high accuracy then our model is best. Yes, accuracy is a great measure but only when you have symmetric datasets where values of false positive and false negatives are almost same. Therefore, you have to look at other parameters to evaluate the performance of your model. For our model, we have got 0.803 which means our model is approx. 80% accurate.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN}$$

Precision - Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. The question that this metric answer is of all passengers that labeled as survived, how many actually survived? High precision relates to the low false positive rate. We have got 0.788 precision which is pretty good.

$$\text{Precision} = \frac{TP}{TP+FP}$$

Recall (Sensitivity) - Recall is the ratio of correctly predicted positive observations to the all observations in actual class - yes. The question recall answers is: Of all the passengers that truly survived, how many did we label? We have got recall of 0.631 which is good for this model as it's above 0.5.

$$\text{Recall} = \frac{TP}{TP+FN}$$

F1 score - F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. Intuitively it is not as easy to understand as accuracy, but F1 is usually more useful than accuracy, especially if you have an uneven class distribution. Accuracy works best if false positives and false negatives have similar cost. If the cost of false positives and false negatives are very different, it's better to look at both Precision and Recall. In our case, F1 score is 0.701.

$$\text{F1 Score} = \frac{2 * (\text{Recall} * \text{Precision})}{(\text{Recall} + \text{Precision})}$$

V. CONCLUSION

We have provided a comprehensive review of research in single-view video anomaly detection. We built an intuitive taxonomy and situated past research works in relation to each other. We also hope this article will serve to clear up some misconceptions among different problem formulations, use of datasets, evaluation protocol and how to compare against methods that use compatible problem formulation and evaluation schema in their assumptions. We now provide some best practices and state some observations on the evolution of the field in terms of overarching trends in representation and modeling as they relate to the increasing size of datasets and increasing compute power of devices.

REFERENCES

[1] Adam, E. Rivlin, I. Shimshoni, and D. Reinitz. Robust real-time unusual event detection using multiple fixed-location monitors. IEEE transactions on pattern analysis and machine intelligence, 30(3):555– 560, 2008.



- [2] S. A. Ahmed, D. P. Dogra, S. Kar, and P. P. Roy. Trajectory-based surveillance analysis: A survey. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–1, 2018.
- [3] AIRS. Traffic response and incident management assisting the river cities. <http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/>, 2004. [Online; accessed 20-Jan-2004].
- [4] Akoçz and M. E. Karşlıgil. Traffic event classification at intersections based on the severity of abnormality. *Machine vision and applications*, 25(3):613–632, 2014.
- [5] P. Babaei. Vehicles behavior analysis for abnormality detection by multi-view monitoring. *International Research Journal of Applied and Basic Sciences*, 9(11):1929–1936, 2015.
- [6] B. Babenko, M.-H. Yang, and S. Belongie. Visual tracking with online multiple instance learning. In *CVPR*, 2009.
- [7] V. Bastani, L. Marcenaro, and C. Regazzoni. Unsupervised trajectory pattern classification using hierarchical dirichlet process mixture hidden markov model. In *MLSP*, 2014.
- [8] V. Bastani, L. Marcenaro, and C. Regazzoni. A particle filter based sequential trajectory classifier for behavior analysis in video surveillance. In *ICIP*, 2015.
- [9] P. Batapati, D. Tran, W. Sheng, M. Liu, and R. Zeng. Video analysis for traffic anomaly detection using support vector machines. In *WCICA*, 2014.
- [10] L. E. Baum and T. Petrie. Statistical inference for probabilistic functions of finite state markov chains. *The annals of mathematical statistics*, 37(6):1554–1563, 1966.
- [11] Y. Benezeth, P.-M. Jodoin, V. Saligrama, and C. Rosenberger. Abnormal events detection based on spatio-temporal co-occurrences. In *CVPR*, 2009.
- [12] S. Bengio. *Statistical machine learning from data: Gaussian mixture models. Lectures*, 2006.
- [13] S. Biswas and R. V. Babu. Short local trajectory based moving anomaly detection. In *ICVGIP*, 2014.
- [14] S. Biswas and R. V. Babu. Sparse representation based anomaly detection with enhanced local dictionaries. In *ICIP*, 2014.
- [15] C. G. Blair and N. M. Robertson. Event-driven dynamic platform selection for power-aware real-time anomaly detection in video. In *VISAPP*, 2014.
- [16] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.
- [17] S. Blunsden and R. B. Fisher. The behave video dataset: ground trothed video for multi-person behavior classification. *Annals of the British Machine Vision Association*, 4(1-12):4, 2010.
- [18] P. V. K. Borges, N. Conci, and A. Cavallaro. Video-based human behavior understanding: A survey. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(11):1993–2008, Nov 2013.
- [19] N. Buch, S. A. Velastin, and J. Orwell. A review of computer vision techniques for the analysis of urban traffic. *IEEE Transactions on Intelligent Transportation Systems*, 12(3):920–939, Sept 2011.
- [20] P. Chakraborty, A. Sharma, and C. Hegde. Freeway traffic incident detection from cameras: A semi-supervised learning approach. In *ITSC*, 2018.
- [21] R. Chalapathy, A. K. Menon, and S. Chawla. Robust, deep and inductive anomaly detection. In *MLKDD*, pages 36–51. Springer, 2017.
- [22] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM Comput. Surv.*, 41(3):15:1– 15:58, July 2009.

