iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** III **Month of publication:** March 2021

DOI: https://doi.org/10.22214/ijraset.2021.33331

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue III Mar 2021- Available at www.ijraset.com

Unwanted Activities Capturing in Crowd using Computer Vision – Survey

Jignesh Gothi[1], Aditya Sinha[2], Saket Kumar Jha[3]
[1]*CSE with Specialization in AI and ML, Vellore Institute of Technology, Chennai, India*
[2]*Director, Centre for Development of Advance Computing, Patna, India*
[3]*Project Engineer, Centre for Development of Advance Computing, Patna, India*

Abstract: ***Current surveillance and control system still require a human super-vision and intervention. India, which has only three Lakh surveillance cameras under human observation, lags in the Surveillance System. This paper will present the capture or observation of suspicious activities and warn the officer that immediate and sufficient action can be taken. Scenes are translated into frames to capture an unusual scene and prepare data as an input for Algorithms. We will then define the frames and translate them into vectors. Filtering a noisy data for each feature vector and if feature values deviate a threshold then it detects a shot edge. Clustering an edge shot and constructing a scene with a limited number of frames. Furthermore, a limited number of frames will reduce the cost of computing. This can detect suspicious objects and automate cameras from human observation to machine vision, both public and private.***
Keywords: Computer Vision, Camera, Scene detection, Scene Tagging, Unusual Activities

I. INTRODUCTION

Companies and communities around the world are experimenting with the use of artificial intelligence to minimize and solve crime and react more rapidly to ongoing crime. The premise behind all of these programs is that crimes are reasonably predictable; it actually takes the ability to sort through a large amount of data to put an end to trends that are beneficial to law enforcement. A few decades ago, this kind of data processing was technologically impossible, but the expectation is that recent advances will take place in machine learning are up to the task.

There is a good reason why businesses and governments alike want to try to use AI in this way. In 2010 , the United States spent more than $80 billion a year on state , local, and federal incarceration. Estimates place the overall expenditure on law enforcement by the United States at over $100 billion a year. A large portion of local government budgets is made up of law enforcement and prisons.

In recent times, the use of AI in providing successful crowd management solutions has been in vogue and has resulted in fruitful results in avoiding urban challenges such as handling mega footfall incidents, emergencies and disasters. During their SG50 Celebrations (50th anniversary of the independence of Singapore), Accenture partnered with the Singapore government and developed a solution to predict crowd activity and possible responses to incidents. 85 percent accuracy in high crowd operation, crowd size estimation and object detection resulted in the solution. The Kumbh Mela Experiment, closer to home, is aimed at predicting the actions of the crowd and the risk of a stampede. Similar solutions for Big Data and AI could assist with advance prediction and response management.

Via Smart Command Centre's, AI technology could provide security with sophisticated surveillance systems that could maintain controls on the movement of citizens, possible instances of crime, and general resident security. Platforms for social media intelligence will provide public safety assistance by collecting social media information and anticipating possible events that could threaten public peace. After the introduction of AI powered protection, the crime rate has decreased by 27 percent in the city of Surat systems.

II. LITERATURE SURVEY

In this section, survey is about converting Human Efforts or Vision into Machine Vision. It has been observed that in India, only three Lakh cameras are installed and all are under human observation. In human observation are made to do efforts to the observe the camera and maintain the law. In human observation, humans can make mistake for finding targets. So by automating the cameras and finding the crime scene and alerting the officer where he can take immediately action towards crime location.

Here Survey shows up till now what are the approaches, methods and algorithms is being used. Here we came to know that many methods have been successfully at national level. It has been seen that crime rates of that nations has been reduced.

1) Crime rates in Sri Lanka have risen significantly in recent years. The police's main priority is to avoid reports of crime in order to improve public safety. Criminals use sophisticated technology, which make the detection of crime cumbersome. In these investigations, police officers expend a great deal of time and effort. A broad range of research is being done to automate crime detection and prediction in the fields of Artificial Intelligence (AI) and Neural Networks. In this paper, they present the E-police system focused on machine learning and deep learning to improve public safety and assist law enforcement. The main aim of the scheme is crime reduction. E-Police is an application that allows police officers to be aware in real time of the events taking place around them. In addition, the framework makes forecasts of potential crimes that are likely to occur in the future, so that measures can be taken to avoid them.

Model	Validation Accuracy	Validation Loss
ResNet50+LSTM	100%	0.0031
InseptionV3+LSTM	74.71%	0.4820
VGG16+LSTM	67.82%	0.5023

2) A detailed analysis is provided in this paper on crime classification and prediction using deep learning architectures. They analyse the efficiency of deep learning algorithms in this field and include suggestions for the design and training of deep learning systems using open data from police reports to predict areas of crime. A comparative analysis of 10 state-of-the-art techniques against 3 different deep learning configurations is performed with a training data time series of crime types per location. We show that the deep learning-based methods consistently outperform the current best-performing methods in our experiments with five publicly accessible datasets. In addition, in the deep learning architectures, they analyze the ability of various parameters and provide insights for configuring them in order to achieve enhanced efficiency in the classification of crime and eventually prediction of crime.

3) There is still an underexplored problem of semantic video indexing. The video search, tracking and surveillance experience will be greatly enhanced by solutions to the issue. This paper covers scene detection and annotation and, in particular, the task of video structure mining using deep features for video indexing. A pipeline consisting of feature extraction and modification, shot clustering and labelling stages is proposed and implemented in the paper. As the source of the functionality, a deep convolutional network is used. For both scene identification and annotation, the pipeline is evaluated using metrics. The results obtained demonstrate high scene detection and approximate annotation efficiency with multiple metrics. In addition, we conducted an overview and review of con-temporary metrics of segmentation and annotation. In real time, the result of this work can be applied to semantic video annotation.

4) In most visual-based surveillance applications and security systems, object detection is the main module. Photos and videos play an important role in providing visual evidence of a scene in crime scene investigation. It helps police officers to reconstruct a scene by identifying items relevant to a particular crime for further study. However, for law enforcement agencies, the job of identifying items of interest is rather repetitive because of the availability of a large amount of data. In this work, they present a real-time framework based on Faster R-CNN (Region-based Convolutional Neural Network), which automatically detects targets that could be found in an indoor environment. They applied this to a subset of ImageNet containing 12 object classes and a Karina dataset to assess the effectiveness of the proposed method. In the Nvidia-TitanX GPU, we achieved an average accuracy of 74.33 percent, and the mean time taken to detect objects per image was 0.12 s.

5) Human oversight and involvement are still needed by current surveillance and control systems. This work introduces a new method of automatic firearm detection in videos suitable for both monitoring and control purposes. We reformulate this problem of detection into the question of reducing false positives and solve it by I constructing the main training data set driven by the outcomes of a deep classifier of Convolutionary Neural Networks (CNN) and ii) testing the best classification model under two methods, the sliding window method and the approach to region proposal. The Faster R-CNN-based model trained on our new database obtains the most promising results. Even in low-quality YouTube videos, the best detector demonstrates high potential and offers satisfactory results as an automatic alarm device. After five consecutive true positives in a time interval of less than 0.2 seconds, the alarm is triggered successfully in 27 scenes out of 30 scenes. A new metric, Alarm Activation Time per Interval (AATpI), is also specified to evaluate the performance of the detection model as an automatic video detection system.

6) To detect motions of interest (abnormal motion) based on local feature modelling within spatio-temporal detectors, we present a motion classification method. Using motion vectors and local detectors, the modelling is finished. The detectors are trained based on labelled samples to learn irregular motion independently. An abnormality score, in both space and time, which is the basis of the final classification, is allocated to each detector. To discriminate against simultaneous occurrences of irregular motion, the spatial relationship between detectors is used. 52 hours of the multi-camera surveillance dataset of the TRECVID 2010 challenge will test the efficiency of the proposed system.

7) Motion detection is a commonly applied technology in dynamic scenes, but a daunting topic in computer vision. Compared with motion detection in static scenes, it is less developed. A novel approach to detecting moving objects in dynamic scenes is discussed in the paper without any prior knowledge about moving objects or dynamic scenes. It is mostly based on the fact that the displacements of the moving object and background characteristics are clearly different, which can be used as a criterion to differentiate between the moving object and the background. To detect feature points, the SIFT (Scale Invariant Feature Transform) algorithm is used, an initial match set is obtained by the Euclidean metric and the nearest neighbour distance ratio law, and a consistency test is performed to obtain robust correspondences with features. A fuzzy c-means clustering algorithm is used with the displacement vectors generated from the robust feature correspondences, and the feature points from moving objects are precisely detected. Using actual video sequences from moving cameras, the feasibility of the suggested approach is illustrated.

8) Perception of the traffic scene (TSP) aims to extract accurate on-road real-time information, which includes three stages: identification of objects of interest, recognition of objects detected, and monitoring of objects in motion. As identification and monitoring often rely on detection performance, the ability to effectively detect objects of interest plays a crucial role in TSP. We focus on three major groups of items in this paper: traffic signs, vehicles, and bicycles. In a single learning-based detection system, they suggest detecting all three important items. The proposed system consists of a dense feature extractor and three major groups of detectors. These features are shared with all detectors until the dense features have been removed. The benefit of using one standard framework is that the speed of detection is much higher, as all dense characteristics need to be tested only once in the testing process. In comparison, for each of these three groups, most previous works have designed unique detectors using distinct features. As part of aggregated channel characteristics, we add spatially pooled characteristics to improve the feature robustness to noises and image deformations. Researchers propose an object subcategorization method as a means of capturing the intraclass variation of objects to further enhance the generalisation performance. In three detection apps: traffic sign detection, car detection, and bicycle detection, they experimentally demonstrate the efficacy and efficiency of the proposed system. The proposed architecture achieves competitive output on multiple benchmark data sets using state-of-the-art approaches.

9) Automation is a method of minimizing human efforts by using technology. We suggest an integrated computer vision-based framework for stock management in supermarkets in this paper. The proposed system would help to minimize the manpower needed in a supermarket by constantly tracking the availability of a supermarket product and automatically reporting valuable information to the individual concerned. The main concept behind the proposed system is to put a few low-cost cameras in the supermarket that will help to capture the videos of the supermarket product racks. Using a structural similarity index (SSIM) based scene change detection technique, the presence of human beings is detected, and an object detection technique is used to count the number of objects present in the same product rack. If the number of things present in a given rack goes below the threshold cap, the authority concerned will receive a short message service (SMS) and/or email. A product identifier (printed) will be maintained just below the product racks to make it more convenient. The product identifier will be detected by an optical character recognition module in the proposed scheme and it will be listed in the SMS or email that will assist the supervisor to arrange the replacement of the products in the racks. The experimental research is conducted by putting sample objects on a rack and the mobile camera is used for tracking purposes as an IP camera with the assistance of the IP webcam android programme. The experimental study shows that in a supermarket setting, the proposed scheme would function reliably.

10) In this research, we advocate the importance of robust local features that enable object shape to be differentiated for detection purposes from other objects. They start with the Histogram of Oriented Gradients (HOG) grid and integrate within it the Scale Invariant Feature Transform (SIFT). The presence of an object is identified in HOG features by the distribution of local intensity gradients or edge directions for various cells. In lieu of computing intensity gradients for these cells, we have computed the SIFT in the proposed process. In this way, the suggested solution not only offers more relevant details than only supplying intensity gradients, but also shows that it tackles the following challenges: (i) invariance of scale; (ii) invariance of rotation; (iii) change of illumination; and (iv) change of points of view. They contrasted the proposed method with other state-

of-the-art object detection methods with qualitative and quantitative experimental evaluation on standard INRIA dataset and demonstrated better performance over them.

III. METHODOLOGY

With the help of Literature Review, we propose a novel approach to video structure mining in this paper, which includes semantics obtained via a deep image classification network. Figure contains an outline of the processing element. The system included of feature estimation, feature ltering, and clustering of shots and scenes.

We use the Places205-AlexNet image classification network as a feature source for the scene detection task for feature estimation. The ILSVRC-winning topology is an AlexNet classic cation network that can be easily trained compared to other deep networks. In subsection of feature estimation, explanations for the use of Places205-trained groups are discussed.

To decrease the impulse noise, we are implementing feature Itering. We are implementing the overlapping relation approach for scene recognition and it is similar and faster to the graph-based method.

A. Feature Estimation

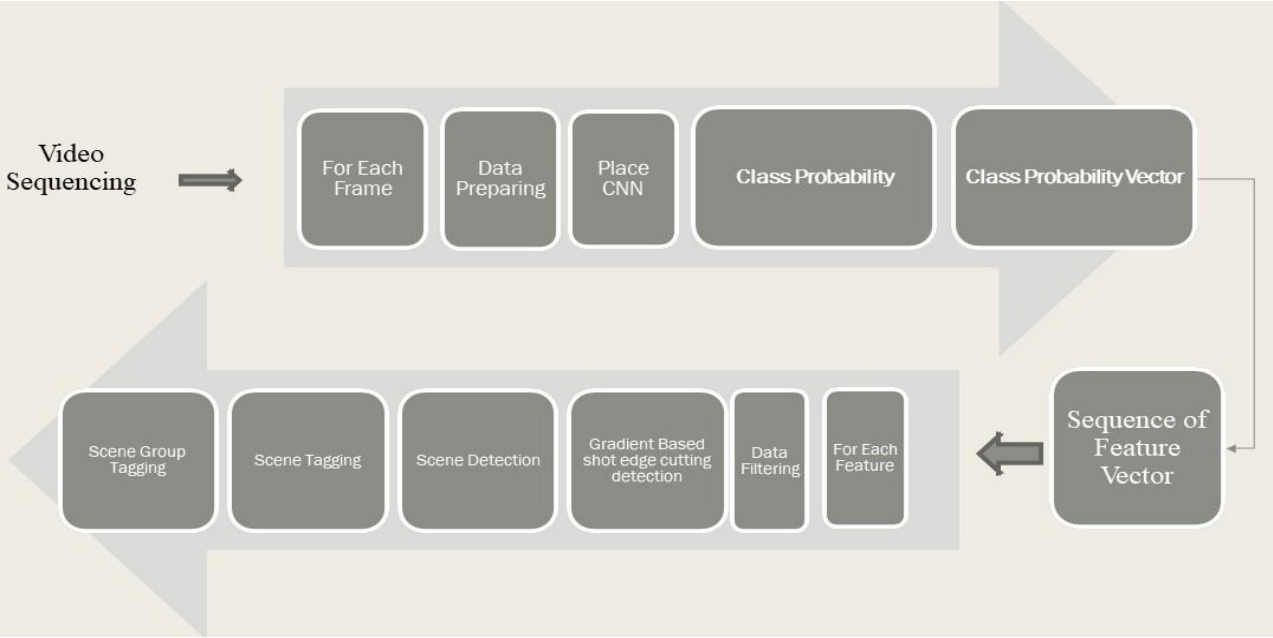
A segment of a motion picture with unified time and space is according to the Film Encyclopedia scene. Researchers treat Unified Time as a continuous series of frames. We propose the definition of a fixed environment for Unified space. Therefore, we agree that knowing the context of the picture is a good way of explaining the location where the scene is shot. Therefore, we will be using a network with a large variety of scene styles qualified for this mission. The dataset for Places205 consists of 205 groups of various types of locations.

B. Shot detection and Feature Filtering

In adjacent frames, we detect shot edges using the deviation of function values. The time-derivative function norm ||dF/dt|| is used to estimate the F deviation of the feature. We detect shot shift if this value crosses the threshold Thn.

C. Scene Recognition

Our pipeline's next step is to cluster shots into scenes. Measure the similarity of the shot as the minimum gap between the frames of two shots. We shall also use this method, as it helps us to establish larger classes of equivalence and minimize over segmentation. Computational costs will reduced by depicting shots with a smaller number of frames. A method of selection of key frames that maximizes the function value range between key frames. Researchers demonstrated that it is necessary to retain just 2-5 percent of frames for effective scene recognition. For our algorithm, we will impose a 5 percent threshold. According to researchers, they prefer over segmentation i.e. in-accurate segmentation rather than under segmentation.



IV. CONCLUSION

It's easy to incorporate or automate a camera with the aid of Computer Vision, where we can recognize suspicious activities. We define the frames for detection and we detect the events using the threshold value. As we have seen, these items have been implemented by many countries to reduce the crime rate of their nation.

REFERENCES

[1] Fahad Daniyal and Andrea Cavallaro, "Abnormal Motion Detection in crowded scenes using local Spatio-Temporal Analysis", IEEE, 2011, 978-1-4577-0539-7.

[2] Swati Nigam, Manish Khare, Rajneesh Kumar Srivastava and Ashish Khare, "An Effective Local Feature Descriptor for Object Detection in Real Scenes", IEEE Conference on Information and Communication Technologies, 2013, 978-1-4673-5758-6/13.

[3] I. S. Jacobs and C. P. Bean, "Examining Deep Learning Architectures for Crime Classification and Prediction," arXiv:1812.00602v1 [cs.LG] 3 Dec 2018.

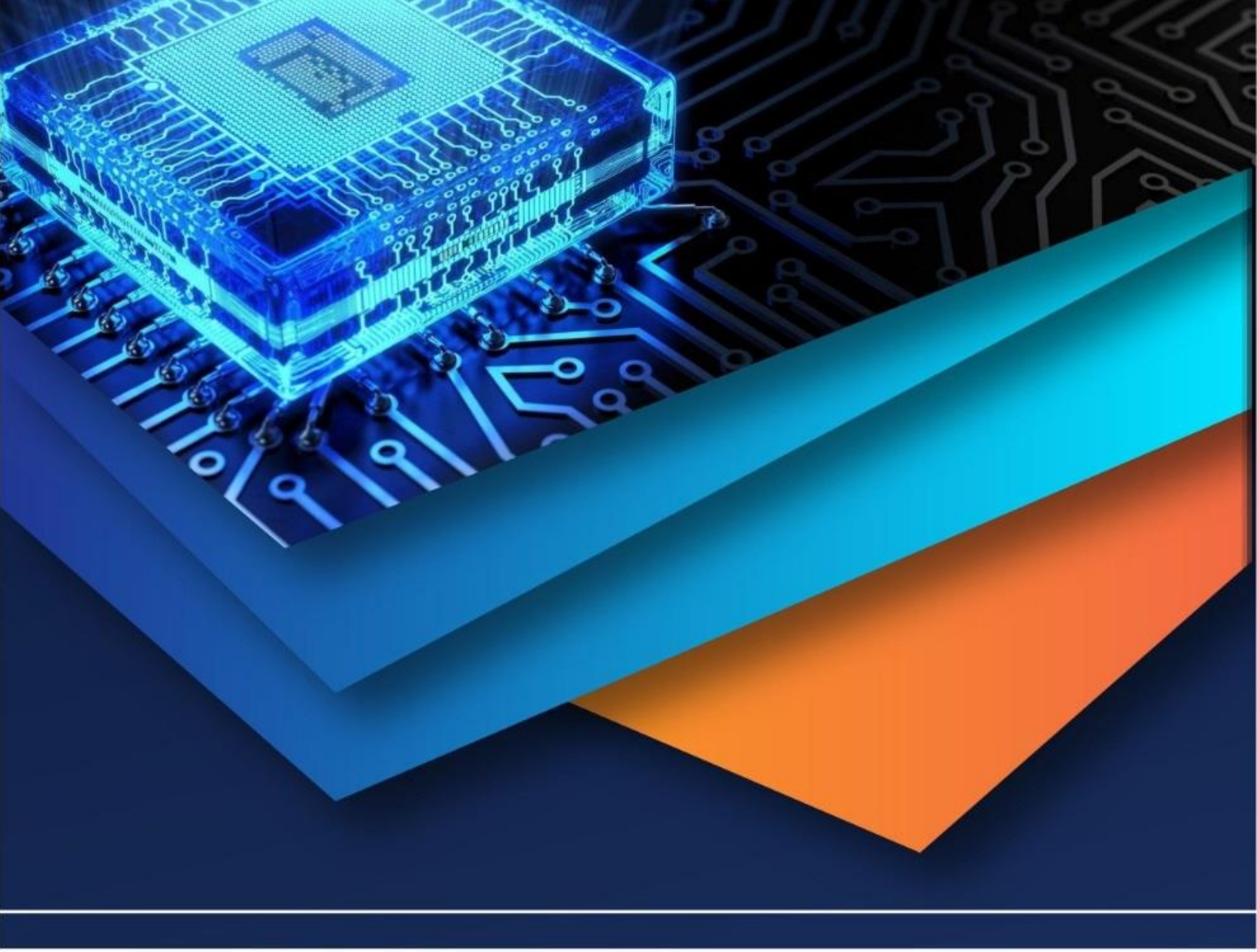
[4] Qichang Hu, Sakrapee Paisitkriangkrai, Chunhua Shen, Anton van den Hengel, and Fatih Porikli, "Fast Detection of Multiple Objects in Traffic Scenes With a Common Detection Framework," IEEE Transactions on Intelligent Transportation Systems, 1524-9050 © 2015 IEEE.

[5] Xiaqiong Yu, Xiangning Chen and Mengnan Gao, "Motion Detection in Dynamic Scenes Based on Fuzzy C-Means Clustering," International Conference on Communication Systems and Network Technologies, 978-0-7695-4692-6/12 $26.00 © 2012 IEEE.

[6] Surajit Saikia1, E. Fidalgo, Enrique Alegre and Laura Fern´andez-Robles, "Object Detection for Crime Scene Evidence Analysis Using Deep Learning," Springer International Publishing AG 2017, S. Battiato et al. (Eds.): ICIAP 2017, Part II, LNCS 10485, pp. 14–24, 2017, https://doi.org/10.1007/978-3-319-68548-9 2.

[7] Sandeep Kumar Yedla, V. M. Manikandan and Panchami V., "Real-time Scene Change Detection with Object Detection for Automated Stock Verification", 2020 5th International Conference on Devices, Circuits and Systems (ICDCS) 978-1-7281-6368-0/20/$31.00 ©2020 IEEE 10.1109/ICDCS48716.2020.243571.

[8] Charuni Rajapakshe, Shashikala Balasooriya, Hirumini Dayarathna, Nethravi Ranaweera, Namalie Walgampaya and Nadeesa Pemadasa ,"Using CNNs RNNs and Machine Learning Algorithms for Real-time Crime Prediction", 2019 International Conference on Advancements in Computing (ICAC), 978-1-7281-4170-1/19/$31.00 ©2019 IEEE.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY