International Journal of Engineering and Management Research
e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-12, Issue-3 (June 2022)
www.ijemr.net
https://doi.org/10.31033/ijemr.12.3.18

Comparative Study of Various Crowd Detection and Classification Methods for Safety Control System

Nida Khan[1] and Dr. Mohd Haroon[2]
[1]PG Student, Department of CSE, Integral University, Lucknow, Uttar Pradesh, INDIA
[2]Associate Professor, Department of CSE, Integral University, Lucknow, Uttar Pradesh, INDIA

[1]Corresponding Author: nidaintegral01@gmail.com

ABSTRACT

A crowd is a distinct collection of people or anything that is involved in a community or society. The phenomenon of a crowd is fairly well known in a wide range of academic fields, including sociology, civil engineering, and physics, amongst others. At this point in time, it has developed into the most active-oriented research and fashionable issue in the field of computer vision. Pre-processing, object detection, and event or behavior identification are the three stages of processing that are traditionally included in crowd analysis. These stages are pre-processing, object detection, and event recognition. Pre-processing, object detection, and event or behaviour identification are the three stages of processing that are traditionally included in crowd analysis. These stages are pre-processing, object detection, and event recognition. This study gives a model of crowd analysis as well as a taxonomy of the most prevalent method to crowd analysis. It may be helpful to researchers and would serve as a good introduction connected to the area of work that has been conducted.

Keywords-- Crowd Analysis, Crowd Detection, Crowd Counting

I. INTRODUCTION

A crowd is a distinct collection of people or anything that is involved in community or society. Tracking the movement of a crowd is quite different from following the movements of individuals inside a crowd. When tracking people, the information is computed at the level of each person being monitored. Crowded situations have become more commonplace in the actual world than they ever have been before [1], due to the growth in population as well as the variety of human activities. It poses significant difficulties in terms of public administration, safety, and security. Humans have the capacity to gather relevant information about the behaviour patterns in the surveillance area, watch the scene in real time for odd events, and give the opportunity for fast intervention. Research in psychophysics demonstrates, however, that humans have significant shortcomings in their capacity to monitor many signals at the same time [2]. Even for a human observer, it is a substantial effort to keep track of the actions of a large number of people in an extremely crowded setting at the same time. In our day-to-day lives, virtually everyone may encounter a crowd at some point. These crowds can be assembled for a variety of reasons, such as a cultural event or a sporting event. Additionally, one may encounter a throng at an airport or train station when travelling. In some circumstances, such as a sporting event, it is possible to escape the crowd; yet, in other circumstances, such as cultural events, it is quite difficult to avoid obtaining the sense of being in a crowd. Therefore, in an area with a high population density, there is always the potential for unwelcome behaviour, like as stampeding, to endanger the lives of humans. For this reason, it is essential to have a system in place for monitoring the crowd and analysing the behaviour of those individuals throughout the course of the allotted amount of time to ensure the safety of those who are present in the crowd. Manually analysing a large group of people is a highly challenging and time-consuming operation. So that we may create a system based on computer vision that is capable of carrying out these duties. There has been a significant amount of study carried out in the subject of crowd behaviour analysis over the last ten to fifteen years with only a moderate degree of success [4], indicating that there is a significant amount of room for expansion in the research field in this particular area. Because of the significance of the issue, the study subfield of computer vision that focuses on monitoring and analyzing the behavior of crowds is now accepting new topics. In the last ten years, several approaches of accomplishing these tasks have been suggested. These methodologies are intended to carry out a variety of tasks for the crowd, some of which include determining the size of the crowd in terms of its numerical strength for the purpose of effective crowd management in real time or for security reasons, forecasting the behavior of the crowd in the future, and other similar activities. Even though quite a few complicated methods have been put into practise for evaluating crowds, there is always room for improvement in terms of methods that can evaluate crowds in real time, particularly for unorganized crowds.

e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-12, Issue-3 (June 2022)
https://doi.org/10.31033/ijemr.12.3.18

1.1 Detection and Machine Learning Concept

The field of artificial intelligence includes the subfield of machine learning (AI). In spite of the fact that machine learning is a subfield of computer science, it is distinct from more conventional methods to computational problem solving. Object detection plays an important part in both the theoretical foundations of computer vision and its implementation in the real world. Object detection includes the tracking, detection, and counting of individual objects. The use of deep learning strategies may be helpful in a number of computer vision applications, including object identification. The world as we know it is now going toward automation, and one of the cutting-edge technologies that is being employed in automation is machine learning. Machine learning is a subfield of artificial intelligence that emulates human behaviour by training itself using data and mathematical formulas in order to perform more accurately. The field of machine learning encompasses the processes of deep learning. Deep learning is a subfield of machine learning that attempts to simulate the functioning of the human brain by using neural networks that have several layers. In the actual world, deep learning may be used in a variety of contexts. Object detection is one of its many applications, and the emphasis of this study is on that particular one. Object detection is a subfield of image processing and computer vision that identifies and locates things inside still photos, films, or real time movies. This may be done using the data included within the image or video. The techniques for detecting objects may be broken down into two distinct categories: the neural approach and the non-neural approach. Item identification using non-neural methods entails first extracting the features from an image and then feeding those features to a regression model in order to predict the position and label on the object in an image. This process is known as feature extraction. The selection of input features, the design of the model structure, and the selection of the learning algorithm are the three most important aspects of the process of developing a neural network system. Each space-time feature cube is labelled by its position, direction, and speed to build our fundamental feature set. This is accomplished by computing the change that occurs in the video of particle motion as it is occurring in the motion region.

II. CROWD ANALYSIS

A crowd's density, position, pace, and color, among other characteristics, are among its most crucial component traits. The information may be gleaned from the computer visions in either an automated or a manual fashion, depending on personal preference. The topology sensor and the typology sensor are the two kinds of sensors that are used throughout the scene capture process. The process of extracting information from a crowd scene should be dependent on the conditions of the environment, such as changes in illumination (the transition from day to night, shadows of background images, and non-static backgrounds such as leaves blown by the wind could be detected as moving object), handling the occlusion, multiple input channels and the amount number of cameras, the changes in motion, and detecting different characteristics, such as hu or lu. This will allow for more accurate information to be obtained from the crowd scene. In most cases, the crowd model is constructed based on the extracted information that either implicitly or explicitly represents the state, while the event detection is carried out with the help of the computational model.

Both of the models have been modified to reflect the newly extracted information.

2.1 Crowd Density Estimation

Estimating the density of the crowd is one of the most difficult aspects of visual surveillance. The safety of public events with huge audiences has always been a primary concern, particularly given the existence of a significant danger of deterioration. Because of their high level of effectiveness in information collection and relatively cheap costs in terms of human resources, approaches involving video analysis are gaining a lot of traction in the field of visual monitoring of public spaces. The flow density of large malls, supermarkets, and places such as subway stations is becoming more and more serious, which brings security risks due to the crowd congestion. This is due to the rapid development of the economy as well as the increasing number of people participating in social activities.

2.2 Crowd Motion Detection

The Background Subtraction Method may be used to determine whether or not there is motion in the crowd. Background Subtraction is a technique that may be used to separate the foreground object from the background object in a series of video frames. This is indicated by the name of the method. It is possible to define a foreground object as an object of attention that not only contributes to the reduction of the quantity of data that has to be processed but also supplies essential information on the activity that is being considered. In many cases, the item that is considered to be in the front of a picture is one that is continuously moving [10]. Objects of interest may be extracted from a scene using a family of methods known as background subtraction, which can be useful for surveillance and other types of applications. The process of building a decent algorithm for background subtraction is fraught with several obstacles. To begin, it has to be resilient against shifting levels of light. Second, it should prevent itself from recognizing any backdrop items that aren't stationary, as well as any shadows thrown by moving objects. A good backdrop model should also be

e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-12, Issue-3 (June 2022)
https://doi.org/10.31033/ijemr.12.3.18

able to respond fast to changes in the background and alter itself to accommodate changes that are happening in the background, such as the movement of a stationary chair from one spot to another. In addition to this, it need to have a high rate of foreground detection, and its processing time for the removal of the backdrop ought to be in real time.

III. PROPOSED METHOD

The proposed system consists of planning and executing the system which secured the answers for the existing system required. Building an Unsupervised Abnormal Crowd Behavior Detection system is part of the assignment, which also involves doing a comprehensive analysis of the system. The goal of the task is to provide expert security for Urban by completing the construction of the system. We built a system that is capable of characterizing normal and abnormal behaviour in crowds by applying a continual image observation system and utilizing SVM, kNN, Neural Network, and linear regression algorithm to assess monitoring of crowded urban areas (see figure 1).

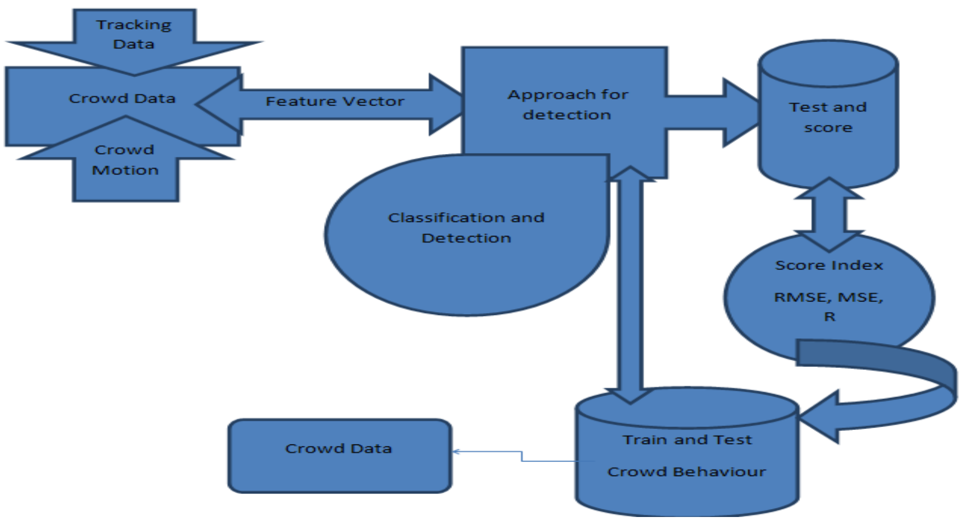

Figure 1: Proposed Model

3.1 Crowd Dataset

The distribution of crowd density in photographs of packed crowds is seldom uniform because of the differences in viewpoint and scene that exist in these photos. Figure 2 provides a selection of photographs for your viewing pleasure. Due of this, it is nonsensical to try to count the individuals in the crowd while simultaneously taking in the whole sight. The divide-count-sum mechanism was implemented into our system after it was updated as a direct result of this issue. After the images have been segmented into patches, a regression model is used in order to map each image segment to the associated local count. This step follows the segmentation of the photographs into patches. In conclusion, the number of global images may be obtained by applying a cumulative computation to the number of these patches. This yields the global image count. Users of image segmentation software may take use of not one, but two unique benefits: To begin, the density of the crowd is spread among the smaller picture patches in a way that is reasonably consistent with itself. Second, the amount of training data that may be made available to the regression model is enhanced when picture segmentation is conducted because of the rise in the amount of data that is segmented from the image. We are now able to construct a regression model that is more robust than it was before because of the benefits that were discussed earlier. Even though there isn't any consistency in the distribution of crowd density, the overall distribution of crowd density has a continuous pattern [4]. This is despite the fact that there isn't any uniformity in the distribution of crowd density. This suggests that picture patches that are contiguous to one another should have densities that are comparable to one another. When we wish to cut the picture up into smaller pieces, we often use overlaps, which strengthen the connection between the various image patches. The introduction of a Markov random field helps to smooth out the rough edges of the estimated count across overlapping image patches, which brings the final result closer to the true density distribution [7]. This helps to rectify any potential estimate errors that could have taken place when counting the picture patches. We make use of a neural network that has all of its connections intact in order to learn a map that goes from the aforementioned attributes to

e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-12, Issue-3 (June 2022)
https://doi.org/10.31033/ijemr.12.3.18

the local count. In addition, in order to extract features from different image patches, we make use of a pre-trained deep residual network. The usage of deep convolutional network features has been beneficial to a wide variety of computer vision applications, including but not limited to image segmentation, object recognition, and picture identification, to name just a few of those applications. This would imply that the learned properties of the deep convolutional network have the potential to be used to a wide range of diverse computer vision applications. The learned features have a greater probability of properly capturing the data [9], when there are more layers in the network. On the other side, you will want more data in order to adequately prepare for a model that is more in-depth. It is not possible to train an exceptionally deep convolutional neural network from the ground up using the datasets that are presently available for crowd counting since they are insufficient.

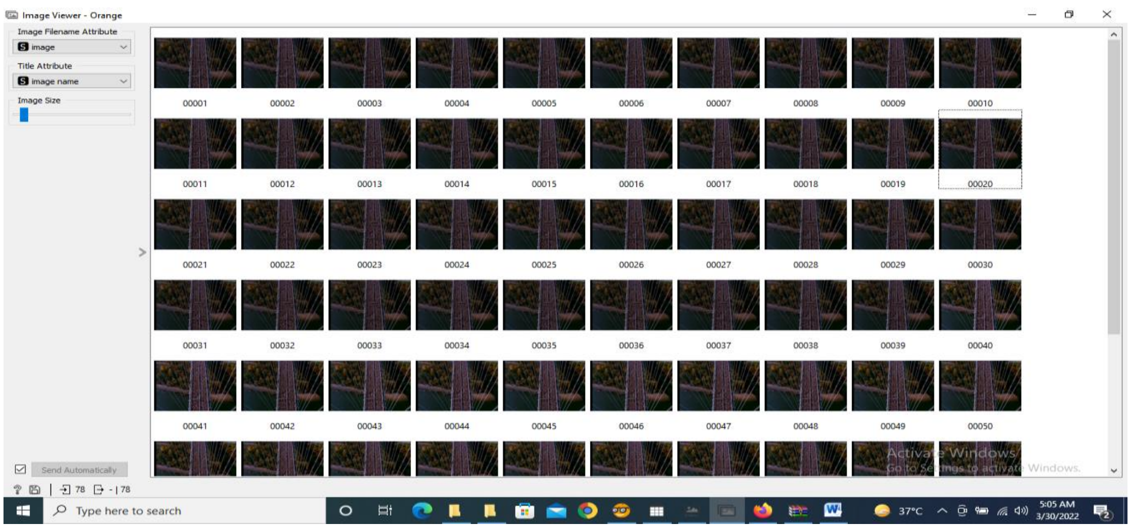

Figure 2: Images

3.2 Data Flow Architecture

Predictive models (figure 4) are assessed using the Test and Score widget, and forecasting based on newly collected information is carried out using the Predictions widget. Test and Score takes a number of different things as input, including data (a data set for assessing models), learners (algorithms to use for training the model), and a preprocessor (which is optional) (for normalization or feature selection).

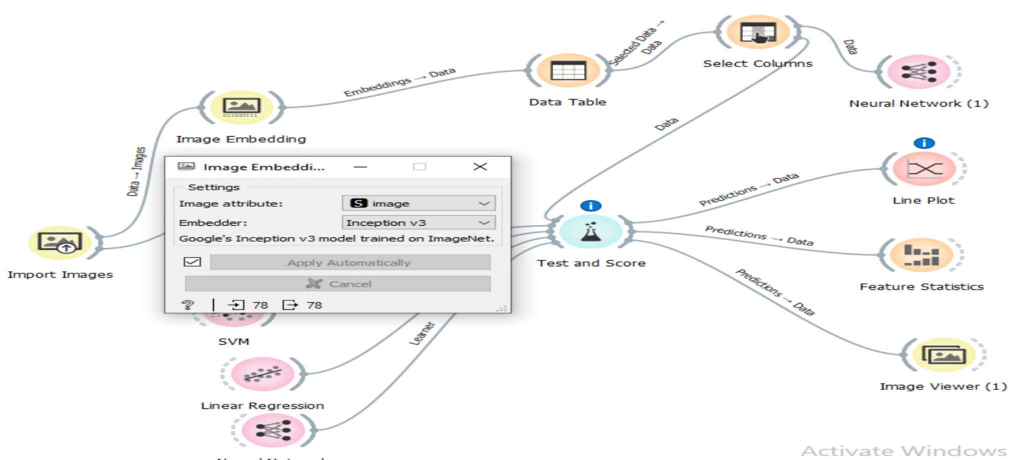

Figure 3: Flow structure

In figure 1.8, we have done the evaluation using the various methods and find the complete changes of crowd.

e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-12, Issue-3 (June 2022)
https://doi.org/10.31033/ijemr.12.3.18

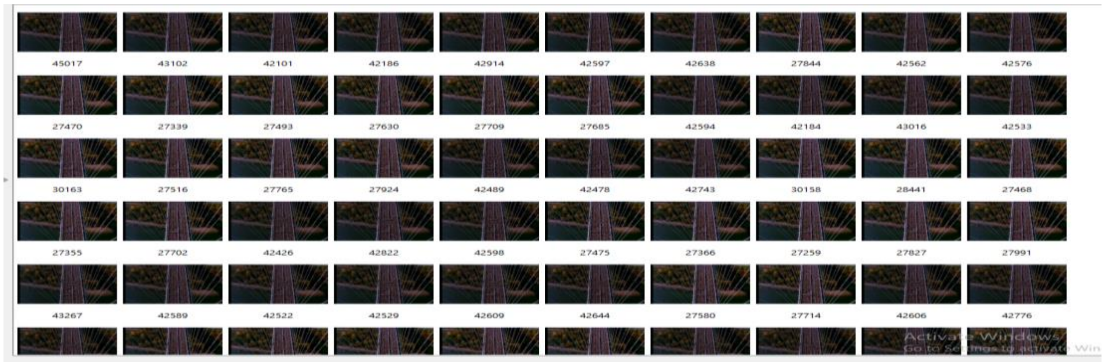

Figure 4: After Evaluation

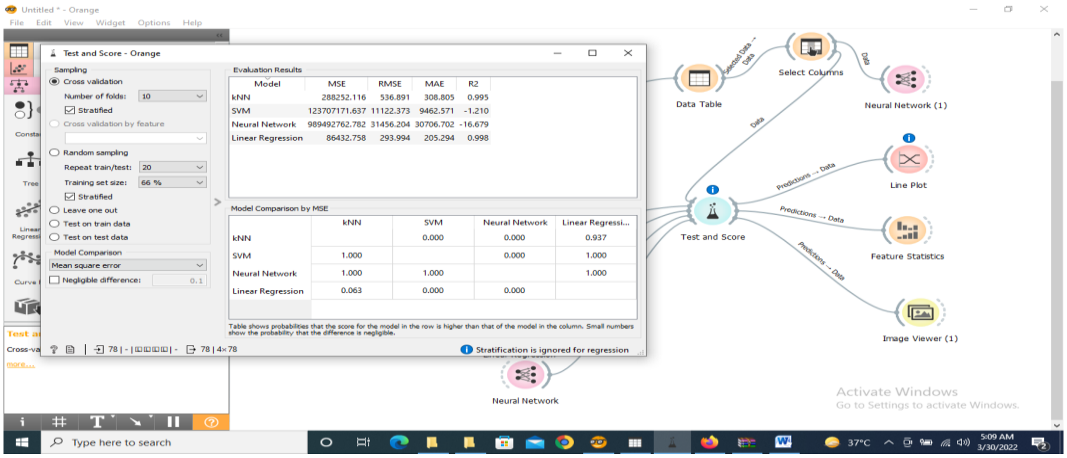

Figure 5: MSE evaluation

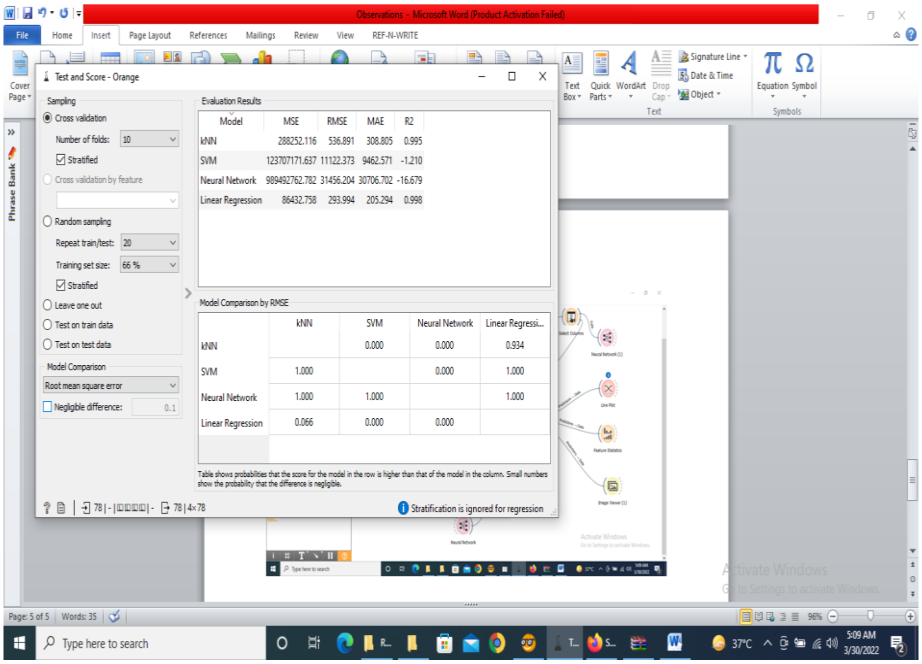

Figure 6: RMSE evaluation

e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-12, Issue-3 (June 2022)
https://doi.org/10.31033/ijemr.12.3.18

In figure 5 and 6, we are illustrated Figure depicts that the MSE (0.96 to 0.99) of was relatively high when compared to kNN and linear regression. This is due to the consideration of density-level classification of image patches with a density-oriented-based repressor approach. In addition, the MSE was rather low when evaluated against the dataset as a whole. This is as a result of taking into account a skip link with scale-oriented training in order to deal with problems of varied magnitude. In comparison to SVM and kNN, the MSE had a very low standard error value. This occurred because a regulated flow of information was taken into account when passing through the convolution and de-convolution layers of the network. As a result, we have reached the conclusion that the use of a task-oriented repressor and convolution results in an improvement in accuracy when predicting the quality of a high-quality density map for datasets that include a dense and varied range of densities. As a result, the low density of datasets may be handled through the use of patch-based augmentation strategies (variable-scale) and a focus on tackling the scale-varying problem in the convolution and de-convolution layers to optimize information flow. Scale-varying issues brought on by the viewpoint view may be addressed. You may find information on each of these tactics in the paragraphs that came before this one.

IV. CONCLUSION

Using competitive machine learning model, this study introduce a novel approach to detect abnormal behaviors in crowd scenes. For the purpose of obtaining information on particle movement and crowd motion, a technique for first extracting features and then describing them has been proposed. This approach applies space-time feature cubes. Then, a detection approach that combines space-time feature cubes and a competitive neural network model is presented as a means of detecting anomalous occurrences in worldwide crowds. This algorithm is intended to identify crowds in global regions. Our system has been proved to be capable of recognizing and localizing anomalous crowd behaviors, as shown by the experimental findings produced via the use of our test video sequences. They classified the methods for evaluating crowd behavior into two distinct categories: object-based and holistic-based methods. In the object-based methodology, a crowd is seen to be a collection of distinct individuals, but in the holistic methodology, the focus is on the organization as a whole rather than on the individual distinctions that exist within it. It is anticipated that each and every individual in a crowd would move with the same characteristics throughout the whole of the investigation while using this methodology. This study investigates the many facets that are associated with crowd modelling and crowd analysis. These are two fields that make use of a wide variety of methods for a wide variety of applications in the real world. This study investigates the many facets that are associated with crowd modelling and crowd analysis. It has been detailed how a complete review of the most recent advancements in methods for counting the number of people has been presented, together with a discussion of both the merits and shortcomings of the various approaches. The findings of this survey provide some insight into the likely trajectory of the monitoring and classification of crowds in the near future. It is vital to have a unified system that is capable of doing any kind of crowd analysis in order to achieve efficient crowd management. This system has to be able to manage a wide range of chaotic circumstances, from those involving small groups of people to those involving a huge number of people. This is done in order to discover the common gaps that are present in the existing procedures and to lay the foundation for future research in this specific subject. The goal of this study is to better understand how to do research in this particular topic. In spite of the fact that researchers have arrived at a conclusion, there are still a number of problems that have not been satisfactorily addressed, which indicates that further study is required.

REFERENCES

[1] Zhan, B, Monekosso, D.N., Remagnino, P., Velastin, S.A. & Xu, L.Q. (2008). Crowd analysis: A survey. Mach. *Vis. Appl.*, *19*, 345–357.
[2] Andrade, E.L., S. Blunsden & R.B. Fisher. (2006). Modelling crowd scenes for event detection. *Proceedings of the 18th International Conference on Pattern Recognition*, *1*, pp. 175-178.
[3] Kim, C. & J.N. Hwang. (2002). Object-based video abstraction for video surveillance systems. *IEEE Trans. Circuits Syst. Video Technol.*, *12*, 1128-1138.
[4] Hazel, G.G. (2000). Multivariate gaussian MRF for multispectral scene segmentation and anomaly detection. *IEEE Trans. Geosci. Remote Sens.*, *38*, 1199-1211.
[5] Mathur, Garima, Devendra Somwanshi & Mahesh M. Bundele. (2018). Intelligent video surveillance based on object tracking. *3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE). IEEE*.
[6] Kumar, Chethan & R. Punitha. (2020). Yolov3 and yolov4: Multiple object detection for surveillance applications. *Third International Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE*.
[7] Wang, Xiaofei, et al. (2014). A high accuracy flow segmentation method in crowded scenes based onstreakline. *Optik-International Journal for Light and Electron Optics 125*(3), 924–929.

e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-12, Issue-3 (June 2022)
https://doi.org/10.31033/ijemr.12.3.18

[8] Cong, Yang, Junsong Yuan & Ji Liu. (2013). Abnormal event detection in crowded scenes usingsparse representation. *Pattern Recognition 46*(7), 1851–1864.
[9] Zhan, Beibei, et al. (2008). Crowd analysis: a survey. *Machine Vision and Applications 19*(5–6), 345–357.
[10] Li, Teng, et al. (2015). Crowded scene analysis: A survey. *Circuits and Systems for Video Technology*, *IEEE Transactions on 25*(3), pp. 367–386.
[11] Lempitsky V. (2010). Zisserman a. learning to count objects in images. *Proceedings of the Advances in Neural Information Processing Systems; Vancouver, BC, Canada*, pp. 1324–1332.
[12] Camille Dupont, Luis Tobıas & Bertrand Luvison. (2011). Crowd-11: A dataset for fine-grained crowd behaviour analysis. In: *IEEE Xplore provided by the Computer Vision Foundation*.