Unusual Activity Detection in Crowd using Deep Learning

[1]Marvin Dabhi, [2]Mukti Shah, [3]Prutha Bharti, [4]Priyanka Puvar, [5]Bhagirath Prajapati

[1]Student, [2]Student, [3]Student, [4]Assistant Professor, [5]Associate Professor

Computer Engineering Department,

A.D. Patel Institute of Technology, Anand, India.

***ABSTRACT*:** Modern technology is making people's life easier, but the safety of life is also the most important thing. Crowded places like public events, stadiums, festival grounds, rally affects not only the comfort level of humans, but it also increases the risk of safety of pedestrians and other civilians. Heavy crowds may lead to major accidents, crowd-crush and causing an overall control loss. To reduce the risk of civilians in heavy crowds, we have worked with technology to manage the crowd. The applications of crowd management are manifold, ranging from crowd counting to human computer interaction. Research articulated here is focused on how to detect any unusual event in crowds at the early stage using modern technology like Deep learning so that it can be handled and managed timely and causes least harm to civilians. The concept of Convolutional Neural Network is utilized for processing of images and videos.

Keywords: **Crowd Management, Deep Learning, CNN, Image Processing, Element Extraction, Frame Division, Farneback.**

I. INTRODUCTION

Several countries in the world are already having difficulties induced by ever-increasing populations, while the frequency of human population growth is genuinely concerning. Cities, streets, communities, and other public places get overcrowded as a result of rapid population growth. The overcrowded places will result in push, mass rush, riots, crowd crushes and overall loss of control.

Crowd management is a public safety strategy by which huge crowds are managed to prevent crowd crushes, criminal damage, battles between drunk and disorderly people or violent clashes from breaking out. In fact, crowd crushes may lead to numerous deaths. Crowd management is often required at large gatherings or public places like street fairs, music concerts, stadiums, street protests, rallies etc. At some gatherings security guards and police are using metal detectors and sniffer dogs to avoid the entry of weapons and drugs into a venue but there are some gatherings where such facilities are not enough to detect unusual activity. Public gatherings like rallies, concerts, street fairs where the possibilities of fights, mass panic, accidents are high and require continuous surveillance. Which is not possible through traditional security systems. With global security issues and the ever increasing need for effective surveillance of public places, more attention has been given to intelligent visual inspection we have considered here.

The aim of our work is to suggest three inspections like tracking the human motion, detecting the direction of the motion and analysing the action. It has become easy to capture and monitor human actions by CCTV cameras around the globe. While video footage capture devices in today's world are more reliable and common, it requires constant human resource intervention to track and evaluate the footage that is not viable. At times or situations like this, intelligent systems with appropriate detection techniques come in handy and proven to be more efficient.

Research of this topic provides number of solutions like detection of unusual activity detection using images, detection through monitoring the screen, human activity reorganization. Different solutions approach different techniques for detecting the human activities. Number of algorithms are available for detecting unusual activities of human. In our work we have tried to analyse the videos of public events using the concept of convolution neural network in deep learning. We took several videos of natural and normal activities of human in crowd as a data set for pre-processing. Our system uses pre-processed data for mapping motion influence. System uses motion influenced vector in feature extraction which extracts feature from the frame and create set of feature vectors. For differentiating usual and unusual activities system performs clustering on the vector set. Vectors having same patterns or having similar patterns belongs to same cluster. Clustering differentiates both type of activity by following the patterns of vector. Which frame or blocks are not part of the usual activity cluster detected as an unusual activity. Detected pixels of frame or blocks are displayed with border. When frame or blocks are detected more than some numeric value alert message is shown on the monitor which can alert the security and management team.

The remaining portion of the paper designed as follows: In section II the workflow of the project is introduced. Section III presents result of the system. Future work of the system is presented in section IV. Finally, section V provides the conclusion.

II. WORKFLOW

The project's aim is to detect the unusual human activity from the crowd to improve the crowd security through the footage necessitates the crowd detection at an earlier stage. If the quality of the footage is smooth enough to identify the crowd, then crowd detection is possible. Detection of crowds also relies on training model we choose to train the dataset. Training the data can be defined as an initial set of raw data used to help models to understand how to apply technologies to learn and produce accurate results. We are approaching the CNN model using deep learning to train the dataset.

Workflow of the proposed work is as shown in the Fig.1. An Input is received from user in form of video file and sent to the trained model which is used for detecting unusual human activities from the video. Once input is received, it will be divided into frames and would be converted RGB to grayscale. Displacement of a pixel between two frames will be counted. When video is split into frames, it is further broken down into equivalent blocks. Threshold value will be defined and pixel displacement between two blocks is measured. If displacement value is greater than threshold value, then it is categorized as affected block and if displacement value is less than threshold value then it is categorized as normal block. Whenever a same affected block is repeatedly detected for more than 10 frames, an alert is generated.

Work is divided in the seven different modules: Pre-processing of input data, Optical flow after pre-processing, Frame Division, Motion Influence of Crowd and Element Extraction, CNN, Training and Testing [1]. Fig.1 introduces the general flow of the human activity detection.

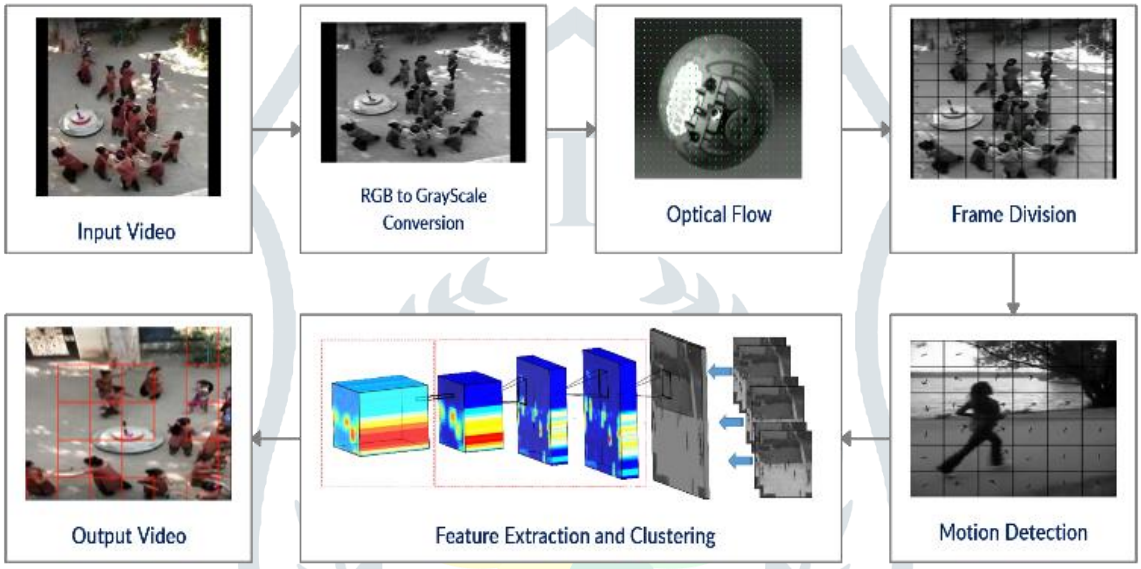

Fig.1 Proposed Architecture

Unusual activities are differentiated in two ways: Local and Global. When unusual event take place within limited area and various patterns that occurs in each frame of the video such as, one person moving in the opposite direction whereas others are moving in one direction, is defined as local activity. Where every pedestrian is moving in the same direction with the same speed of motion and suddenly the change occurs in the moving speed of all pedestrians such as, in a rally the crowd drives at the same speed in one direction and unexpectedly an animal joins the crowd and the crowd starts running suddenly is known as Global Activity [2]. Our research is more focused on global activity of the crowd.

2.1 Pre-Processing of Input Data

We provide video as an input which is subject to pre-processing. A video is viewed as a series of images called frames which are sequentially processed. Initially, RGB frames are transformed to a grey scale as grayscale images are entirely sufficient for many tasks and faster in processing. Fig.2 (a) and Fig.2 (b) show RGB and grayscale screenshot of video taken for pre-processing the data.

Fig.2 (a) Screenshot of input video in RGB form

Fig.2 (b) Screenshot of input video in Grayscale form

2.2 Optical flow after pre processing

Optical flow is the artifacts of motion, structures and borders in a video frame induced by relative movement between an object and the image. Optical is used to process all the pixels from the frame image. For each frame in the video optical flow is computed. We are using the farneback algorithm in the system as it processes all the pixels of image. Farneback algorithm is a dense technique for processing the frames [3]. Dense techniques are more accurate than sparse technique. A solution is to determine movement or displacement area from only two frames and attempt to adjust for background movements. The optical flow in OpenCV measures a complex optical flow that use the Gunnar Farneback's algorithm [3] [4].

2.3 Frame Division

After processing on the pixel of the frames, the system starts to divide the frames into several units called blocks. By dividing a screen in to the units after estimating optical flow within the same frame for every pixel, we divide the screen image into A by B equal units without any loss in which we can index the blocks by {P1, P2..., PAB}. If the size of the frame is 240 X 320 divided into 48 blocks where the size of each block is 20 X 20. Fig.3 describes that how the frame is divided into the blocks. It is the screenshot of the frame division of pre-processed video.

Fig.3 Frame division of pre-processed video

After separating the screen into units, we calculate each unit's optical flow by calculating the optical flow average of all the pixels that define a unit.

2.4 Motion Influence and Element Extraction

After separating a screen into units, the system moves forward to detect the motion of the crowd. Various factors like hurdles along the road, nearby pedestrians, animals etc may influence the direction of movement within a mass gathering.

We can define the feature of interaction as the influence of motion. We believe that two factors determine the units under effects on which a motion body will impact: Motion direction and momentum. The faster an element moves; the more adjacent blocks are affected under the object's impact. Neighbouring blocks have a greater impact than blocks afar. Fig. 4. Describes that how motion of human is detected. This figure is the example of the human motion detection.

Fig.4. Motion Detection

Algorithm: [8]:

Take Input as motion vector set

Output should be motion influence map vector

Step 1: Initially, motion influence map of any block is set to 0 at the starting of every frame.

Step 2: Compute threshold value for every moving object in every block in frame.

Step 3: If vector value of moving object and block is not equal then go to step 4.

Step 4: Calculate Euclidian distance between the origin and the motion vector.

Step 5: Validate threshold value if it is greater than Euclidian distance then go to step 6.

Step 6: Calculate angle between an origin and motion vector.

Step 7: Find out the direction of motion vector

Step 8: If vector is in satisfactory state, the motion influence weight with vector position will be determined.

Step 9: End

A unit in where a suspicious behaviour of human takes place, together with its adjacent block, has unique vectors of motion effect. [1] [9]. In addition, since any behaviour is detected by several subsequent frames a feature vector is extracted identified by n x n blocks. The system starts making mega blocks after extracting elements. Creating Mega blocks frames are classified into the blocks which does not overlap, one of each is a collection of many blocks of movement influence. The motion influence value of megablock is equal to the sum of the movement influence values of all the little blocks that form a bigger block. Extracting elements which are divided into mega blocks after the recent frame ‘t’. An 8 x t dimensional concatenated elements vector is extracted from all frames for each mega block. For instance, we consider a big block of all frames and combine their element vectors to construct a concatenated block element vector. [1]. Now the system will perform clustering on the mega blocks. Clustering is performed using the functions spatio-temporal and fix the centre values. To train the data model, we use video clips of natural and common activities. The mega block centre values thus design the patterns of normal behaviour that might occur in concerned area. Fig.5 Illustrates construction of motion influence map with different scenarios.

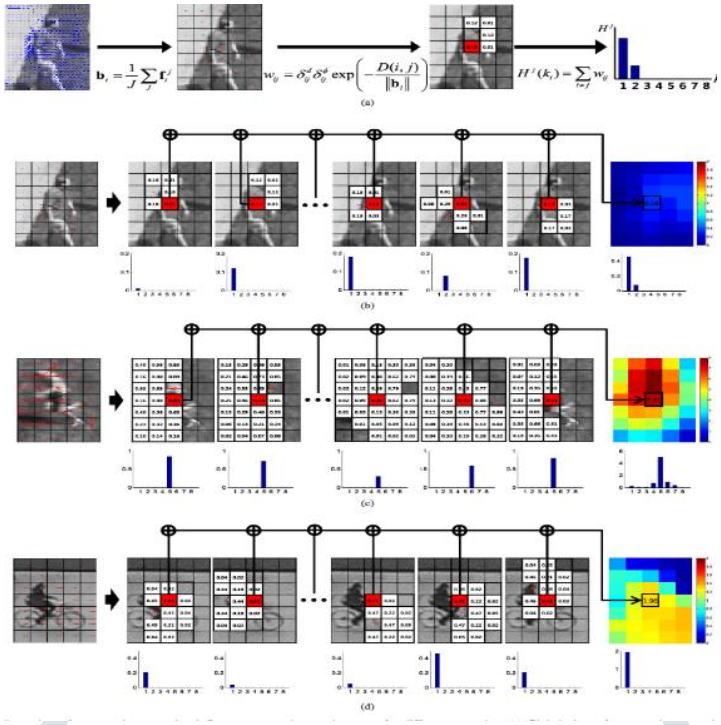

Fig.5. illustration of constructing a motion influence map and exemplar maps for different scenarios. [5]

2.5 CNN – Convolutional Neural Network

Deep learning algorithms such as artificial neural networks, convolution neural networks have been introduced to fields where values equal to or equivalent to human experience have been obtained. A convolutional neural network (CNN) in deep learning is a subset of deep neural networks, most generally applied to visual imaging research. CNNs are multilayer variants of regularized perceptron. CNN consists of an input layer, output layer and many hidden layers. Fig.6 describes the proposed CNN model workflow with respect to frames considered in this article. CNN model detects the feature in both spatial and temporal direction.

We apply the idea of CNN to the input images to create the element maps. While programming the CNN, it takes the input with number of images x number of heights x number of width x number of depths. After passing through the convolutional layer the image becomes simplified to an element map with number of images x number of heights x number of depths. Then the result is passed to a classifier to detect the object contained [6].

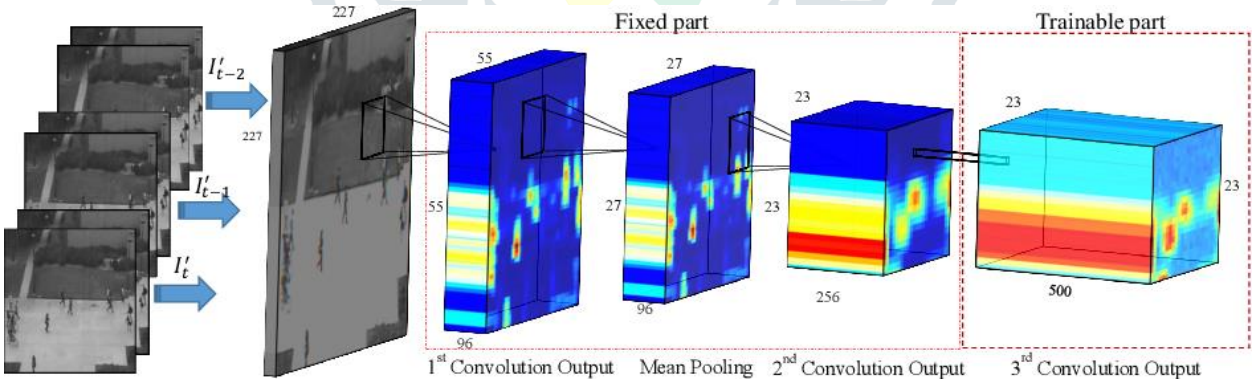

Fig.6 Proposed CNN structure [6]

2.6 Training and Testing

Available data are typically split into three sets: a training, validation and test set. A training set is used to train a network where loss values are calculated via forward propagation. And learnable parameters are updated with backpropagation. Backpropagation algorithm is the approach widely used to train neural networks, where the concept of loss and the algorithm for optimizing gradient descent play important roles. During the training cycle, a validation collection is used to track the model performance. Ideally, a test set is only used once at the very end to assess the efficiency of the final model, which is selected with training and validation sets on the testing phase [1] [7]. We have taken various video of normal activities in the crowd as a raw data for training. After the videos were collected, we placed them in a folder. We have taken video footage of the clashing rally for testing the data. After extracting the spatio-temporal element vectors for all mega blocks, we create a minimum distance matrix over the mega blocks. In which, in the corresponding mega block, the value of an element is determined by the minimum Euclidean distance between the element vector of the current test frame and the mega block centre values.

The smaller the value of an element the less likely the block will have an anomalous activity. On the other hand, if there is a higher value in the minimal-distance matrix, we can assume that in consecutive frames unusual events occur. In the minimum-distance matrix, therefore, we consider the highest value to be the frame representative element value. If the maximum value of the minimum distance matrix is greater than the threshold then the current frame is marked as unusual [1] [8].

III. RESULT

This system can be useful in detecting any abnormal activity in the crowd that may take a large form if not attended at early stage. Detecting a human behaviour or activity is generally achieved with the aid of deep learning and neural networks. This work will detect a human being's motion or speed and if it exceeds the threshold value warning will appear on the screen.

We believe that it is necessary to understand human behaviour, which involves observing human actions and interaction among themselves. Differentiating between usual and anomalous human behaviour is an essential part of the system.

Fig. 7 Suspicious activity frame is detected

If any type of movement occurs on the screen it will be detected by small frames. If the screen detects 8 or more than 8 frames which include any suspicious behaviour of the crowd then pop-up message saying that “Unusual activity” is displayed in order to alert the management. We took a video of rally clashing which contains the maximum suspicious activity. The blocks which are pop with a red border as shown in Fig. 7 detects the suspicious activity and Fig. 8 indicates the alert message that suspicious activity is detected.

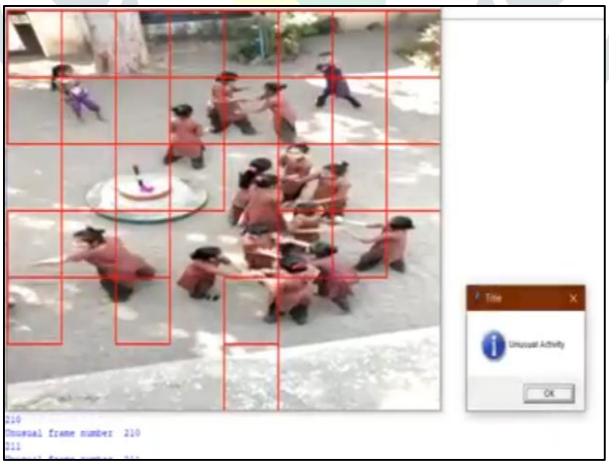

Fig. 8 Alert message pop-up

This system can be very much useful for the events where constant monitoring is required.

IV. FUTURE WORK

In the future we hope to improve the system’s performance. The current system displays alert message only on the computer screen. After getting alert message, supervisor who monitors the screen continuously send the message to the security team or management team. This process taken long time. In order to reduce the time of alerting process this system can be extended by providing direct message to the security or management team’s mobile phone. The alert message also can be improved by giving the exact location of where the activity has been detected. Mobile based application can be utilized to provide mobility to enhance the current system. Mobile based application provides the mobility to our system. Through mobile based application, security team or management team can monitor the crowd activity at anywhere and anytime.

V. CONCLUSION

Human behaviour analysis and on-going activity identification has become a revolutionary field for study. Advancement in technology allowed the identification of complex human activities. Automatic human activity identification has many applications, such as early recognition of unusual human activity, public safety at busy areas, prevention of terrorist attacks, smart surveillance.

We have identified, classified and monitored human behaviours in this study. Detecting any motion in a complex scene is a challenging job in terms of weather changes and detection of a shadow. There were different monitoring strategies for individuals and groups of people. After researching some of the strategies, we chose to go along with a hybrid approach to overcome the obstacle of tracked entities, losing the target due to rapid motions and cluttering in the background. We find the approach that provides the best result for periodic gestures and behaviour.

REFERENCES

[1] D. Lee, H. Suk, S. Park and S. Lee, "*Motion Influence Map for Unusual Human Activity Detection and Localization in Crowded Scenes*," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 25, no. 10, pp. 1612-1623, Oct. 2015, doi: 10.1109/TCSVT.2015.2395752.

[2] Igor R. de Almeida, Vinicius J. Cassol, Norman I. Badler, Soraia Raupp Musse, Cláudio Rosito Jung, "*Detection of Global and Local Motion Changes in Human Crowds*", *Circuits and Systems for Video Technology IEEE Transactions on*, vol. 27, no. 3, pp. 603-612, 2017.

[3] Farnebäck, Gunnar. "*Two-frame motion estimation based on polynomial expansion.*" *Scandinavian conference on Image analysis*. Springer, Berlin, Heidelberg, 2003.

[4] de Boer, Jasper, and Mathieu Kalksma. "*Choosing between optical flow algorithms for UAV position change measurement.*" *12th SC@ RUG 2014-2015* (2015): 69.

[5]https://www.semanticscholar.org/paper/Motion-Influence-Map-for-Unusual-Human-Activity-and-Lee-Suk/3850dc2d460ea734b7ae8e466923e9269fe0fcff/figure/9

[6] Sabokrou, Mohammad, et al. "*Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes.*" *Computer Vision and Image Understanding* 172 (2018): 88-97.

[7]https://towardsdatascience.com/train-validation-and-test-sets72cb40cba9e7

[8] http://ijtimes.com/papers/finished_papers/IJTIMESV05I07150730155609N.pdf

[9] Li, Weixin, Vijay Mahadevan, and Nuno Vasconcelos. "Anomaly detection and localization in crowded scenes." *IEEE transactions on pattern analysis and machine intelligence* 36.1 (2013): 18-32.