iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 Issue: VII Month of publication: July 2021

DOI: https://doi.org/10.22214/ijraset.2021.36368

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 VII July 2021- Available at www.ijraset.com

Video Based Suspicious Human Behaviour Recognition System

Dr. Chetana Prakash[1], Aneesabanu[2], Deeksha J Udasi[3], Gousiya Banu I[4], Chandana M[5]
[1, 2, 3, 4, 5] *Department Computer Science and Engineering, BIET Davanagere*

Abstract: In recent years, the number of surveillance cameras installed to monitor private and public spaces has increased rapidly. The demand has raised for smarter video surveillance of public and private spaces using intelligent vision systems which can differentiate between 'suspicious' and 'unsuspicious' behaviours according to the human observer. Generally, the video streams are constantly recorded or monitored by operators. In these cases, an intelligent system can give more accurate performance than a human. We have proposed a method called motion influence map under machine learning for representing human activities. Optical-flow is computed for each pixel in a frame that are processed sequentially. The key feature of the proposed motion influence map is that it effectively reflects the motion characteristics such as movement speed, movement direction, and size of the objects or subjects and their interactions within a frame sequence. It further extracts frames of high motion influence values and compares with the testing frames to automatically detect suspicious activities.

Keywords: Optical Flow, Motion Influence Map, Suspicious behaviour

I. INTRODUCTION

In recent years, automatic human activity recognition has drawn much attention in the field of video analysis technology due to the growing demands from many applications, such as surveillance environments, entertainment environments and healthcare systems. In a surveillance environment, the automatic detection of abnormal activities can be used to alert the related authority of potential criminal or dangerous behaviours, such as automatic reporting of a person with a bag loitering at an airport or station. Similarly, in an entertainment environment, the activity recognition can improve the human computer interaction (HCI), such as the automatic recognition of different player's actions during a tennis game so as to create an avatar in the computer to play tennis for the player. Furthermore, in a healthcare system, the activity recognition can help the rehabilitation of patients, such as the automatic recognition of patient's action to facilitate the rehabilitation processes. There have been numerous research efforts reported for various applications based on human activity recognition, more specifically, home abnormal activity, ballet activity, tennis activity, soccer activity, human gestures, sport activity , human interaction, pedestrian traffic and simple actions, and healthcare applications.

Human action detection, motion tracking, scene modelling, and behaviour understanding (human activity recognition and discovery of activity patterns) have received a lot of attention in the computer-vision and machine-learning communities. Applications have been in—but not limited to—video surveillance, human–computer interfaces, and multimedia semantic annotation and indexing. Intelligent visual surveillance has got more research attention and funding due to increased global security concerns and an ever increasing need for effective monitoring of public places such as airports, railway stations, shopping malls, crowded sports arenas, military installations, etc., or for use in smart healthcare facilities such as daily activity monitoring and fall detection in old people's homes. Often times, the objective is to detect, recognize, or learn interesting events which contextually may be defined as "suspicious event", "irregular behaviour", "uncommon behaviour", "unusual activity/event/behaviour" , "abnormal behaviour", "anomaly" , etc.

Among three video surveillance research directions namely detection and tracking, human motion analysis, and activity analysis (parsing temporal sequences of object observations to produce high-level descriptions of agent actions and multi agent interactions) "activity analysis will be the most important area of future research in video surveillance." This projection appears no less true today as research publications in this field over the last decade show. The use of closed-circuit television (CCTV) cameras to capture and monitor scenes by human agents has become ubiquitous. Although video footage capturing devices are more affordable and popular in today's world, available human resources to monitor and analyse the footage are quite limited and sometimes not cheap. In many situations where surveillance cameras are used, it is common to find poor monitoring due to human factors like fatigue. The CCTV operators suffer boredom because in most cases, nothing "strange" or something that catches the attention occurs in the scene.

545

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 VII July 2021- Available at www.ijraset.com

A. Problem Statement

As the importance of security increases in human life, a great number of surveillance cameras have been installed in private and public places. However, the plethora of video sequences available is overwhelming the human resources monitoring them. They perform well but are limited to detecting only those specific types of predefined anomalies. To overcome this problem, there has been significant interest in a smart surveillance system that can automatically detect unusual or abnormal activities.

B. Objectives

The main goal of our work is to detection of suspicious human activities in the videos. In order to attain the goal, the following objectives have been framed:

1) To collect the datasets in the form of videos and perform pre-process.
2) Identify the heed points in the videos which acts as features and extracted through motion influence map generation technique.
3) To develop a highly coherence abnormal human activity detection system which are the behaviour representation and modelling.
4) To calculate the metrics for performance evaluation with our datasets.

C. Proposed System

We describe a method for representing motion characteristics for the detection and localization of unusual activities within a crowded scene. We considered two types of unusual activities: local and global. Local unusual activities occur within a relatively small area. Different motion patterns may appear in a portion of the frame, such as the unique appearance of non-human objects or the fast movement of a person when most of the other pedestrians are walking slowly. Global unusual activities occur across the frame, for example, when every pedestrian within a scene starts to run suddenly to escape from the scene. Given a sequence of frames, the motion information at both the pixel-level and block-level is computed sequentially. Based on the block-level motion information, the motion influence energy is computed and a motion influence map is then constructed from the energies in each frame. The proposed motion influence map represents both the spatial and temporal characteristics within a single feature matrix. For the classification, we divide the motion influence map into an uniform grid, and perform the k-means clustering for each region. The distances between the centre of the clusters and each extracted spatio-temporal motion influence feature are used as the feature values for unusual activity detection at the frame-level. Once a frame is classified as unusual, we further localize the exact position of the unusual activity at the pixel-level.

II. LITERATURE SURVEY

In 2007, G. Lavee, L. Khan, and B. Thuraisingham proposes a framework to aid video analysts in detecting suspicious activity within the tremendous amounts of video data that exists in today's world of omnipresent surveillance video. Ideas and techniques for closing the semantic gap between low-level machine readable features of video data and high-level events seen by a human observer are discussed. An evaluation of the event classification and detection technique is presented and a future experiment to refine this technique is proposed.

In 2004, H. Zhong, J. Shi, andM. Visontai present an unsupervised technique for detecting unusual activity in a large video set using many simple features. No complex activity models and no supervised feature selections are used. They divide the video into equal length segments and classify the extracted features into prototypes, from which a prototype–segment co-occurrence matrix is computed. Motivated by a similar problem in document-keyword analysis, they seek a correspondence relationship between prototypes and video segments which satisfies the transitive closure constraint. They show that an important sub-family of correspondence functions can be reduced to co-embedding prototypes and segments to N-D Euclidean space. They prove that an efficient, globally optimal algorithm exists for the co-embedding problem. Experiments on various real-life videos have validated our approach.

In 2009, R. Mehran, A. Oyama, and M. Shah introduce a novel method to detect and localize abnormal behaviors in crowd videos using Social Force model. For this purpose, a grid of particles is placed over the image and it is adverted with the space-time average of optical flow. By treating the moving particles as individuals, their interaction forces are estimated using social force model. The interaction force is then mapped into the image plane to obtain Force Flow for every pixel in every frame. Randomly selected spatial-temporal volumes of Force Flow are used to model the normal behavior of the crowd. They classify frames as normal and abnormal by using a bag of words approach. The regions of anomalies in the abnormal frames are localized using interaction forces. The experiments are conducted on a publicly available dataset from University of Minnesota for escape panic scenarios and a challenging dataset of crowd videos taken from the web.

546

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 VII July 2021- Available at www.ijraset.com

The experiments show that the proposed method captures the dynamics of the crowd behavior successfully. In addition, they have shown that the social force approach outperforms similar approaches based on pure optical flow.

In 2010, R. Poppe made survey on vision-based human action recognition is the process of labelling image sequences with action labels. Robust solutions to this problem have applications in domains such as visual surveillance, video retrieval and human–computer interaction. The task is challenging due to variations in motion performance, recording settings and inter-personal differences. In this survey, they explicitly address these challenges. They provide a detailed overview of current advances in the field. Image representations and the subsequent classification process are discussed separately to focus on the novelties of recent research. Moreover, they discuss limitations of the state of the art and outline promising directions of research.

In 2008, J. Yin, Q. Yang, and J. J. Pan present a novel two-phase approach for detecting abnormal activities based on wireless sensors attached to a human body. Detecting abnormal activities is a particular important task in security monitoring and healthcare applications of sensor networks, among many others. Traditional approaches to this problem suffer from a high false positive rate, particularly when the collected sensor data are biased towards normal data while the abnormal events are rare. Therefore, there is a lack of training data for many traditional data mining methods to be applied. To solve this problem, our approach first employs a one-class support vector machine (SVM) that is trained on commonly available normal activities, which filters out the activities that have a very high probability of being normal. They then derive abnormal activity models from a general normal model via a kernel nonlinear regression (KNLR) to reduce false positive rate in an unsupervised manner. They show that our approach provides a good trade-off between abnormality detection rate and false alarm rate, and allows abnormal activity models to be automatically derived without the need to explicitly label the abnormal training data, which are scarce. They demonstrate the effectiveness of our approach using real data collected from a sensor network that is deployed in a realistic setting.

In 2015, R. Hamid, A. Johnson, S. Batta, A. Bobick, C. Isbell, and G. Coleman present a novel representation and method for detecting and explaining anomalous activities in a video stream. Drawing from natural language processing, we introduce a representation of activities as bags of event n-grams, where we analyze the global structural information of activities using their local event statistics. They demonstrate how maximal cliques in an undirected edge-weighted graph of activities, can be used in an unsupervised manner, to discover regular sub-classes of an activity class. Based on these discovered sub-classes, we formulate a definition of anomalous activities and present a way to detect them. Finally, they characterize each discovered sub-class in terms of its “most representative member,” and present an information theoretic method to explain the detected anomalies in a human-interpretable form.

In 2011, D. Ryan, S. Denman, C. Fookes, and S. Sridharan propose a novel visual representation called textures of optical flow. The proposed representation measures the uniformity of a flow field in order to detect anomalous objects such as bicycles, vehicles and skateboarders; and can be combined with spatial information to detect other forms of abnormality. They demonstrate that the proposed approach outperforms state-of-the-art anomaly detection algorithms on a large, publicly-available dataset.

III. SYSTEM DESIGN

System design thought as the application of theory of the systems for the development of the project. System design defines the architecture, data flow, use case, class, sequence and activity diagrams of the project development.

A. System Architecture

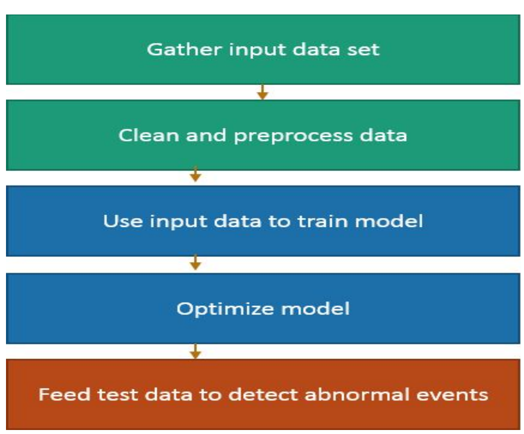


Fig. 1 Architecture Diagram

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 VII July 2021- Available at www.ijraset.com

The proposed methodology consists of following phases for developing abnormal human activity system respectively.

1) *Data set Collection*: We use UCSD and UMN dataset consists of 11 video clips of crowded escape scenarios from three different indoor and outdoor scenes. It includes 7,740 frames in total, where the frame size is 320240. Each video clip starts with people walking around the scene. In this case, walking was considered a normal activity. In each clip, each of the subjects makes a sudden quick running movement away from the scene which was considered unusual in this dataset. We take these video file as our input video for detecting abnormal human activities. The video file extension was .avi (audio video Interleave).
2) *Data Pre-processing:* The video file is given as a contribution to the framework, which is exposed to pre-handling. A video is treated as succession of pictures called outlines and these edges are handled consecutively. An RGB outline is first changed over to gray scale. A gray scaled picture comprises of just the force data of the picture as opposed to the clear hues. RGB vector is 3 dimensional while gray scaled vector is one dimensional.
3) *Feature Extraction:* we extracted features of the implementation phase map module in motion. Extraction work on the movement impact map, a square where uncommon action happens alongside its neighboring squares, has unique vector development impact. In this way, as an activity is caught by different successive edges, over the latest t number of casings, we remove a component vector from a cuboid distinguished by n/n squares. Extricating Features Following the ongoing ' t ' number of casings partitioned into uber hinders, a 8 t-dimensional linked capacity vector is separated over all edges for each super square.
4) *K-means Clustering:* For extracted features, we apply k-means clustering algorithm. We cluster for each mega block utilizing the spatio-fleeting highlights, and set the focuses as codewords. That is, we have K codewords for the (I, j) super square, {w (i,j) k} k=1. Here, we recall that we just use video clasps of typical exercises in our preparation arrange. Henceforth, a super square's codewords model the examples of regular exercises that may happen in the area concerned.
5) *Frame Level Detection:* In this, the lower the estimation of a thing, the more outlandish an abnormal movement is to happen in the individual square, the edge level identification of uncommon exercises in a base separation lattice. Then again, if a higher worth happens in the negligible separation lattice, we can expect that there are surprising occasions in t back to back edges. Furthermore, we locate the most elevated an incentive in the base separation lattice as the agent estimation of the casing. In the event that the base separation framework's most noteworthy worth is more prominent than the limit, at that point we distinguish the present casing as uncommon.
6) *Pixel Level detection:* Discovery of unpredictable exercises Pixel level Once an edge is recognized as unordinary, we look at the estimation of each uber square's base separation network with the edge esteem. In the event that the worth is more noteworthy than the edge, we mark the square as irregular.

B. Dataflow Diagram

Data flow diagram also referred as bubble graph. This diagram is useful for representing the system for all degree of constructions. The figure is differentiated into parts which show maximizing data path & practical aspect. The below figure 2 shows the dataflow of the proposed system.

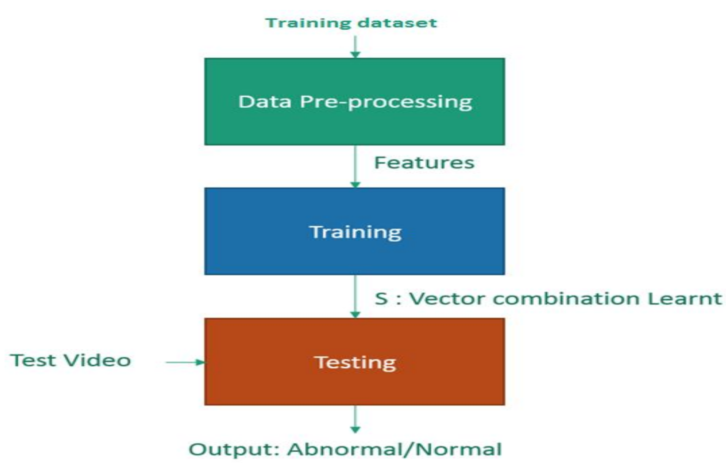


Fig. 2 Dataflow Diagram

548

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 VII July 2021- Available at www.ijraset.com

IV. IMPLEMENTATION

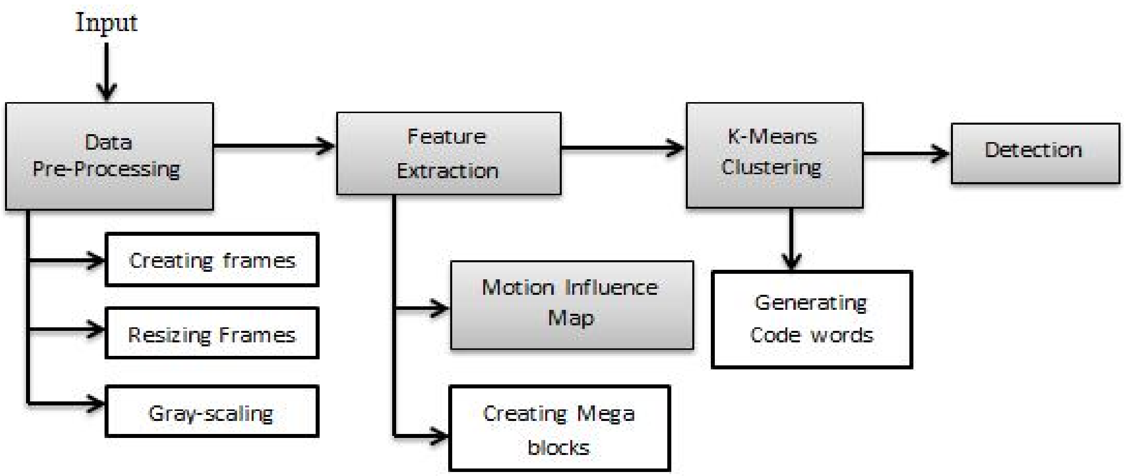


Fig 3: Implementation of proposed system for human abnormal detection in video frames

The implementation is divided into 5 modules optic flow of blocks, motion influence generator, create megablocks, training and testing.

In this section, a method for representing motion characteristics is described for the detection and localization of unusual activities within a crowded scene. Here, we should note that we considered two types of unusual activities:

1) local and 2) global.

Local unusual activities occur within a relatively small area. Different motion patterns may appear in a portion of the frame, such as the unique appearance of nonhuman objects or the fast movement of a person when most of the other pedestrians are walking slowly. Global unusual activities occur across the frame, for example, when every pedestrian within a scene starts to run suddenly to escape from the scene.

The video file is given as an input to the system, which is subjected to pre-processing. A video is treated as sequence of images called frames and these frames are processed sequentially.

A. Optical-Flow of blocks

After the pre-processing step, for each frame in the video, optical-flow is computed for each pixel in a frame. Optical flow is the pattern of apparent motion of objects, surfaces, and edges in a visual scene caused by the relative motion between an observer and the scene. Optical-flow is a vector of the form (r , θ), where, r represents the magnitude of each pixel and θ represents the direction in which the each pixel has moved relative to the corresponding pixel in the previous frames. The calcOpticalFlowFarneback( ) function in openCV computes a dense optical flow.

1) *Dividing a frame into blocks:* After computing the optical flows for every pixel within a frame, we partition the frame into M by N uniform blocks without loss of generality, where the blocks can be indexed by {B1, B2, . . . , BMN}.
2) *Calculating Optical-Flow of each block:* After dividing the frames into blocks, we compute optical-flow of each block by computing the average of optical-flows of all the pixels constituting a block. Figure 4 gives the formula for calculating the optical-flow of a block. Here, bi denotes an optical flow of the ith block, J is the number of pixels in a block, and f ji denotes an optical flow of the jth pixel in the ith block. Optical-flow of a block is a vector (r , θ) which represents how much each block has moved and in which direction compared to the corresponding block in the previous frames.

B. Motion Influence Generator

1) *Motion Influence Map:* The movement direction of a pedestrian within a crowd can be influenced by various factors, such as obstacles along the path, nearby pedestrians, and moving carts. We call this interaction characteristic as the motion influence. We assume that the blocks under influence to which a moving object can affect are determined by two factors:

a) The motion direction

b) The motion speed.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

The faster an object moves, the more neighbouring blocks that are under the influence of the object. Neighbouring blocks have a higher influence than distant blocks.

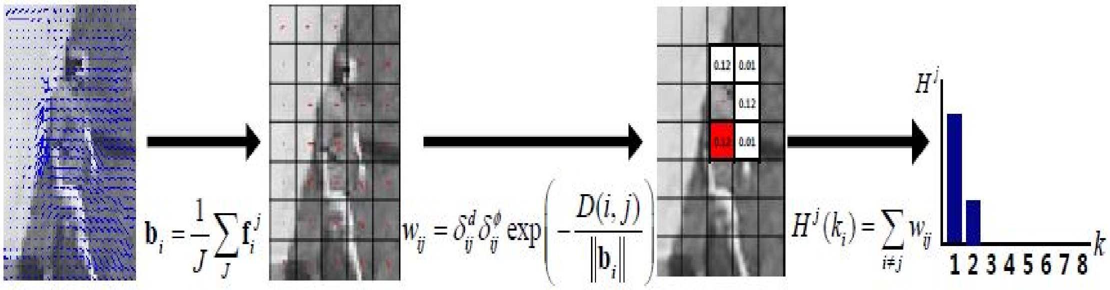


Fig 4 (first)optical flow in a pixel-level, (second) motion vector in a block-level, (third and fourth)computing a motion influence weight and the corresponding a motion influence vector for red-colored block.

2) Algorithm for creating a motion influence map

```
INPUT: B ← motion vector set, S ← block size, K ← a set of blocks in a frame
OUTPUT: H ← motion influence map
H j (j   K) is set to zero at the beginning of each frame  ∈
for all i   K do
Td = bi × S;
Fi/2 = bi + ? 2 ; −Fi/2 = bi − ? 2 ;
for all j   K do
if i = j then
Calculate the Euclidean distance D(i, j) between
bi and bj
if D(i, j) < Td then
Calculate the angle ?ij between bi and bj
if − Fi 2 < ?ij < Fi 2 then
H j ( bi) = H j ( bi) + exp− D(i,j)bi
end if
end if
end if
end for
end for
```

C. Create Megablocks

In the motion influence map, a block in which an unusual activity occurs, along with its neighbouring blocks, has unique motion influence vectors. Furthermore, since an activity is captured by multiple consecutive frames, we extract a feature vector from a cuboid defined by n × n blocks over the most recent t number of frames. After the recent 't' number of frames are divided into Megablocks, for each megablock, an 8 × t-dimensional concatenated feature vector is extracted across all the frames. For example, we take mega block (1,1) of all the frames ('t' number of frames) and concatenate their feature vectors, to create a concatenated feature vector for block (1,1).

D. Training and Testing

1) *Training* **:** For each mega block, we perform clustering using the spatio-temporal features and set the centers as codewords. That is, for the (i , j)th mega block, we have K codewords, {w(i, j) k }K k=1. Here, we should note that in our training stage, we use only video clips of normal activities. Therefore, the codewords of a mega block model the patterns of usual activities that can occur in the respective area.
2) *Testing:* Now that we have generated the codewords for normal activities, it is time to test the generated model with a test dataset which contains unusual activities. In the testing state, after extracting the spatio-temporal feature vectors for all mega blocks, we construct a minimum distance matrix E over the mega blocks, in which the value of an element is defined by the minimum Euclidean distance between a feature vector of the current test frame and the codewords in the corresponding mega block.

550

V. RESULTS

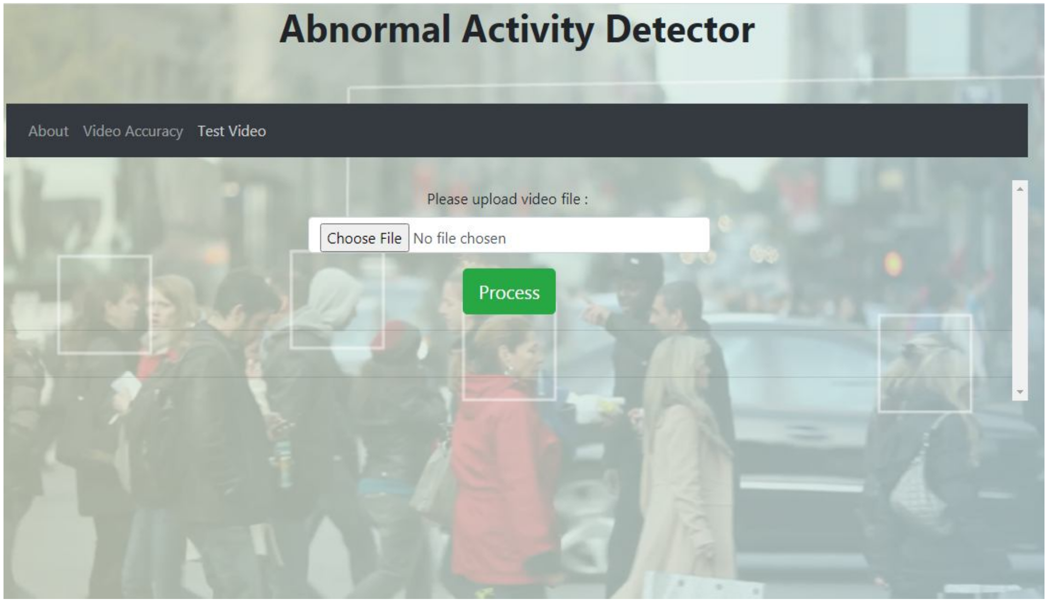


Fig 5 Front end for training and testing process

Figure 5 shows the front end web application consists of train classifier button for training process. For testing purpose we can upload test videos using browse and process button to start testing process.

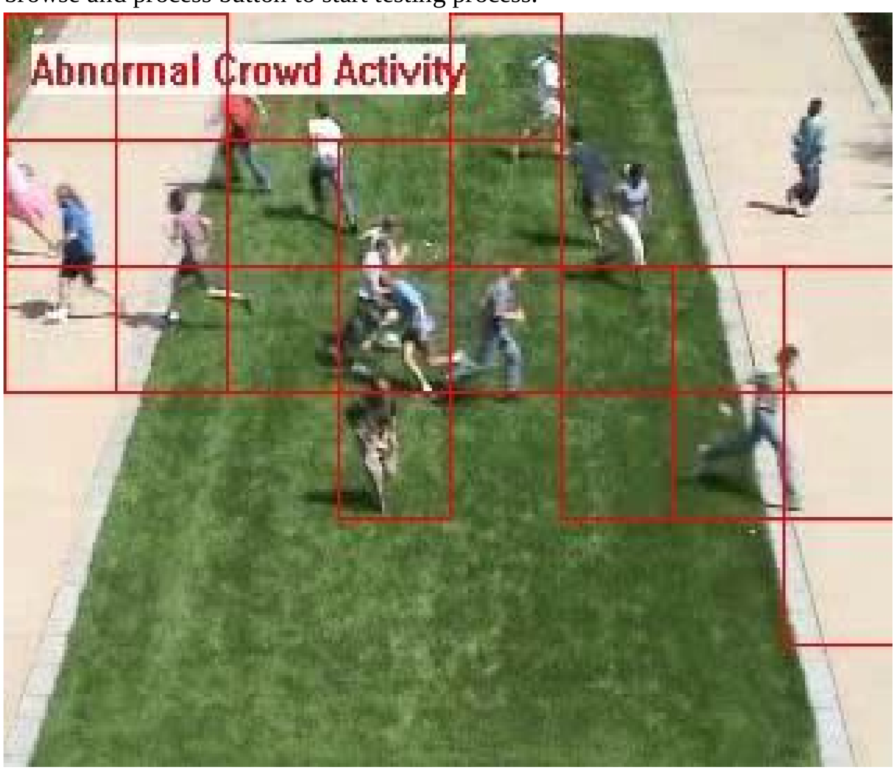


Fig 6 Unusual activities frame is detected

Figure 6 shows the output of our proposed system which consists of unusual activity frame in which red rectangle shows the pixel level abnormal activity identification.

551

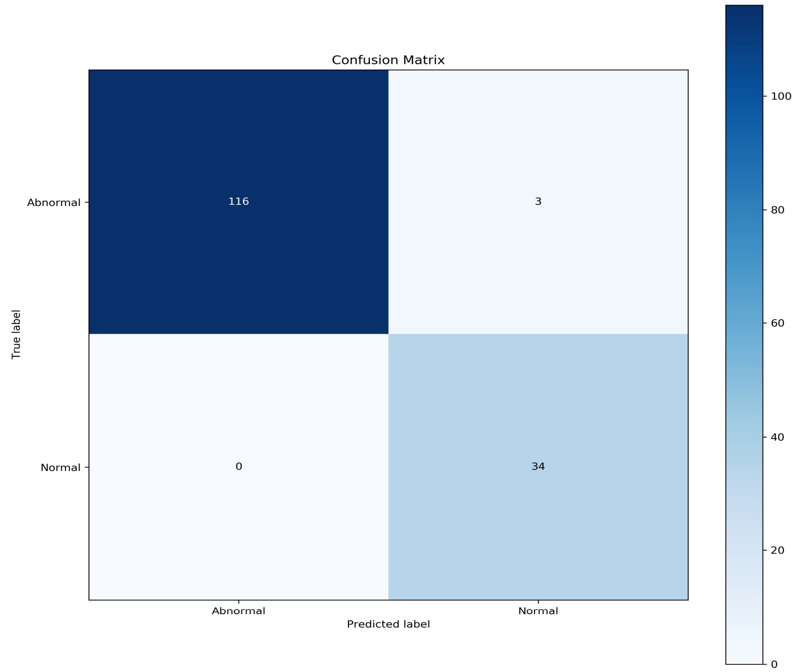


Fig 7 Confusion Matrix

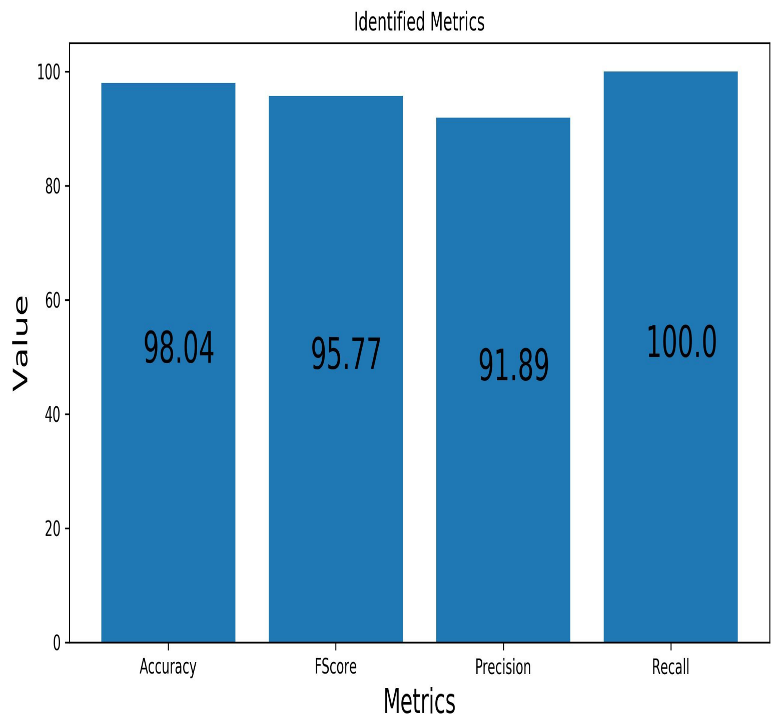


Fig 8 Performance Analysis

We evaluate the performance using Precision, Recall and F-Score. Figure 8 shows the performance metrics results. We compute the confusion matrix which shows that true positivity is very high and accuracy of the system is nearly 98.04. Below table 1 shows the performance analysis values with performance metrics.

Table1.Performance analysis of unusual human activity detection in video

Metrics	Percentage
Accuracy	98
Precision	83.3
Fscore	74.83
Recall	99

The formulas used to find Accuracy, Precision, FScore and Recall :
Accuracy: Accuracy is one metric for evaluating classification models. Informally, accuracy is the fraction of predictions our model got right. Formally, accuracy has the following definition:

Accuracy = Number of correct predictions / Total number of predictions
Accuracy = (TP +TN)/total.

Precision: Precision is the ratio of correctly predicted positive observations to the total predicted positive observation

Precision = TP/Prediction.

FScore: F-score, also called the F1-score, is a measure of a model's accuracy on a dataset The F-score is a way of combining the precision and recall of the model, and it is defined as the harmonic mean of the model’s precision and recall.

FScore = TP/Actual detection.

Recall**:** Recall is calculated as the number of true positives divided by the total number of true positives and false negatives. The result is a value between 0.0 for no recall and 1.0 for full or perfect recall

Recall = (2*precision*recall) / (precision + recall)

Where, TP= True Positive detections.

TN= True Negative detection

A. Comparison

Table 2: Comparison Table

Model	Accuracy
AMDT+LDP	87.10
Motion influence map	98.33

As we observed Abnormal Event Detection in Crowd Scene Based On the Unsupervised Learning - July 2020. They have used LDP based prediction to build the model and got the accuracy of 87.10%. Here we have used Motion influence map for prediction and we got accuracy of 98.33%.

553

VI. CONCLUSION

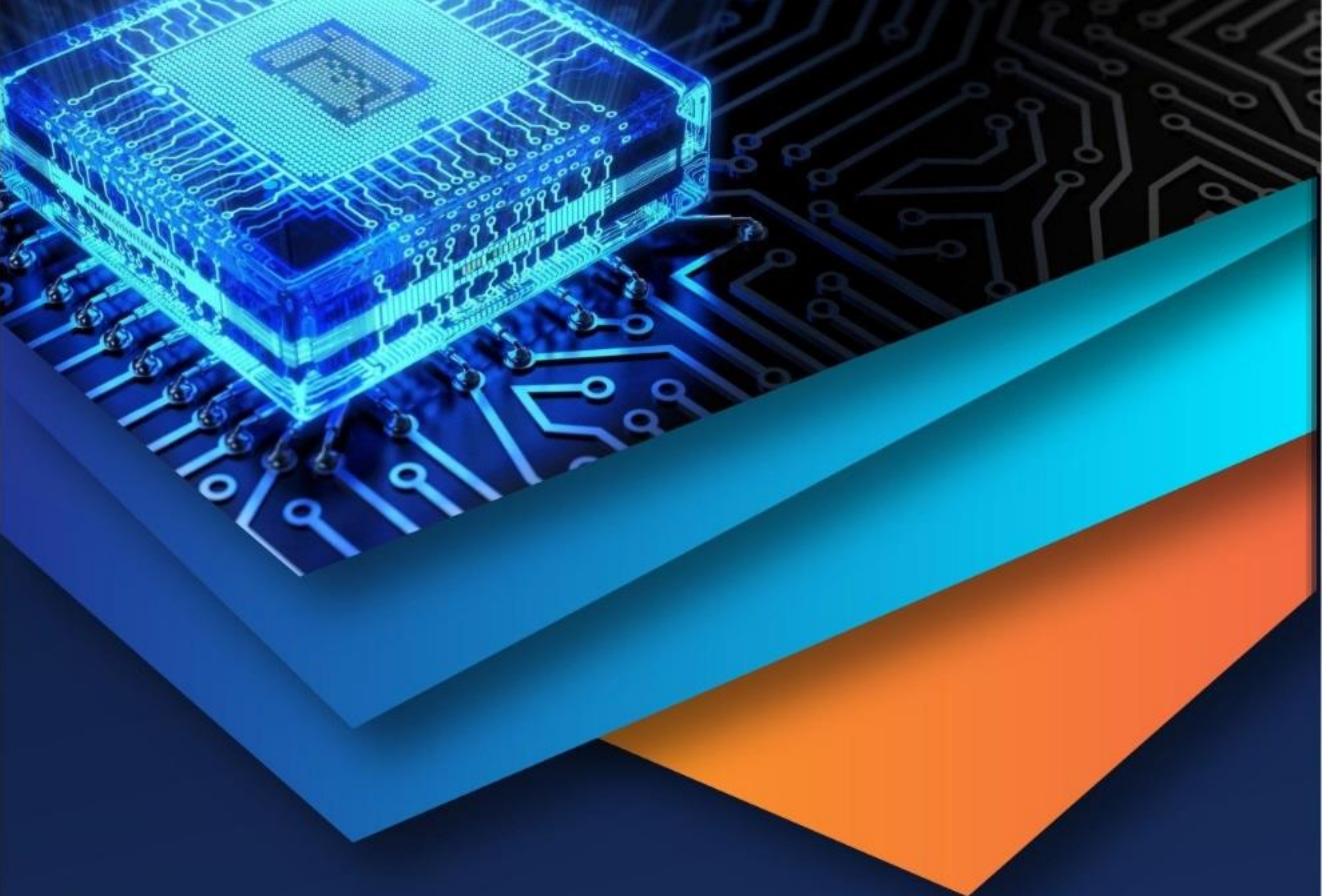
With the growing number of surveillance cameras installed in private and public areas, there has been a demand for the automatic and intelligent analysis of video sequences using computers. Unusual event or activity detection in videos has recently been of great interest in the area of vision based surveillance. In this paper, we proposed a method for representing the motion characteristics within a frame to detect and localize unusual human activities in a videos. Owing to the representational power of the proposed motion influence map for both space and time, we can classify a frame as usual or unusual, and localize the areas of unusual activities within a frame. For a real application, a smart surveillance system needs to efficiently detect both local and global unusual activities within a unified framework. In our experimental studies on two public datasets, i.e., the UMN and UCSD datasets, we validated the effectiveness of the proposed method, which outperformed other competing methods from the literature.

The main focus of this work is to detect unusual activities within surveillance video scenes, for which the cameras usually cover a wide area, resulting in small objects being present in the scene without significant perspective changes. Also our experiments were limited to a fixed viewpoint, and there is a limitation in the applicability of the approach for surveillance cameras with pan, zoom, or tilt functionality. At this moment the proposed method deals only with static cameras. However, it can be easily extended to PTZ cameras using localization results.

REFERENCES

[1] G. Lavee, L. Khan, and B. Thuraisingham, “A framework for a video analysis tool for suspicious event detection,” Multimedia Tools Appl., vol. 35, pp. 109–123, 2007.
[2] H. Zhong, J. Shi, andM. Visontai, “Detecting unusual activity in video,” in Proc. 2004 IEEE Comput. Vis. Pattern Recog., 2004, vol. 2, pp. II- 819–II-826.
[3] R. Mehran, A. Oyama, and M. Shah, “Abnormal crowd behavior detection using social force model,” in Proc. IEEE Conf. Computer Vision Pattern Recog., 2009, pp. 935–942.
[4] R. Poppe, “A survey on vision-based human action recognition,” Image Vision Comput., vol. 28, pp. 976–990, 2010.
[5] J. Yin, Q. Yang, and J. J. Pan, “Sensor-based abnormal human-activity detection,” IEEE Trans. Knowledge Data Eng., vol. 20, no. 8, pp. 1082– 1090, Aug. 2008.
[6] R. Hamid, A. Johnson, S. Batta, A. Bobick, C. Isbell, and G. Coleman, “Detection and explanation of anomalous activities: Representing activities as bags of event n-grams,” in Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recog., vol. 1, pp. 1031–1038.
[7] D. Ryan, S. Denman, C. Fookes, and S. Sridharan, “Textures of optical flow for real-time anomaly detection in crowds,” in Proc. 8th IEEE Int. Conf. Adv. Video Signal Based Surveillance, 2011, pp. 230–235.

554



10.22214/IJRASET

INDEX COPERNICUS
45.98

ISRA JIF
IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY