

An Intelligent Video Surveillance System: Moving Object Behavior Analysis

Surbhi Kapoor¹, P. K. Pateria¹ and G. Singh²

¹Department of Computer Science and Engineering, Lovely Professional University, Phagwara - 144411, Punjab, India; surbhi31892@gmail.com, pushpendra.14623@lpu.co.in

² School of Electronics and Electrical Engineering, Lovely Professional University, Phagwara - 144411, Punjab, India; gursharanjeet.singh@lpu.co.in

Abstract

Objectives: The system aims at developing an intelligent video surveillance system which analyses the behavior of moving object and classifies it as standing, walking (normal behavior) and bending (abnormal behavior). The proposed system is also capable of detecting the persons moving into the restricted region providing the user with an ease to select the area of its own choice, thereby eliminating the problem of static region of interest. **Methods/Statistical Analysis:** In this paper we have detected the moving objects and tracked them using Kalman Filter on the basis of id assigned to each moving object. The direction of the moving object is labeled as right, left, up and down. The experiments are conducted in the corridors and pathways at Lovely Professional University on various persons with different heights. Neural network is used to validate our work and helps us in setting an experimental threshold which decides when the person bends or walks. This eliminates the need of any special detector which estimates the human pose, thereby reducing the cost. The neural network produces an efficiency above 95% which is quite promising. The implemented system follows a new approach of eradicating the problem of static region of interest and helps us to generate an alert when a person enters into an illegal entry zone. **Findings:** The objective of the proposed work is to turn the passive cameras into active cameras. The system finds out the direction of the moving object with an accuracy lying between 86.9% and 100%. The existing systems deploy special detectors for pose estimation making the system very costly. The implemented system eradicates the need of such detectors. Neural network validates the abnormal bending behavior producing an efficiency above 95%. The proposed approach eliminates the problem of static region by allowing the user to define its own region of interest before the process starts. This region of interest is regarded as restricted region. When the person enters into the restricted region, the system generates an alert and marks it as illegal entry. This will reduce the time of the humans to monitor every single camera installed in the campus as the system will generate an alert in case of an illegal entry which is regarded as an abnormal event. There are three events in the dataset for illegal entry and the system detects all the three events correctly. We can also track the trajectory of every individual which can be useful if any mischievous activity is performed by that individual. **Application/Improvements:** With intelligent and automated systems, performance can be increased and cost can be reduced. The manpower can be reduced, thus increasing accuracy. The implemented system has its shortcomings as if it leaves the frame and enters again, the system will assign it a new id. The accuracy of calculating the direction can be improved.

Keywords: Background Subtraction, Kalman Filter, Region of Interest, Surveillance, Trajectory

1. Introduction

In the era of advanced technology, everything is changing from the manual system to the automatic system having some kind of artificial intelligence which astonishes the people from time to time. From security to surveillance, sports to gaming, medical to astronomy, intelligent system has established its roots. Video surveillance has a broad spectrum of applications which proves to be very useful in the current scenario for example the surveillance cameras act as a deterrent to crime and apprehends a crime when it occurs. It is very difficult for a human to keep a watch on every single moment captured in the cameras. This leads to wastage of resources. Moreover humans are not available every time and can miss a critical event happening in the environment. These are the disadvantages of passive cameras which may prove to be dangerous. The researchers are actively participating in solving this problem by turning these cameras from passive into active cameras. The purpose of making the cameras active is to increase the performance and reduce the cost. The manpower will no longer be needed after making the system intelligent and automated, making the security guards free from this tiring job.

In the automated system, detecting an abnormal event is a challenging task. Abnormal event detection can be classified into pattern recognition based technique and machine learning based technique. In the former approach, the type of abnormal event is known in advance whereas in latter approach the normal behavior is learned by the system and object is detected if its behavior is different from the normal behavior as fed in the system¹. The objective of the proposed work is to turn the passive cameras into active cameras. The system aims at developing an intelligent video surveillance system by automatically analyzing the behavior of the moving object. The first step is to detect the moving object and obtain its trajectory by tracking its motion via Kalman filter. After the trajectories are obtained, the behavior of the moving object is analysed as normal walking, standing (normal behavior) and bending (abnormal behavior). The system does not require any special detector for analyzing the human posture of bending. This is the advantage over the existing systems. The method is discussed in the further section of the paper. There are some places in the environment for

example in college campus, some areas are restricted and people are not allowed to move into that region. The proposed system is capable of detecting the persons moving into the restricted region providing the user with an ease to select the area of its own choice, thereby eliminating the problem of static region of interest.

The rest of the paper is organized as follows. Section 2 presents the literature review on detecting the abnormal event in various situations. Section 3 throws light on the proposed approach and section 4 presents the performance and evaluations. The paper is concluded in section 5.

2. Related Work

S.Lee and R. Nevatia proposed an approach to detect the abnormal activities that take place in outdoor environment. They assumed that the human trajectories are obtained from the baseline tracking system. It is very common observation that people do not form line while walking. Either they walk alone or in groups, but formation of line is very rare and comes under suspicion. They detected the formation of line through spatial proximity and two thresholds depending upon the prior knowledge of the event. They also used the voting method to solve the temporal constraint². Cross scene abnormal event method was proposed which uses the bag of words model along with Scale Invariant Feature Transform (SIFT) and Support Vector Machine (SVM) classifier. This method has an edge over existing methods that earlier methods were applicable on learned scenes whereas this method works in unlearned scenes too³. The given method comprises of two steps namely feature coding and spatial pooling. They detected the events such as people creating panic in the crowded scenes or detection of vehicles in the area where vehicles are not allowed. The experiment was conducted on Upper Motor Neurone (UMN) and UC San Diego (UCSD) database, achieving promising results⁴. Contributing to the detection of panic events, H. Tabia gave an approach in which a probabilistic model is made for normal situations. Abnormal situations are detected by making a comparison between the actual observation and the stored templates. The probabilistic model is used to make an educated guess regarding the hasty changes for each frame in the video sequence⁵.

Crowd safety is also an important area of concern from the security point of view. Researchers have revealed that crowd disaster is caused by the energy released beyond a certain level by the people present in crowded regions. Therefore, proper management of crowd is required to prevent the disaster. H.Yin et al. calculated the crowd energy taking into consideration the three basic energies i.e. kinetic energy, internal energy and potential energy of the pedestrians and evaluated the performance for Beijing subway. They concluded that Crowd energy is positively correlated with the number of persons and the intensity of collision among them⁶. J.Ma and W.Song considered the flow pattern of pedestrians to count for the crowd safety. They presented an approach called automatic clustering to determine the abnormal flow pattern of the pedestrians and obtained the human trajectories via mean shift tracking⁷.

3. Proposed work

The implemented system consists of the following phases. The phases will be discussed in detail in the further section.

1. To detect and track a moving objects in a video.
2. To count the number of moving objects in a video.
3. To identify whether the person is walking or bending.
4. To select a region of interest dynamically.
5. To detect the person crossing region of interest indicating an abnormal behavior.

3.1 Object Detection and Tracking

Detection of moving is the first step in video surveillance system. Foreground objects are detected using background subtraction based on Gaussian mixture model⁸. Mixture of Gaussian approach models intensity of every single pixel as a mixture of Gaussians. A comparison is made between the current frame and the background frame to check which pixels are a part of foreground and background. The result of the process is a foreground mask. A bounding box is drawn for each foreground object. Morphological operations i.e. `imopen()`, `imclose()` and `imfill()` are used to remove the noise and fill in the holes. Once we obtain the foreground objects, Kalman filter plays its role.

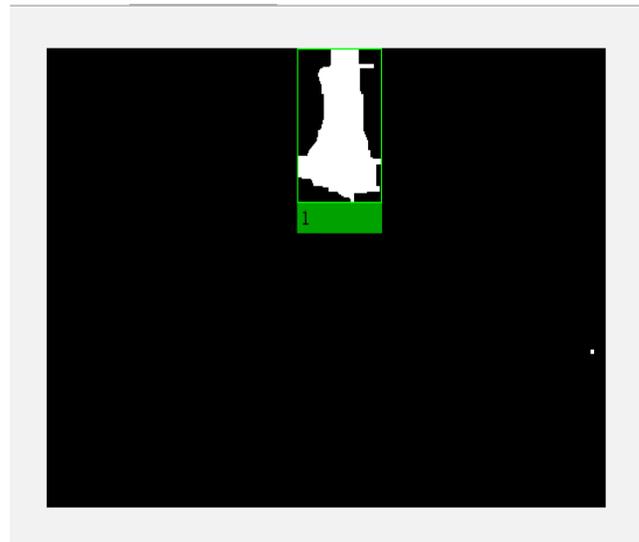
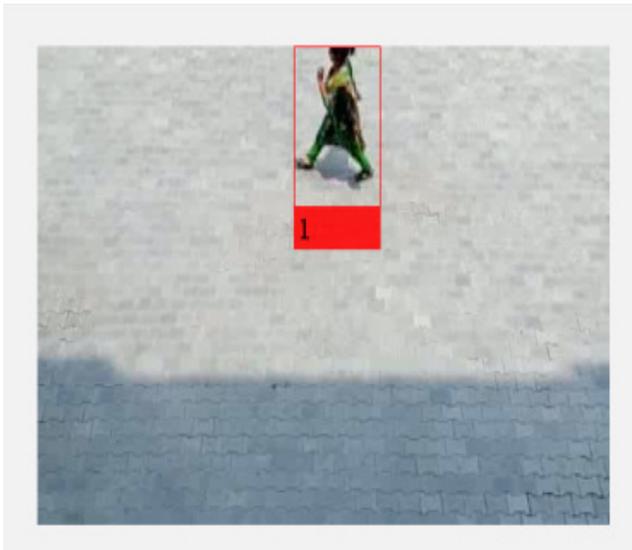


Figure 1. EEG sResults of background subtraction.

Object tracking keeps the track of the motion of moving object and its trajectory. Kalman Filter^{9,10} can be applied in the situation where there is some vague information regarding the dynamic system and an intelligent guess is made about the system's next move. Kalman Filter predicts the next step of the object by using a transition matrix. Initially we have a state vector which stores the current and the previous state of the system. Kalman

filter requires no memory because it just needs just the preceding state and no more extra state which makes it computation very easy. Sometimes, there are some factors like person may slip down or it collides, which may result in the wrong prediction. These uncertainty factors are stored in a control vector. The method is quite intelligent to handle those changes. A correction is made by the system itself if some uncertainty occurs and is added to

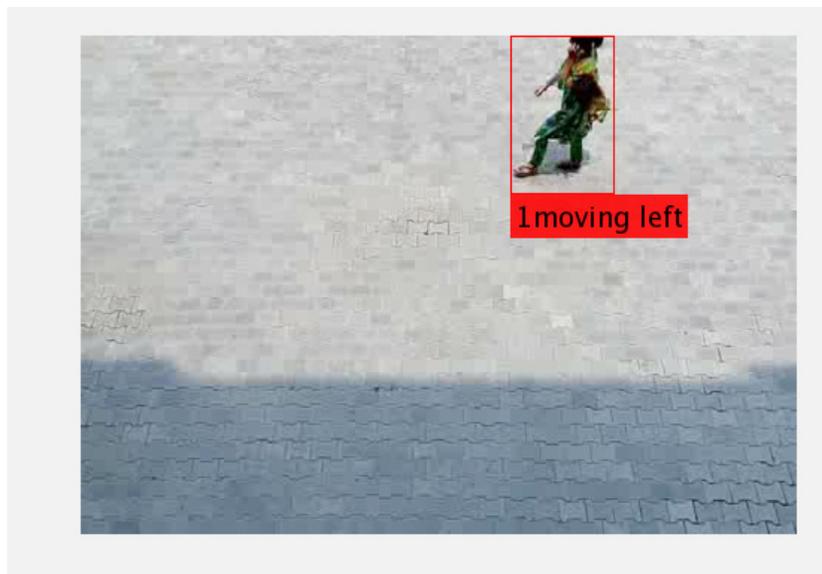


Figure 2. Person moving in left direction.

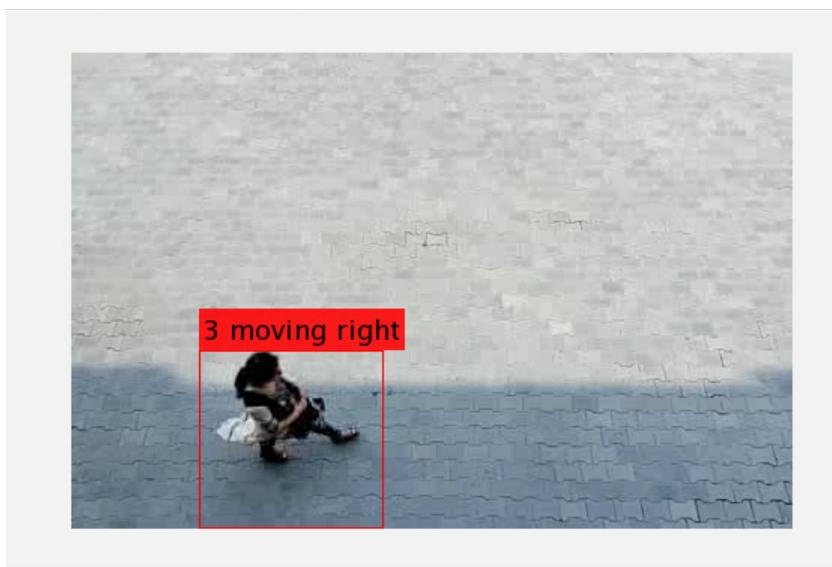


Figure 3. Person moving in right direction.

the prediction. Every person is given a unique id when it enters the frame and the id is retained till it remains in the frame. If the person is invisible for too long, the system deletes that particular id. The id serves as the primary key for tracking. So it is very important to preserve it. A .dat file is maintained at the back end which stores each and every location of the moving object. The content of file are x and y coordinate along with the height and width of the bounding box. The system automatically creates the files at the back end by the name of the id of the object as soon as it enters the frame. The system also finds the direction of the moving object by observing the positions stored in .dat file and label them as right, left, up or down. If the current coordinates of x is less than the previous coordinates, the person is moving in left direction. Otherwise the person is moving in right direction. If the current coordinates of y is less than the previous coordinates, the person is moving upwards, otherwise downwards.

The people exhibits different behavior while moving in the pathway like walking, running, bending etc. A person can perform any criminal activity while moving or bending. So it is very important to detect that behavior

at very initial stage. Existing systems use special detector, space time based approaches¹¹ and kinect sensor¹², to detect different poses which can be very expensive to install. The implemented system detects when the person is walking or when it bends without the use of any special detector making it cost effective. The implemented system assumes that if the height of the moving object is less than an experimental threshold, the person bends.

Otherwise, the person is walking or standing. Height of the moving object varies from person to person. So it is very difficult to set a particular threshold. We cannot assume any threshold which decides when a person bends, so we calculated a threshold which satisfies all the videos which we took. Many experiments are conducted on dataset containing different heights of persons like 6', 5'5", 5'3", 5' etc. Artificial neural network plays its role in setting a threshold. Neural network is created and trained for first half of the dataset and tested for remaining half of the dataset. The height and width of the moving object is fed to the system as input. The target is either 0 or 1. The average of the height of each person is considered. If the height of the person is less than average height of

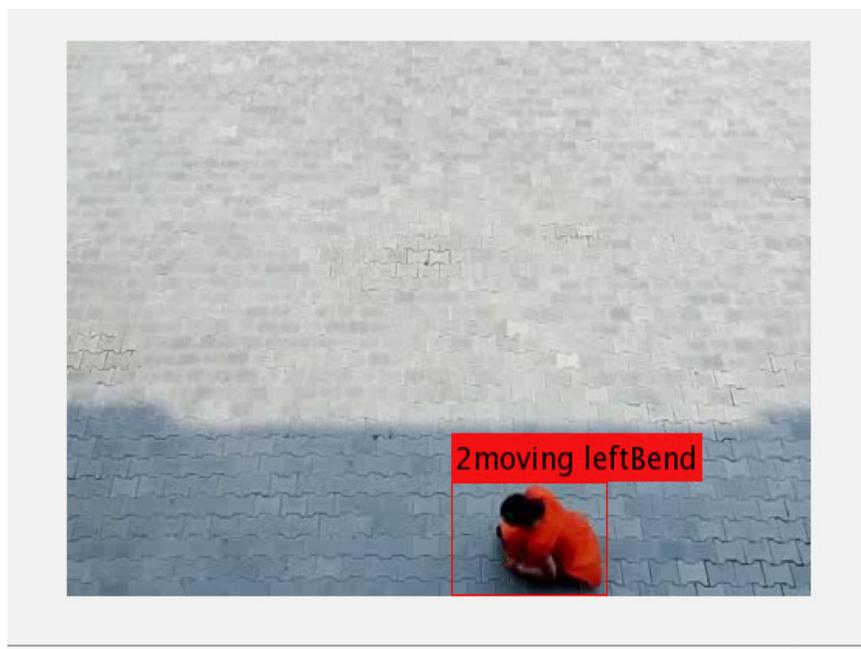


Figure 4. Person moving in left direction bends.

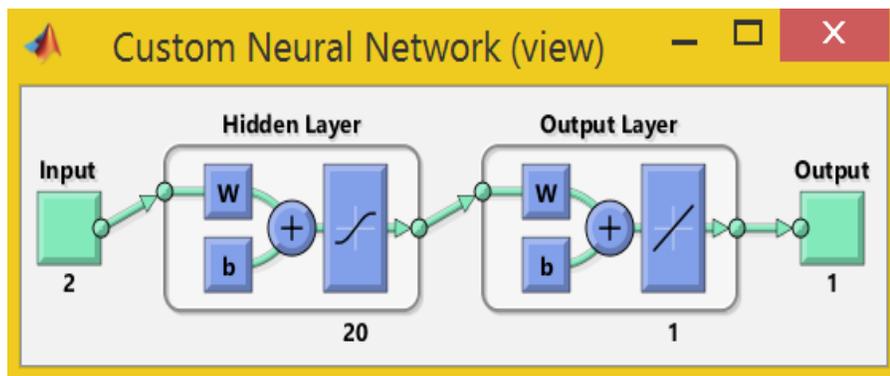


Figure 5. Creation of neural network.

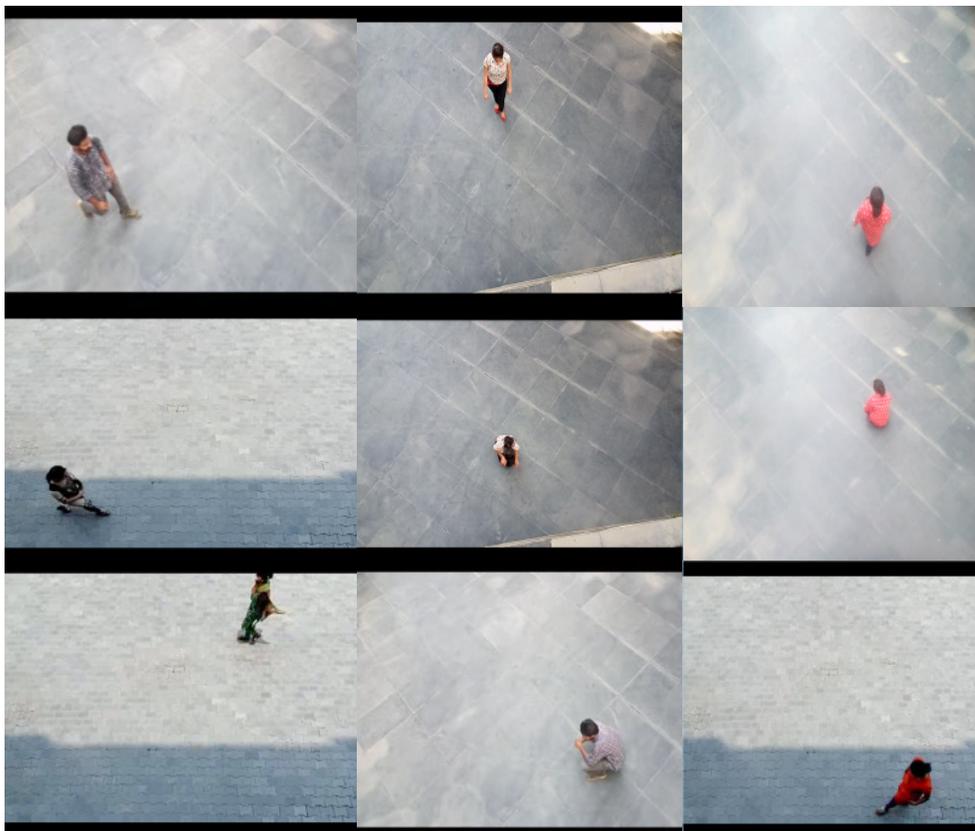


Figure 6. Examples of dataset.

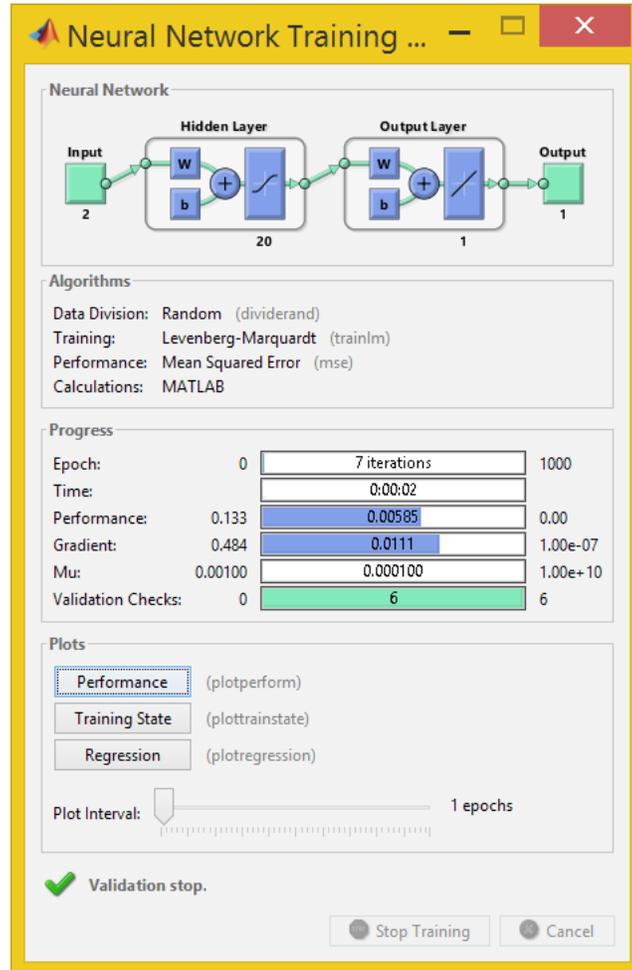


Figure 7. Training of neural network of id 3.

the person, the output is 0 indicating the person bends. Otherwise the output is set to 1 which indicates the person is either walking or standing.

The procedure for training and testing the neural network is given in the following algorithm.

1. Load data file for all the moving objects.
2. Plot the trajectory for each individual.
3. Define input parameters for height and width.
4. Calculate the average for each individual for height and width.
5. Create an output matrix of zeros.
6. For $i=1$ to end
 - a. If the input parameters of height of each individual is greater than the respective average
 - b. Set the output to 1.

- c. End.
- 7. Define target based on input and output.
- 8. Design and train the neural network based on input and target parameters which is the half of dataset.
- 9. Test the neural network on another half of the dataset.
- 10. Check the final output after testing the neural network.

- 11. Normalize the output by changing the values which are greater than 0.5 to 1, otherwise 0.
- 12. Calculate the efficiency for the neural network.

3.2 Illegal Entry Detection

People usually walk through the permissible areas. There are some critical areas in which people are not allowed to enter. If someone enters into that region, it is considered as abnormal event. Before the process starts, the video is read and segmented into frames. The first frame which is the background frame is extracted and displayed. The background frame is shown in figure 8.



Figure 8. Extracted background frame.

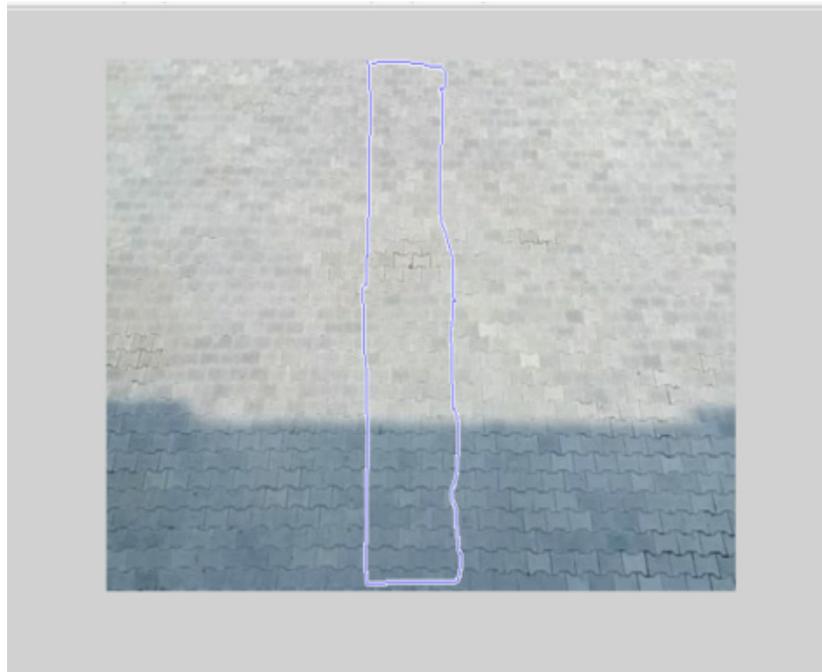


Figure 9. Region drawn by the user.



Figure 10. Binary mask generated.

Now the user selects a particular region which can be regarded as a sensitive region. The reason why the user is given the freedom to select the region is that if the algorithm is fed with the static region, it would be restricted to a single camera only. The advantage of the user defined region is that the algorithm can be applied to any camera. The user just needs to select a region according to his or her situation at the initial stage of the process. Once the region is selected, a binary mask is generated and the position of the region is stored. Some researchers have generated the reason of interest on the basis of repeated regular pattern of objects¹³. The algorithm continues with the process of tracking. The system checks whether the position of the person matches with the position of the region. The alert will be generated if the position matches. The experiment was conducted on the pathways of Lovely Professional University. The dataset contained three events of illegal entry and the algorithm detected all the three events correctly. The user can draw the mask anywhere in the scene. This will not make the camera stick to one place. The alert generated by the system will make the security personnel aware that someone is entering into the restricted area. The system faces a problem due to changes in weather. Improved background subtraction produces better results in case of weather problems.¹⁴

The procedure for illegal entry detection is given in the following algorithm.

1. Read the video.
2. Extract the first video frame and display it.
3. Draw a region with free hand on the extracted frame using `imfreehand()`. Drawing a region will create a binary mask.
4. Display the binary mask.
5. Get the position of the region.
6. Perform the steps as mentioned in Moving object detection and Tracking.
7. If the position of the moving object is equal to the position of the region.
8. Display illegal entry.

4. Performance and Evaluations

The experiments are conducted in the pathways and corridors of Lovely Professional University. The speed for image acquisition is 30 frames per second. Many experiments are conducted with the persons having heights 6', 5'5", 5'3", 5'. Three events are considered in the experiments i.e. walking, bending and illegal entry. The direction of the moving person is observed manually which is regarded as ground truth and compared with the output obtained via

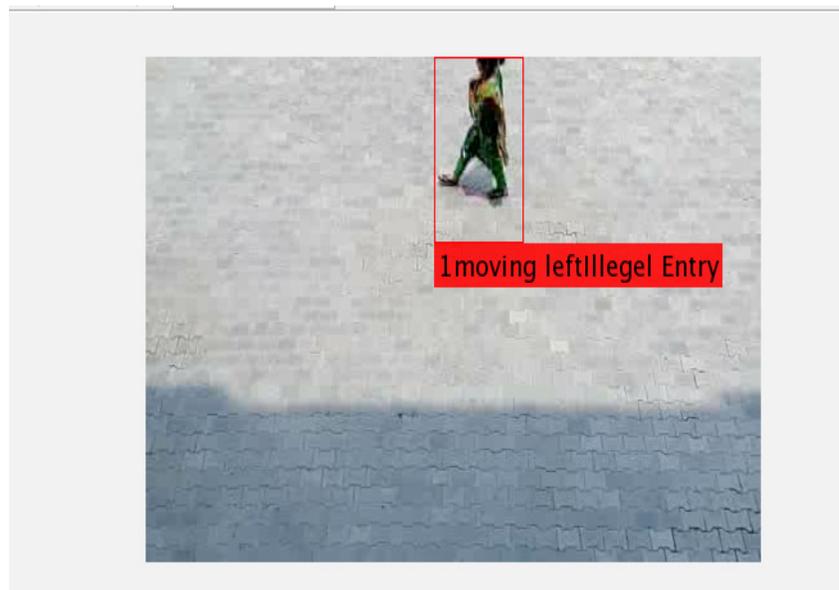


Figure 11. Person moving in restricted region.

Table 1. Accuracy obtained in calculating direction

Id	Total frames travelled by person (T)	Frames correctly classified (C)	Frames not correctly classified	Accuracy
1	130	113	17	86.9%
2	221	201	20	90.9%
3	130	130	0	100%

the proposed method. The algorithm is implemented in real conditions where people are coming from different directions and bends. The technique sometimes produces wrong directions i.e. the method shows the message of left direction even if it is moving in right direction. We have calculated the accuracy using the following formula.

$$\text{Accuracy} = (C/T) \times 100$$

Where C is the numbers of frames in which the direction calculated by our method is correct and T is the total number of frames travelled by the person.

It is very clear from the table that the method yields an accuracy above 85% which is quite promising.

Neural network is used to set an experimental threshold which decides when the person is walking or when it bends. The Figure 5 shows the created neural network. Two inputs are fed into the network i.e. height and width which produces a single output which is either 0 or 1. Neural network is trained with these inputs as shown below. Number of hidden neurons in this case is 20. It can be increased above 20. More numbers of neurons in the hidden layer makes the network flexible. It increases the size of the layer. If the hidden layer becomes too large, it

might cause the problem to be under-characterized and the network must optimize more parameters than there are data vectors to constrain these parameters. The system is generalized. The concept generalized means the system should produce good results when it is tested with the inputs which are not used in the training of the neural network. To make the system generalized we have divided the dataset into two halves. The neural network is trained with the first half of the dataset. The other half is used for testing the neural network. We have calculated the size of output which is in the form of $1 \times N$ matrix where 1 indicates the number of rows and N indicates the number of columns. The obtained output of the neural network falls between 0 and 1. So we have normalized the output by setting a threshold as 0.5. The output which is greater than 0.5 is set to 1, otherwise 0. The obtained output is matched with the target output and counter is incremented every time the match occurs. The efficiency is calculated based on this count. The trajectory can also be plotted for each individual. The trajectory tells the path followed by the individual in the whole video. It is used for analyzing the movement of the moving object. The formula for efficiency is given as:

$$\text{Efficiency} = (\text{count}/N) \times 100$$

Table 2. Efficiency obtained by neural network

Id	Efficiency
1	96%
2	98%
3	100%

Where count is number of times our output matches with the target and N is the size of the output.

Neural network is created for each individual moving object. So the efficiency is computed for each id. The following table shows the efficiency obtained by our method.

The efficiency for three objects in the video lies above 95%.

The graph for performance is plotted for each moving object.

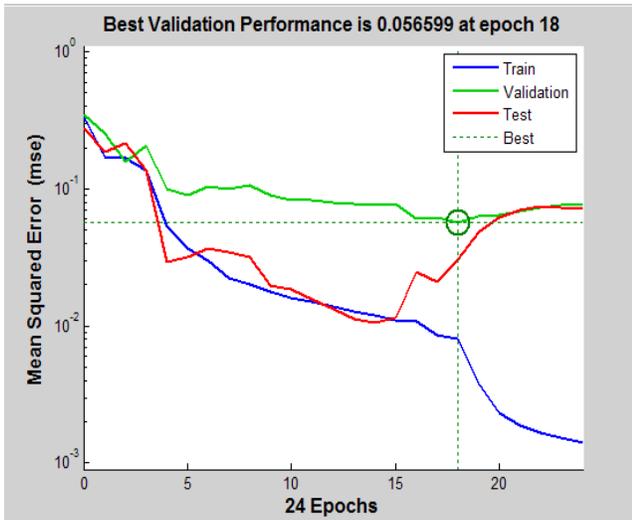


Figure 12. Performance plot for id 1.

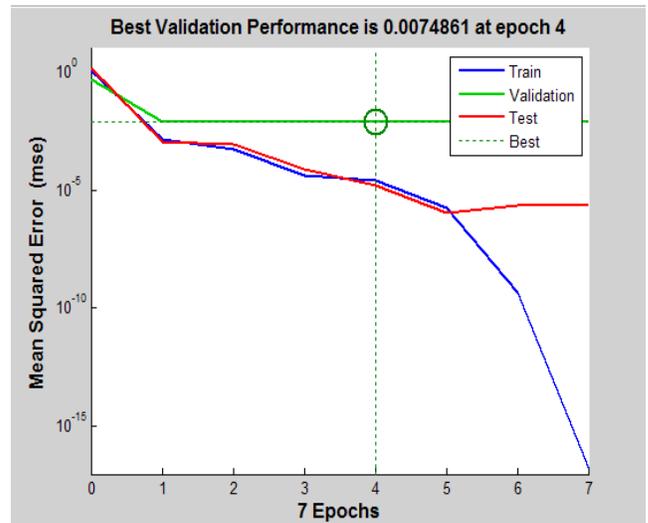


Figure 14. Performance plot for id 3.

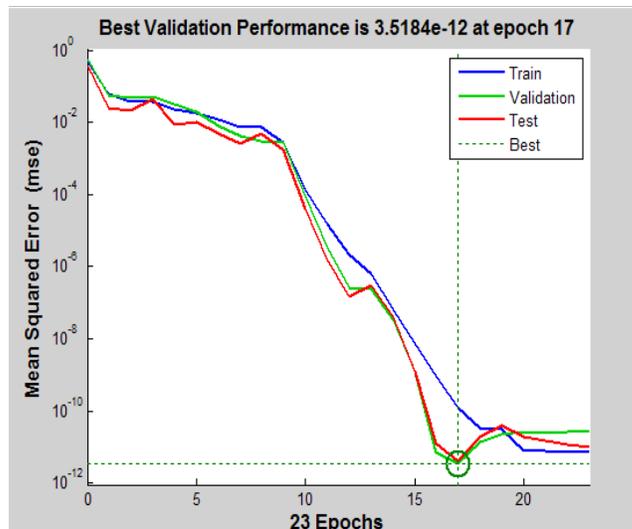


Figure 13. Performance plot for id 2.

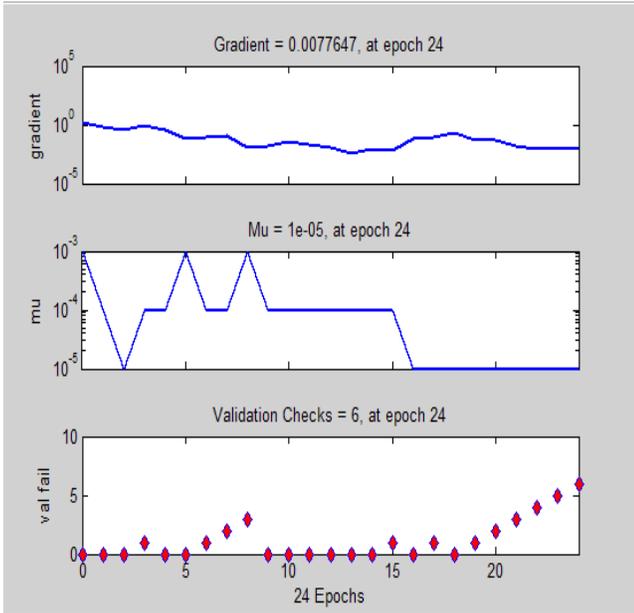


Figure 15. Training state for id 1.

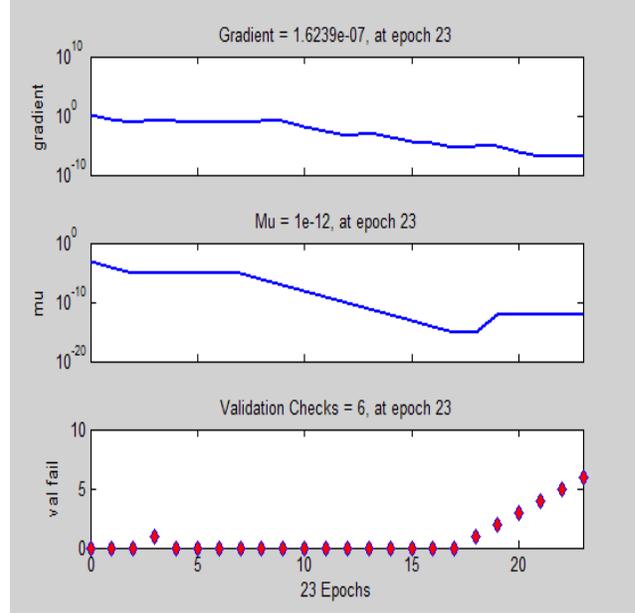


Figure 16. Training state for id 2.

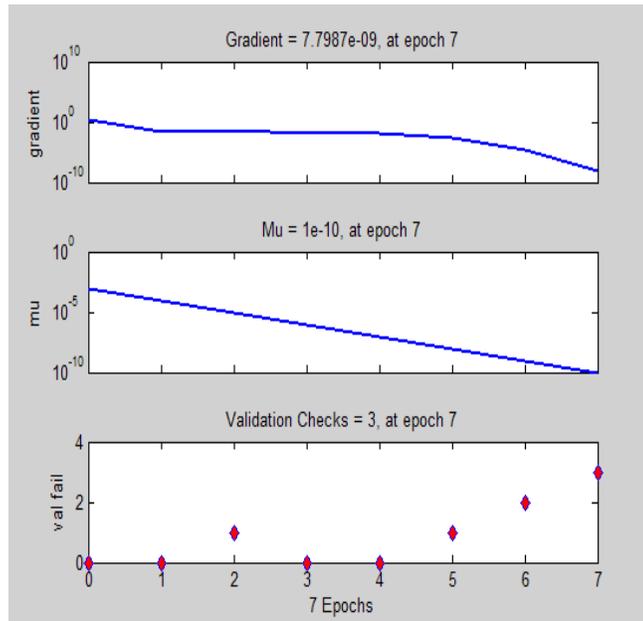


Figure 17. Training state for id 3.

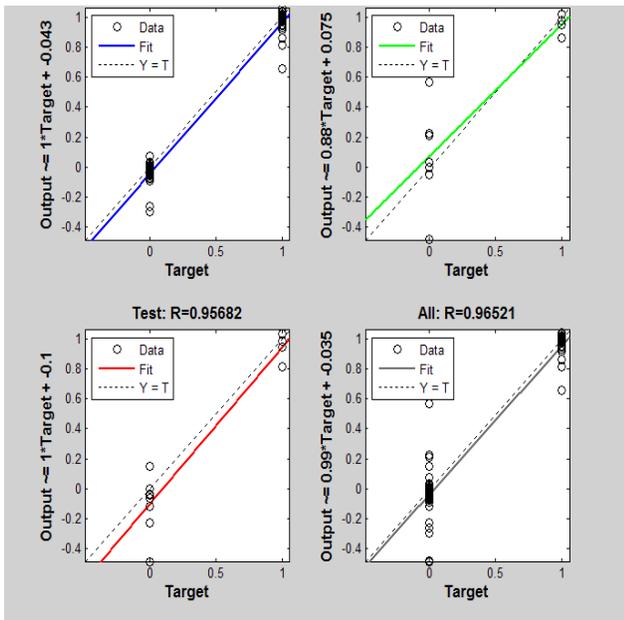


Figure 18. Regression for id 1.

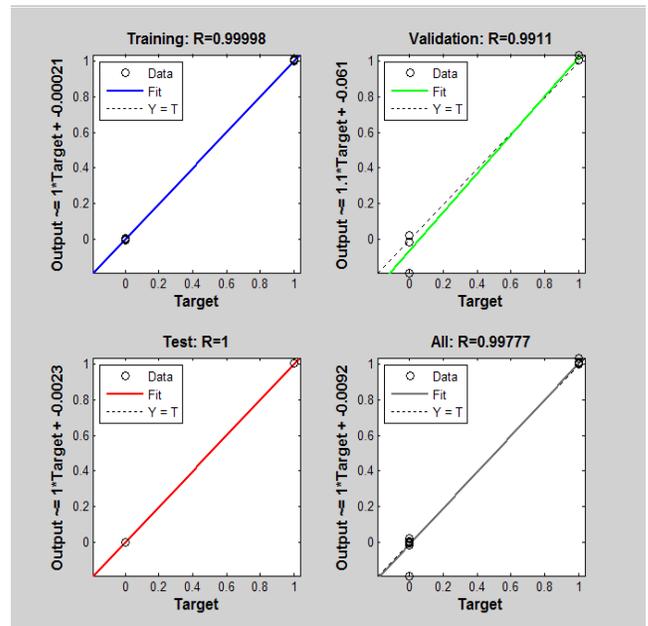


Figure 19. Regression plot for id 2.

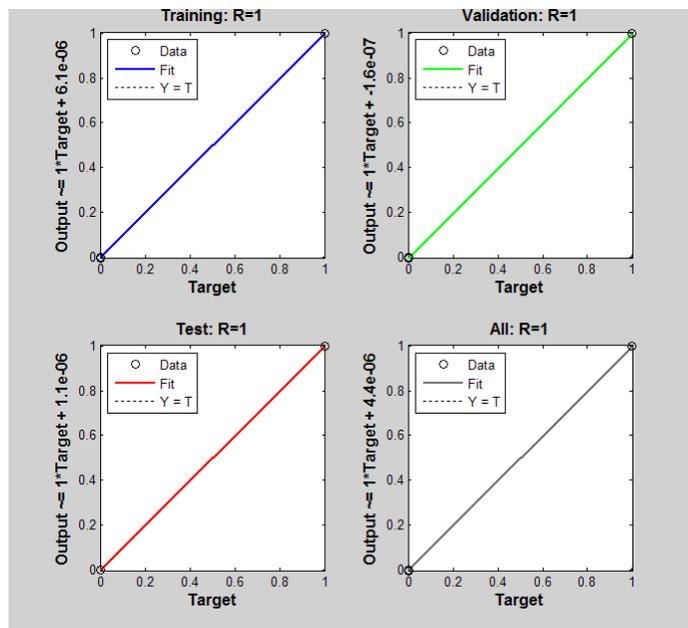


Figure 20. Regression plot for id 3.

Generally, the error decreases as the number of epochs increases but sometimes, the error starts increasing on the validation due to the fact that the network overfits the training data. The dashed line in the performance plot indicates the best performance that is the iteration at which validation reached a minimum. Id 1 took 24 iterations for which the performance plot is shown in Figure 12. The error for the validation in the initial epochs is less but increases as the epochs increased. In the case of testing, the red curve showed less error as compared to validation curve but had a sudden increase in the error with the increasing epochs. The Figure 13 shows the performance plot for id 2. The training continued for 23 epochs. The validation curve and test curve are quite similar. Therefore, the system does not face any difficulty while training. Overfitting can arise if the test curve increases before the increase in validation curve. Performance plot for id 3 is shown in Figure 14. The training took 7 epochs. The performance plot concludes that training was perfect as the validation curve is similar with the dashed line which represents the best solution i.e. the iteration at which validation reached a minimum.

The plot for training state tells some other training statistics. The plot shows three values i.e. Gradient, mu and validation fails. The value of gradient tells that the system reached the bottom of the local minima of the objective function. Val fail represents the validation fails. These are the iterations where validation mean square error increases. Consecutive fails represents overtraining of system. If there are consecutive 6 fails, training is automatically stopped by Matlab.

The following regression plots represent the outputs of network corresponding to targets for validation, training and testing. If the data falls along a 45 degree line, it is regarded as a perfect fit. In this case, the outputs obtained from the network are equal to the target outputs. In all the dataset, all the R values fall above 0.93 which is considered as a good fit. The results can be more precise by retaining the network. Retaining the network changes the initial weights and bias of the network and results in better results. Figure18 shows the regression plot for id 1. All the R values are above 0.93 except 1 in which R value is 0.88. It is not considered as a perfect fit. In this case

error first decreases and thereafter starts increasing. The efficiency for this case is 96%.

Id 2 yields an efficiency of 98% and all the R values are above 0.93. The validation line is slightly misplaced from the best fit. In all the other cases, the output values exactly fit the target values. The results are shown in Figure19. In case of id 3, the output values are exactly equal to the target values yielding an efficiency of 100%. It is regarded as the best fit.

5. Conclusions

This paper focuses on analysing the human behavior which is a vital issue in this era of research. Our technique detected the moving person in the video via background subtraction and tracks the moving object via Kalman Filter. The given method also finds out the direction in which it is moving with accuracy as 86.9%, 90.9% and 100%. Many experiments are conducted on the dataset consisting of persons with different heights i.e. 6', 5'5", 5'3", 5'. Neural network is used to set an experimental threshold which decides when the person is waling and when it bends. The graph for performance, training state and regression are plotted. The approach yields efficiency above 95%.

The proposed approach eliminates the problem of static region by allowing the user to define its own region of interest before the process starts. This region of interest is regarded as restricted region. When the person enters into the restricted region, the system generates an alert as illegal entry. This will reduce the time of the humans to monitor every single camera installed in the campus as the system will generate an alert in case of an illegal entry which is regarded as an abnormal event. There are three events in the dataset for illegal entry and the system detects all the three events correctly. We can also track the trajectory of every individual which can be useful if any mischievous activity is performed that individual. This method can be useful in the case of crowded scenes where there is a danger of an abnormal event. We can monitor the posture of the person as bending or walking or standing.

6. Future Scope

The proposed approach is yet not suitable in case where there are moving leaves, water and trees. The algorithm also considers the moving leaves as moving object which is a wrong detection. In the future scope, researchers can try to eliminate the problems occurred due to these factors.

7. References

- Javanbakhti S, Zinger S. Fast abnormal event detection from video surveillance. *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (ICCV)*. 2012 Jan 1. p. 1.
- Lee SC, Nevatia R. Hierarchical abnormal event detection by real time and semi-real time multi-tasking video surveillance system. *Machine Vision and Applications*. 2014 Jan 1;25(1):133–43.
- Kapoor S, Pateriya PK. A study on various methods used in automatic video surveillance. In *Cloud System and Big Data Engineering (Confluence)*, 2016 6th International Conference 2016 Jan (pp. 413-416). IEEE.
- Hung TY, Lu J, Tan YP. Cross-scene abnormal event detection. 2013 IEEE International Symposium on Circuits and Systems (ISCAS2013). IEEE. 2013 May 19. p.2844–7...
- Tabia H, Gouiffes M, Lacassagne L. Motion modeling for abnormal event detection in crowd scenes. *International symposium on signal, Images, Video and communications*, University of Valenciennes, France. 2012.
- Yin H, Li D, Zheng X. An energy based method to measure the crowd safety. *Transportation Research Procedia*. 2014 Dec 31; 2:691–6.
- Ma J, Song W. Automatic clustering method of abnormal crowd flow pattern detection. *Procedia engineering*. 2013 Dec 31; 62:509–18.
- Bouttefroy PL, Bouzerdoum A, Phung SL, Beghdadi A. On the analysis of background subtraction techniques using Gaussian mixture models. 2010 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE. 2010 Mar 14. p. 4042–5.
- Faragher R. Understanding the basis of the Kalman filter via a simple and intuitive derivation. *IEEE Signal processing magazine*. 2012 Sep 1; 29(5):128–32.
- Kleeman L. Understanding and applying Kalman filtering. *Proceedings of the Second Workshop on Perceptive Systems*, Curtin University of Technology, Perth Western Australia. 1996 Jan.
- Anuradha K, Sairam N. Spatio-temporal Based Approaches for Human Action Recognition in Static and Dynamic Background: A Survey. *Indian Journal of Science and Technology*. 2016 Feb 9; 9(5).
- Madhubala JS, Umamakeswari A. A vision based fall detection system for elderly people. *Indian Journal of Science and Technology*. 2015 May 1; 8:167–75.
- Pateriya PK. An Automated Attendance System: A Technique using Background Subtraction and Color Image Processing of RGB Planes. *Indian Journal of Science and Technology*. 2016 Jul 27; 9(28).
- Thakur N, Wasson V. Obstacle Detection from Still Images using Improved Background Subtraction Method. *Indian Journal of Science and Technology*. 2016 Aug 24; 9(31).