An Efficient Method for Suspicious Activity Detection

S. S. Gurav, B. B. Godbole

Abstract: ***This paper contributes on suspicious activity detection from length of video with less complex processing algorithm. The proposed method in this paper is easy to implement and robust enough to monitor different suspicious activities such as sudden seating, standing up, hiding from midway path, entry from midway. The suspicious frame detection is a novel approach and then confirmation is done using SURF based descriptor matching for speedy processing requirements. The results obtained in terms of tracking window and detection capability are satisfactory.***

Keywords: **Suspicious activity, surveillance video, SURF descriptors, *monitoring human activity.***

I. INTRODUCTION

Human behavior monitoring in surveillance video is urgent need of the time to control the crime and assure the safety for the society. The continuous manual monitoring on display screen is prone to human errors and missing of important activities when crime event occur. The monitoring is required to be assisted with automated function which can provide reliability while assuring the safety needs. The robust and reliable working algorithms are essential target that has to be focused while developing the activity monitoring algorithms. The implementation feasibility and complexity of algorithms are two main inter dependent aspects as video data handling requires large storage spaces. The execution time versus crime happening time should also be optimum such that within time analysis will be with security personal for further activities to be carried out. The algorithms for human monitoring along with activity monitoring are complex so far made available by the researchers. This paper focuses on simple but effective suspicious activity monitoring method useful for surveillance video systems

II. LITERATURE SURVEY

Xiaojing Chen, Le An, and Bir Bhanu, [1], have given a method for tracking multiple objects in multi camera scenario. The color histogram, pertained brightness function is the methods used for tracking the object of interest. The performance of the system degrades with respect to number of objects to be tracked. Rohit Agarwal et al [2] have given a method for crowd density estimation using three modalities, viz., carbon dioxide level, sound intensity level, and received signal strength. The clustering algorithms for the feature types such as temporal, spatial, and spatio-temporal are used. The fusion of information with context is used for efficient crowd monitoring. The acoustic effects are considered while deploying the sound capturing sensors and respective pre-processing functions are used to suppress the noises and reverberation effects.

Md. Shakowat Zaman Sarker et al [3], have segmentation method for foreground and background estimation applicable for object tracking. The inclusion of watershed for multi object segmentation and foregraound identification is done. The region adjacency list (RAL) is built for the region merging process. The merging effects are observed with different test inputs. Seema Kamath [4], have given camera calibration method. The state updates and measurement variations are corrected using probabilistic error correcting method. After calibration, LBP based object tracking is used. Loris Bazzani [5], have given a method for specific people identification from a group of people. The candidate matching method is used for object tracking in single and non-overlapping multi camera videos. The candidate matching is done using descriptors based features matching. Atif ILYAS [6], given object re identification in multi camera videos. A codebook based foreground and background segmentation method is used. The descriptor based object features matching method is used for object identification. The compactness and robustness are main advantages of the algorithm. Branko Markoski , ai [7], have given method for basketball player identification. The head leg arms of the person are segmented and LBP, HOG features are extracted. The features are boosted using adaboost method. The SVM based matching shows good results for player identification.

III. PROPOSED METHOD

The person tracking algorithm to be developed is based following assumptions

1. Tracking should not lose the track while camera to camera shift
2. The scale and rotation in human objects in particular frame should not degrade the performance of the tracking
3. The tracking should be less complex such that execution time of tracking should be optimum.
4. The test video sequences are original and have not gone under morphing or any other modifications such as forging.

The proposed work consist of suspicious activity detection ith various processing stages as shown in figure 1.

Manuscript published on 30 September 2019.
***Correspondence Author(s)**

Mr. S. S. Gurav*, Department of E&TC, Sharad Institute of Technology, College of Engineering Yadrav, Ichalkaranji.
Email: sgurav8849@gmail.com
Dr. B. B. Godbole, Professor, Department of Electronics Karmaveer Bhaurao Patil, College of Engineering & Polytechnic, Satara

At first we consider single camera video file as a test input. The human objects in each frame are to be identified first before tracking them. For this, the input video sequence is converted into frames to treat each frame as image while processing. Consider there is N number of frames in the video sequence which is having frame rate as 25 frames per second. Let Fr be the set of the frames obtained from input video by converting video to frame.

$$F_r = \{ f_1, f_2 , \dots , f_N\} \quad \dots (1)$$

Where f1, f2,…,f_N are the frames. Each frame in set Fr may or may not contain the human object. Also, it may happen that, start to end in set of frames there may not be presence of the person even if person was present in few starting frames and few ending frames.

Each second in video sequence is responsible to provide 25 frames. As, normal human being (healthy but not sports person case) can walk at the highest speed of 25 to 30 kmph in average scenarios, in each second window of video sequence the human object should be present once it has entered into the viewing angle of the camera. There is need to save number of iterations while processing the frames to get faster output without having loss in terms of results. For this periodically distant frames from the set obtained in (1) are to be selected. At actual, there can be maximum two frames considerable as key frames when frame rate is considered, but to be on better results side we select one frame after every five frames from set given in (1). This way, for example, when there will be need to process 10 seconds long video, instead of processing 250 frames, only 50 frames processing will be required as per our strategy.

Hence, in nuts and shell, total N/5 number of frames will be required to be processed. Also, out of these N/5 frames only those frames will be required to be processed which contain actual human object. The frame which does not contain any living object and hence empty and constant background is considered as a marker frame. For the sake of verification of presence of object we first use simple subtraction process of all frames from this marker frame. To make this process and reduce the processing complexity we first convert all the frames into grayscale which will provide two dimensional matrices while processing given by F_{gr} in (2).

$$F_{gr} = \{ fg_1, fg_2 , \dots , fg_M\} \quad \dots (2)$$

Where M= N/5 in (2) and in further equation representing total count of items in each set considered and calculated. The set of frames obtained after subtraction process is given by F_{sr} as given in (3).

$$F_{sr} = \{ fs_1, fs_2 , \dots , fs_M\} \quad \dots (3)$$

For each frame from set in (2) when exactly matches with that of marker frame, the subtraction frame obtained in (3) will contain almost zero value for all pixels. If there is object present such as human then there will be non zero pixels present in the subtraction frame in (3). For verifying this, the mean summation of all pixel values present in each frame in (3) is calculated given by set Fav, average value set as given in (4).

$$F_{av} = \{ fa_1, fa_2 , \dots , fa_M\} \quad \dots (4)$$

Where,

$$F_{ai} = \frac{\sum_{x=1}^{R}\sum_{y=1}^{C} p(x,y)}{R \times C} \quad \dots (5)$$

Where,
R = number of rows in frame
C = number of columns in frame
x = x location (row number) of pixel P
y = y location (column number) of pixel P
Fai = each member of set Fav in (4)
R x C= total number of pixels.
i = 1 to M

The values obtained in set F_{av} will possess either zero value if the scene is similar to that of marker frame else some other value. The values other than zero indicate presence of foreground objects and need to be processed. The number of item in set (4) which are to be processed, the frames from set (2) are then selected for processing. Here for the sake of worst case scenario, considering all the frames in sequence (2) which possesses the object being present is considerable.

Once the frame to be processed is identified, the next phase of work is identifying the object as human. For this process, the haar cascade method developed by viola jones [2] is considered. The haar cascade based processing provides the resulting window of human object in the frame in rapid manner. Hence identifying that presence of human being or other objects is simplistic task performed here which will provide respective window of human being in terms of coordinates and hence for all frames in (2).

The obtained result in window or bounding box in frame consist of four values such that starting point coordinate that is row number and column number and width and height of the widow. Let the obtained set of windows be represented by Fw as given in (6).

$$Fw= \{W_1,W_2,\dots,W_M\} \quad \dots (6)$$

Where,

$$Wi= [ri\ ci\ wdi\ hti] \quad \dots (7)$$

Where,
ri = row count of starting point of window
ci = column count of starting point of window
wdi = width of window
hti = height of window
and i =1 to M for M frames.

The window obtained in (7) can be considered as bounding box. The resulting bounding box of each frame is considered for further matching of each human being in each frame in (2).

The bounding box matching is done by extracting SURF (scaled up robust features (modified SIFT (scale invariant feature transform))) descriptors of each bounding box window. The reasons for considering SURF descriptors are,

1. There may be change in scale of human object due changes in distance from camera to human object
2. There may be change in object in terms of rotation in vertical axis and also in horizontal axis of human object

Most of the time, while performing match of two images, the changes in scale and orientation degrade the performance and hence SURF plays the important role of keeping results of matching optimum even in case of scale and orientation changes.

The process of matching using SURF descriptors is possible by three types of distances, viz., Euclidian distance, Bhattacharya distance and Manhattan distance. Euclidian distance is considerable for estimating similarity score between two sets and hence considered here.

The scores obtained while matching these bounding box windows are gathered together in a set F_d given in (8).

$$F_d = \{D1, D2, \ldots, D_M\} \qquad \ldots(8)$$

As far as same human being is present in all frames, each distance score in set Fd will possess similar value with very little tolerance. This results in identification of frame number where particular human being is not present in particular frame and hence given sequence of video.

IV. RESULTS AND ANALYSIS

In the work of suspicious activity detection, two methods are proposed for suspicious activity detection. The first phase of work depends on the array set F_{av} obtained in (4). The comparative analysis of this array amongst neighboring values can provide variation in the target mean values which can provide set of frames for suspicious activity. To obtain comparative vector, difference of each value in F_{av} with its immediate neighbor is taken using equation (9). The resulting vector F_{avs} can provide idea about suspicious activity.

$$F_{avs} = \{ fav_1, fav_2, \ldots, fav_M\}$$

Where,

$$Fav_i = Fa_i - Fa_{i+1} \qquad \ldots(9)$$

The frames sample considered for this detection method is shown in figure 2 and 3. The figure 2 shows frame with normal human activity of walking and frame in figure 3 shows the suspicious activity of the human. The resulting graph of mean values of all frames obtained as per equation (4) is shown in figure 4. The values that are suddenly changing show the change in object size which is nothing but total motion pixels occupying small window in the specific region. The frames with such count are chosen for SURF matching to verify the suspicious activity detection as part of second phase of work of suspicious activity detection. The graph of SURF matching points is shown in figure 5.

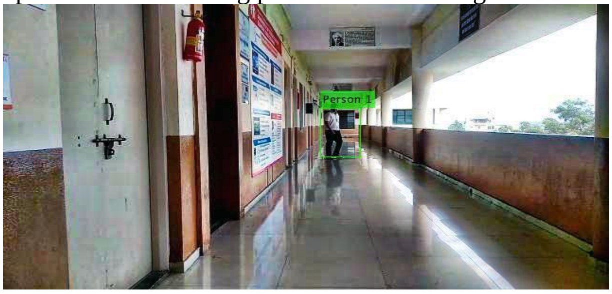

Figure 2: Normal activity of human

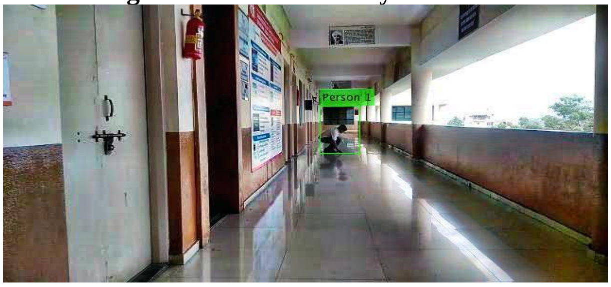

Figure 3: Suspicious activity of human

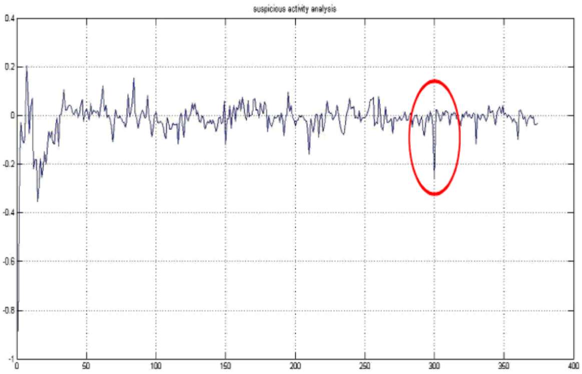

Figure 4: Graph of vector in equation (5) indicating suspicious activity frames region in video

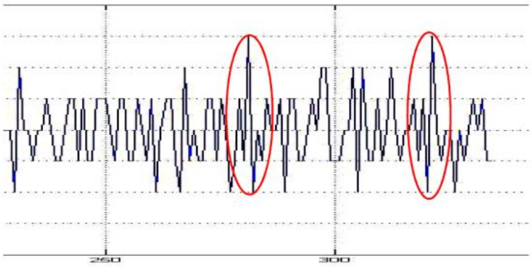

Figure 5: Two peak SURF matching points set vector differences marked where suspicious activity starts and ends

From figure 5 it can be understood that, there are two peaks which indicate the start and end of the suspicious activity. The peak differences are because of sudden change in body appearances of the human in the frame where it seats at first peak and stands up again in second peak. After this video the same principle of working is used for detecting suspicious activity in which following combinations are verified for detection.

1. A human being suddenly vanishes from the frame to enter into one of the room.
2. A human being suddenly seats and stands up.
3. A human being suddenly jumps on a small wall aside where other times no presence of human expected.
4. A human being walks in random motion instead of straight.

Analysis:

The performance evaluation comprises two types of analysis. First one consist of tracking accuracy of proposed algorithm and second consist of detection capability of various combinations of suspicious activities. The first type analysis is compared with histogram tracker while second analysis is done using manual suspicious activity monitoring versus detection accuracy of the algorithm.

The performance evaluation of second type of analysis is done by counting total number of suspicious activity frames using manual counting method and using proposed algorithm based method.

$$Accuracy = \frac{Total\ number\ of\ detected\ frames\ for\ suspicious\ activity}{Total\ number\ of\ actual\ frames\ of\ suspicious\ activity}$$

Table I: comparative study of precision vs location error

Video	Histogram tracker method	Proposed method
Video 1	0.13	0.03
Video 2	0.14	0.04
Video 3	0.112	0.024
Video 4	0.16	0.04
Video 5	0.18	0.07
Video 6	0.146	0.053
Video 7	0.11	0.03

The observed values are tabulated as in table I. The precision is evaluated with respect to ground truth bounding box values and location error of the bounding box. The proposed method shows considerably less errors compared to histogram based tracker.

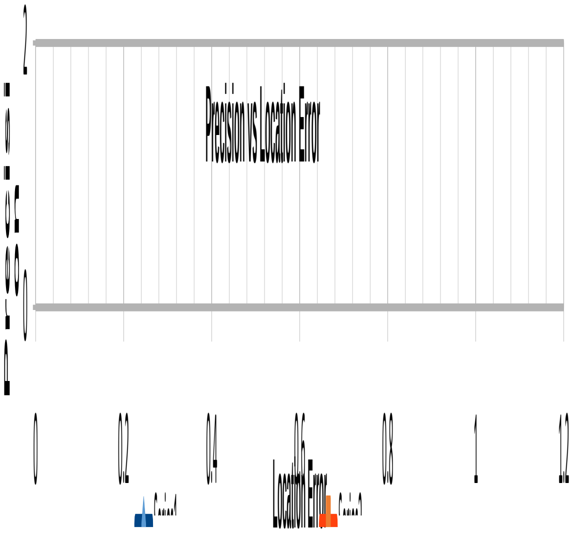

Figure 6: Precision versus location error comparative analysis

Annotations:

1. From this analysis it can be observed that, tracking of object depends on various frame contents in which ROI object and its surrounding area affects the performance of the tracking.
2. Due effect of surrounding objects correct estimation of histogram for ROI gets affected and hence error is seen in tracking window.
3. The pattern in graph can be seen matching as most error effect is similar but much less error is seen in effect due to prosed method.

Table II: comparative study of precision vs location error

Video	Actual Suspicious Frames	Detected Suspicious Frames	Accuracy
Video 1	32	30	0.935
Video 2	28	24	0.95
Video 3	41	38	0.965
Video 4	34	33	0.942
Video 5	36	32	0.955
Video 6	29	27	0.96
Video 7	30	28	0.94

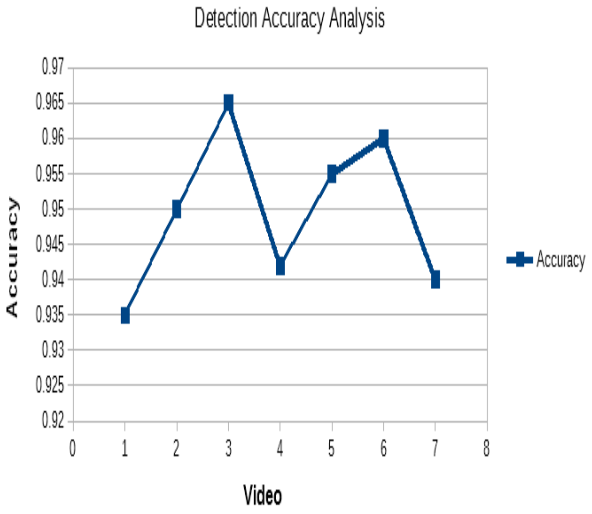

Figure 7: Accuracy of detection analysis

Observations:

1. While analyzing accuracy it has been observed that, detection of start and end of suspicious activity in entire video sequence is important task.
2. The suspicious activity detection depends on total number of SURF points matching and hence descriptor matching based principle.
3. The orientation and scale of the object does not matter on detection capability due to use of SURF points.
4. As right from start only those frames are processed which contain object of interest, the complexity of algorithm is at lowest as per requirement of application.

V. CONCLUSION:

This paper focuses on development of novel method for suspicious activity of human being from surveillance video dataset. The activity is considered as suspicious when human in video suddenly seats, stands up, enters into one of the rooms thereby gets hidden from camera, enters from mid frame where no entry is expected. The evaluation of results in terms of tracking window size and accuracy of detection of frames as suspicious to that of actual number of frames is done. The results are satisfactory for the proposed method and class of activities considered.

REFERENCES

1. Xiaojing Chen, Zen Qin, Le An, BirBhanu“An Online learnedElementary Grouping model for multi-target tracking”, University of California, Riverside, In CVPR 2014.
2. RohitAgarwal, Sudhir Kumar, Rajesh M. Hegde“Algorithms for Crowd Surveillance Using Passive Acoustic Sensors Over a Multimodal Sensor Network”, IEEE Sensors Journal, Vol. 15, No. 3March 2015.
3. Md. ShakowatZamanSarker, Tan Wooi Haw and RajasvaranLogeswaran“Morphological based technique for image segmentation” International Journal of Information Technology, Vol. 14 No. 1.
4. SeemaKamath“Distributed algorithms for camera network localization and multiple targets tracking using mobile robots” Master of Science Thesis, RPI Troy, New York, Nov 2007.
5. Loris Bazzani“Beyond Multi-target Tracking Statistical Pattern Analysis of People and Groups”, UniversitydegliStudi di Verona Department of Informatics.” May3, 2012.

6. Numerod'ordre, Annee"Object Tracking and Re-identification in Multi-Camera Environments",Thesis University Lumiere Lyon, 17 June 2011.
7. BrankoMarkoski, ZdravkoIvankovic, LadislavRatgeber, PredragPecev,DraganaGlusac"Application of AdaBoost Algorithm in Basketball Player Detection" University of Novi Sad,ActaPolytechnicaHungarica Vol. 12, No. 1, 2015.

AUTHORS PROFILE

Mr. S. S. Gurav, completed his bachelor degree from in ETC from Shivaji university in the year 2010. He has completed Master of Engineering from Shivaji University in the year 2013. Currently he is PhD student in Shivaji University. His area of specialization is video processing.

Dr. B. B. Godbole, Professor,Department of Electronics
Karmaveer Bhaurao Patil,College of Engineering & Polytechnic,Satara.

Retrieval Number: J96630881019/19©BEIESP
DOI: 10.35940/ijitee.J9663.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)