REAL TIME ABANDONED BAG DETECTION USING OPENCV

Atul Borkar, Dr. M. S. Nagmode, Dhaval Pimplaskar

Abstract— *Abandoned Object Detection is one of the important tasks in video surveillance system. This paper proposes a work related to automatic detection of abandoned and unknown objects using background subtraction, morphological operation and centroid method. The aim of the approach is to automatically recognize activities around restricted area to improve safety and security of the servicing area. The system takes as input from the camera, tracking and recognition results and fuses these into object estimation. A proposed algorithm for object tracking in video, based on image segmentation is proposed. With the image segmentation all objects in video can be detected whether they are moving or not by using segmentation results of successive frames. Consequently, the proposed algorithm can be applied to multiple movements.The algorithm was tested on real time video surveillance system and it produces very low false alarms and missing detection. This approach definitely provides security and detects the moving object in real time video sequence and live video streaming.*

     Qualitative and quantitative results in terms of Detection Rate (DR), False Alarm Rate (FAR), success rate and average processing time per frame are given. The proposed algorithms are compared with the established methods based on simple difference and background subtraction.

Index Terms— AOD-abandoned object detection, *OpenCv, action detection, suspicious event detection, histogram equalization, DR*, FAR, success Rate.

——————————— ◆ ———————————

1 INTRODUCTION

Recent terrorist attacks have highlighted the need of video surveillance at public places. Also early event detection is required to avoid disasters like fire. In last two decades, visual surveillance has attracted more and more researchers because of its tremendous application prospects. In most of the current setups, the recorded video has to be analyzed to reconstruct an event after the related alarming situation has been detected. Because of the sequential scanning of entire video time lapse between occurrence of event and action required to correct the flaw increases considerably. Therefore there is a need for the development of smart surveillance system which will be able to automatically detect potentially dangerous situations. Researchers are working towards making the video surveillance systems more versatile, by developing fast, reliable and robust algorithms for moving object detection, classification, tracking and activity analysis [1].

In general an abandoned object is an object which is left at a particular place under surveillance and unattended over a period of time. Second, it should remain static in recent frames or for some time t. Detecting abandoned object is a very important in places like airports, railway stations, big shopping malls etc. where there is potentially high security threat. AOD is one of highly challenging task in video surveillance systems, lot of research is carried out to enhance and automate the surveillance system.

One of the major and important tasks in video surveillance system is to detect abandoned objects. The biggest challenge in abandoned Object Detection is classifying an object as abandoned object, the object which was not present previously in the scene may be an abandoned object and mostly the object becomes an abandoned object when it is carried by a person previously and it is unattended for a particular period of time which creates the potential threat [2].

Visual surveillance is an important computer vision research problem. As more and more surveillance cameras appear around us, the demand for automatic methods for video analysis is increasing. Such methods have broad applications including surveillance for safety in public transportation, public areas, and in schools and hospitals. Automatic surveillance is also essential in the fight against terrorism. In recent times there is lot of research done in field of abandoned object detection system for video surveillance systems with the human controlled or monitored CCTV systems. To improve the quality and the effectiveness of system various algorithms and techniques are suggested and implemented by researchers in various ways. But due to their complexity issues, the implementation was not so accurate using languages like Matlab, Java and the processing speeds of such a language are quit slowly. In this work the OpenCV tool is used with Microsoft visual studio 2008. The detailed algorithm is provided. In this research work absolute differencing method is used for background subtraction and histogram equalization is used for adjusting image intensities to enhance contrast. For security purpose it has become important to have in place efficient threat detection systems that can automatically detect and recognize potentially dangerous situations, and alert the authorities to take appropriate action by raising alarm on right time. When a person left the bag in restricted area, the system itself analyzes and determines its most likely bag position, where the position is defined as the location where the bag into the scene is left unattended. Through successive frames, the system keeps a lookout for the bag position and person position whose presence in the scene defines the status of the bag, and decides the appropriate course of action [3]. Automatic suspicious bag detection systems can assist security personnel by providing better situational awareness, enabling them to re-

——————————————————

- *Atul Borkar is currently pursuing masters degree in VLSI & Embedded in Pune University, India, E-mail:aborkarmit@gmail.com*
- *Dr. M. S. Nagmode received his Ph.D. from Pune University in Image processing. Currently he is working as Professor, in Department of E&TC , MITCOE, Pune, India, E-mail:manoj.nagmode@mitcoe.edu.in*
- *Dhaval Pimplaskar is currently pursuing masters degree in VLSI & Embedded in Pune University, India, E-mail:dhaval.87@hotmail.com*

spond to critical situations more efficiently.

2 LITERATURE SURVEY

Lots of work has been proposed for Abandoned Object Detection. The author of [4] proposed the method for detecting items of luggage left unattended at a busy train station in the UK. In this scenario, if an item of luggage is left unattended for more than 30s, an alarm should be raised. This is a challenging problem for automatic systems, as it requires two key elements: the ability to reliably detect luggage items and the ability to reliably determine the owner of the luggage and if they have left the item unattended.

Almost all current methods for static suspicious object detection are aimed at finding abandoned objects using a static camera in a public place, e.g., commercial center metro station or airport hall. Spengler and Schiele propose a tracking / surveillance system to automatically detect abandoned objects and draw the operator's attention to such events [1].

The object video surveillance system [5] keeps track of background regions which are stored right before they are covered by an abandoned object. This approach fails when the static objects stays long enough in the scene, which makes the matching more difficult due to differences in Lighting.

Jianling et al [6] uses different background modeling technique with GMM for different scene type. In this approach, pixel variance of foreground in recent frames are found and pixels with robust variances lower than threshold is taken as suspected regions and connected component analysis is done and the abandoned object is detected using classifiers. To overcome this problem, Yingli et al [7] proposed a method that uses more than one GMM to describe the statistics of each pixel. When an abandoned object is detected, GMM model A with high learning rate would show the static object faster than GMM model B with low learning rate. In method [8], texture information is integrated into the foreground mask computation to enhance Mixture of Gaussian method and region growing is used. Another approach is proposed which uses two backgrounds as in [previous] and uses radial reach filter to enhance the foreground mask.

The method which doesn't use background modeling is very less. A method is proposed in [9] in which unattended and stolen object is detected based on the fusion of information derived from three fast detectors. However, it exploits different types of information like shape, similarity, Contour similarity, background similarity.

In this approach, we introduce a method to detect the abandoned objects in real time environment. For detection of object we are using absolute difference method [10]. Tracking of the object is analyzed with reference of centroid of foreground object.

3 SYSTEM OVERVIEW

Fig.1. shows the system block diagram. The system has four main modules: (a) background subtraction; (b) morphological operation; (c) object tracking; (d) action detection. First background subtraction is performed to detect any new object that may have entered the scene. After that we determine which

person left bag in the scene. The final module differentiates between removed and abandoned objects. The system notifies the user of an abandoned object by raising an alarm.

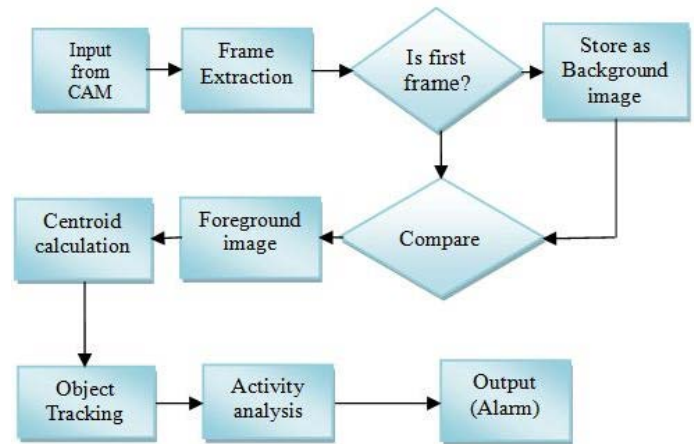The following sections will elaborate on each of these steps.

Fig. 1 Block diagram of the proposed system

The intelligent surveillance system consists of object detection, object features extraction, tracking and activity detection system is shown in figure.1. The proposed approach of whole system makes use of the observation. This system is able to distinguish moving and stopped foreground objects from the static background scene, track the objects and detect the unusual activity.

The computational complexity and the constant factors of the algorithms are important for video surveillance system. The selected algorithms for various problems in computer vision are affected by computational run time performance and their quality. Furthermore, our system uses the stationary camera. We initialize the system by giving the video imagery from a static camera where surveillance is provided.

3.1 BACKGROUND SUBTRACTION

Background subtraction is a critical part of object detection systems as its out came is fed to higher level processes such as object recognition and tracking and these processes rely heavily on the accuracy of background subtraction techniques. This method is particularly a commonly used technique for segmentation in static camera. It attempts to detect moving regions by subtracting the current image pixel -by- pixel from a reference background image. The pixels where difference is above a threshold are classified as foreground. After getting a foreground object, some morphological post processing operations such as dilation and erosion are performed to reduce the effects of noise and enhance the detected regions.

3.2 ABSOLUTE DIFFERENCING METHOD

We are taking absolute difference between background and foreground image. We use a simple adaptive background subtraction technique where we threshold the absolute difference between a

new image at time t, V(x, y, t+1), and a 'reference background image' V(x, y, t) containing only the background. This technique is very sensitive to changing background conditions.

Frame difference (absolute) at time $t + 1$ is

$$D(t+1) = |V(x, y, t+1) - V(x, y, t)| \quad \ldots\ldots\ldots\ldots\ldots (1)$$

Where, V is the velocity function of the moving object

x is the height of the foreground object

y is the width of the foreground object;

t is the time at which event occur.

The background is assumed to be the frame at time *t*. This difference image would only show some intensity for the pixel locations which have changed in the two frames. Threshold is put on this image to improve the subtraction. The equation for that given below

$$|V(x, y, t) - V(x, y, t+1)| > Th \quad \ldots\ldots\ldots\ldots\ldots (2)$$

Where, Th is threshold value

Higher the velocity of the moving object greater will be the threshold value we have to set.

3.3 HISTOGRAM EQUALIZATION

Histogram equalization is a method in image processing of contrast adjustment using the image's histogram. This method usually increases the global contrast of many images, especially when the usable data of the image is represented by close contrast values. Through this adjustment, the intensities can be better distributed on the histogram. This allows for areas of lower local contrast to gain a higher contrast. Histogram equalization accomplishes this by effectively spreading out the most frequent intensity values. The method is useful in images with backgrounds and foregrounds that are both bright or both dark. A key advantage of the method is that it is a fairly straightforward technique and an invertible operator. So in theory, if the histogram equalization function is known, then the original histogram can be recovered. The calculation is not computationally intensive. A disadvantage of the method is that it is indiscriminate. It may increase the contrast of background noise, while decreasing the usable signal [13].

Let's assume that a given image has a continuous range of intensity levels from 0 to 1 and let p(r) be the probability density function (PDF) of the intensity levels. We proceed to perform the transformation:

$$S = T(r) = \int p(w)\,dw \quad \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

Gonzalez and Woods [12] show that the output of such a system will have a uniform PDF at the output $P_o(s)$, thus

$$p_o(s) = \begin{cases} 1. For. 0 \le s \le 1 \\ 0. Otherwise \end{cases}$$

The result, then, will consist on an image whose intensity is distributed equally throughout the range, or, in other words, a high contrast image. From the equation, it is clear that T(r) is simply, the cumulative distribution function (CDF) of the system. If instead of continuous signals, we are working with discrete intensity levels, then p (rj) is really the normalized histogram of the input image with j=0, 1, 2...L being the discrete intensity levels and the transformation T (rk) is then known as the Histogram Equalization. Since we are working with discrete values, integration becomes summation and the transform function is then

$$s_k = T(r_k) = \sum_{j=1}^{k} p_r(r_j) = \sum_{j=1}^{k} \frac{n_j}{n} \quad \ldots\ldots\ldots\ldots (4)$$

Where, nj is the number of pixels with the intensity level j and n the total number of pixels.

Due to the discrete nature in the system, the output will not be completely uniform, although its dynamic range will increase dramatically. The result after histogram equalization is shown below
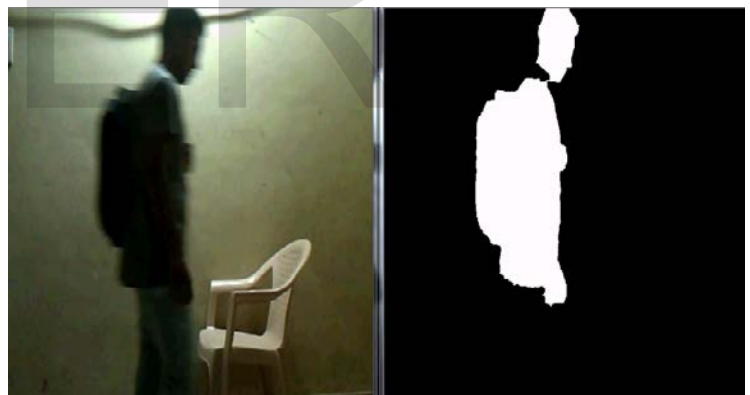

Fig.2 Results after Histogram equalization

3.4 MORPHOLOGICAL OPERATION

After histogram equalization, the morphological operation such as dilation and erosion are performing to get proper boundaries of foreground object.

The erosion of the binary image *A* by the structuring element *B* is defined by

$$A \ominus B = \{z \in E | B_z \subseteq A\} \quad \ldots\ldots\ldots\ldots\ldots (5)$$

Where, B_z, is the translation of B by the vector z, i.e.

$$B_z = \{b + z \,|\, b \in B\} \forall_z \in E \,\ldots\ldots\ldots\, (6)$$

When the structuring element B has a center (e.g., B is a disk or a square), and this center is located on the origin of E, then the erosion of A by B can be understood as the locus of points reached by the center of B, when B moves inside A.

The erosion of A by B is also given by the expression

$$A \ominus B = \bigcap_{b \in B} A_{-b} \,\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\, (7)$$

The dilation of A by the structuring element B is defined by

$$A \oplus B = \bigcup_{b \in B} A_b \,\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\, (8)$$

The dilation is commutative, also given by

$$A \oplus B = B \oplus A = \bigcup_{a \in A} B_a \,\ldots\ldots\ldots\ldots\ldots\, (9)$$

The dilation can also be obtained by

$$A \oplus B = \{z \in E \,|\, (B^s)_z \cap A \neq \emptyset\} \,\ldots\ldots\ldots\, (10)$$

Where, B^s denotes the symmetric of B, that is

$$B^s = \{x \in E \,|\, -x \in B\} \,\ldots\ldots\ldots\ldots\ldots\ldots\, (11)$$
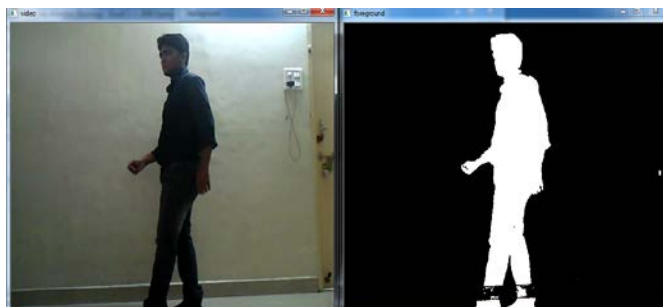

Fig. 3 Background subtraction before erosion

Fig.4 Background subtraction after erosion

3.5 EXTRACTING OBJECT FEATURE

After segmentation of foreground regions, we extract the features of corresponding objects from the current image scene. These features are the size (S_i) and center of mass (C_e) of the object. In order to estimate the size of the object we just calculate the number of pixels of foreground and to calculate the center of mass $C_e = (xC_i, yC_i)$ of an object O, use the equation given in [6].

$$xC_i = \frac{\sum_i^k x_i}{k}, \quad yC_i = \frac{\sum_i^k y_i}{k} \,\ldots\ldots\ldots\ldots\, (12)$$

Where, x_i, is the location of white pixel on x Co-ordinate,
y_i is the location of white pixel on y Co-ordinate

3.6 OBJECT TRACKING

The objective of object tracking is to construct a correspondence between objects in consecutive frames. Detection of objects for tracking in frame by frame is a significant and difficult problem. It is a crucial part for video surveillance system since without tracking the object, the system could not extract the information about objects and further higher level event analysis steps would be difficult. On the other hand, inaccurate segmentation of foreground objects due to occlusions, shadow, and reflectance makes tracking a difficult and active research problem.

An object level tracking algorithm is used in our video surveillance system. We don't track the object parts such as limbs of human, but track the object as a whole from frame to frame. In tracking step, the extracted information is adequate for most of the video surveillance applications.

3.7 ACTIVITY RECOGNITION

After successfully tracking the moving objects from one frame to another in a video, the problem of recognizing an event from scene follows naturally. Activity recognition involves action recognition and description. Activity recognition can guide the development of object motion analysis systems. It is the most important area of future research in motion analysis. Activity recognition is to analyze the human motion, and give high level description of actions.

The basic problem of activity detection is how to learn the reference action sequences, and how to effectively interpret events.

Terms appearing throughout the literature, such as "activity", "behavior", "action", "scenario", "gesture", and "event" are used to describe the same concepts. These concepts

have an ambiguous definition in literature. In this section the aim is to clarify these terms and propose a specific terminology which we will use to describe specific work. Activity detection could either be an application in sports, where the activity could be defined as a goal. It could also be an application in the surveillance area like abandoning or carrying of bag or accident detection through traffic surveillance video or person following someone with irregular speed.

In this activity, the person is moving along with a bag. If the person left that bag in the scene and moving forward or backward, then find out the distance between two objects. If the distance between them is greater than certain threshold value then we alert the guard by visual/audible alert because suspicious activity found. Otherwise if the person moving along with bag then the activity results called as normal activity.

4 ALGORITHM

The process is in real time so provide the video stream from live Cam. Here we are using inbuilt web cam. The frame is grabbing from the video. If the frame capture is first frame then store it as background image and new frames are subtracted from it. After background subtraction we get raw foreground image contain only moving object. For preprocessing, foreground image is fed to erosion and dilation operation. This processing removes the noise and gives us exact boundaries of foreground object. Applying threshold to image, if difference between these is greater than threshold value then foreground object is detected otherwise no foreground object is found.

After getting the foreground object, measure the height width of foreground object and find out the centroid of the image. Object is tracking with the reference of centroid. With the help of centroid we can find out the position of the object in the scene.

The most important step in system is to analyze the activity. For that we have to initialize the minimum and maximum limit of the x-coordinate. When person left an object in the scene and move away from it then calculate the distance between two objects. If the distance between these two is greater than some threshold value then alerts an authorized person with visual/ audible alert.

The experimentation is done using OpenCV 2.3 with Microsoft visual studio on windows platform.
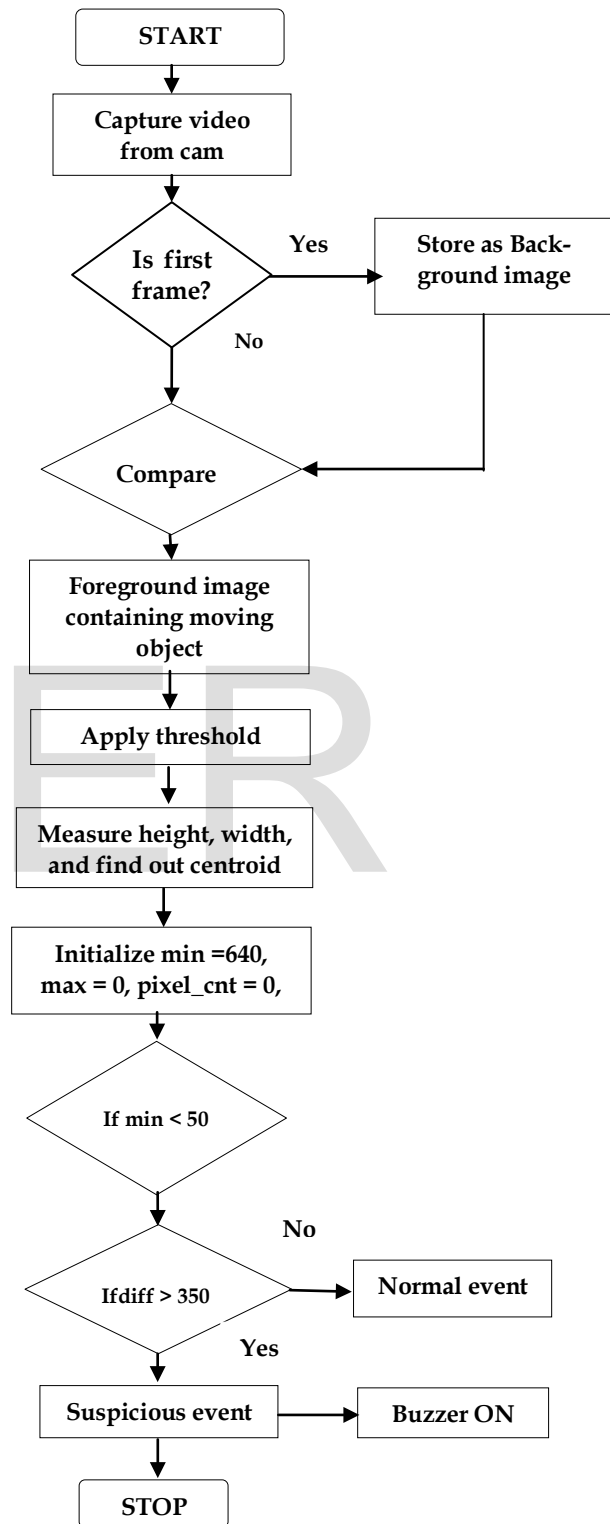

Fig.5 algorithm of proposed system
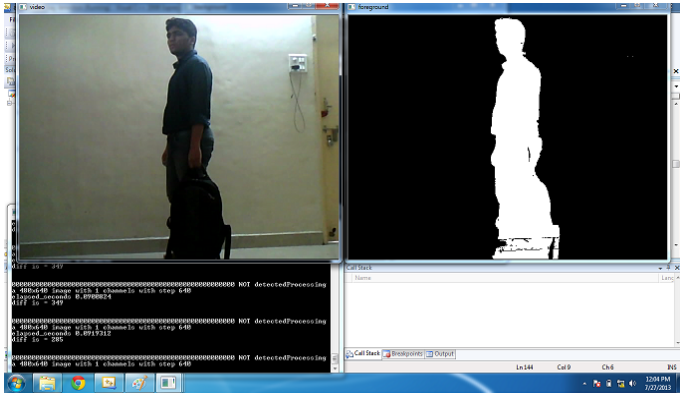
5 RESULT

Qualitative analysis:

Fig. 6 Normal activity detected when Person moving with a bag and doesn't left the bag

Here a person moving along with the bag and he doesn't left the bag in the scene so it is normal activity. That activity results shown by 00000000000000000000 and printing not detected.
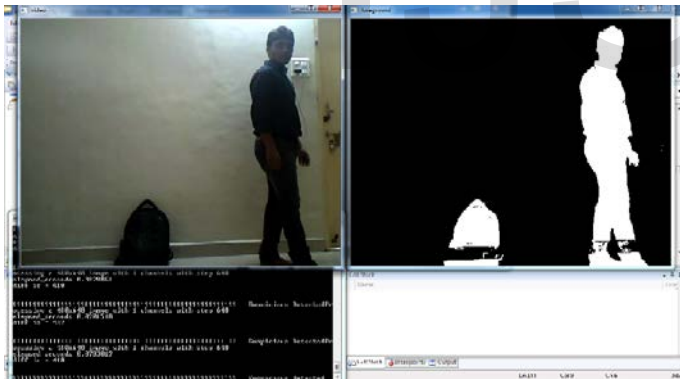

Fig.7 Result of suspicious activity detected for abandoned bag detection

Here a person moving along with the bag. He left the bag in the scene and moving away from the bag, so it is suspicious activity. That activity results shown by 11111111111111111111111 and printing suspicious detected with blinking the screen.

Quantitative analysis:

Quantitative analysis is done using two metrics viz. Detection Rate (DR) and False Alarm Rate (FAR). These metrics are calculated based on following parameters-

TP (*true positive*): detected regions that correspond to suspicious bag detection.

FP (*false positive*): detected regions that do not correspond to suspicious bag detection. (also known as false alarms).

FN (*false negative*): suspicious bag not detected (also known as misses).

These scalars are combined to define the following metrics:

$$DR = TP / (TP + FN) \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots \text{ (13)}$$

$$FAR = FP / (TP+FP) \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots \text{ (14)}$$

$$\text{Success Rate (\%)} = DR / (DR + FAR) \quad \dots\dots\dots \text{ (15)}$$

Table 1
Processing time

Sr. No	Different Environment	Processing Time of each frame (ms)
1	Environment 1	26
2	Environment 2	21.3
3	Environment 3	23
4	Environment 4	22.2
5	Environment 5	18.4

Results of abandoned bag detection for different environment based on above metrics.

Table 2
DR, FAR, success rate

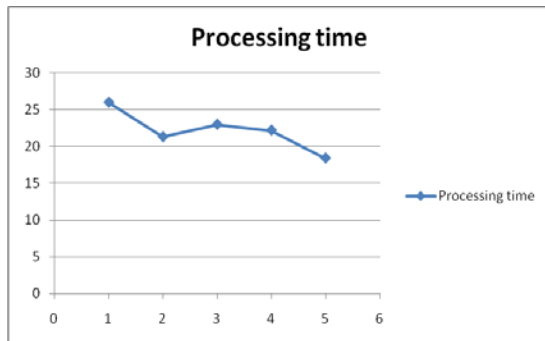| Environment \ Parameter	1	2	3	4	5
TP	54	54	54	60	60
TN	1	2	4	3	3
FP	0	2	3	1	3
FN	1	0	3	1	1
DR	98.18	100	94.73	98.36	98.36
FAR	0	3.57	5.26	1.64	4.76
Success Rate	100	96.55	94.73	98.36	95.38

Fig. 8 Processing time graph

6. CONCLUSION

In this paper, we presented methods for detecting irregularity in a video surveillance system. We implemented the system which automatically detected suspicious activity. The proposed object tracking method successfully tracks the whole body objects in consecutive frames but handling Occlusion problem is difficult.

Our system is designed for unusual activity detection task for single person in the real time. The implementation of this approach runs at 25-30 frames per second. The application is implemented in C++ using OpenCV library in Windows environment with a single camera view. The methods we presented for video surveillance system show promising results for activity.

REFERENCES

[1] Dr. M. S. Nagmode, Aditi S. Jahagirdar, Atul L. Borkar, Dhaval S. Pimplaskar, "*Moving Object detection and tracking based on correlation and wavelet Transform Technique to optimize processing time*" in CIIT International Journal of digital image processing, volume 5, No 1, January 2013, ISSN 0974-9691.

[2] Suresh.P and Dr.Mohan.S "*Abandoned Objects Detection in Video Surveillance System: A Survey*" in International Conference on Computer Science and Information Technology (ICCSIT'2011) Pattaya Dec. 2011, pp. 104-107.

[3] M. Bhargava, C-C. Chen, M.S. Ryoo, and J.K. Aggarwal, "*Detection of Abandoned Objects in Crowded Environments*", in Proceedings of IEEE Conference on Advanced Video and Signal Based Surveillance,2007, pp. 271 – 276

[4] K. Smith, P. Quelhas and D. Gatica-Perez, "*Detecting Abandoned Luggage Items in a Public Space*", *9th PETS, CVPR 2006, IEEE*, pp. 75-82

[5] P. L. Venetianer, Z. Zhang, W. Yin, A. J. Liptop, "*Stationary Target Detection Using the ObjectVideo Surveillance System*", IEEE International Conference on Advanced Video and Signal based Surveillance, London , UK , September 2007.

[6] Jianting,Wen,Haifeng Gong , Xia Zhang and Wenze Hu, "*Generative Model For Abandoned Object Detection*",IEEE International Conference on Image Processing, 2009.

[7] Liangliang Cao, YingLi Tian, Zicheng Liu, Benjamin Yao, Zhengyou Zhang and Thomas S. Huang, "*action detection using multiple spatial-temporal interest point features*", ICME 2010.

[8] Ferrando.S, Gera.G, Regazzoni.C, "*Classification of Unattended and Stolen Object in Video-Surveillance System*", in IEEE International Conference on Video and Signal Based Surveillance, 2006. AVSS '06

[9] J. San Miguel and J. Martnez, "*Robust unattended and stolen object detection by fusing simple algorithms,*" in *Proc. IEEE Int. Conf. on Advanced Video and Signal Based Surveillance (AVSS 08)*, 2008, pp. 18–25.

[10] Niraj Gadhe, Dr:B.K . Lande, VJTI," *Intelligent System for detecting, Modeling, Classification of human behavior using image processing, machine vision and OpenCV*" in IOSR Journal of Engineering (IOSRJEN), Vol. 2 Issue 2, Feb.2012.

[11] Gary Bradski and Adrian Kaehler, "Learning OpenCV Computer Vision with the OpenCV Library", O'Reilly,2008.

[12] Gonzalez, "Digital Image Processing", Pearson Education India, 2009.

[13] Robert Laganiere, "OpenCV 2 Computer vision Application Programming Cookbook" PACKT publication, May 2011.