ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

Abandoned Object Detection in Video Surveillance

Suyash Jagtap[1] , Saurabh Parkhi[2], Vishal Pithe[3], Shubham Menkudle[4], Pharande N[5]

B.E., Dept. of Computer, Sinhgad College of Engineering, Pune, India[1234]

Assistant Professor, Dept. of Computer, Sinhgad College of Engineering, Pune, India[5]

ABSTRACT: Surveillance cameras are cheap and ubiquitous. The advent of smart surveillance cameras with higher processing capabilities has now made it possible to design video surveillance systems which can contribute to the safety of people in the home and in public places such as shopping malls, airports, railways stations and etc. Terrorist attacks have become a critical threat of public safety especially, explosive attacks with abandonedpackages are repeatedly concentrated on such public places. A key function in such a surveillance system is the understanding of human behaviour in relation with objects left abandonedin public places.

The proposed system consists of three processing steps: (i) Object extraction, involving a new background subtraction algorithm based on combination of periodic background models with shadow removal and quick lighting change adaptation, (ii) Extracted objects classification as stationary or dynamic objects, and (iii) Classified objects investigation by using running average about the static foreground masks to calculate a confidence score for the decision making about event.

An automated video surveillance system for robustly and effectively detecting of abandoned objects is increasing the worldwide attention in many contexts, especially, in the consumer world of applications. In these systems, it should be a sufficiently high accuracy enabling a real-time performance. Thus, a prime goal of automated visual surveillance is to obtain a live description of what is happening in a monitored area and take appropriate action.

KEYWORDS: Abandoned object, surveillance, consumer video surveillance, intelligent analyzer, multiple background model.

I. INTRODUCTION

Object detection has evolved over the years. There have been many advances in the technology of object detection and classification. Object detection and classification is a very key part in today's surveillance systems. Technology can be classified into mainly three generations namely 1GSS, 2GSS, 3GSS.

1. 1GSS: This generation used analog systems for Image processing and the retrieval of data was a tedious job
2. 2GSS: This generation used both analog and digital systems which improved the speed of object detection
3. 3GSS: This generation used full-fledged digital systems for image processing and object classification. This proved the retrieval of the data to be fast and a central repository reduced the cost of the systems

The proposed system targets abandoned object detection and its database management in a video surveillance system for security purposes and marketing strategies.

II. MOTIVATION

This work is motivated by the ever increasing application of image processing and high capacity storage devices used in places such as military services, traffic surveillance etc. There is increasing demand for data analysis in the field of security systems, banking environment, forensics etc.

III. CONSUMER REQUIREMENTS ON ABANDONED OBJECT DETECTION IN VIDEO SURVEILLANCE

We live in a consumer surveillance society. In all the rich countries of the world everyday life is suffused with surveillance encounters, not merely from dawn to dusk but 24/7. In a world becoming ever more attuned to potential security threats, the detection of abandoned baggage is a key capability of any surveillance system. This leads to the consumer requirements to include an immediate identification of neglected baggage and simultaneous assessment of the circumstances of its abandonment, which can signify the difference between effective control and potential chaos. The main benefits for consumers from abandoned object detection in video surveillanceare:

1. Identifies suspicious, abandoned baggage within moments of its abandonment,
2. Allows fast recognition of the moment of abandonment to determine whether a threat exists.

Thus, we can see that the impacts of abandonedobject surveillance video analyzers on consumer world of applications are enormous especially when security and safety are concerned. In this aspect, the major challenges for consumer applications are as follows.

1. The stationary and non-stationary objects detection and analysis results should have sufficient accuracy for consumer acceptance and expectation.
2. High-processing efficiency achieving (near) real-time operation with low-cost consumer video cameras.
3. A conversion of visual results to a real world space can facilitate the analysis of special events such as very still person existence, cases of robbery, stolen and potential suicide bombing.

IV. PROBLEM DEFINITION

Detect abandoned object accurately from the video using an effective method and focus on detecting only the abandoned object and discarding other things extracted from the video like human being.

V. RELATED WORK

To address the challenging problem of accurately analysing the stationary and non-stationary objects detection and achieving high-level event analysis from monocular video sequences, the system should provide analysis at different semantic levels. A joint analysis tool is required to bridge the gaps between the pixel-level, object-level and event-level analysis and classifications. Our system has been designed such that it incorporates multiple levels of background models and motion analysis from the object-level onwards. The system can be utilized in surveillance applications with analysis results at all four levels.

For the second challenge, we have organized an evaluation of our method by partly embedding it in a new experimental real-time video content-analysis system. The evaluation has proved its efficiency, as it achieves a near real-time performance. More over resource management is applied to optimize the content-analysis processing, even when not all resource requirements of all components can be satisfied at the same time. Regarding to the third challenge we introduce a real world application scheme for scene understanding. The location and posture of persons are visualized in a virtual world after investigating context knowledge. The accurate and realistic reconstruction in a virtual space can significantly contribute to the scene understanding, like crime-evidence collection and healthcare behaviour analysis. Therefore, it is interesting to extend scene-reconstruction functionality in advanced surveillance applications, as consumers require more semantic results than only the conventional visualization that most of existing systems normally provide.

VI. PROPSED SYSTEM MECHANISM

A. Preprocessing: The periodic background modelling along with stochastic likelihood image and moving object detection are implemented. Each image within the video covering an individual human body and static objects are segmented to extract the 'blobs' representing foreground objects. In this processing, the periodic concept based backgrounds with periods of Short Length (SL) and Long Length (LL) are automatically built and updated by temporal statistical analysis. The main motivation is that the recently changed pixels that stay static after they changed can be distinguished from the actual background pixels and the pixels corresponding to the moving regions by analysing the intensity variance in different temporal scales. We employ the mixture of the periodic models along with Stochastically

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

Varied (SV) likelihood image background and update them based on stable history maps and difference history maps. After motion detection, a shadow removing procedure is performed on each image in order to discard shadow points that, generally, deform the shape of the moving objects. The intensity and texture information are integrated to remove shadows and to make the algorithm working for quick lighting changes.

B. Stationary Object Detection Processes: A matching algorithm is employed to detect if the object is abandonedlong enough to trigger the alert. Moreover a mixture of multiple statistical models is used to analyse the foreground as moving objects, abandonedobjects, or removed objects (ghosts), and still person while detecting the backgrounds. Different thresholds are used to obtain the foreground mask (for moving objects) and the static region mask (for stationary objects).

C. Classifying Process for Object Type: For the stationary region mask, a segmentation method is developed to detect the type of the static region (abandonedor removed or still), significantly outperforming previous techniques. Only those abandoned/removed objects that meet the user defined alert requirements will trigger the alerts. With the method proposed in this paper, our system can be more robust to illumination changes and dynamic background, and it can also work very well even if the images of the video are in low quality. In addition, the rule based classifier is used to distinguish the abandonedobject and the still-standing persons, which is a problem that is not solved in previous approaches.

VII. SYSTEM ARCHITECTURE

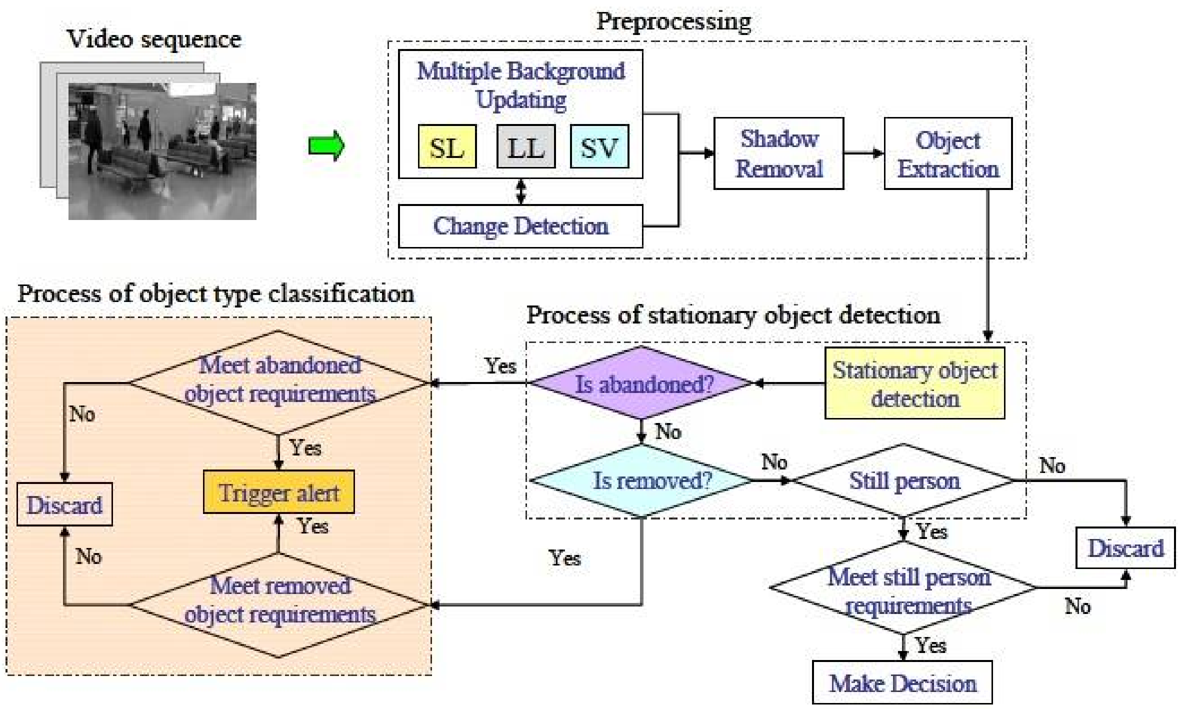

Fig.1 System Architecture

VIII. PROPOSED ALGORITHM

Most of the existing surveillance systems start with a period of empty scenes to facilitate the construction of the original background. This way of starting a surveillance system is hardly applicable to real world phenomena especially consumer world. In order to facilitate the consumer application areas such as crime-evidence collection, we develop a mathematical periodic concept for background maintenance and subtraction as a labelling problem in a series of images. Specifically, we establish three reference background models. They are named as:

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

1. Short Length periodic updated background model (SL),
2. Long Length periodic updated Background model (LL),
3. Stochastically Varied likelihood image model (SV).

For the first two backgrounds SL and LL the user can adjust the time interval between the update of reference background frames to adapt different needs and environments, furthermore, both the backgrounds update dynamically, the first one is updated frequently while the second one has a slower update rate according to the change of the environments. We then aggregate the frame-wise motion statistics into a stochastically varied likelihood image by updating the pixel-wise values at each frame. The periodic concept is originally used in theory of Markov chain for classification of states and it is now becoming a very successful tool in solving queuing and waiting line problems. But to the best of our knowledge we have not found any literature for applying this concept to image processing technologies and human behaviour analysis. It is completely new and earns novelty in this aspect. In our case the main motivation is that the recently changed pixels that stay static after they changed can be distinguished from the actual background pixels and the pixels corresponding to the moving regions by analysing the intensity variance in different temporal scales. At any given time, any given pixel is not only one element of a particular pixel process, but also one element of image. Contextual constraint of both temporal and spatial is necessary in the robust labelling. To model the temporal and spatial contextual information, our model for background has two components. One component processes images at pixel level and the other processes images at frame level. In pixel level process a background is determined by maintaining the most consistent states of each pixel within a certain time. With such background, the changed pixels which do not fit the requirement are obtained, also pixel colour, pixel intensity information is used for background process. Similarly, moving objects, lighting changes, and reflections on floors and walls need to clear up efficiently with only stationary objects remaining in the scene. Moreover, to avoid exhausted scanning of all possible bounding boxes, we first introduce two criteria to screen out a small number of suspected regions. To become an abandonedobject, two conditions should be satisfied. First, it should be a foreground object. Second, it should remain static in recent frames. This means that by comparing the original background with the moving foreground regions, we can hypothesize whether a pixel corresponds to an abandoneditem or not. On the other hand, an item stolen is original part of the background, when it is taken off from the scene, we could also determine whether the pixel belongs to a stolen object by the same principle. However, the background image cannot always maintain a static state, it must update with the changing circumstances.

Algorithm 1: **Periodically Updated Background Models**

The first frame of the inputting video image is initialized as SL and LL respectively in our application, and an improved adaptive background updating method is applied by constructing two maps of pixel history. The first map is Stable history Map (SM) which represents the number of times a pixel is stable in consecutive frames.

For the n^{th}image frame in a video sequence, a pixel is said to be stable if $| I_n(x, y) - I_{n-1}(x, y) | < Th_s$ and unstable of otherwise. Here,$Th_s$is a pre-defined threshold. We can then define the updating rule for SM as follows.

$$SM = \begin{cases} SM + 1 & \text{if } | I_n(x, y) - I_{n-1}(x, y) | < Th_s, \\ 0 & \text{otherwise} \end{cases} \qquad (1)$$

The initial value for each pixel in SM is set to 0. If a pixel is in the object plane, it is marked as unstable and set its value to 0. The second map is a difference history map (DM), which represents the number of times a pixel is significantly different from the background in consecutive frames. It is the condition for a stationary object becoming a part of background. By this definition, we observed that DM = n –SM where n stands for the total number frames in the sequence. The initial value for each pixel in DM is 0. If the pixel belongs to the object plane, its value increases by 1. Based on the information from both maps and taking the still object and uncovered background situation into account, the backgrounds adaptively updated frame-by-frame by:

$$SL_n(x, y) = \begin{cases} I_n(x, y) & \text{if } SM(x, y) > Th_f \text{ and } DM(x, y) > Th_f \\ SL_{n-1}(x, y) & \text{if } SM(x, y) > Th_f \text{ and } DM(x, y) = 0 \\ (1-\alpha)SL_{n-1}(x, y) + \alpha I_n(x, y) & \text{if } SM(x, y) = 0 \end{cases} \qquad (2)$$

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

$SL_n(x, y)$ and $SL_{n-1}(x, y)$ represent the short periodic updated backgrounds pixel value at position (x, y) in current and previous frames. In the same way, $LL_n(x, y)$ and $LL_{n-1}(x, y)$ represent the long length periodic updated background at position (x, y) and the corresponding updating rules are,

$$LL_n(x, y) = \begin{cases} SL_n(x, y) & \text{if } SM(x, y)>Th_f \text{ and } DM(x, y)>Th_f \\ LL_{n-1}(x, y) & \text{if } SM(x, y)>Th_f \text{ and } DM(x, y)=0 \\ (1-\beta)LL_{n-1}(x, y) + \beta SL_n(x, y) & \text{if } SM(x, y)=0 \end{cases} \quad (3)$$

Where, Th_f is a predefined threshold value and α, β are the learning rate of two backgrounds.
At every frame, we estimate the short length periodic foreground (SF) and long length periodic foreground (LF) by comparing the current frame I by the background models SL and LL. We obtain two binary foreground maps SF and LF where $F(x, y) = 1$ indicates the pixel (x, y) is changed. The LF shows the variations in the scene that were not there before including the moving objects, temporarily static objects, moving shadows, noise, and illumination changes that the background models fail to adapt. The foreground SF contains the moving objects, noise, etc. However, it does not show the temporarily static regions that we want to detect.
According to the updating rules, even if the foreground changes at a fast pace, it will not affect the background, but if the foreground is stationary, it will gradually merges into the background. In this way, we prevent the background model to be polluted by pixel which is logically not belonging to the background scene. Moreover, we could see that the intensity of each pixel of SL or LL has great connection with the corresponding foreground. Furthermore, the following inferences are made.

1. $SF_n(x, y) = 1$ and $LF_n(x, y) = 1$, which means (x, y) is a pixel indicate that there is a new moving object come into the scene, and it does not belong to any backgrounds. In this case, it can be seen that SL adapts itself to the relatively consistent changes, but it does not learn temporary color changes due to motion of the objects. Thus, such a pixel is marked as $SF(x, y) = 1$ in the short periodic foreground. Since the LL is updated less frequently, a temporary change cannot alter SL. The pixel is also marked as $F(x, y) = 1$ in the long periodic foreground mask.
2. $SF_n(x, y) = 1$ and $LF_n(x, y) = 0$, where pixel (x, y) is a part of the detected object which is changed now, then sets back to its original value. And an assumption that it takes more time to adapt the original background to the detected object than the change period is made. In case a pixel that was a part of the scene background is occluded for sometimes and then uncovered, the long periodic foreground will still be zero, $LF(x, y) = 0$. LL is updated less frequently hence it is not responsive enough to adapt to the new color during the occlusion. Yet, SL is responsive and adapts itself during the occlusion, which causes $SF(x, y) = 1$.
3. $SF_n(x, y) = 0$ and $LF_n(x, y) = 1$, (x, y) is a scene background pixel that was occluded before. A stationary pixel will be blended into SL i.e. $SF(x, y) = 0$ if it stays stationary long enough. Assuming this duration is not prolonged to blend the pixel in the scene background. As a result, the long periodic updated foreground will be one, $LF(x, y) = 1$. This is expected for the left behind items.
4. $SF_n(x, y) = 0$ and $LF_n(x, y) = 0$, which shows (x, y) is a pixel equal to both background pixel, this means that there is no change in the scene. In this condition, only background updating operate.

Algorithm 2: **Stochastically Varied Likelihood Image Model (SV)**
Although the relationship between two backgrounds and their relative foreground has been discussed in previous, but the case $SF_n(x, y) = 0$ and $LF_n(x, y) = 1$ is of great essential for detection. Under this condition, a pixel (x, y) may correspond to a static object, in the cause of the changed pixel already blended in SF_n, but not prolonged enough to blend in LF_n. Thus we will construct a stochastically varied likelihood image based on the two previous updating models as follows.

$$SV(x, y) = \begin{cases} SV(x, y) + 1 & \text{if } P(SF(x, y) = 1 \cap LF(x, y) = 0) = 1, \\ SF(x, y) - k & \text{if } P(SF(x, y) \neq 1 \cup LF(x, y) \neq 0) = 1, \\ 0 & \text{if } P(SV(x, y) < 0) = 1, \\ maxSV(x, y) & \text{otherwise} \end{cases}$$

Where max and k are positive numbers and $P(.)$ is the probability measure of an event.
The likelihood image enables removing noise in the detection process. It also controls the minimum time required to assign a static pixel as an abandoneditem. For each pixel, the likelihood image collects the evidence of being an

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

abandoneditem. Whenever this evidence elevates up to a preset level, we mark the pixel as an abandoneditem pixel and raise an alarm flag. The evidence threshold max is defined in term of the number of frames and it can be chosen depending on the desired responsiveness and noise characteristics of the system. In case the foreground detection process produces noisy results, higher values of max should be preferred. High values of max decrease the false alarm rate. On the other hand, higher the preset level gets longer the minimum duration a pixel takes to be classified as a part of an abandoneditem.

Algorithm 3:**Shadow Removing**
After the background subtraction only the blobs whose area is greater than a certain threshold are maintained.
Unfortunately each preserved blob contains not only the relative moving object but also its own shadows. The presence of shadows is a great problem for a motion detection system, because they alter real size and dimension of the objects. This problem is more complex in indoor contexts, where shadows are emphasized by the presence of many reflective objects; in addition shadows can be detected in every direction, on the floor, on the walls but also on the ceiling, so typical shadow removing algorithms that assume shadows in a plane orthogonal with the human plane, cannot be used. To prevent all these problems, correct shapes of the objects must be extracted and to do that a shadow removing algorithm is implemented. The shadow removing approach described here starts from the assumption that a shadow is a uniform decrease of the illumination of a part of an image due to the interposition of an object with respect to a bright point-like illumination source. From this assumption, we can note that shadows move with their own objects but also that they do not have a fixed texture, as real objects do: they are half-transparent regions which retain the representation of the underlying background surface pattern. Therefore, our aim is to examine the parts of the image that have been detected as moving regions from the previous segmentation step but with a texture substantially unchanged with respect to the corresponding background.
Formally, we evaluate, for each candidate point (x, y) the ratio as,
$R = I_n(x, y)/B_n(x, y)$ where $I_n(x, y)$and $B_n(x, y)$ are the intensity value the pixels (x, y) in the current image and in the background image, respectively. After this, pixels with uniform ratio will be removed. The output of this phase provides an image with the real shape of the detected objects, without noise or shadows.

IX. ABANDONED OBJECT ANALYZING PROCESS

In video surveillance one of the most important applications is to distinguish the abandonedor removed object from still person. In order to do so, we subdivide extracted objects moving object was classified into one of four types, Temporary Static Object (TS), Moving Person (MP), Still Person (SP),AbandonedObject (AO), and Unknown (U), using a simple rule-based classifier for the real-time process. It uses features such as the velocity of a blob, and exponent running average.
To classify, we used three critical assumptions:
1. Abandonedobject does not move by itself,
2. Abandonedobject has an owner and
3. The size of the abandonedobject is probably smaller than a person.

If objects were detected, they were initially classified as Unknown. Then, using the velocity of the moving object, the Unknown was classified as Person or AO. That is to say, if Unknown moved at a velocity higher than that of the threshold value, $Th_v$for several consecutive frames, it was identified as a Moving Person. If Unknown's velocity was below the threshold velocity TL_v, it was classified as (TS). If Unknown is identified as TS, AO and Still Person were distinguished by using the Exponent Running Average (ERA). If ERA is greater than a predefined threshold value Th_e, the TS is classified as still person and otherwise it will be abandonedobject.

X. EXPERIMENTAL RESULTS

Experiments were carried out in a public transportation environment. The test video sequences used are taken by our normal video camera at railway station. The environments are also randomly chosen. No special background conditions have been imposed. We have taken five video sequences in crowded environments in which some are in front of check-in gate. It also contains complex scenarios with multiple people sitting, standing and walking at variable speeds. Some are sitting in very still position. This type of environment is very common in our daily life. Even though most of existing methods so far do not take this type of realistic situations into account, the proposed method can handle

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

successfully these cases. We also have considered partial occlusion and sometimes completely occlude in a specified moment. All videos have instances of various shapes of abandonedobjects and still people. These are taken from different venues. Each video sequence tests a different viewpoint. So, we tested totally 20 video sequences with various public areas in real time environments. The images used here 320x240 pixels resolution. The frame rate is 10 fps.

We have also compared the proposed method with some traditional methods. We make a list of the method compared in our experiment with (i) single background model, (ii) dual background model and (iii) multiple background models (the proposed method). One scene includes at least 2000 frames, and its first 10 frames are used for initialization. According to our experimental results, the single background model and dual background models cannot handle the background changes, but the multiple background models with stochastically updated background reinforcement could detect object regions accurately compared with the other traditional methods.

From our experimental works, we also observed that the single background model is sensitive to the short-term illumination changes. It results in erroneous detection of the ground surface, the wall and so on. On the other hand, the dual background model is robust for the short-term illumination changes, but it detects not only the object regions but also surrounding pixels of the objects. Considering the characteristics of these two models, the advantages of the single background model matches the disadvantages of the dual background model, and vice versa. Both traditional models cannot detect the object location frame exactly but our multiple background approach has high advantage in this aspect which is the most important factor for abandonedobject detection problems. Moreover, our method works well without making any restrictions for the initialization. So, our method is useful for surveillance applications even though the pure background image is not available.

XI. CONCLUSION

We have proposed an innovative periodic concept based framework that enables multifunctional abandonedobjects in a human surveillance system for consumer use. This system can also be applied for detecting special events such as recording a burglary, robbery or monitoring school zone safety problems, for school children, thereby contributing to the safety of people in the home and schools. Moreover, it achieves an object recognition accuracy rate of about 95% to enable a realistic implementation of a surveillance system, or further analysis of human behaviour. The occlusion problem is also thoroughly tackled and successfully dealt with various aspects. The proposed object detection method works surprisingly well with well-known public datasets. Due to its simplicity the computational effort is kept low and no training steps are required.

REFERENCES

1. Monnet, A. Mittal, N. Paragios, and V. Ramesh, “Background Modeling and Subtraction of Dynamic Scenes,” Proc. of IEEE Int. Conf. on Computer Vision (ICCV), Nice, France, pp. 1305-1312, Oct. 2003.
2. Q-Z. Wu and B-S. Jeng, “Background subtraction based on logarithmic intensities,” Patter Recognition Letters, vol. 23, no. 13, pp. 1529-1536, Nov. 2002.
3. T. B. Moeslund, A. Hilton, and V. Kruger, “A survey of advances in vision-based human motion capture and analysis,” Computer Vision and Image Understanding, vol. 104, pp. 90-126, 2006.
4. A. Mittal, N. Paragios, “Motion-based Background Subtraction using adaptive kernel Density Estimation,” Proc. of CVPR, pp 302- 309, 2004.
5. P. Spagnolo, A. Caroppo, M. Leo, T. Martiriggiano, T. D’Orazio, “An Abandoned/Removed Objects Detection Algorithm and Its Evaluation on PETS Datasets”, Proc. of IEEE Conf. on Video and Signal Based Surveillance (AVSS’06), pp. 17-21, 2006.
6. F. Porikli, “Detection of temporarily static regions by processing video at different frame rates”, Proc. of IEEE Conf. on Advanced Video and Signal Based Surveillance (AVSS’07), pp. 236-241, 2007.

BIOGRAPHY

Suyash Jagtap, Saurabh Parkhi, Vishal Pithe, Shubham Menkudle are students in Sinhgad College Of Engineering, Computer Department, Savitribai Phule Pune University. Our research interests are Image Processing etc.