Detection of unattended and stolen objects in videos

Sharon Femi. P [1], Thaiyalnayaki. K[2]

Department of Information Technology

Sri Venkateswara College of Engineering, Chennai, India

Abstract— **This research work presents an efficient approach of detecting unattended or stolen objects in live videos based on background subtraction and foreground analysis. The most common algorithm for performing background subtraction is the Gaussian Mixture model (GMM). An improved Multi- Gaussian Adaptive background model is employed for background subtraction to determine the static region. A simple split and merge method is used to detect the static region from which the static objects are identified. The time and presence of static objects, which may be either unattended or stolen, are informed by sending a mail and SMS to the security officials. Also, Haralick's texture operators are employed for images to identify objects under low contrast situations. The system is efficient to run in real time and produce good results.**

***Keywords*— Background subtraction, Mixture of Gaussians, Static region, stolen object, unattended object.**

I. INTRODUCTION

The need for video surveillance has grown tremendously in many areas to maintain social control, recognize and monitor threats and prevent, investigate criminal activity. In addition to security applications, video surveillance is also used to measure traffic flow, detect accidents in highways and military applications [2]. It alerts the security officers of a burglary in progress or a suspicious individual loitering in a restricted area helping to avoid threat. Detection of objects plays an important role in surveillance system. The objects that are introduced in the foreground have to be detected instantaneously in time, in order to avoid dangerous situations. Identifying moving objects from a video sequence is a rudiment task for many computer-vision applications [3]. A common approach is to perform background subtraction, which detects the foreground objects from the portion of video frame that differs from the background model.

Background modelling is used in numerous applications to model the background and detect foreground objects in the scene as in video surveillance. It is the key step of background subtraction methods with the use of static cameras. The simplest background modelling involves acquiring the background image with no moving object so that image subtraction can be done to determine the moving objects. But, the problem is that the background cannot be obtained when dynamic changes occur under situations like illumination changes, camera jitter and movement in the background. The movement in the background may be either objects being introduced or removed from the scene. A good background model should react to quick changes in background and adapt itself so as to accommodate changes occurring in the background. To be robust and adaptable, many background modelling methods have been developed among which, the background subtraction models should have a good foreground detection rate and should be capable of operating in real time.

The system developed starts by detecting the objects placed idle for some time. The security officials are alerted of the situation to take the necessary action. This makes it easy for the security officials to bring the situation to notice thus preventing any security threat from occurring. The system captures the videos using static cameras.

An unattended object is a static object that is not in the scene before and stolen object is the object that was in the scene before but not present anymore. To detect the unattended and stolen objects, the static objects should be determined first. The static objects are the changes in the scene that remain in the same position for relatively long period of time.

The rest of this paper is organized as follows: Chapter 2 presents the related work. Chapter 3 presents the proposed work. Chapter 4 presents the background subtraction. Chapter 5 presents the static region and object detection method. Chapter 6 presents the statistical analysis of texture. Chapter 7 presents the experimental results in live videos.

II. RELATED WORK

Stationary objects in multiple object tracking [6], detects the foreground objects with several moving objects and is inspired by human's visual cognition processes. It relies on tracking information to detect drop-off events. This system produced larger errors under bright lighting conditions.

Magno *et al* [4] employs an unobtrusive embedded platform. This method uses a wireless video sensor to detect the abandoned object. The system employs multimodal sensor integration which saves power consumption. The objective is to develop a multimodal video sensor with low power and low cost to detect abandoned objects. This uses new algorithms for energy efficient image processing without giving up the flexibility of in-field configuration. In spite of using a video sensor, the number of false positives is 13% of the total detected objects.

Singh *et al* [1] uses a dual-time background subtraction algorithm to dynamically update two sets of background. This method is dynamic, adaptive, non-probabilistic and intuitive in nature. It uses pixel color/ intensity information for background processing. The binary image is divided into a number of legitimate blobs. Once the blobs are generated, the system applies an algorithm for tracking of the abandoned objects. The system is robust to variations in lighting conditions and the number of people in the scene. The system does not classify stable objects as unattended and removed objects.

Kong *et al* [3], detects nonflat abandoned objects by comparing a reference and target video sequences. The system uses GPS information to align the videos to find the frame pairs. The camera is mounted on a moving platform to scan along a specified trajectory for nonflat abandoned objects. The difficulty of the system is to cope with moving objects, presence of shadows, lighting conditions.

In robust detection of abandoned and removed objects in complex surveillance videos [7], the method detects abandoned and removed objects using GMM algorithm. The type of static regions is determined by a method that exploits context information. A matching method is used to detect the abandoned and removed object and it outperforms the edge based techniques. A person-detection process is integrated to differentiate static objects from stationary people. The system is robust to quick-lightning changes and occlusions. The accuracy of the detection is influenced by the size of the object, light conditions, and contrast situations.

III. PROPOSED SYSTEM

In this section, we provide a solution for detecting unattended and stolen object in videos. Figure 1 shows the proposed system diagram. The system includes the following modules. i) Background subtraction ii) Static region detection iii) statistical analysis of texture iv) unattended or stolen object alert detection.

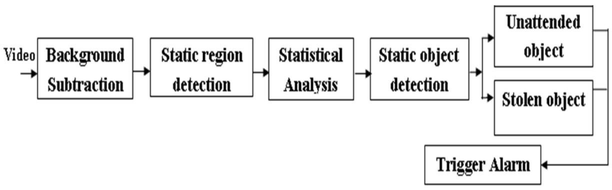

Fig 1: Proposed Detection System

The work introduced in this paper involves the following contributions.

i. Mixture of Gaussians are employed to detect the moving objects while subtracting the background.
ii. The frames are extracted from the live video and background subtracted video at the rate of 1 frame / 3 seconds.
iii. The extracted background subtracted frames are processed to determine the mismatched frames using a simple 'split and merge' algorithm.
iv. Haralick's texture features are determined for the frame so as to identify the static object under low contrast situation.
v. An alert is triggered by sending a mail and SMS on detection of unattended or stolen object in videos.

IV. BACKGROUND SUBTRACTION

In this section, we explain the background subtraction method employed to determine the static region. Background subtraction is performed using Mixture of Gaussians (MoG) method.

Mixture of Gaussians implements a classic multivariate Gaussian mixture model where every pixel is represented by a mixture of four Gaussian distributions. The modelling of the Gaussians is based on the Mahalanobis distance between the source and background model pixels. This model is designed to handle multimodal backgrounds with moving objects and illumination changes.

In Mixture of Gaussians, each pixel is characterized by its intensity in the RGB color model [7]. The various steps involved are explained in the following sections.

A. Pixel Characterization

The probability of each pixel value X_t is calculated as

$$P(X_t)=\sum_{i=1}^{k} w_{i,t} * \eta\left(X_t, \mu_{i,t}, \Sigma_{i,t}\right) \quad (1)$$

where,

k is the number of Gaussians (value may be 3 -5)
$w_{i,t}$ is the weight associated to the Gaussian i at time t
X_t is the pixel value at time t
μ is the mean value of the ith Gaussian distribution
Σ is the covariance matrix

η is the Gaussian probability density function defined as below:

$$\eta\left(X_t, \mu, \Sigma\right) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(X_t-\mu_t)^T \Sigma^{-1}(X_t-\mu_t)} \quad (2)$$

n is the dimension of the intensity at the pixel X.

Each pixel has the same covariance matrix and is of the form $\Sigma_{i,t} = \sigma_{i,t}^2 I$ and thus each pixel is characterized by a mixture of k Gaussians.

B. Parameter Initialization

The various parameters involved in Mixture of Gaussians are k, Σ, $w_{i,t}$. In our system, k is set to 4, Σ is initialized to 50 and $w_{i,t}$ is initialized as in equation 3.

$$w_{i,t} = (1 - \alpha) w_{i,t} + \alpha \quad (3)$$

where $\boldsymbol{\alpha}$ is the learning rate set to 0.001. The mean and covariance matrix of the Gaussian at each pixel is continuously updated.

C. Foreground Detection

Initial foreground detection is made and the parameters are updated. Initial foreground detection is made by determining the ratio $r = w / \sigma$ and order the Gaussians following this ratio. The first B distributions are considered as the background model, where

$$B = arg\, min_b \sum_{i=1}^{b} w_{i,t} > T \quad (4)$$

This ensures that a high weight with a weak variance refers to a background pixel. The other distributions are considered to present a foreground distribution. The pixels at each frame are classified as foreground or background by calculating the Mahalanobis distance between the source and background model pixels, and comparing this distance to a threshold.

When a new frame enters at time t+1, a matching test is performed for every pixel. The Mahalanobis distance between the source and background model pixels are calculated using the formula,

$$Dist = sqrt((X_{i+1} - \mu_{i,t})^T - \Sigma_{i,t}^{-1}(X_{t+1} - \mu_{i,t})) < k\sigma_{i,t} \quad (5)$$

where,
k and T are the threshold set to 2.5 and 0.5 respectively.

Two cases may occur as a result of the matching test:
Case 1: Match found with one of the k Gaussians.
In this case, if the Gaussian identified is one among the B distributions, the pixel is classified as background, else it is foreground pixel.

Case 2: No match with any of the k Gaussians
In this case, the pixel is identified as foreground. To proceed for the next foreground detection, the parameters must be updated.

Two cases occur in the foreground detection as below:
Case a: A match found with one of the k Gaussians.
The updation of values for the matched component is as follows

$$w_{i,t} = (1 - \alpha)w_{i,t} + \alpha, \quad (6)$$

where α is the constant learning rate

$$\mu_{i,t+1} = (1-\rho)\,\mu_{i,t} + \rho X_{t+1} \quad (7)$$
$$\sigma_{i,t+1}^2 = (1-\rho)\sigma_{i,t}^2 + \rho(X_{t+1} - \mu_{i,t+1}).(X_{t+1} - \mu_{i,t+1})^T \quad (8)$$

where

$$\rho = \alpha.\eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \quad (9)$$

For the unmatched component, the μ and Σ remains unchanged and only the weight is updated as

$$w_{j,t+1} = (1-\alpha)\, w_{j,t} \quad (10)$$

Case b: No match with any of the k Gaussians
In this case, the distribution k is replaced with the parameters
$w_{k,t+1}$= low prior weight (11)
$\mu_{i,t+1} = X_{t+1}$ (12)
$\sigma_{k,t+1}^2$= large initial variance (13)

Once the parameters maintenance is made, foreground detection can be made and so on. The blind update employed by the method makes it less sensitive to initial conditions but tends to integrate stationary foreground objects into the background.

V. STATIC REGION AND OBJECT DETECTION

Static region is the region that has recently changed in the scene. To determine the static region from the background subtracted video, we propose an algorithm to determine the mismatched frames. Mismatched frames are the frames that contain a recent change in the scene and it may correspond to static object. The background subtracted video has to be converted to frames for the proper functioning of the algorithm. Thus the frames extracted from the background subtracted video consist of either black pixel of white pixel.

The algorithm works by processing *n* number of frames simultaneously by parallel execution at any given time. The various steps involved in the algorithm are

i. Each frame is divided to sub-blocks of size, say *k*.
ii. The pixel count of each of the *k* sub-block has to be determined
iii. If the number of white pixels of the sub-block is greater than a threshold value, that sub-block is considered as “white block”, else the sub-block is considered as a “black block”.
iv. The identified number of white blocks of a frame are stored in a buffer.
v. If the consecutive values of the buffer remain the same, then the corresponding frames are to be considered as mismatched frames.
vi. An alert is triggered by sending mail and SMS to indicate the presence of abandoned or stolen object.

The process is repeated for the next *n* frames of the background subtracted video till the end of the video so as to identify the mismatched frames.

VI. STATISTICAL ANALYSIS

To detect the objects that is of the same colour as that of the background situations in an image, we determine the following Haralick’s texture operators [8]

i. Angular Second Moment(ASM) $=\sum_i \sum_j \{p(i,j)\}^2$
ii. Entropy = $-\sum_i \sum_j p(i,j) \log p(i,j)$
iii. Homogeneity=$\sum_i \sum_j \{\frac{p(i,j)}{(1+|i-j|)}\}$
iv. Mean=$(\sum_i \sum_j P_{i,j})/ij$
v. Variance=V=$\sum_i \sum_j (i-M)^2 p(i,j)$

The image is divided to four equal quadrants. The above Haralick’s texture operators are calculated for the four quadrants and compare the values of each parameter. The quadrant with the minimum value for all the operators may contain an unattended object of the same texture.

VII. EXPERIMENTAL RESULTS

This section presents the results of the experiments conducted for the proposed method. The system is developed using Visual Studio and runs for live videos taken in the webcam to identify the static object. The system is able to detect the abandoned and stolen object with limited delay as live videos are monitored instead of datasets.

The importance is the system is able to identify the static object if the lightening effects changes, which can be seen from the output presented in Fig 2.

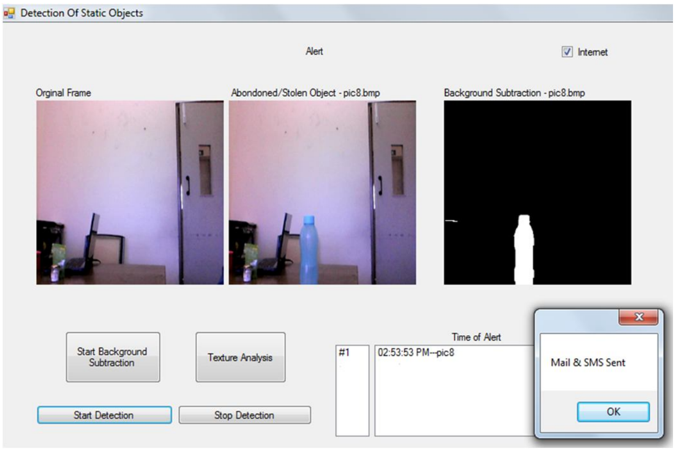

Fig 2: Detection of static object

The detection system also tries to detect the objects of same colour under low contrast situations in images using Haralick's texture operators. The mean, variance, IDM, ASM, entropy and homogeneity are calculated and if a change is noted in different quadrants, an alarm is raised to inform the presence of an object of the same colour. The results are shown in Fig 3.

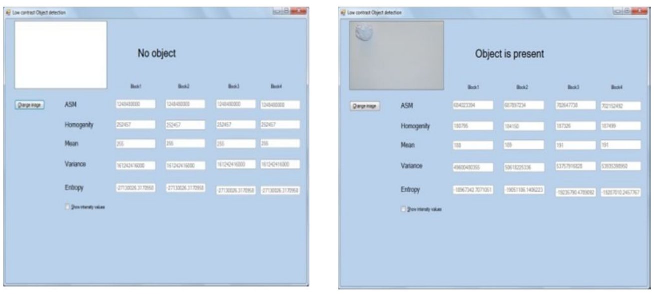

Fig 3. Statistical analysis of images

However, it is observed that the developed system suffers from shadow effects. The false negatives and false positives are avoided and the detection system developed is able to produce true positives for the live videos with limited delay.

VIII. CONCLUSIONS

Thus a framework to detect the unattended and stolen object in live videos has been implemented successfully. The mixture of Gaussians BGS method is used to detect both background and static foregrounds. The static region is determined using the simple split and merge algorithm. The static region is either classified as unattended or stolen object and the notification is given to the security officials by mail and SMS regarding the time of the presence of static object. Also, a trial for detecting objects under low contrast conditions are carried out using Haralick's texture operators and the results are shown. The testing results, based on live videos, have proved that this approach can be employed in video surveillance applications.

REFERENCES

[1] A.Singh, S.Sawan, M.Hanmadhu, V.K.Madasu, B.C.Lovell 2009 An abandoned object detection system based on dual background segmentation in Proc. IEEE Conf. Adv. Video and Signal Based Surveillance.

[2] E. Auvinet, E. Grossmann, C. Rougier, M. Dahmane, and J. Meunier 2006 Left-luggage detection using homographies and simple heuristics in PETS 2006, pp. 51–58.

[3] Hui Kong, Jean Ponce 2010 Detecting Abandoned Objects with a Moving Camera in Proc. IEEE Transactions on Image Processing,Vol 19,2010.

[4] M.Magno, F.Tombari, D.Brunelli, L. Di Stefano and L.Benini 2009 Multimodal abandoned/removed object detection for low power video surveillance systems in Proc AVSBS09, pp 188-193.

[5] N. Bird, S. Atev, N. Caramelli, R. Martin, O. Masoud, and N. Papanikolopoulos 2006 Real time, online detection of abandoned objects in public areas in Proc. IEEE Int. Conf. Robot. Automation, pp. 3775–3780.

[6] S.Guler, J.A. Silverstein, and I.H.Pushee 2007 Stationary objects in multiple object tracking in Proc. IEEE Conf. Adv. Video Signal Based Surveillance, London, U.K.

[7] Y Tian, Rogerio Schmidt Feris, Haowei Liu, A Hampapur and Ming-Ting Sun 2010 Robust Detection of Abandoned and Removed Objects in Complex Surveillance Videos in Proc. IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews, Vol. 41.

[8] Robert M.Haralick,K.Shanmugam and Its'hak Dinstein 1973 Textural Features for Image Classification in IEEE transactions on Systems, Man and Cybernetics November 1973, pp 610-621.