A Novel Probabilistic Video Analysis for Stationary Object Detection in Video Surveillance Systems

Thi Thi Zin, *Member, IAENG*, Pyke Tin, Takashi Toriu, and Hiromitsu Hama

Abstract— **In this paper, we propose a novel probabilistic approach for detecting and analyzing stationary objects driven visual events in video surveillance systems. This approach is based on a newly developed background modeling technique and an adaptive statistical sequential analysis method. For background modeling part, we use the concepts of periodic Markov chain theory producing a new background subtraction method in computer vision systems. We then develop an object classification algorithm which can not only classify the objects as stationary or dynamic but also eliminate the unnecessary examination tasks of the entire background regions. Finally, this paper introduces a sequential analysis model based on exponent running average measure to analyze object involved events such as whether it is either abandoned or very still person. In order to confirm our proposed method we present some experimental results tested on our own video sequences taken in international airports and some public areas in a big city. We have found that the results are very promising in terms of robustness and effectiveness.**

Index Terms—**stationary object, background models, video surveillance, exponent running average**

I. Introduction

IN the recent years, video surveillance systems have become an extremely active research area due to a sharp increasing in the levels of terrorist attacks on crowded public places, like airports, stations, subways, entrances buildings, sporting events, and other public venues. Terrorist attacks have also a critical threat of public safety; especially, explosive attacks with abandoned/removed or unattended objects/ packages are repeatedly concentrated on the public places. Hence, establishing a surveillance system with high-tech appliances to against terrorism is a critical issue nowadays. This has led to motivation for the development of a strong and precise automatic processing system, an essential tool for safety and security in both public and private sectors. Video surveillance systems aim to provide automatic analysis tools that may help the supervisor personnel in order to focus his/ her attention when a dangerous or strange event takes place.

In this context, the detection of stationary objects is receiving a special attention because it is a critical analysis stage in applications like the detection of abandoned objects or parked vehicles frequently used in the surveillance of public areas. Additionally, the recognition of stationary objects in crowded unconstrained contexts is a challenging task. Issues related to occlusions (by moving or stationary objects), appearance variations (e.g., color composition, shape) as people move relatively to the camera, lighting changes, speed and density structure of moving objects should be taken into account. Thus an essential component of a video surveillance system is the capability of correctly and accurately detecting suspicious objects and people involved in crowded areas. So that the system can be able to help the monitoring personals to immediately find a dangerous or strange event takes place in the monitored area.

Moreover, automatic analysis and interpretation tools are required to obtain the real time demands in a video surveillance system. For this purpose, immediate detection of suspicious packages or objects is vital to the safety of innocent citizens in the current age of terrorists who often use primitive home-made explosive devices. Thus, solving the problem of detecting stationary objects (also referred to as abandoned, static, left, or immobile objects) is currently one of the most promising research topics for public security and video surveillance systems. Furthermore the problem is increasing the worldwide attention in many contexts, especially for its application in crowded environments potentially at risk which therefore require particular controls to guarantee security.

However, high-level video interpretation tasks related to surveillance are usually completely performed by human operators, who have to process large amounts of visual information presented to them through one or more monitors. Various scenes are often guarded simultaneously by a single operator. It is widely known that, if the operator is exposed to this type of work for several hours his attention decreases, thus probability of missing dangerous situations increase. It is crucially important to support human conduct with semiautomatic surveillance systems in order to inform the supervisor in case of abandoned object and to focus his

Manuscript received Aug 13, 2012. This work is partially supported by SCOPE: Strategic Information and Communications R&D Promotion Program (122307004).

Thi Thi Zin is with Graduate School of Engineering, Osaka City University, 3-3-138 Sugimoto, Sumiyoshi-ku, Osaka 558-8585, Japan (Phone & Fax: 81-6-6605-3096, e-mail: thithi@ip.info.eng.osaka-cu.ac.jp).

Pyke Tin was with University of Computer Studies, Yangon, Myanmar. He is now with Graduate School of Engineering, Osaka City University. (e-mail: pyketin@ip.info.eng.osaka-cu.ac.jp).

T. Toriu is with Graduate School of Engineering, Osaka City University. (e-mail: toriu@info.eng.osaka-cu.ac.jp).

H. Hama was with Graduate School of Engineering, Osaka City University. He is now with R&D Center of 3G Search Engine, Incubator, Research Center for Industry Innovation, Osaka City University, Osaka, Japan (e-mail: hama@ado.osaka-cu.ac.jp).

attention on the event [1]-[2]. Abandoned/stationary object detection is the task of locating objects that are left behind in a scene. Often these objects are quite small (compared to the people at least) and are frequently occluded by other people or vehicles moving about the scene.

In the literature, several methods have been found describing on abandoned object detection and their applications to public safety and security problem. They can be categorized into two approaches: one is based on tracking methodology [3]-[6] and the other is based on detection approach [7]-[12]. The tracking-based methods encounter the problems of merging, splitting, entering, leaving, occlusion, and correspondence. These problems are not easy to solve in many cases since it is difficult to track all the objects precisely in crowded situations. On the contrary, the detection-based methods do not need to handle the complicated problems associated within the tracking-based methods, and only the abandoned objects that are not there initially should be of concern.

Although many researchers have paid lots of attentions on above research aspects, however, still few papers can be found that they changed the classic tracking and people detection problem by applying the static foreground regions [13]-[14]. In this aspect, the existing methods can be divided into two categories according to their use of one or more background subtraction models. For example, a statistical model of the background is used to detect foreground regions and to eliminate object shadows [15]. On the other hand, two backgrounds modeling techniques in order to detect stationary objects have also been appeared in the literature [16]-[17]. By using samples taken on the basis of frame rates the two backgrounds are established. But this approach fails to develop a mechanism that can correctly classify the events whether it is abandoned or very till person. It is worthwhile to note that most of existing surveillance systems do not work well when the initial background contain object left behind in the scene.

We also observe that each approach has merits and demerits depending on the assumptions of characteristics of the background and the illumination [18]-[19]. Moreover, there exists a class of problems that traditional single foreground/background detection methods still cannot solve. For instance, left behind objects, such as suitcases, packages, etc. are needed to be paid a special attention. They are static; therefore, they should be labeled as background. On the other hand, they should not be ignored as they do not belong to the original scene background. Therefore, to achieve more robust object detection, or to acquire more effective background model, we should combine adaptively background models having different characteristics.

In this paper, we present a new method that use periodic backgrounds and does not require object tracking. Our method does not require object initialization, tracking, or offline training. It accurately segments objects even if they are fully occluded for a certain period of times. The system is able to deal with people who stop and sit for extended periods of time and not regularly detect them as abandoned objects. A sequential analysis is introduced to classify detected objects as either an abandoned object or a still person. This paper extends and modifies to achieve more fruitful results for our previous work presented in [20].

The organization of the remaining parts of paper is as follows. Section II presents a condensed overview of the various background subtraction approaches used for stationary objects and motion detection. Section III contains our proposed stationary object detection method and the classification of detected object types. Section IV covers some experimental results on standard datasets as well as our real-world surveillance scenarios. Finally, concluding remarks and discussions are presented in section V.

II. Related Work

In this section, we describe some related works for categorizing the stationary foreground detection methods based on background-subtraction techniques (see Fig. 1). Depending on the use of one or more background subtraction method [21-28], these techniques can be categorized as one model approach and two models approach.

A. One Model based Background Subtraction

In the one model approach, three types of techniques are involved. They are

(i) standard background technique,
(ii) model property analysis technique, and
(iii) sub-sampled analysis technique.

Standard background technique

The standard background technique describes the methods that employ typical background subtraction techniques followed by another type of analysis (e.g. tracking). In the typical background segmentation stage a Gaussian Mixture Model (GMM) is usually employed along with a blob tracking analysis stage. This tracking analysis is based on finding the correspondence between the blobs identified in two consecutive frames. Some rules, as color, shape, distance or object size are used in this module to perform the tracking analysis. Fig. 2 depicts the processing scheme followed by the selected approach.

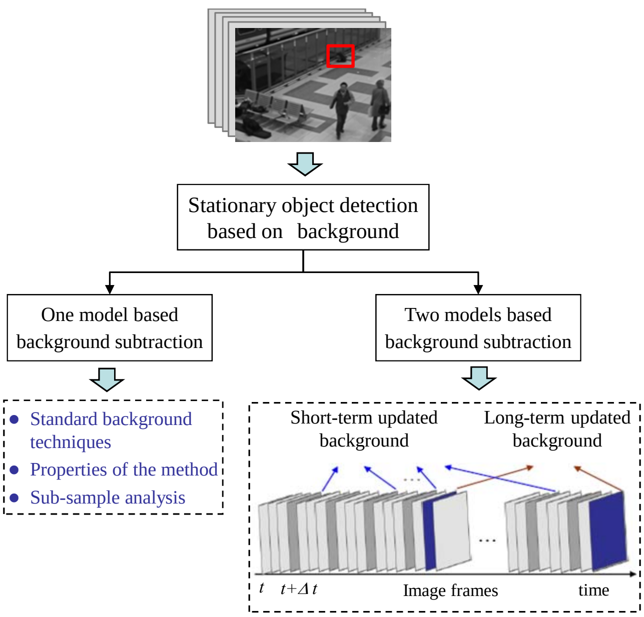


Fig. 1. Stationary foreground detection methods based on background subtraction techniques.

(Advance online publication: 28 August 2012)

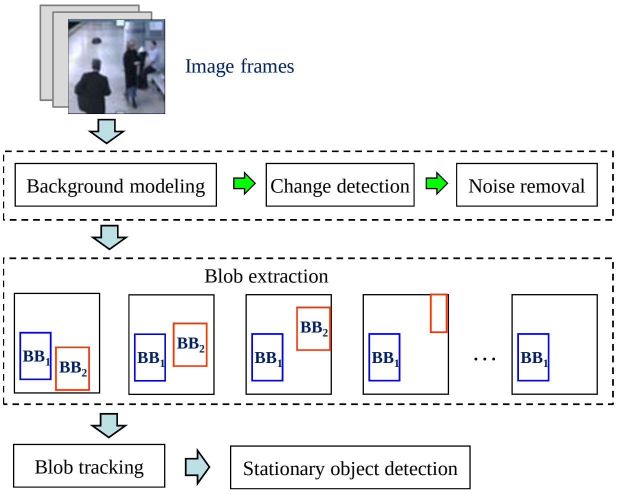


Fig. 2. Illustration of stationary object detection procedure.

Model property analysis technique

The model property analysis technique usually focuses on the use of the GMM for detecting foreground objects and inspecting the properties of that model to detect stationary objects. The stationary object detection is based on the observation of the transition states between the new Gaussian distributions created (for the new foreground pixels detected) and their transition to the dominant background state. Maximum of three Gaussians distributions are used in the GMM model resulting property analysis diagram shown in Fig. 3. This approach describes a set of necessary conditions and corresponding observations on the diagram to detect stationary objects imposing time stability, spatial stability and enough distribution weight constraints.

Sub-sampled analysis technique

In this sub-sampled analysis technique as described in [21], we first compute the intersection of a number of background subtracted frames which are sampled over a period of time. Since the abandoned objects are assumed to be static foreground blobs, they will be contained in this intersection set. Here, each pixel is assumed to obey the probability law of Gaussian distribution while a background subtraction process is performing. In addition, the gradual intensity changes of each pixel are taken into account for a weight function. Then a number of sample foreground marks are taken from the last frames to be analyzed. Let the number of samples be n, in this case we assume n = 6 and the sample frames are denoted by $F_1, \ldots, F_6$.

We also denote $F_k(i, j)$, $k = 1, \ldots, 6$ and $B(i, j)$ as the foreground and background for the pixel (i, j) respectively. We then define that $F_k(i, j)$ is a foreground pixel if and only if

$$|F_k(i,j) - B(i,j)| > w(i,j)\sigma(i,j) \text{ for } k = 1, \ldots, 6,$$

where $w(i, j)$ is a weight function due to the gradual intensity changes and $\sigma(i, j)$ is the standard deviation of the image. It is observed that the weight function $w(i, j)$ is directly proportional to the values of i. It means that when the value of i decreases then $w(i, j)$ decreases and it becomes larger for large value of i. After this, the sample foreground marks are binarized and symbolized by $M_k(i, j)$ for $k = 1, \ldots, 6$, such that $M_k(i, j) = 1$ if $F_k(i, j)$ is a foreground pixel and $M_k(i, j) = 0$ for otherwise. Then these foreground marks are convolved to obtain the intersection set **S** as the static foreground object. In mathematical terms, the set **S** can be expressed as $\mathbf{S} = M_1 * M_2 * \ldots * M_6$, where '*' is a convolution operator. Since the binarization allows the intersection set **S** to be taken as point-wise multiplication over all sampled foreground marks, **S** indicates a region that should be very likely to correspond with stationary objects. These processes are illustrated in Fig. 4.

B. Two-Model based Background Subtraction

In this category, a detection stage using on the application of two background subtraction methods at different frame rates is considered in [13], [21]. The two models are based on the GMM employing one model for short-term detection (updating it every frame) and another for long-term detection (updating it every n frames). Short-term Background (SB) is adapted faster and the scene changes are introduced more quickly on it. On the other hand, Long-term Background (LB) is adapted to the changes of the scene at a lower learning rate. Then, the foreground masks of the two models are computed at every frame and a combination of them is performed (see Fig. 5).

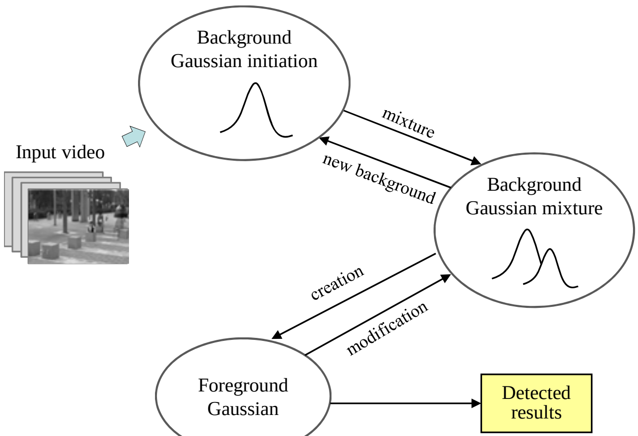


Fig. 3. Model property analysis diagram.

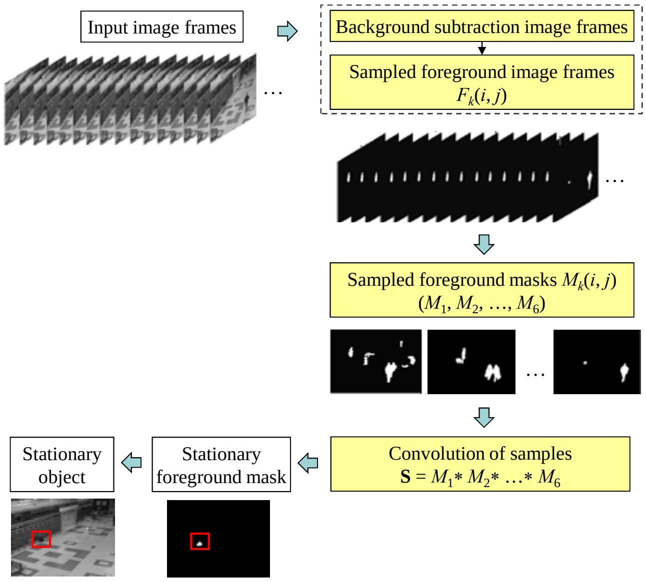


Fig. 4. Illustration of sub-sample analysis.

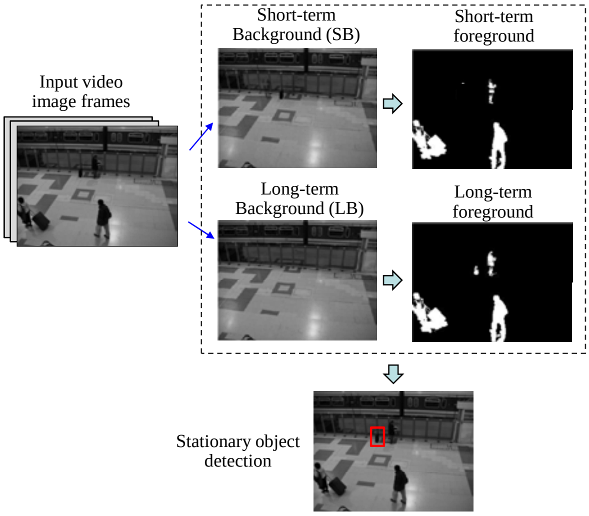


Fig. 5. Illustration of Two-Model based Approach.

The first frame of the incoming video is initialized SB. Subsequently, the intensity of each pixel of this background is compared with the corresponding pixel of the next frame (after every 0.4 seconds). If it is less, then the intensity of that pixel of background is incremented by one unit, otherwise it is decremented by one unit. In case of equality, the pixel intensities remain unchanged. This way, even if the foreground is changing at a fast pace, it will not affect the background but if the foreground is stationary, it gradually merges into the background.

To investigate all those objects which are stationary for a long period of time (and thus have gradually merged into the background), it is necessary to maintain another set of background images called LB. Here, all those pixels which do not belong to the prospective static objects set are made equal to that of SB. This is done at an interval of every 20 seconds.

Difference of the two backgrounds is represented as a binary image with the white portion representing foreground (blobs).

III. Proposed Method

In this section, we describe a novel solution to detect abandoned, removed objects and still person. Fig. 6 shows our proposed system architecture which contains three components:

(a) multiple background subtraction and moving object detection,

(b) stationary object detection process, and

(c) classifying process for object type.

In the motion detection process, the multiple backgrounds are updated by using 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 analyzing the intensity variance in different temporal scales. We employ the mixture of pixel processing models along with stochastic background and update them based on stable indicator set and difference indicator set. Then the motion detection process is immediately followed by the shadow removal process to discard shadow pixels. For shadow removal process, we employ both intensity and texture information. Thus, the process can work well for quick lighting changes.

Moreover a mixture of multiple statistical models is used to analyze the foreground object for further classification. In this analysis, four object types such as moving objects, abandoned objects, removed objects (ghosts), and still person are to be classified. Different thresholds are used to obtain the foreground mask for moving objects and the static region mask for stationary objects. 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, a statistical analysis classifier is used to distinguish the left-baggage and the still-standing persons, which is a problem that is not solved in previous approaches.

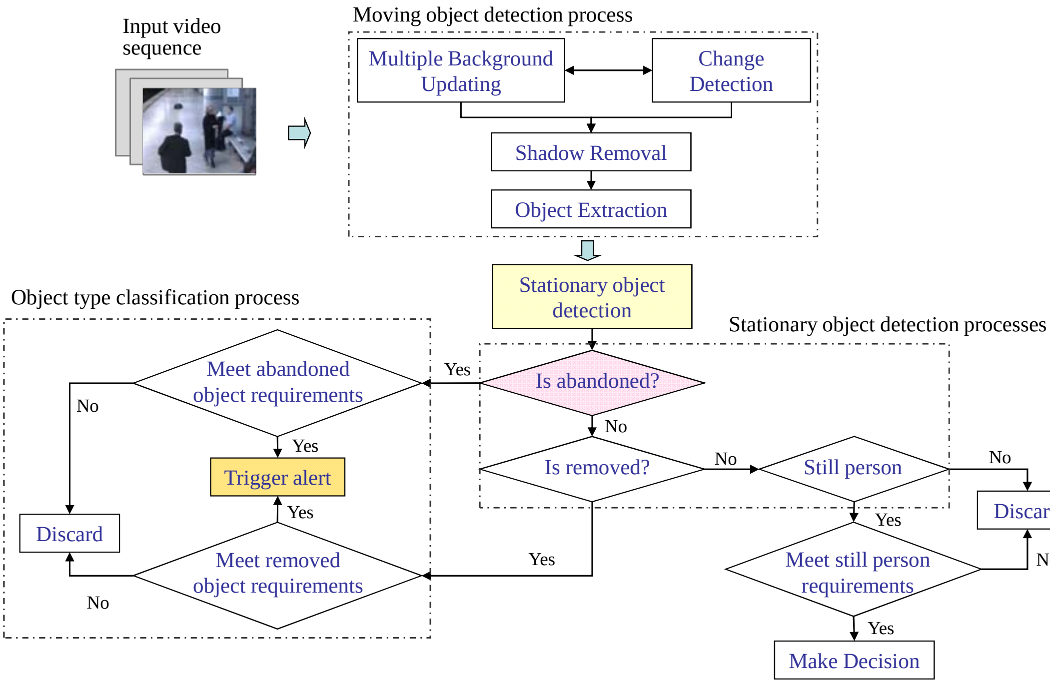


Fig. 6. Abandoned Object Detection System Architecture.

A. Novel Background Modeling

Generally speaking, most of the surveillance system starts with a period of empty scenes to facilitate the construction of the original background. In our approach, this constraint is not required. Mathematically, background maintenance and subtraction can be formulated as a labeling problem in a series of images. 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 labeling. 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 color, 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. However, no matter which kinds of background models are applied for object detection, because of the updating rate, the pixels belonging to temporary static object may be mistaken as a part of the background or the moving regions.

So, we consider a single background model is not sufficient to separate the temporarily static pixels from the scene background, and then a new background subtraction method based on three backgrounds is developed. 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 abandoned object, 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 abandoned item 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.

Hence three reference background models are established. They are named as:

1) Frequently-updated Background (FB) model,
2) Occasionally-updated Background (OB) model,
3) Stochastically-updated Background (SB) model.

For the first two backgrounds, FB and OB, 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 stochastic likelihood image by probability based updating the pixel-wise values at each frame.

Updating schemes for two backgrounds

The first frame of the inputting video image is initialized as FB and OB respectively in our application, and an improved adaptive background updating method is applied by constructing two indicator sets based on the time series of pixel history.

The first indicator set, Stable Indicator (SI) set represents the number of frequencies that a pixel is in stable state between two consecutive frames. Formally, the stable indicator set is defined as

$$SI(x,y) = SI(x,y) + \delta_1(x,y), \tag{1}$$

where $\delta_1(x,y)=1$ if the absolute difference between two consecutive frames is less than a predefined threshold value and $\delta_1(x,y)=0$ for otherwise. The initial value for each pixel in SI is set to zero.

The second indicator set, Difference Indicator (DI) set indicates the number of frequencies that a pixel is significantly different from the background between two consecutive frames. This indicator set is to define a condition or conditions to be satisfied for a stationary object to become a part of background. Specially, the difference indicator set is defined as

$$DI(x,y)=DI(x,y)+\delta_2(x,y), \tag{2}$$

where $\delta_2(x,y)=\begin{cases}1 & \text{if } \delta_1(x,y)=0,\\ 0 & \text{if } \delta_1(x,y)=1.\end{cases}$

The initial value for each pixel in DI is 0. If the pixel belongs to the object plane, its value increases by 1.

By using the stated two indicator sets along with taking the existence of still object and uncovered background into account, we define the background updating schemes as follows:

If $SI(x,y)>Th_f$ and $DI(x,y)>Th_f$,

$$FB_n(x,y)=I_n(x,y),$$
$$OB_n(x,y)=FB_n(x,y).$$

If $SI(x,y)>Th_f$ and $DI(x,y)=0$,

$$FB_n(x,y)=FB_{n-1}(x,y),$$
$$OB_n(x,y)=OB_{n-1}(x,y).$$

If if $SI(x,y)=0$,

$$FB_n(x,y)=(1-\alpha)FB_{n-1}(x,y)+\alpha\, I_n(x,y),$$
$$OB_n(x,y)=(1-\beta)OB_{n-1}(x,y)+\beta\, FB_n(x,y).$$

$I_n(x, y)$ is the pixel value in current frame and α, β is the learning rate of two backgrounds. The pixel values of the frequently updated backgrounds in two consecutive frames represent $FB_n(x, y)$ and $FB_{n-1}(x, y)$. In the same way, $OB_n(x, y)$ and $OB_{n-1}(x, y)$ represent the occasionally updated backgrounds, respectively.

By using the stated two background updated rules, we estimate the corresponding foregrounds. The resultant binary foreground maps are named as Frequently-updated Foreground (FF) and Occasionally-updated Foreground (OF) correspond to FB and OB, respectively. 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. This fact makes the background model not including the pixels which do not belong to the background scene. Moreover, we could see that the intensity of each pixel of FB or OB has great connection with the corresponding foreground.

In this aspect, the combined system of two foreground marks (FF, OF) can be in one of four states:

S_1= (1,1), S_2= (1,0), S_3= (0,1), and S_4=(0,0).

(i) *Case I*: The system is in state S_1= (1,1). In this case, we can interpret this situation as a new moving object come into the scene. When a new moving object comes into the scene, due to the stability of changes in FB, the object motion does not affect significantly on the model. Thus, we have the frequent foreground mask as one, i.e. $FF(x, y)$ = 1 for pixel (x, y). On the other hand, since the OB is updated less frequently a temporary change does not affect that much on OF. Thus, we have $OF(x, y)$ = 1.

(ii) *Case II*: The system is in state S_2= (1,0). This case can be interpreted as an uncovered background. This means that the corresponding pixel is occluded over a certain period of times and then uncovered from occlusion scene. This situation makes the occasional foreground remains zero. Thus, we have $OF(x, y)$ = 0. However, during the occlusion period, the frequent foreground can update itself so that $FF(x, y)$ = 1 which lead to the system to be in state S_2= (1,0).

(iii) Case III: The system is in state S_3= (0,1). In this case, we observe that a static pixel is driven into the frequently updated background which makes the corresponding foreground $FF(x, y)$ = 0. But, if it may not be so long to mark the pixel as a scene background, then it will make the pixel as the occasional foreground mark such that $OF(x, y)$ = 1. Thus, in this situation, we note that the detected pixel as a part of the left behind object. In other words, this case will lead to a potential candidate abandoned object for further analysis. This is the case which we will thoroughly investigate to confirm whether the detected object is abandoned or removed or still person.

(iv) Case IV: The system is in state S_4=(0,0). This state shows that the pixel values are not changed in both occasional and frequent foregrounds. This means that there is no change in the scene backgrounds.

In summary, the main advantage of this two background model is that it can accurately segment object and realize in low-computational load. Moreover the current background is robust to the sudden change and adapts to the changes in the scene are blended more rapidly. In contrast, the occasional background is more stable than the frequent background. So, stationary object detection could be easily obtained by observing the difference between $FF(x, y)$ and $OF(x, y)$. These processes are illustrated in Table 1. However, as described in Case III, it is necessary to make further improvement for better and accurate detection results of an abandoned objects.

Stochastically Updated Background Model

Although the relationship between two backgrounds and their relative foregrounds has been discussed in previous, but the case $FF_n(x, y)$ = 0 and $OF_n(x, y)$ = 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 FF_n, but not prolonged enough to blend in OF_n. Thus we will construct a stochastically updated background model which gives the stochastic foreground likelihood image SF with respect to SB. In order to do so, we denote the stochastic foreground likelihood image at time n by SF_n and the event E represents the simultaneous co-occurrence of $FF_n(x, y)$ = 0 and $OF_n(x, y)$ = 1. The probability measure of E is denoted as $P(E)$ and Th_1 and Th_2 are predefined thresholds. We then define the stochastically updating rule for SF_n as follows:

$$SF_n(x,y)=\begin{cases} SF_{n-1}(x,y)+1 & \text{if } P(E)\geq Th_1, \\ SF_{n-1}(x,y)-1 & \text{if } Th_2 \leq P(E) < Th_1, \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

According to updating rule in (3), we obtain the likelihood foreground image which is able to remove noise while detecting the objects. It can also keep the time requirement to assign a static pixel as an abandoned item to be minimized. The likelihood image also collects the evidence information of an object being to be abandoned so that the classification process gives accurate results. We then analyze the collected evidence scores by comparing with a preset level. This preset level is defined based on the number of frames and the characteristics of noise structures in the system. It is to be noted that the more noisy results in the foreground detection process the higher preset level values. In such case, it will take longer duration for a pixel to be classified as an abandoned object pixel. Construction of three backgrounds is illustrated in Fig. 7.

Table I. Analysis of potential events.

Image frames	Occasional foreground (OF)	Frequent foreground (FF)	Potential event
	Object present	Object present	Moving object
	Object absent	Object present	Uncovered background
	Object present	Object absent	Candidate abandoned object
	Object absent	Object absent	Scene background

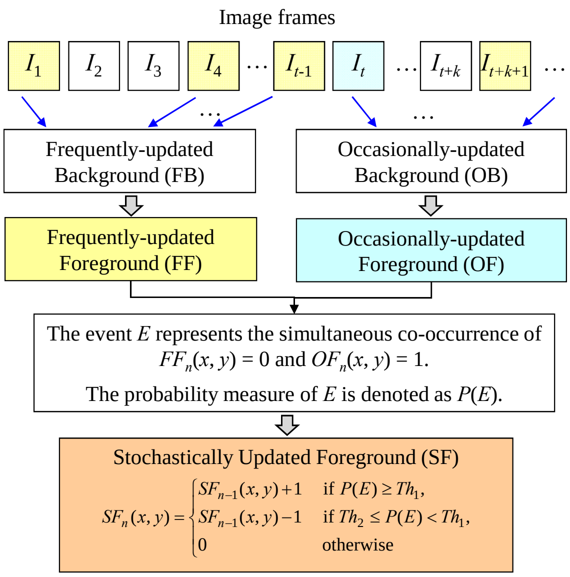


Fig. 7. Frequent, occasional and stochastic background models.

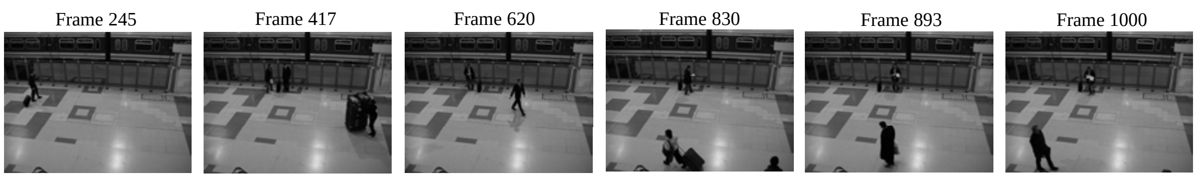
It is worthwhile to note that the classification of likelihood image is solely dependent one and only one parameter. We also observe that the backgrounds and the combination of backgrounds are independent of stochastic image predefined parameter values. Consequently, it is not necessary to make any particular constraints for initializing the background modeling process. This property makes our method more robust and efficient detection even for the video sequences taken by using ordinary consumer cameras in complex environments. The power of our method is tested by using the video sequences taken in public transportation areas such as airports and train stations in real time. Since we can set a suitable proportional value of evidence score and likelihood parameter through normal observations, the robustness of detection process is significantly and well confirmed in our experiments.

A sample of stationary object detection is illustrated in Fig. 8. Some examples of image frames in our input video sequence are shown in Fig.8(a). The effectiveness of using stochastically updated foreground is confirmed through the experimental results as shown in Fig.8(b). Even though the static human regions are included in the detection result of using FF and OF, the results of SF gives only the accurate region of the stationary object.

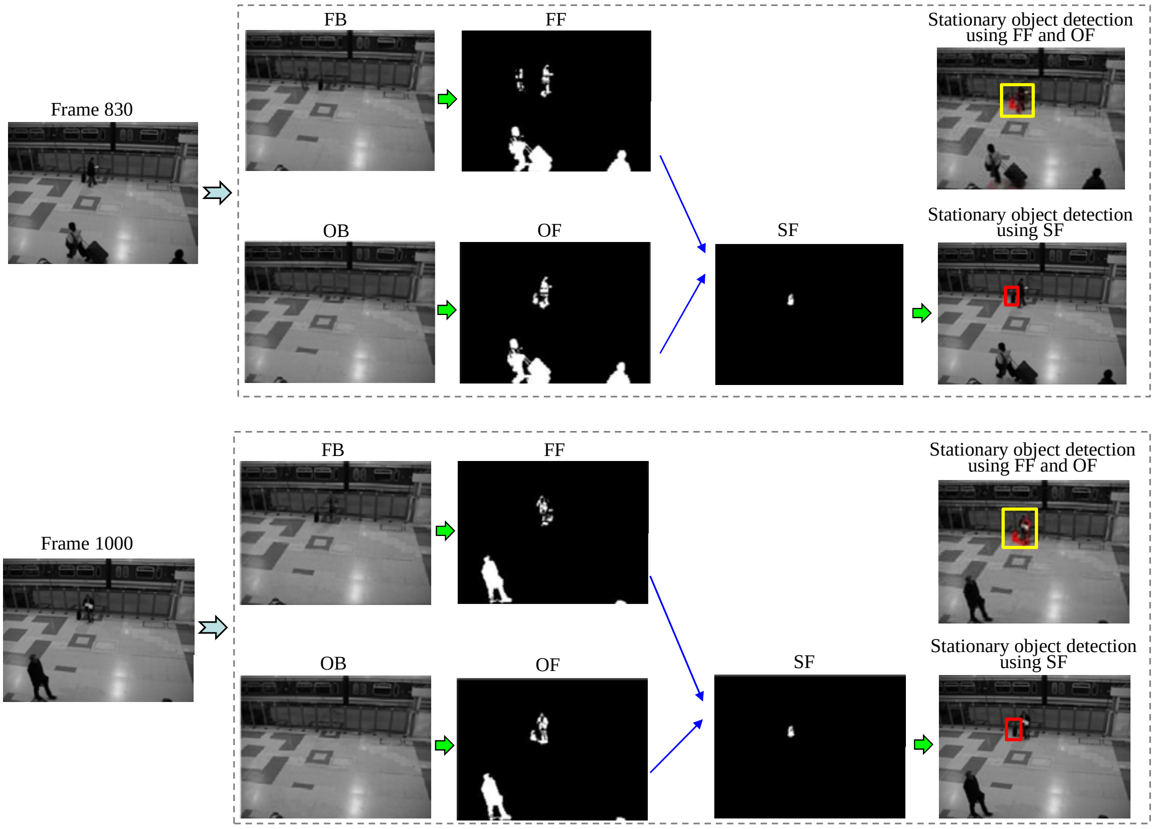
Shadow removing

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 using similar method described in [28].

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



(a) some frames of input video sequence



(b) updating background models and the stationary object detection (at frame 830 and frame 1000)

Fig. 8. The result of stationary object detection.

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.

To do it, we look for moving points whose intensity ratios are similar; differently, moving points belonging to true foreground regions will have different ratios.

In addition, these values will be lower than 1, because of the minor light that illuminates the shadow regions. Formally, we evaluate, for each candidate point (x, y) the ratio as $R = \frac{I_n(x,y)}{B_n(x,y)}$ 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 process provides an image with the real shape of the detected objects, without noise or shadows.

B. Abandoned Object Detection Process

In video surveillance one of the most important applications is to distinguish the abandoned or removed object from still person. In order to do so, the extracted moving objects are to be classified into one of five types; Temporary Static object (TS), Moving Person (MP), Still Person (SP), Abandoned Object (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:

- Abandoned object does not move by itself,
- Abandoned object has an owner and
- the size of the AO 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 MP 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 Th_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 abandoned object. Fig. 9 shows the five types of objects and their thresholds. Let $V(X)$ be the velocity of blob X.

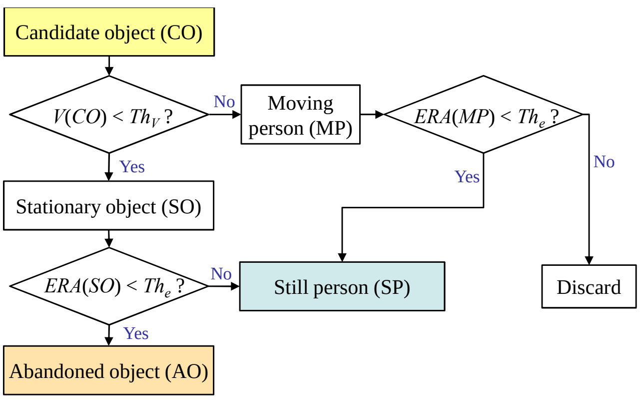


Fig. 9. Object classification of extracted foreground.

IV. Experimental Results

In order to confirm the efficiency of our proposed method, we conducted some experiments by using our own video sequences taken in international airports and in the university campus. In performing experiments we have used only normal video camera and have not imposed any restrictions on the initial background scenes. For the international airport scenario, we have taken ten video sequences in crowded environment including check-in counter scene and arrival lobby. In addition, the scenes of normal people sitting, standing and walking at various patterns are also included. Moreover, some people are sitting in very still position. These scenarios can be found in our daily life of real world environments. This type of realistic conditions has not been taken into account in other existing methods. However, our proposed method can handle such a real world environment successfully and robustly.

We have also considered partial occlusion and sometimes completely occluded for a certain period of times. Moreover, we have used various shapes and types of abandoned objects and various poses of still people. The experiment videos are taken at different venues in various illumination conditions. Furthermore, we have also tested our method by using PETS2006 datasets which are taken at railway station. Altogether, our experimental results are obtained by using 25 video sequences. The images used here have 320×240 pixels (QVGA) resolution.

Fig. 10 shows the detection results for our own video sequences in outdoor scene. In this figure, the original image frames are shown in the first column. Their related frequently-updated background and foreground, FB and FF, are described in the second and third columns. Similarly, OB and OF are shown in the fourth and fifth column, respectively. The stochastically updated foreground SF and the detected stationary object regions (the red rectangle) are shown in the last columns, respectively.

We also present some detection results by using our own video sequence taken in indoor scene. The video sequences are taken at a place near check-in counter in international airport. Some sample image frames are shown in Fig.11(a). In these images, it can be seen that there are significant reflection in the floor areas. This causes the detection problem more complex. In such cases, the use of only two backgrounds cannot handle to achieve an accurate result as seen in Fig.11(d). By using our proposed stochastic background model, we can overcome such kind of problems and can detect the abandoned object accurately without shadow and noise as shown in Fig.11(e). The detected result on the image is described in Fig.11(f). Some more examples of the experimental results are shown in Fig. 12. In this figure, the first two rows show the experimental results of our own video sequences and the last row shows the results obtained by using PETS2006 datasets video sequence.

Now we shall present some comparison results of our proposed method with some conventional background modeling methods, namely single background model and dual background model. The experimental findings are described as follows:

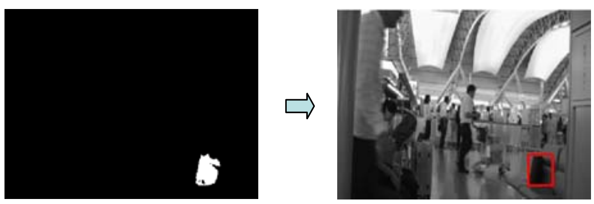
(i) The single background model and dual background models cannot handle the drastic background changes while our proposed multiple backgrounds model with the support of stochastically updated background can detect the abandoned objects accurately.
(ii) The single background model is sensitive to the short term illumination changes so that it cannot handle situation of reflected ground surface, the wall and so on.
(iii) Even though the dual background model can cope with such short term illumination changes, it detects object regions as well as unnecessary surrounding pixels.

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 abandoned object 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.

Fig. 10. Detail procedure of abandoned object detection (our own video sequence in outdoor environment).

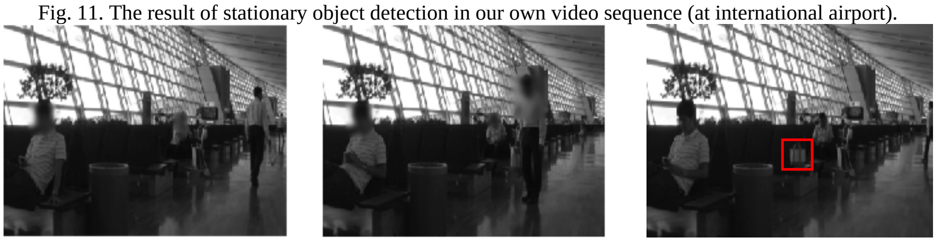
(a) some frames of input video sequence

(b) Frequently-updated Background (FB) (c) Occasionally-updated Background (OB) (d) the difference region between FF and OF



(e) Stochastic Foreground (SF) (f) the detected stationary object

Fig. 11. The result of stationary object detection in our own video sequence (at international airport).



(a-i) brings a bag (a-ii) leaves the bag (a-iii) abandoned object

(b-i) carrying a bag (b-ii) leaves the bag (b-iii) left-bag

(c-i) bag under shadowed area (c-ii) multiple people walking (c-iii) abandoned object

Fig. 12. Examples of detected images: (a, b) sequences of our own datasets and (c) sequence of PETS 2006 datasets.

V. Conclusions

We have presented a computationally efficient and robust method to detect abandoned object in public areas. This method uses three backgrounds that are learned by processing the input video at different frame rates. After the detection of foreground regions, a shadow re-moving algorithm has been implemented in order to clean the real shape of the detected objects. The proposed object detection method works surprisingly well in crowded environments and can handle with illustration changes. It can also detect the very small abandoned objects contained in low quality videos. Due to its simplicity the computational effort is kept low and no training steps are required. Finally, we can discriminate effectively between abandoned or still person by using a simple rule-based algorithm. The reliability of the proposed framework is shown by the experimental tests performed in big public transportation areas.

Acknowledgment

The authors thank to the reviewers for their valuable comments to improve the paper. We also thank Scope project members and the students of Physical Electronics and Informatics, for their participation in producing tested videos.

References

[1] M. Valera, S.A Velastin, "Intelligent distributed surveillance systems: a review," *IEE Proceedings on Vision, Image and Signal Processing*, 152(2), pp. 192–204, 2005.

[2] Y. Tian, R.S. Feris, H. Liu, A. Hampapur, and M. Sun, "Robust Detection of Abandoned and Removed Objects in Complex Surveillance Videos," *IEEE Trans. on Systems, Man, and Cybernetics, Part C*, vol. 41, no. 5, pp. 565–576, Sep. 2011.

[3] M. S. Nagmode, A. J. Madhuri, M. S. Ashok, "A Novel approach to Detect and Track Moving Object using Partitioning and Normalized Cross Correlation," *International Journal ICGST-GVIP*, ISSN: 1687-398X, vol. 9, no. 4, pp. 49–56, Aug. 2009.

[4] S. Ferrando, G. Gera, C. Regazzoni, "Classification of Unattended and Stolen Objects in Video Surveillance System," *Proc. of IEEE Intl. Conf. on Advanced Video and Signal-Based Surveillance* (AVSS'06), Sydney, Australia, Nov. 2006.

[5] N. Krahnstoever, P. Tu, T. Sebastian, A. Perera, and R. Collins, "Mutli-view detection and tracking of travelers and luggage in mass transit environments," *in Proc. of 9th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance* & CVPR, New York, USA, 2006, pp. 67–74.

[6] S. Cheng, X. Luo, and S. M.Bhandarkar, "A multiscale parametric background model for stationary foreground object detection," *in IEEE Workshop on Motion and Video Computing*, 2007.

[7] H. Kong, J-Y. Audibert and J. Ponce, "Detecting Abandoned Objects with a Moving Camera," *IEEE Trans. on Image Processing*, vol. 19, no. 8, pp.2201–2210, 2010.

[8] S. Kwak, G. Bae, H. Byun: "Abandoned luggage detection using a finite state automaton in surveillance video," *Optical Engineering*, vol. 49, no. 2, pp. 027007–027007-10, Feb. 2010.

[9] M. Bhargava, C-C. Chen, M. S. Ryoo, J. K. Aggarwal, "Detection of object abandonment using temporal logic," *Machine Vision and Applications*, vol. 20, no. 5, pp.271–281, 2009.

[10] X. Li, C. Zhang, D. Zhang: "Abandoned Objects Detection Using Double Illumination Invariant Foreground Masks," *Proc. of 20th Intl. Conf. on Pattern Recognition* (ICPR'10), Istanbul, Turkey, 2010, pp. 436–439.

[11] M. Beynon, D. Hook, M. Seibert, A. Peacock, and D. Dudgeon, "Detecting Abandoned Packages in a Multi-camera Video Surveillance System," *Proc. of IEEE Intl. Conf. on Advanced Video and Signal-Based Surveillance* (AVSS'03), Miami, FL, USA, 2003, pp.221–228.

[12] P.L. Venetianer, Z. Zhang, W. Yin, AJ. Liptop, "Stationary Target Detection Using the Object Video Surveillance System," *Proc. of IEEE Intl. Conf. on Advanced Video and Signal-Based Surveillance (AVSS'07)*, London, UK, Sep. 2007, pp. 242–247.

[13] F. Porikli, Y. Ivanov, and T. Haga, "Robust abandoned object detection using dual foregrounds," *in EURASIP Journal on Advances in Signal Processing*, vol. 2008, no. 1, 11 pages, 2008.

[14] R. Miezianko and D. Pokrajac, "Detecting and recognizing abandoned objects in crowded environments," *in Proc. of Computer Vision System*, 2008, pp. 241–250.

[15] P. Spagnolo et al, "An Abandoned/Removed Objects Detection Algorithm and Its Evaluation on PETS Datasets," *Proc. of IEEE Intl. Conf. on Advanced Video and Signal-Based Surveillance* (AVSS'06), Sydney, Australia, Nov. 2006, pp. 17–21.

[16] F. Porikli, "Detection of temporarily static regions by processing video at different frame rates," *Proc. of IEEE Intl. Conf. on Advanced Video and Signal-Based Surveillance (AVSS'07)*, London, UK, Sep. 2007, pp. 236–241.

[17] J. Zhang, L. Zhang, and H.M. Tai, "Efficient video object segmentation using adaptive background registration and edge based change detection techniques," *IEEE International Conference on Multimedia and Expo* (ICME 04), Jun. 2004, Vol. 2, pp. 1467–1470.

[18] Thi Thi Zin, Pyke Tin, H. Hama and T. Toriu, "Unattended Object Intelligent Analyzer for Consumer Video Surveillance", *IEEE Trans. on Consumer Electronics*, Vol. 57, No. 2, pp. 549-557, May. 2011.

[19] Thi Thi Zin, H. Hama, Pyke Tin and T. Toriu, "Evidence Fusion Method for Abandoned Object Detection", *ICIC Express Letters (Part B: Applications): An International Journal of Research and Surveys,* Vol. 2, No. 3, pp.535–540, Jun. 2011.

[20] Thi Thi Zin, Pyke Tin, T. Toriu and H. Hama, "A Probability-based Model for Detecting Abandoned Objects in Video Surveillance Systems," *Lecture Notes in Engineering and Computer Science: Proc. of The World Congress on Engineering 2012* (WCE 2012), London, U.K, 4-6 Jul. 2012, vol. 2, pp. 1246-1251.
[21] A. Bayona, J. C. SanMiguel, J. M. Martínez, "Comparative evaluation of stationary foreground object detection algorithms based on background subtraction techniques," *in Proc. of the IEEE Int. Conf. on Advanced Video and Signal based Surveillance* (AVSS'2009), Geneva, Italy, Sep. 2009. pp. 25–30.
[22] I. Haritaoglu, D. Harwood and L.S. Davis, "W4: Real-time surveillance of people and their activities," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 22, no. 8, pp. 809–830, Aug. 2000.
[23] N. M. Oliver, B. Rosario, and A. P. Pentland, "A Bayesian computer vision system for modeling human interactions," *IEEE Trans. Patt. Anal.Mach. Intell.*, vol. 22, no. 8, pp. 831–843, Aug. 2000.
[24] J. Martínez, J. Herrero, C. Orrite, "Automatic Left luggage Detection and Tracking using a Multi-camera UKF," *in Proc. of 9th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, New York, USA, 2006, pp. 59–66.
[25] C. Stauffer, W.E.L. Grimson, "Adaptive background mixture models for real-time tracking," *Proc. of Intl. Conf. on Computer Vision and Pattern Recognition* (CVPR'99), Ft. Collins, CO, USA, Jun. 1999, vol. 2, pp. 2246–2252.
[26] San Miguel J.C.; Martinez, J.M. "Robust unattended and stolen object detection by fusing simple algorithms," *Proc. of 5th IEEE Intl. Conf. on Advanced Video and Signal based Surveillance* (AVSS2008), Santa Fe, USA, Sep, 2008, pp. 18–25.
[27] Y. Tian, "Robust and efficient foreground analysis for real time video surveillance," *Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition* (CVPR 2005), San Diego, USA, Jun. 2005, pp. 1182–1187.
[28] T. Martiriggiano, A. Caroppo, M. Leo, P. Spagnolo, and T. D'Orazio, "An Innovative Approach for Abandoned or Removed Objects Detection", *Proc. of Intl. Symposium on Communications, Control, and Signal Processing* (ICSSCP 2006), Marrakech, Morocco, Mar. 2006.