A Probability-based Model for Detecting Abandoned Objects in Video Surveillance Systems

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

***Abstract*—Detection of suspicious packages or abandoned objects is one of the most important tasks in video surveillance systems. Some recent terrorist attacks involving explosive packages left behind in many contexts such as airports, rail stations and etc. illustrate the importance of this problem. In this paper, we propose a probability-based model for robustly and efficiently detecting abandoned objects in complex environments. Specifically, we develop a new probability-based background subtraction algorithm based on combination of multiple background models for motion detection. In addition, several improvements are implemented to the background subtraction method for shadow removal and quick lighting change adaptation. We then analyze the extracted objects to classify as static or dynamic objects. After the analysis, we employ the statistical running average of the static foreground masks for event type decision making either abandoned or very still person. Finally, the robustness and efficiency of the method are tested on our video sequences and PETS2006 datasets.**

***Index Terms*—probability-based model, abandoned object, video surveillance, still person**

I. Introduction

In recent years, as a sharp increase in terrorist attacks on crowded public places, like airports, stations, subways, entrance to buildings, and other public venues, video surveillance systems have been more demanded. In this aspect, immediate and reliable detection of suspicious packages or abandoned objects is vital to the safety of innocent citizens since explosive devices are usually left unattended. Abandoned object detection is the task of locating objects that are left 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. For solving these problems, there are mainly two types of methods in the literature. The first type of methods is based on tracking [1-4]. The second type of methods use background model to detect suspected region [5-11]. 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. And it is difficult to track all the objects precisely in crowded situations. On the contrary, methods based on background subtraction, using an appropriate threshold procedure on the difference between each image of the sequence and a model image of the background can provide the best compromise between performance and reliability.

In last few years, many background subtraction methods were proposed in the literature. Most of them used different types of background models and updating schemes. For example, [4, 5] used a mixture of Gaussian distribution. Frame-to-frame analysis was employed in [12, 13] and a sub-sample analysis was applied in [14-18].

In many surveillance scenarios, the objects which are contained in the initial background are removed from the scene. The problem of correctly classifying a foreground object to be abandoned or still person is a major concern in background modeling. But most of existing systems ignore this important problem. In this paper, we propose a new method that use probability-based multiple backgrounds. Our method does not require object initialization, tracking, or offline training. It accurately segments objects even if they are fully occluded. 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 logic-based system is introduced to classify detected objects as either an abandoned object or a still person.

The rest of the paper is organized as follows: In section II, we describe an overview of the proposed system. In section III, we develop a probability-based model along with multiple backgrounds for stationary region detection. The classification procedure of detected object types is presented in section IV. Section V covers some experimental results on standard datasets as well as our real-world surveillance scenarios. Finally, concluding remarks and discussions are presented in section VI.

II. System Overview

Fig. 1 shows our proposed system architecture. In this system, we first establish probability-based multiple backgrounds and update them by using statistical analysis.

Manuscript received March 23, 2012; revised April 15, 2012. This work was supported in part by the SCOPE: Strategic Information and Communications R&D Promotion Program (102307010).

Thi Thi Zin is with the Graduate School of Engineering, Osaka City University, 3-3-138 Sugimoto, Sumiyoshi-ku, Osaka 558-8585, JAPAN, (corresponding Author, Tel: & 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 a visiting Professor with the Graduate School of Engineering, Osaka City University, JAPAN.

T. Toriu is with the Graduate School of Engineering, Osaka City University, JAPAN.

H. Hama was with the Graduate School of Engineering, Osaka City University, JAPAN. He is now in the Research Center for Industry Innovation, Osaka City University, JAPAN.

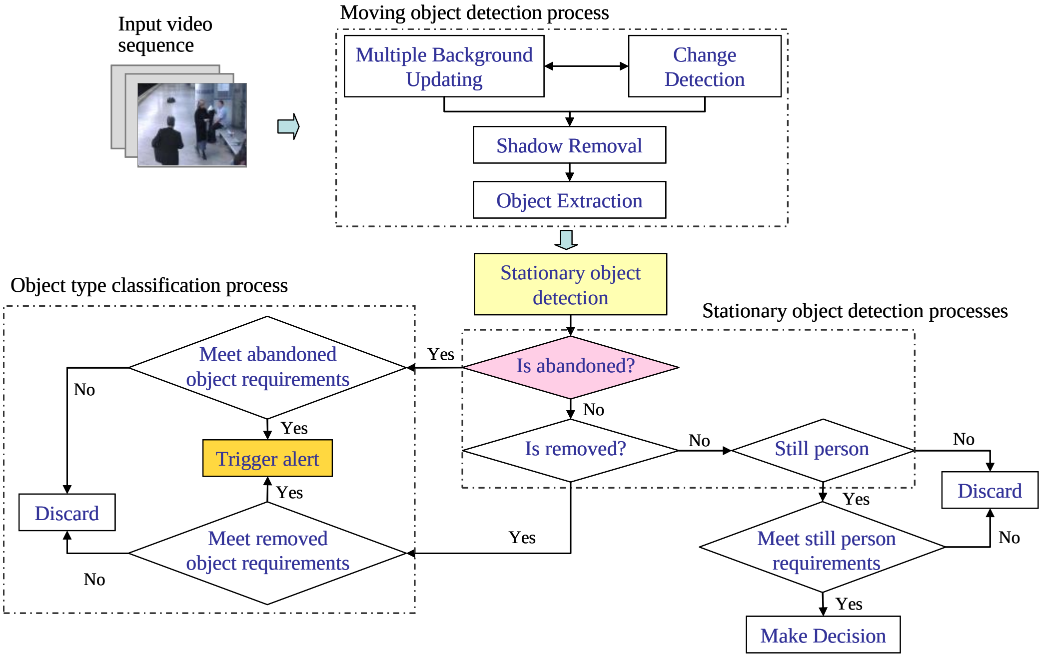


Fig. 1. Overview of proposed method.

We assume that the recently changed pixels that stay static after they changed can be distinguished from the actual background pixels by analyzing the intensity variance in different temporal scales. Specifically, we develop and employ three background models and update them by introducing the stable history maps and difference history maps. We then employ a shadow removing method after motion detection. In the shadow removal process, we integrate both intensity and texture information to handle quick lighting changes.

We then analyze the foregrounds by using statistical models to discriminate the extracted objects as moving objects and stationary objects. We have used different thresholds in obtaining moving objects and stationary objects. For stationary objects, we further develop a segmentation method to detect abandoned or still person. It significantly outperforms previous techniques. In addition, the logical rule based classifier is used to distinguish the abandoned objects and the still-standing persons, which is a problem that is not solved in previous approaches.

III. A Probability-based 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. Specifically we develop a new probability-based background subtraction method using three backgrounds. They are named as:

(i) Frequently-updated Background model (FB)
(ii) Occasionally-updated Background model (OB) and
(iii) Probability-based Background model (PB).

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 scheme at each frame.

A. 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 maps of pixel history. The first map, stable history map (SM), represents the number of times a pixel is stable in consecutive frames and defined as

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

where I_n is the n^{th} frame in the video sequence, Th_p and Th_f are the pre-defined thresholds. The initial value for each pixel in SM is set to zero. 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 still object becoming a part of background.

$$DM(x,y)=\begin{cases} DM(x,y)+1 & \text{if } |I_n(x,y)-I_{n-1}(x,y)|>Th_p, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

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:

ISBN: 978-988-19252-1-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$$FB_n(x,y)=\begin{cases} I_n(x,y) & \text{if } SM(x,y)>Th_f \text{ and } DM(x,y)>Th_f, \\ FB_{n-1}(x,y) & \text{if } SM(x,y)>Th_f \text{ and } DM(x,y)=0, \\ (1-\alpha)FB_{n-1}(x,y)+\alpha I_n(x,y) & \text{if } SM(x,y)=0. \end{cases} \tag{3}$$

$FB_n(x, y)$ and $FB_{n-1}(x, y)$ represent the frequently updated backgrounds pixel value at position (x, y) in current and previous frames. In the same way, $OB_n(x, y)$ and $OB_{n-1}(x, y)$ represent the occasionally updated background at position (x, y) and the corresponding updating rules are

$$OB_n(x,y)=\begin{cases} FB_n(x,y) & \text{if } SM(x,y)>Th_f \text{ and } DM(x,y)>Th_f, \\ OB_{n-1}(x,y) & \text{if } SM(x,y)>Th_f \text{ and } DM(x,y)=0, \\ (1-\beta)OB_{n-1}(x,y)+\beta FB_n(x,y) & \text{if } SM(x,y)=0, \end{cases} \tag{4}$$

where $I_n(x, y)$ is the pixel value at position(x, y) in current frame and α, β is the learning rate of two backgrounds.

At every frame, we estimate the frequent foreground (FF) and occasional foreground (OF) by comparing the current frame I by the background models FB and OB. The OF 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 frequent foreground FF 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, a pixel which is logically not belonging to the background scene cannot pollute the background model. Moreover, the following four cases are analyzed and illustrated in Fig. 2.

Case I: A new moving object come into the scene, and it does not belong to any backgrounds. This states that object is presented in both foregrounds.

Case II: A part of the detected object which is changed then sets back to its original value. The object appears in the occasional foreground but disappears in the frequent foreground.

Case III: A scene background pixel that was occluded before. The object is absent in the occasional foreground but appeared in the frequent foreground.

Case IV: A pixel equal to both background pixel, this means that there is no change in the scene. This case states that object is absent in both foregrounds.

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

Fig. 2. Analysis of potential events.

B. Probability-based Background Model

Under *Case III* 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 probability-based background model which gives the probability-based foreground likelihood image PF with respect to PB. We denote the likelihood image at time n by PF_n and the event E represents the simultaneous co-occurrence of *Condition III*. 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 PF_n as follows:

$$PF_n(x,y)=\begin{cases} PF_{n-1}(x,y)+1 & \text{if } P(E)\ge Th_1, \\ PF_{n-1}(x,y)-1 & \text{if } Th_2 \le P(E)<Th_1, \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

The likelihood image enables removing noise in the detection process. It also controls the minimum time required to assign a static pixel as an abandoned item. For each pixel, the likelihood image collects the evidence of being an abandoned item. In our model, we only use a single parameter for the likelihood image. Neither of the backgrounds and their mixture models depends on the likelihood stochastic image preset 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.

Fig. 3(a) shows some examples of image frames in our input video sequence. Frequently-updated background (FB) and corresponding occasionally-updated background (OB) are displayed in Fig. 3(b) and Fig. 3(c), respectively. Based on these two backgrounds, we construct the probability-based foreground (PF) for detection of stationary object in Fig. 3(d). By using PF, we got the candidate abandoned object region as shown in Fig. 3(e).

C. Shadow Removing

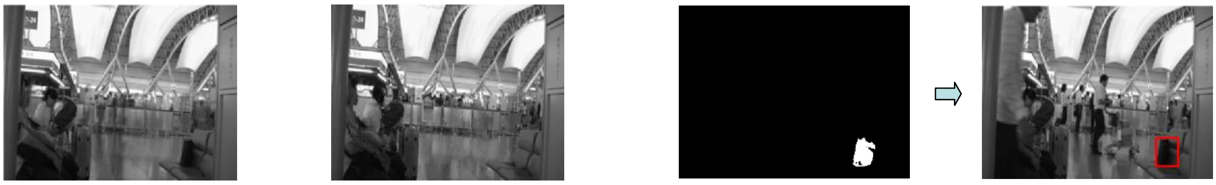
In the shadow removing process, we assume that a shadow is uniformly decreased according to the illumination source [19]. Thus, shadows can move with their own objects. They do not have a fixed texture although the real objects have. To do it, we look for moving points whose intensity ratios are similar; differently, moving points belonging to true foreground regions will have different ratios. 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 phase provides an image with the real shape of the detected objects, without noise or shadows.

ISBN: 978-988-19252-1-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

(a) some frames of input video sequence



(b) Frequently-updated background (c) Occasionally-updated background (d) probability-based foreground (e) the detected stationary object

Fig. 3. The result of stationary object detection.

IV. Abandoned Object Detection Process

First, we subdivide extracted objects into one of five types, Stationary Object (SO), Moving Person (MP), Still Person (SP), Abandoned Object (AO), and Candidate Object (CO). We employ a simple rule-based classifier for the real-time process. If objects were detected, they were initially classified as CO. Then, using the velocity of the moving object, the CO was classified as MP or AO. That is to say, if CO moved at a velocity higher than that of the threshold value, Th_v for several consecutive frames, it was identified as a MP. If CO's velocity was below the threshold velocity Th_v, it was classified as SO. If CO is identified as SO, AO and SP were distinguished by using the Exponent Running Average (ERA). In this case, ERA is calculated by using definition described in Eq. (6) for a given dataset $x_1, \dots x_n$ (say).

Definition: $e^{\alpha} = 2/(n+1)$ where n is number of data or number of images. We define

$$ERA(t) = \alpha\, x_t + (1-\alpha)ERA(t-1) \text{ for } t > 2, \tag{6}$$

where $ERA(1)$ is undefined and $ERA(2)$ can be initialized in many ways. For example, $ERA(2)$ will be x_1 or average of first 4 or 5 observations or simple moving average of a few terms. Once $ERA(2)$ is initialized, then $ERA(t)$ can be evaluated iteratively by using Eq. (6). If ERA is greater than a predefined threshold value Th_e, the SO is classified as SP and otherwise it will be AO. We also denote the velocity of type X by $V(X)$. Fig. 4 shows the five types of objects and their thresholds.

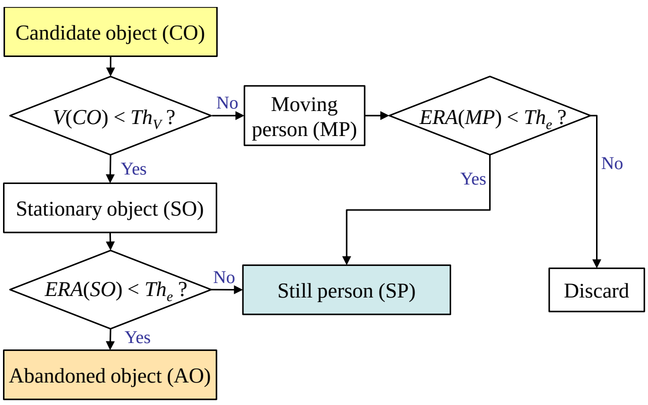


Fig. 4. Object classification of extracted foreground.

V. Experimental Results

In this section, we present the experimental results to evaluate the effectiveness of the proposed method. This is tested our own video sequences taken with normal video cameras in International Airports. 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 successfully these cases. Here, in the experimental performances, some of our group members had played as the performers. We also have considered partial occlusion and sometimes completely occlude in a specified moment. All videos have instances of various shapes of abandoned objects and still people. These are taken from different venues. Each video sequence tests a different viewpoint. Moreover, we have already confirmed the performances of our method using PETS2006 datasets which are taken at railway station. So, we tested totally 20 video sequences with various public transportation areas in real time environments. The images used here 320×240 pixels (QVGA) resolution.

Fig. 5 shows the detail procedure of abandoned object detection results. The first row shows the input frames. In the second row and third row, the corresponding backgrounds (FB, OB) and their extracted foregrounds (FF, OF), respectively are described. By using these foregrounds, the respective probability-based foregrounds are shown in the fourth row. The final row describes the detected abandoned object regions shown in the red rectangular box. As visible, after the candidate object was detected as an abandoned object, temporary occlusions due to the moving people do not cause the system to fail. When we observe the input frames number 570 and 1079, these are big illumination changes. These changes have made severe effects on the two background models of FB and OB. But our proposed probability-based model has been overcome these effects as shown in the fourth row. More detection results for our own video sequences and PETS 2006 datasets are shown in Fig. 6.

ISBN: 978-988-19252-1-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

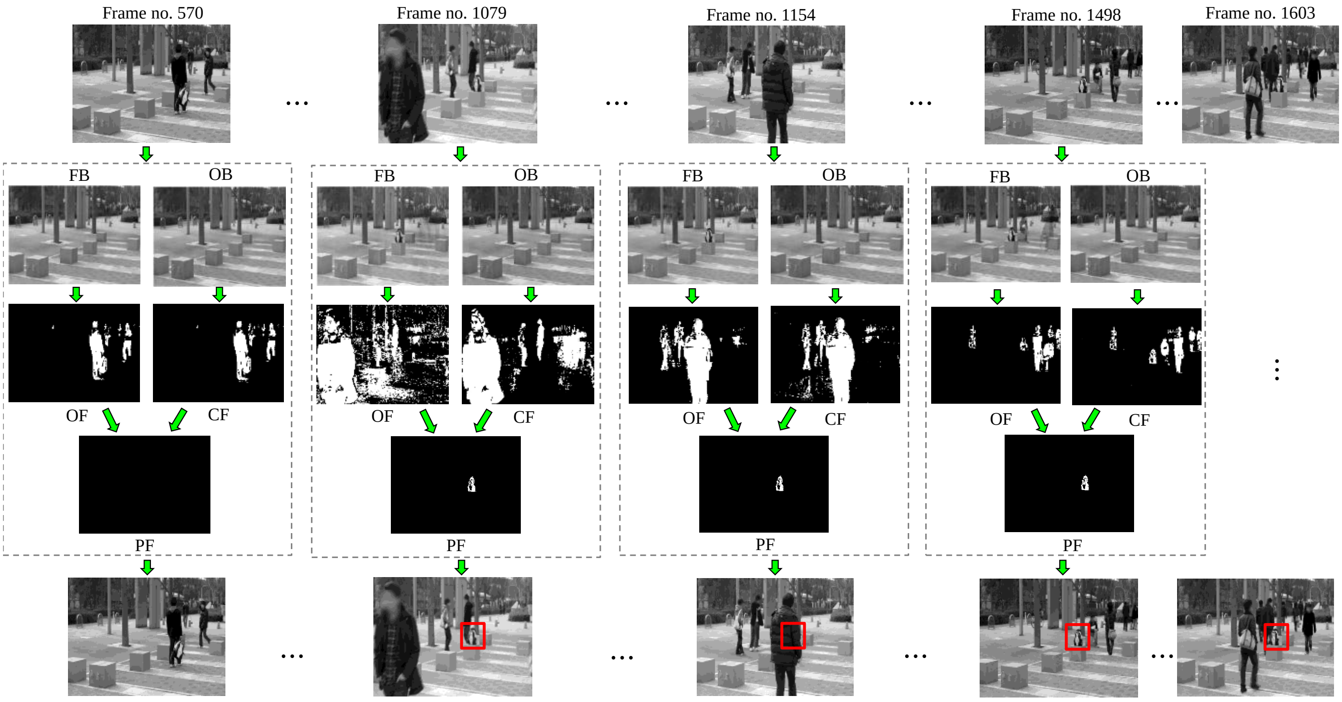


Fig. 5. Detail procedure of abandoned object detection (our own video sequence in outdoor environment).

(a) video sequence of our own dataset (in International Airport)

(b) video sequence of PETS2006 dataset

Fig. 6. The results of abandoned object detection.

We also have 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 probability-based 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 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. Comparison results are summarized in Table I.

ISBN: 978-988-19252-1-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Table I Comparison results

Background model	Comparison			
	Background changes	Illumination changes	Initialization	Object detection
Single	cannot handle	sensitive	require	not exact
Double	cannot handle	no sensitive	require	not exact
Proposed	can handle	no sensitive	not require	exact

VI. 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 rule based object detection 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.

REFERENCES

[1] M. Valera, S.A Velastin, "Intelligent distributed surveillance systems: a review", *IEE Proceedings - Vision, Image and Signal Processing*, 152(2), pp. 192-204, 2005.
[2] M. S. Nagmode, A. J. Madhuri, M. S. Ashok, "A Novel approach to Detect and Track Moving Object using Partitioning and Normalized Cross Correlation", *ICGST-GVIP Journal*, 9(4), Aug. 2009.
[3] 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)*, pp.21, Nov. 2006.
[4] 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 9th PETS, CVPR*, pp. 67-74, 2006.
[5] 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.
[6] Sooyeong Kwak, Guntae Bae, Hyeran Byun: "Abandoned luggage detection using a finite state automaton in surveillance video", *Optical Engineering*, (49)2, 027007, Feb. 2010.
[7] M. Bhargava, C-C. Chen, M. S. Ryoo, J. K. Aggarwal, "Detection of object abandonment using temporal logic", *Machine Vision and Applications*, 20, pp.271–281, 2009.
[8] X. Li, C. Zhang, D. Zhang: "Abandoned Objects Detection Using Double Illumination Invariant Foreground Masks", *International Conference on Pattern Recognition*, 2010.
[9] A. Bayona, J. C. SanMiguel, J. M. Martínez, "Comparative evaluation of stationary foreground object detection algorithms based on background subtraction techniques", *in Proceedings of the 2009 IEEE Int. Conf. on Advanced Video and Signal based Surveillance, AVSS'2009*, Geneva, Italy, pp. 25-30, Sep. 2009.
[10] 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, pp. 242-247, Sep. 2007.
[11] 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.
[12] F. Porikli, Y. Ivanov, and T. Haga, "Robust abandoned object detection using dual foregrounds," *in EURASIP Journal on Advances in Signal Processing*, 2008.
[13] R. Miezianko and D. Pokrajac, "Detecting and recognizing abandoned objects in crowded environments," *in Proc. of Computer Vision System*, pp. 241–250, 2008.
[14] P. Spagnolo et al, "An Abandoned/Removed Objects Detection Algorithm and Its Evaluation on PETS Datasets", *Proc. of IEEE Conf. on Video and Signal Based Surveillance*, pp. 17-21, 2006.
[15] 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*, pp. 236-241, 2007.
[16] F. Porikli, Y Ivanov and T. Haga, "Robust Abandoned Object Detection Using Dual Foregrounds", *EURASIP Journal on Advances in Signal Processing*, Vol. 2008, Article ID 197875, 11 pages, 2008, doi:10.1155/2008/197875.
[17] Thi Thi Zin, H. Hama, Pyke Tin and T. Toriu, "Evidence Fusion Method for Abandoned Object Detection", *In the 5th Intl. Conf. on Innovative Computing, Information and Control (ICICIC-2010)*, Xi'an, China, Dec. 2010.
[18] Thi Thi Zin, Pyke Tin, H. Hama and T. Toriu, "Unattended Object Intelligent Analyzer for Consumer Video Surveillance", *IEEE Transactions on Consumer Electronics*, Vol. 57, No. 2, pp. 549-557, May. 2011.
[19] T. Martiriggiano, A. Caroppo, M. Leo, P. Spagnolo, and T. D'Orazio, "An Innovative Approach for Abandoned or Removed Objects Detection", *Proc. of ISCCSP2006*, Marrakech, Morocco, Mar. 2006.

ISBN: 978-988-19252-1-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)