Two-Stage Unattended Object Detection Method with Proposals

Nam Trung Pham, Karianto Leman, Jie Zhang, Isaac Pek
Institute for Infocomm Research
Singapore
e-mail: {ntpham, karianto, zhangj, ipek}@i2r.a-star.edu.sg

***Abstract*—Unattended object detection is a crucial task in visual surveillance systems. However, it is challenging in handling false alarms and miss detection rate. In this paper, a two-stage method for the unattended object detection is proposed where the first stage tries to detect all possible unattended objects and prevent miss detections by considering attributes of objects such as staticness, foregroundness, and abandonment. This stage is called the unattended object proposal stage. In the second stage, our method reduces false alarms with candidates obtaining from the first stage by using a deep learning similarity matching between candidates and the background model. With the capability of reducing false alarms and miss detections, our method can be applied in large-scale deployment systems for unattended object detection.**

Keywords-unattended object; detection; tracking; matching

I. INTRODUCTION

Recently, there is a large demand for video analytics in visual surveillance due to a huge amount of video data from cameras needing to be processed. Although humans are good at detecting events in videos, it is impossible to hire a large number of individuals to analyze the big video data. Hence, it is necessary to have video analytic systems. One of the crucial tasks in video analytic systems is to detect unattended objects for security purposes, especially in public areas such as airports, shopping malls, and railway stations. When an object, normally a suitcase, a bag or a box, is unattended in public areas, it constitutes a security threat to people in the scene because the unattended object may contain dangerous items like explosives from terrorists. However, detecting unattended objects in unconstrained environments is not an easy task. This is because of challenges in classifying which objects are unattended amongst a large number of objects in the scene. Two important tasks for an unattended object detection system are to handle miss detections and reduce false alarms in the detection process.

The purpose of the unattended object detection is to detect static foreground objects which do not move much for a long time and no one in the scene cares about them. Some techniques applied tracking methods to find unattended objects. Objects are detected and tracked from the single background model [1, 2, 3, 4, 5, 6]. There are two challenges for these methods in the crowded scenes. The first challenge is maintaining the background model for static objects. When an object is static in the scene, it is easy to be absorbed into the background model with the fast model updating rate. On the other hand, there are many false alarms in detections if using the background model with the slow model updating rate. The second challenge is to handle heavy occlusions when tracking objects under crowded situations.

To be more efficient in detecting static foreground objects, some other methods utilized two types of background models: short-term and long-term background models [7, 8]. The short-term background model has an important role in reducing false alarms from the background and the long-term background model tries to reduce the miss rate of unattended object detections. However, in these methods, when unattended objects are occluded by persons for a period of time, these objects may not be detected. This is because blobs of these objects cannot be obtained by using both short-term and long-term background models.

Some other techniques applied classification methods to detect unattended objects [9, 10]. In [9], authors defined some attributes for an unattended object such as staticness, foregroundness, and abandonment. Staticness is defined as the degree of immobility of an object in the video. Foregroundness refers to the distinctiveness of the object relative to the background. Abandonment expresses the isolation of objects after in vicinity of some other entities. These three attributes were used to classify whether an object is unattended. Here, instead of using these three attributes in the classification process, we use staticness and abandonment of objects in a stage and foregroundness of objects are applied in another stage of the whole process of the unattended object detection.

In this paper, we propose a two-stage method to detect unattended objects in visual surveillance. The first stage is called the proposal stage for the unattended object detection. In this stage, staticness and abandonment attributes are used to identify candidates. First, short-term and long-term background models are applied to extract objects from background. Then, a method for tracking will help to calculate the staticness of candidate objects. Our tracking method is based on tracklet association and it can track unattended objects under crowded situations with heavy occlusions. Abandonment of candidates is defined by the number of persons moving near the object. In the second stage, a classification background model will be built to help classify whether candidates in the first stage are unattended. The classification is done by the deep learning method in [11]. In summary, our method uses three background models in two stages and incorporates with other methods such as person detection, tracking and deep learning to decrease the miss detections in Stage 1 and reduce false alarms in Stage 2.

978-1-5386-0969-9/17/$31.00

II. SYSTEM DESCRIPTION

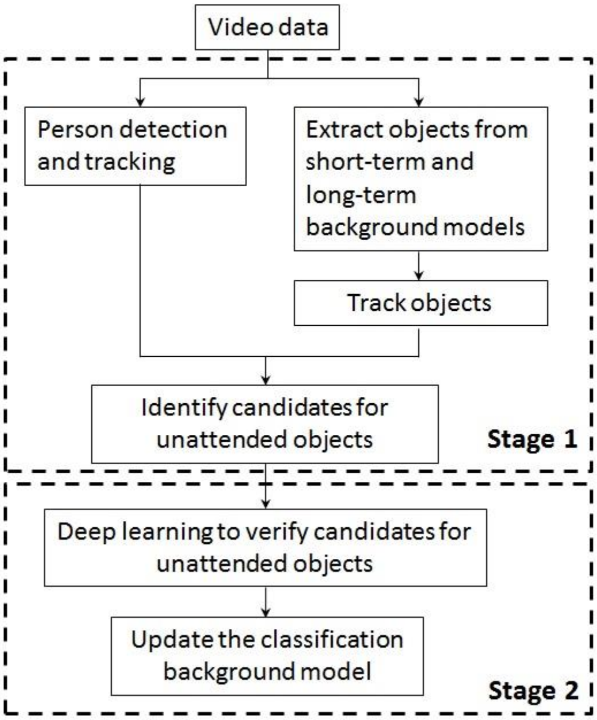

Figure 1. Our unattended object detection system.

As mentioned in the previous section, our unattended object detection includes two stages: the proposal stage for unattended object candidates (Stage 1) and the verification stage (Stage 2), shown in Figure 1. In the first stage, our system tries to detect all possible candidates for unattended objects. An unattended object is defined as the static foreground object in the scene. When an unattended object appears in the scene, it should be captured by both short-term and long-term background models. After a certain period of time, the short-term background model will not be able to detect this object because it is absorbed into the background model by a fast updating rate. However, it still appears in the long-term background model. Hence, if the system uses the short-term background model only, it cannot detect the unattended object after a certain period of time. On the other hand, if the system solely uses the long-term background model, there will be many false alarms due to noises from slow updating rate of the background model. Here, the appearance of static objects is detected by using both the short-term and the long-term background model. Then, a tracking method will follow these objects with assistance from the long-term background model only. However, occlusions between unattended objects and crowd of people will cause challenges to follow these objects. This is because unattended objects do not appear in both the short-term and the long-term background models under heavy occlusion conditions. This challenge is solved by applying the tracklet association technique in tracking. If the static object appears or is followed by tracking for a long time in the scene, it is possible an unattended object. In the last step of this stage, person detection and tracking methods are applied to measure the abandonment of objects. It is reasonable to identify the static object to be unattended when no person is loitering in the nearby area. We can conclude that no one cares about the static object, and it is abandoned. Then, a list of unattended object candidates is created.

However, some false alarms may be created due to challenging of detecting objects from the short-term and the long-term background model. For examples, a moving leaf from a tree can be a candidate of unattended objects in the first stage. Hence, in the second stage, our system applies a classification method to determine whether it is a genuine unattended object or a false alarm. A classification background model is built by using the median filter for the background model. Image patches from candidates in Stage 1 will be matched with the classification background model by using the deep learning image patch matching in [11] to confirm whether it is an unattended object.

III. PROPOSALS FOR UNATTENDED OBJECT DETECTION

A. Person Detection and Tracking

Recently, deep learning has been successfully applied in many areas in visual surveillance, especially object detection. Detection methods based on deep learning can work in different scenarios without much considering in adjusting parameters. This makes video analytic algorithms more efficient and more deployable in large scale systems. In this paper, the person detection employed from the method described in [12]. The success of object detections based on deep learning also makes tracking methods more robust in many applications. In particular, object tracking based on detections [13, 14, 15] can be used to track people in the crowd more efficiently with deep learning object detections. If a static object is not near any person who is loitering in the scene, this object can be considered to be an unattended object candidate. An example of our tracking system is in Figure 2.

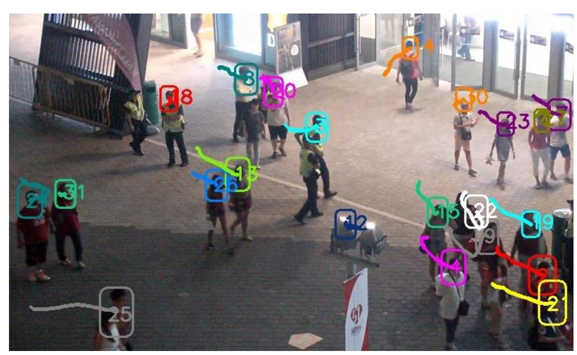

Figure 2. Person detection and tracking in our system.

B. Detect and Track Static Objects

Candidates of unattended objects are detected by using short-term and long-term background models. The method to model background images is described in [16]. The short-term background model is defined as the background model with a slow update rate β = 0.0001. A static object will be absorbed into the short-term background model after 20 frames. The long-term background model is the background model obtained from [16] with background update rate β = 0.01. If an object appear in both short-term and long-term background models more than 5 frames, this object will start to be tracked.

The tracking method for static objects is based on detections in long-term background model. The method is similar to the method described in [13, 14] with an add-on layer for tracklet association. The tracklet association will help to recover the tracking of objects after it is occluded. Hence, our method can maintain the tracking continuity of static objects even among crowd of people. If an object does not move much for a long time and no person loitering near the object, this object will be added in the candidate list. This list is called the proposal list of unattended objects. Each object in the proposal list has two attributes: staticness and abandonment. The proposal list may have some false alarms due to wrong detections from short-term and long-term background models. An example to detect unattended object is shown in Figure 3.

Figure 3. An example of detecting static objects from AVSS-AB [17]: (a) foreground mask from the short-term background model, (b) foreground mask from the long-term background model, (c) static object detection and tracking.

C. *Identify Candidates of Unattended Objects in the Proposal Stage*

After detect and track static objects, candidates of unattended object detection will be identified by some rules:

R1. Candidates should be tracked for a long time (larger than 300 frames)

R2. Candidates should appear in both short-term and long-term background models at the beginning and be absorbed in short-term background model later.

R3. Candidates should be near some persons at the beginning and they have some frames that are not near any person.

Rules R1 and R2 measured the staticness of candidates. If an object is static, it should be absorbed in short-term background model. Rule R3 measured the abandonment of candidates. Candidates in the proposal stage will be verified again in the Stage 2 to confirm they are unattended objects.

IV. VERIFY CANDIDATES OF UNATTENDED OBJECTS

As the proposal list of unattended objects can have some false alarms, it is necessary to have a verification step. First, a classification background model is built from a median filter of learning images. This classification background model is updated every half an hour. Then, for each candidate in the proposal list, an image patch will be extracted and match with the classification background model by deep learning in [11] with Siamese network (Figure 4). There are two branches of networks. Each of them takes one input image patch to propagate through layers. The outputs of these branches will be concatenated and give the decision score. The score will indicate whether candidates are unattended objects.

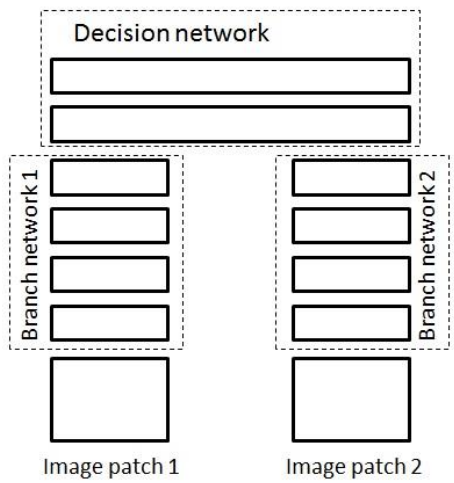

Figure 4. Image patch matching by deep learning.

V. EXPERIMENTAL RESULTS

Our system is evaluated by some public data set: AVSS-AB [17], PETS2006 [18], and our own dataset. In AVSS-AB, there are 3 cases of unattended objects that are recorded at the train platform. This data is a part of i-LIDS data [19]. PETS2006 includes 7 different scenarios captured by 4 cameras. To ensure the performance, we also have mocked up 36 scenarios. The performance measurement is based on recall and precision rate as in Eq. (1).

$$\begin{aligned} recall &= \frac{\#truedetecion}{\#events} \\ precision &= \frac{\#truedetecion}{\#truedetecion + \#falsealarms} \end{aligned} \tag{1}$$

The comparisons of results are shown in Table I. As illustrated in Table 1, our method successfully detected all of unattended object events and had high precision. While other methods tried to detect events and reduce false alarms at the same time, our method separated the detection into two stages: capture all events in the first stage and reducing false alarms as much as possible in the second stage. We have deployed our system in several public events to verify the performance of our system. The challenge in our system is to identify the background model for the second stage.

TABLE I. PERFORMANCE OF UNATTENDED OBJECT DETECTION

Methods	PETS2006		AVSS-AB		SAA-UO	
	P	R	P	R	P	R
[9]	0.95	0.8	0.91	1.0	n.a	n.a
[6]	0.6	1.0	0.1	1.0	n.a	n.a
[20]	0.5	1.0	0.03	1.0	n.a	n.a
[21]	0.83	0.83	0.5	1.0	n.a	n.a
[7]	1.0	1.0	1.0	1.0	n.a	n.a
Our method	1.0	1.0	1.0	1.0	0.87	1.0

VI. CONCLUSIONS

The paper described a two-stage method for unattended object detection. The method was validated on both public data and custom-made videos. Moreover, it was deployed to analyze the video data of the whole day with high performance. The advantage of our method is in the two-stage system. The first stage can find potential candidates of unattended objects. Most of unattended objects in the experimental data set were detected in the first stage. On the other hand, the second stage aimed to reduce the false alarms. Hence, its performance has low false alarm and high detection rate. Moreover, the method does not need to have much customizing in parameter settings. Performance analysis in the experimental result section and the flexibility of the method showed that it is a promising method for large scale deployment.

REFERENCES

[1] K. Smith, P. Quelhas and D. Gatica-Perez, "Detecting abandoned luggage items in a public space," in PETS, 2006.

[2] S. Ferrando, G. Gera, M. Massa, C. Regazzoni, "A new method for real time abandoned object detection and owner tracking," in ICIP, 2006.

[3] M. Spengler and B. Schiele, "Automatic detection and tracking of abandoned objects," in CVPR, 2005.

[4] J. Pan, Q. Fan, and S. Pankanti, "Robust abandoned object detection using region-level analysis," in ICIP, 2011.

[5] Y.-L. Tian, R. Feris, and A. Hampapur, "Robust detection of abandoned and removed objects in complex surveillance videos," IEEE Trans. Sys., Man, and Cyber., Part C: Appl. Rev., vol. 41, no. 5, 2011.

[6] S. Guler and K. Farrow, "Abandoned object detection in crowded places," in PETS, 2006.

[7] K. Lin, S.-C. Chen, C.-S. Chen, D.-T. Lin, and Y.-P. Hung, "Abandoned object detection via temporal consistency modeling and back-tracing verification for visual surveillance," IEEE Trans. Information Forensics and Security, vol. 10, no. 7, 2015.

[8] F. Porikli, Y, Ivanov, and T. Haga, "Robust abandoned object detection using dual foregrounds," EURASIP J. Adv. Sig. Proc., 2008.

[9] Q. Fan, P. Gabbur, and S. Pankanti, "Relative attributes for large-scale abandoned object detection," in ICCV, 2013.

[10] M. Bhargava, C. -C. Chen, M. S. Ryoo, and J. K. Aggarwal, "Detection of abandoned objects in crowded environments," in AVSS, 2007.

[11] S. Zagoruyko, N. Komodakis, "Learning to compare image patches via convolution neural networks," in CVPR, 2015.

[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detction with region proposal networks," in NIPS, 2015.

[13] Z. Li, L. Yuan, and R. Nevatia, "Global data association for multi-object tracking using network flows," in CVPR, 2008.

[14] H. Pirsiavash, D. Ramanan, and C. Fowlkes, "Globally-optimal greedy algorithms for tracking a variable number of objects," in CVPR, 2011.

[15] A. Andriyenko and K. Schindler, "Global optimal multi-target tracking on a hexagonal lattice," in ECCV, 2010.

[16] T. W. Chua, K. Leman, and Y. Wang, "Adaptive texture-color based background subtraction for video surveillance," in ICIP, 2012.

[17] AVSS-AB. In www.eecs.qmul.ac.uk/andrea/avss2007.html.

[18] PETS 2006. In www.cvg.rdg.ac.uk/PETS2006/data.html.

[19] i-LIDS. In ftp://motinas.elec.qmul.ac.uk/pub/iLids.

[20] R. Mathew, Z. Yu, and J. Zhang, "Detecting new stable objects surveillance video," in Multimedia Signal Processing, 2005.

[21] Q. Fan and S. Pankanti, "Robust foreground and abandonment analysis for large-scale abandoned object detection," in AVSS, 2012.