A Method for Detecting Long Term Left Baggage based on Heat Map

Pasquale Foggia, Antonio Greco, Alessia Saggese and Mario Vento*
Dept. of Computer Eng. and Electrical Eng. and Applied Mathematics, University of Salerno, Via Giovanni Paolo II, 132, Fisciano (SA), Italy

Keywords: Left Bag, Abandoned Luggage, Stopped Object Detection.

Abstract: In this paper we propose a method able to identify the presence of objects remaining motionless in the scene for a long time by analyzing the videos acquired by surveillance cameras. Our approach combines a background subtraction strategy with an enhanced tracking algorithm. The main contributions of this paper are the following: first, spatio-temporal information is implicitly encoded into a heat map; furthermore, differently from state of the art methodologies, the background is not updated by only evaluating the instantaneous movement of the objects, but instead by taking into account their whole history encoded in the heat map. The experimentation has been carried out over two standard datasets and the obtained results have been compared with state of the art approaches, confirming the effectiveness and the robustness of our system.

1 INTRODUCTION

In the last years the research community has shown a great interest toward the problem of detecting stopped objects. This is mainly due to the wide range of applicative fields where such technology may be profitably used, ranging from left baggages detection in metro stations and airports to the detection of garbage illegally dumped along the streets.

Although it is not possible to partition the existing methods into clear-cut, non-overlapping categories, two different typologies of approaches can be identified, namely tracking-based and background-subtraction based. The methods belonging to the former category first locate a foreground object when it is moving and then analyze its trajectory so as to check if it becomes a stationary object. On the other hand, the approaches based on the second category take advantage of a properly defined background model and on foreground extraction techniques to detect stopped objects.

As for the tracking-based approaches, in (Guler et al., 2007) a 4-level tracking method inspired by the human visual attention model is proposed: the considered levels are peripheral tracker, vision tunnels, scene description layer and stationary object layer. An object is considered stationary if its dwell time within the same region exceeds a given threshold, chosen by the human operator during the configuration step. In (Bhargava et al., 2007) the authors propose a backtracking approach to detect abandoned luggage in crowded scenes. The attention is not only focused on baggage, but also on the owner: in fact, the system detects a stopped bag and performs backtracking to keep track of the owner, generating an alarm only if he does not retrieve the luggage within 60 seconds. The joint analysis of the bag and its owner is also exploited in (Acampora et al., 2012), where moving objects trajectories are analyzed by a Time Delay Neural Network and a decision about the event of interest is taken by exploiting a set of manually defined fuzzy rules. In (Bevilacqua and Vaccari, 2007) the authors focus on the problem of stationary vehicles detection, by proposing a tracking algorithm based on corner point detection. Occlusions are solved with a SOM neural network and the trajectories are smoothed with a moving average, so as to partially eliminate noise caused by tracking algorithms and to facilitate stopped delay measurement: for each vehicle, its trajectory is analyzed; if it is stable in a certain region and for a long time interval, then it is considered stopped.

The above mentioned algorithms are very intuitive and achieve a very high performance in sterile environments, where only a few people populate the scene. However, their main limitation lies in the fact that they cannot be effectively used in crowded scenes: in fact, they are very sensitive to occlusions, which usually prevent the system from tracking ob-

*IAPR Fellow

Foggia P., Greco A., Saggese A. and Vento M..
A Method for Detecting Long Term Left Baggage based on Heat Map.
DOI: 10.5220/0005306803850391
In *Proceedings of the 10th International Conference on Computer Vision Theory and Applications* (VISAPP-2015), pages 385-391
ISBN: 978-989-758-090-1

jects and then from extracting trajectories in a reliable way. On the other hand, methods based on background subtraction are able to overcome such limitation and in general to achieve better performance. This is due to the fact that these methods perform a temporal analysis of foreground objects by sophisticated background modeling strategies, and thus common problems such as occlusions cannot influence the performance of this kind of systems.

For instance, the authors in (Maddalena and Petrosino, 2013) propose a general framework, called stopped foreground subtraction (SFS), that is independent of the specific background modeling and foreground extraction methods: in fact, they created a model of the stopped objects and use it to classify a new stationary one. Furthermore, they propose a background updating method based on neural networks, called 3DSOBS, which proved its effectiveness if compared with the traditional MOG (Stauffer and Grimson, 1999). In (Porikli et al., 2008) the algorithm maintains two different backgrounds, updated with different speeds by using the traditional MOG: a short term and a long term one. Thus, an evidence image is computed by considering the pixels whose change rate ranges between the short term and the long term background updating rate; this image is finally used to detect stopped objects. The method proposed in (Boragno et al., 2007) is developed on the Ipsotek VI platform and is able to detect vehicles parked in prohibited areas. The algorithm is based on 3 steps: motion detection through block matching, stopped object detection using MOG and object classification into trucks, cars, pedestrians and packages.

Starting from the spatio-temporal information obtained from the background analysis, several methods also include a tracking phase in order to extract more useful information about the objects in the scene, so as to increase the detection capabilities of the system as well as to decrease their false alarm rate. Such algorithms are usually referred to as hybrid approaches

For instance in (Smitha and Palanisamy, 2012) a parked vehicles detection algorithm is described; it uses a simple background subtraction technique and a region-based tracking to identify the stop of a vehicle. The method proposed in (Singh et al., 2009) is based on a dual background subtraction technique to detect stationary objects and on a tracking algorithm, optimized to solve occlusions, able to reduce the number of false positives introduced by the pure background subtraction approach. In fact, the system raises an alarm only if the object hit count exceeds a given threshold, i.e. if it remains stationary for a certain time interval. In (Venetianer et al., 2007) the authors propose a method for abandoned objects or stopped vehicles detection developed on the ObjectVideo commercial platform. The algorithm consists of 4 steps: background subtraction, blob detection, tracking and stopped object detection. The last step is performed bringing almost immediately the stationary object in the background and retaining the original background: the temporal evolution of the comparison between the current frame and the original background is used to detect the event of interest. A more sophisticated approach is proposed in (Tian et al., 2011), which performs stopped object detection through 4 steps: MOG-based background subtraction, abandoned and removed object classification, object classification (human, vehicle, package) and tracking.

In (Albiol et al., 2011) the authors propose a parked vehicle detection algorithm, based on corner detection, which creates a set of spatiotemporal maps, used to understand what is happening in the scene and to extract information such as the number of available stalls, the number of parked vehicles in prohibited areas, the mean stopping times, the queue length and so on.

The analysis of the literature performed up to now makes it evident that the most successful methods are those able to provide an accurate background subtraction. However, in order to achieve a good performance, it is advisable to use a tracking algorithm, possibly robust to occlusions, in order to evaluate the movement of an object in the scene in a more sophisticated way. Starting from these considerations, in this paper we propose a hybrid solution that involves the use of a novel background updating technique based on a spatio-temporal analysis and of a tracking algorithm based on objects similarity evaluation (Di Lascio et al., 2012)(Foggia et al., 2013). The background image is modeled using adaptive selective updating, whose adaptive weights depend on the time spent by the particular pixel inside the scene. In particular, we generate a grayscale heat map, where the intensity of each pixel grows proportionally with its persistence time in the foreground mask. The main idea is that pixels corresponding to objects crossing the scene have not to influence the background, and thus their updating weight should be very low. On the contrary, pixels belonging to objects stopped in the scene for a long time should enter the background in a very fast way, so their updating rate should be very high.

Furthermore, differently from traditional methods, the tracking is performed on the heat map instead of the foreground mask, so as to make the proposed approach insensitive to occlusion problems; indeed, any persons who temporarily exclude a left object from the camera view can not accumulate a sufficient dwell time to enter the heat map and thus is not tracked by

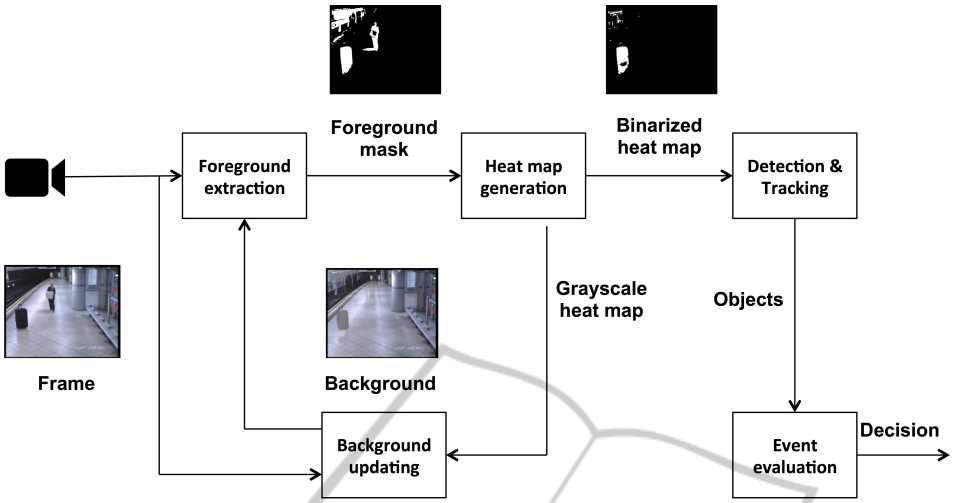


Figure 1: Overview of the proposed approach: once it has extracted the foreground mask and generated the heat map, the system keeps track of each abandoned object and raises an alarm if its dwell time exceeds a threshold.

the system. In this way, the proposed algorithms is able to detect abandoned objects by keeping track of those having a stopping time greater than a fixed time, chosen by the human operator during the configuration step.

2 THE PROPOSED METHOD

An overview of the proposed approach is shown in Figure 1. The pixels corresponding to moving objects are extracted (Foreground extraction) and are used to update the heat map (Heat map generation), so that the longer a pixel remains in the foreground mask, the brighter it appears in the heat map. The heat map and the foreground mask are used by the background updating module, which adjusts foreground pixels updating weights according to the corresponding intensity on the heat map; thus, an object enters slowly the background when its intensity is low but, gradually, it enters faster if its persistence time increases. Note that such strategy allows the system to easily discard spurious objects due to the detection step as well as persons that do not stop in the scene and that do not need to be further analyzed. Finally, the system performs on the heat map the detection and the tracking: if an object is found and it stops for a long time in the scene, then the system will raise an alarm.

2.1 Heat Map Generation

The objects moving in the scene at the current frame t are encoded by the so called foreground mask (see Figure 2(c)), obtained by traditional background subtraction algorithms.

Let be $D_t(x,y)$ the distance between the the current image I_t and the background updated up to the previous frame B_{t-1} in the generic pixel (x,y):

$$D_t(x,y) = |I_t(x,y) - B_{t-1}(x,y)| \quad (1)$$

The foreground mask can be computed as follows:

$$F_t(x,y) = \begin{cases} 1 & if\ D_t(x,y) \geq \tau_{fm} \\ 0 & otherwise \end{cases} \quad (2)$$

being in our experiments τ_{fm} set to 25.

Note that F_t does not encode any information about the temporal variation, that is in this kind of applications a very important and not negligible feature. For this reason, we also introduce an heat map H_t, able to encode in a single image a temporal analysis (see Figure 2(d)). In more details, the heat map is a grayscale image whose generic pixel is updated by a weighted moving average:

$$H_t(x,y) = \alpha \cdot F_t(x,y) + (1-\alpha) \cdot H_{t-1}(x,y), \quad (3)$$

where α is the heat map updating weight, whose value depends on the latency time chosen by the user during the configuration step. In this way, we are able to obtain a kind of transparency overlay of F_t on H_t. Starting from H_t, its binarized version H_t^{bin} is computed:

$$H_t^{bin}(x,y) = \begin{cases} 1 & if\ H_t(x,y) \geq \tau_{hm} \\ 0 & otherwise \end{cases} \quad (4)$$

Thus, the system can start verifying if an object is abandoned only once he has entered into H_t^{bin}, so significantly reducing the computational effort required by the successive detection and tracking steps.

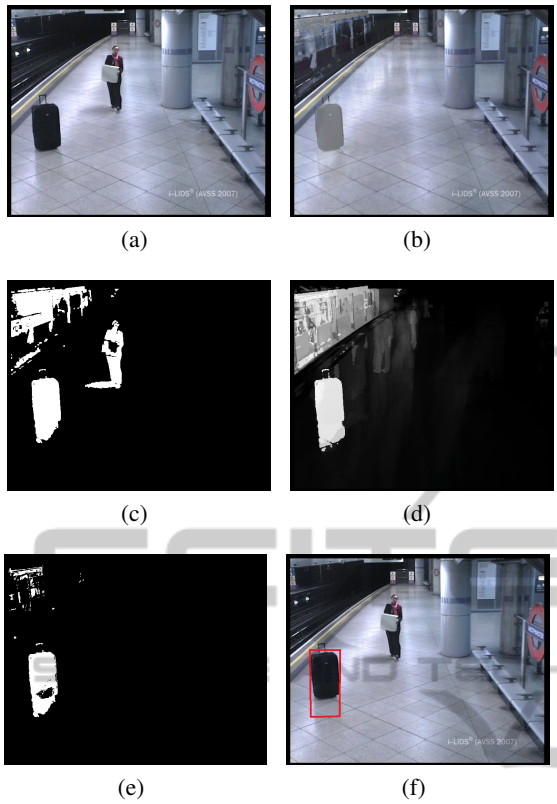


Figure 2: An example of the different modules involved in the stopped object detection: (a) current frame, (b) background, (c) foreground mask, (d) heat map, (e) binarized heat map, (f) detected objects overlaid on the current frame. Note that the event has been detected, so the object can start entering the background.

2.2 Background Updating

One of the main contributions of the proposed approach pertains to the definition of a novel background subtraction algorithm. In fact, traditionally the background is only updated by evaluating the foreground mask at the current frame. On the other hand, our aim is to control the entering time of the objects moving in the scene depending on their motion, and then on their history. The key idea is that the lower is the time spent in the scene by an object, the higher is the time required for updating the background.In order to achieve this aim, we decided to selectively update the background depending on both the heat map and the foreground mask. In fact, the weights for pixels corresponding to moving objects are dynamically updated, depending on the time spent by the object in that particular position.

In particular, let H_t^s be the static heat map, that is the heat map evaluated only on the pixels corresponding to moving objects at the current frame and computed as follows:

$$H_t^s(x,y) = H_t(x,y) \cdot F_t(x,y) \tag{5}$$

Starting from H_t^s, we can evaluate the background B_t as follows:

$$B_t(x,y) = \begin{cases} \alpha_B \cdot I_t(x,y) + \overline{\alpha_B} \cdot B_{t-1} \\ \qquad\qquad if\ H_t^s(x,y) = 0 \\ \alpha_F(x,y) \cdot I_t(x,y) + \overline{\alpha_F(x,y)} \cdot B_{t-1} \\ \qquad\qquad otherwise \end{cases} \tag{6}$$

being $\overline{\alpha_F(x,y)} = 1 - \alpha_F(x,y)$. α_B and $\alpha_F(x,y)$ represent the updating weights for the background and for the foreground pixels, respectively. Note that α_F is not fixed, but instead it depends on the particular pixel (x,y). In fact, it is related to the corresponding intensity value on the heat map:

$$\alpha_F(x,y) = \alpha_H \cdot H_t^s(x,y) \tag{7}$$

As for the α values, they strongly depend on the minimum dwell time chosen by the human operator during the configuration step. In particular, we consider that a stopped object has to enter the background as soon as the related event is detected.

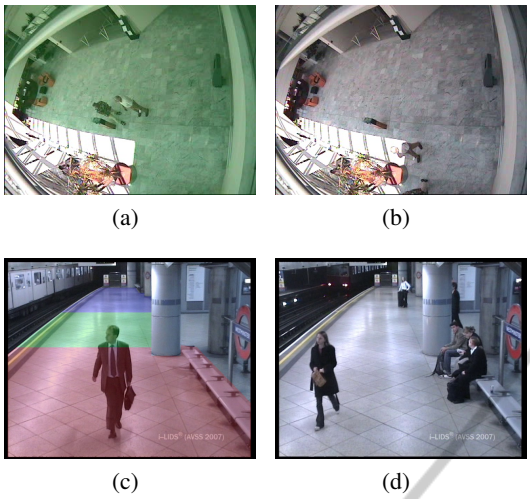
2.3 Event Evaluation

Differently from other state of the art approaches, the detection of connected components and the tracking of the objects is performed directly on the binarized heat map. In particular, the tracking algorithm proposed in (Di Lascio et al., 2013) has been used for our experimentations. The main advantage deriving from the choice lies in the fact that only those objects stopped for a long time are involved in the tracking process, so making the system particularly robust with respect to occlusions as well as especially suited for working in real and crowded environments.

Objects are finally analyzed and their permanence time in the heat map is properly evaluated. Once an event of interest has been detected, an alarm is raised to the human operator.

3 EXPERIMENTAL RESULTS

The proposed approach has been tested over two standard and widely adopted datasets, namely the CAVIAR Dataset (CAVIAR, 2003) and the Imagery Library for Intelligent Detection Systems Abandoned Baggage Dataset (i LIDS, 2007).



(a) (b) (c) (d)

Figure 3: A few images extracted from the datasets used for testing the proposed approach: CAVIAR (a,b) and i-LIDS (c,d). On the left (a,c) the regions of interest are overlaid.

The former is composed by 26 videos acquired in an indoor environment and showing different scenarios: in particular, 5 videos contain a left bag event, while the remaining 21 contain other situations typically occurring in real scenarios, such as a person walking alone, people meeting with others, fighting and so on. In this way, the proposed approach can be evaluated not only in terms of events correctly recognized, but also in terms of situations wrongly recognized as left bag events (false positive). The resolution is half-resolution PAL standard (384x288, 25 fps) and compressed using MPEG2; the total video sequence duration is about 17 minutes.

The i-LIDS Abandoned Baggage dataset (hereinafter i-LIDS) consists of 3 video with different difficulty levels, namely easy, medium and hard. It has been used during i-LIDS bag and vehicle detection challenge, hosted by AVSS 2007. Each video has a resolution of 720x570 and has been acquired at 25 fps; the total length of the videos is about 10 minutes. The scene captured by the camera shows a railway where a person leaves a baggage unattended for 60 seconds. According to the competition rules, the detection area can be divided into 3 zones (near, mid and far), so as to set different parameters depending on the distance from the baggage.

A few examples for both the datasets are shown in Figure 3, while the obtained results are reported in Tables 1 and 2 for CAVIAR and i-LIDS datasets, respectively.

In particular, as for the CAVIAR dataset, the results for each typology of event is reported; the table can be read as follows: TD indicates that a left bag event has been correctly detected (true detected),

Table 1: Results obtained on the CAVIAR dataset, in terms of True Detected (TD) and False Detected (FD).

CAVIAR Dataset			
	Video	**TD**	**FD**
Left Bag	Left Bag	1/1	0
	Left Bag At Chair	1/1	0
	Left Bag Behind Chair	0/1	0
	Left Bag Picked Up	1/1	0
	Left Box	1/1	0
Walk	Walk 1	-	1
	Walk 2	-	0
	Walk 3	-	0
Browse	Browse 1	-	0
	Browse 2	-	2
	Browse 3	-	0
	Browse 4	-	0
	Browse While Waiting 1	-	0
	Browse While Waiting 2	-	0
Rest	Rest Fall On Floor	-	2
	Rest In Chair	-	0
	Rest Slump On Floor	-	1
	Rest Wiggle On Floor	-	3
Meet	Meet Crowd	-	0
	Meet Split 3rd Guy	-	0
	Meet Walk Split	-	0
	Meet Walk Together 1	-	0
Fight	Fight Chase	-	2
	Fight One Man Down	-	3
	Fight Run Away 1	-	0
	Fight Run Away 2	-	0

while FD indicates that something in the scene has been wrongly detected as a left bag event (false detected). For instance, the first row can be read as follows: in the *Left Bag* video, one event out of one has been correctly recognized (so implying that there are not missing events) and no false positives have been detected. We can note that abandoned baggages have been successfully detected in 4 videos out of 5 of the Left Bag dataset sequences, as also shown in Figure 4; the missed event is due to the fact that the bag is hidden behind the chair, so it is not possible with any traditional algorithm based on either tracking or background subtraction methodologies to discover this kind of event. We can also note that a few false alarms have been detected, for instance in the videos *Fight One Man Down* and *Fight Chase*. It is mainly due to the fact that in such scenarios there are people stopping on the wall for a very long time, therefore the system raises an alarm since no information about the typology of the objects is provided. It is evident that the introduction of a classification step, able to distinguish, for instance, bags from persons, may avoid the generation of these errors. Note that classification step is made possible in the proposed approach by the tracking algorithm, which allows to keep track of the objects and eventually of their re-

Table 2: Results obtained over the i-LIDS dataset and compared with state of the art approaches.

i-LIDS Abandoned Baggage						
	AB Easy		AB Medium		AB Hard	
Method	**TD**	**FD**	**TD**	**FD**	**TD**	**FD**
Proposed method	1/1	0	1/1	0	1/1	1
(Maddalena and Petrosino, 2013)	1/1	0	1/1	0	1/1	1
(Evangelio et al., 2011)	1/1	0	1/1	5	1/1	6
(Pan et al., 2011)	1/1	0	1/1	0	1/1	0
(Tian et al., 2011)	1/1	0	1/1	0	1/1	1

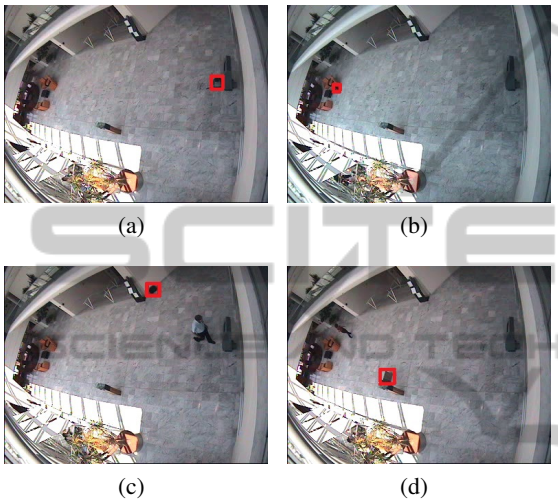


Figure 4: The baggage found in the CAVIAR dataset over the videos (a) Left Bag, (b) Left Bag At Chair, (c) Left Bag Picked Up and (d) Left Box.

Figure 5: The baggages identified in the i-LIDS dataset in videos (a) AB Easy, (b) AB Medium and (c) AB Hard. In (d) the false detected object is reported.

lated properties, such as the class it belongs to.

As for the i-LIDS dataset, the obtained results are shown in the first row of Table 2. We can note that the event of interest is recognized in all the videos (Easy, Medium and Hard) and only one false positive has been detected in the Hard videos. Figures 5(a-c) show the detected event. It is worth noting that the baggage in the Hard sequence has been correctly detected (Figure 5(c)) although the occlusion occurring in the previous frames. On the other hand, Figure 5(d) shows the false positive detected by the system, due to the stopped legs of the person sitting on the bench. As said for the CAVIAR dataset, also in this case the introduction of a classification step would make it possible to distinguish bags from non bags objects. In order to further confirm the effectiveness of the proposed approach, a proper comparison with state of the art methods has been carried out. As shown in Table 2, the achieved results are comparable both in terms of true detected and false detected. Not that the only method which does not detect any events of interest is proposed by (Pan et al., 2011): it is due to the fact that it includes a post processing step in order to analyze the detected events of interest and further reduce the number of false positives arisen by the system.

4 CONCLUSIONS

In this paper we proposed a novel approach based on an advanced background subtraction algorithm for detecting stopped objects in crowded environments. The experimentation has been conducted over two standard datasets, namely the CAVIAR and the i-LIDS datasets, and the obtained results, compared with state of the art approaches, confirm that the system is able to reliably detect the events also in presence of occluding objects, typically affecting this kind of algorithms. Furthermore, the high recognition rate is not paid in terms of false positives generated by the system, that are still low even if compared with state of the art approaches. Future works include the possibility to apply a classification algorithm to the detected objects stopped in the scene for a long time, so as to distinguish between persons and inanimate objects;

this would make possible a further reduction of the number of false positives detected by the system and thus would improve the usability for a human operator.

ACKNOWLEDGEMENTS

This research has been partially supported by A.I.Tech s.r.l. (http://www.aitech-solutions.eu).

REFERENCES

Acampora, G., Foggia, P., Saggese, A., and Vento, M. (2012). Combining Neural Networks and Fuzzy Systems for Human Behavior Understanding. *2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance*, pages 88–93.

Albiol, A., Sanchis, L., and Mossi, J. M. (2011). Detection of parked vehicles using spatiotemporal maps. *Intelligent Transportation Systems, IEEE Transactions on*, 12(4):1277–1291.

Bevilacqua, A. and Vaccari, S. (2007). Real time detection of stopped vehicles in traffic scenes. In *Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on*, pages 266–270. IEEE.

Bhargava, M., Chen, C.-C., Ryoo, M. S., and Aggarwal, J. K. (2007). Detection of abandoned objects in crowded environments. In *IEEE AVSS*, pages 271–276. IEEE.

Boragno, S., Boghossian, B., Black, J., Makris, D., and Velastin, S. (2007). A dsp-based system for the detection of vehicles parked in prohibited areas. In *Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on*, pages 260–265. IEEE.

CAVIAR (2003). Caviar test case scenarios. http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/.

Di Lascio, R., Foggia, P., Percannella, G., Saggese, A., and Vento, M. (2013). A real time algorithm for people tracking using contextual reasoning. *Computer Vision and Image Understanding*, 117(8):892–908.

Di Lascio, R., Foggia, P., Saggese, A., and Vento, M. (2012). Tracking interacting objects in complex situations by using contextual reasoning. In *VISAPP (2)*, pages 104–113.

Evangelio, R. H., Patzold, M., and Sikora, T. (2011). A system for automatic and interactive detection of static objects. In *Person-Oriented Vision (POV), 2011 IEEE Workshop on*, pages 27–32. IEEE.

Foggia, P., Percannella, G., Saggese, A., and Vento, M. (2013). Real-time tracking of single people and groups simultaneously by contextual graph-based reasoning dealing complex occlusions. In *Performance Evaluation of Tracking and Surveillance (PETS), 2013 IEEE International Workshop on*, pages 29–36.

Guler, S., Silverstein, J. A., and Pushee, I. H. (2007). Stationary objects in multiple object tracking. In *Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on*, pages 248–253. IEEE.

i LIDS (2007). Abandoned baggage dataset. ftp://motinas.elec.qmul.ac.uk/pub/iLids/.

Maddalena, L. and Petrosino, A. (2013). Stopped object detection by learning foreground model in videos. *IEEE transactions on neural networks and learning systems*, 24(5):723–735.

Pan, J., Fan, Q., and Pankanti, S. (2011). Robust abandoned object detection using region-level analysis. In *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pages 3597–3600. IEEE.

Porikli, F., Ivanov, Y., and Haga, T. (2008). Robust abandoned object detection using dual foregrounds. *EURASIP Journal on Advances in Signal Processing*, 2008:30.

Singh, A., Sawan, S., Hanmandlu, M., Madasu, V. K., and Lovell, B. C. (2009). An abandoned object detection system based on dual background segmentation. In *Advanced Video and Signal Based Surveillance, 2009. AVSS'09. Sixth IEEE International Conference on*, pages 352–357. IEEE.

Smitha, H. and Palanisamy, V. (2012). Detection of stationary foreground objects in region of interest from traffic video sequences. *International Journal of Computer Science Issues*, 9(2):194–199.

Stauffer, C. and Grimson, W. E. L. (1999). Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 2. IEEE.

Tian, Y., Feris, R. S., Liu, H., Hampapur, A., and Sun, M.-T. (2011). Robust detection of abandoned and removed objects in complex surveillance videos. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 41(5):565–576.

Venetianer, P. L., Zhang, Z., Yin, W., and Lipton, A. J. (2007). Stationary target detection using the objectvideo surveillance system. In *Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on*, pages 242–247. IEEE.