Classification of Abandoned & Unattended Objects, Identification of Their Owner with Threat Assessment for Visual Surveillance

Harsh Agarwal, Gursimar Singh, Mohammed Arshad Siddiqui
PDPM Indian Institute of Information Technology Design & Manufacturing, Jabalpur, India

Motivation

The world is witnessing terrorism and security risk at a global level. There has been a significant rise in such attacks in recent years. In last five years, India alone has been a victim of over 1400 attacks. A large number of these attacks occur in a crowded social setting such as railway station or airport resulting in substantial number of casualties. Such attacks use explosives that are hidden inside bags and suitcases which are left at public places and go unnoticed.

Year	Blasts	Causalities
2016	337	591
2015	268	574
2014	190	370

Table 1. Bomb Blasts in India.

Abstract

We propose a model that can classify abandoned and unattended objects separately in a visual surveillance feed and then backtrack to identify the owner as well as mark his/her last known location. Our model differentiates between multiple threat levels associated with the time for which the object was left abandoned. The threat level of an abandoned object is subsequently raised once a certain predefined time has passed without any action.
Our model is divided into three major modules namely static foreground segmentation, abandoned object classification and backtracking to identify the owner. We use Gaussian Mixture Model as proposed by [1] for background modelling along with pixel based finite state machine for static foreground detection. Humans are identified using a combination of state of the art MobileNet-SSD [2] and HOG-SVM [3] human classifier.
Object is classified into attended/unattended/abandoned/ based on the real world distance between the object and the owner found using statistical approach on a perspective camera model [4].

Long-Term and Short-Term Model

Two different learning rates are used to create two Gaussian Mixture Models namely Long Term and Short Term model. In Short Term model new static elements are labelled as background while in the long term model they are labelled as foreground. Difference of both the models give the static foreground pixels.

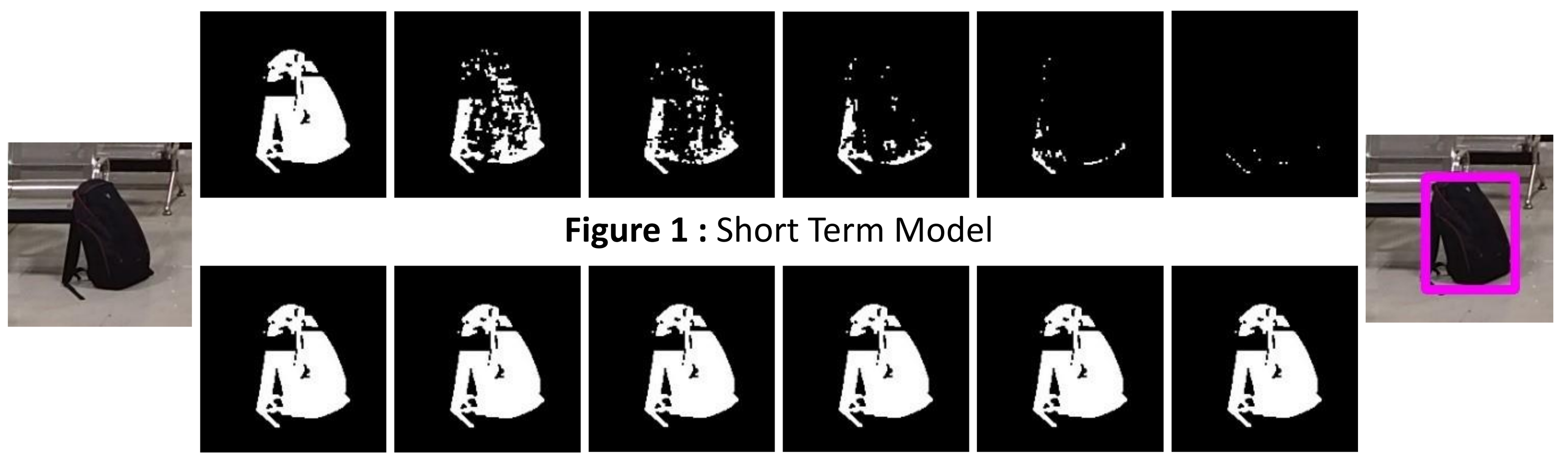


Figure 1 : Short Term Model

Figure 2 : Long Term Model

To filter the pixels which remain static in subsequent frames we propose a Pixel-based Finite State Machine (PFSM). The state of each pixel is defined as:

$$S(i) = F_{ST}(i)\,F_{LT}(i)$$

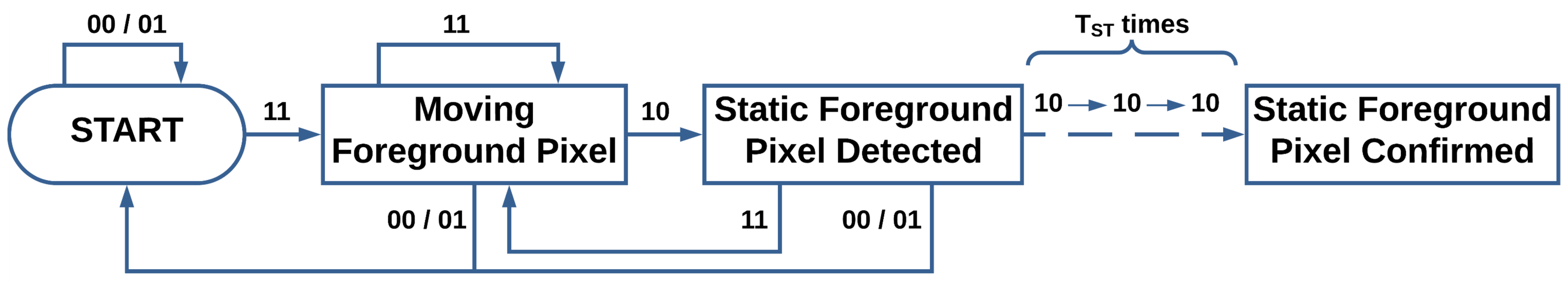


Figure 3 : PFSM Module

Real World distance estimation

In general a well calibrated camera or multi-camera setup is required to obtain precise 3D locations of the objects, which cannot be expected in the real-world situations.
Here we use a statistical approach on a perspective single camera model. We assume that the object is on the ground and camera tilt is low.
Below formulas are used for calculating 'x' and 'z' coordinates of the object

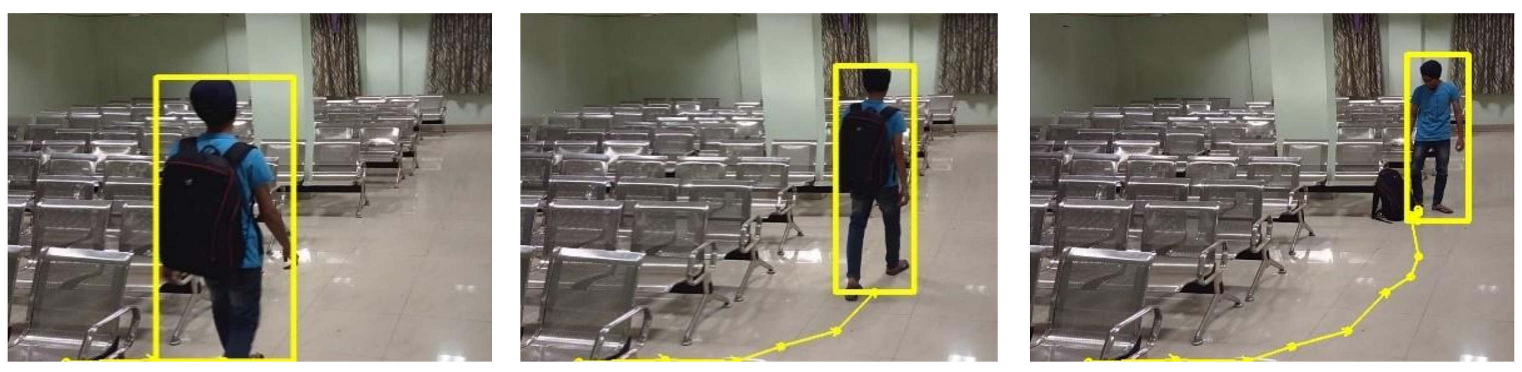
$$x = \frac{z\,(u_b - u_c\cos\theta_x)}{f}$$

$$z = \frac{f y_c}{v_b - v_c\cos\theta_x + f\sin\theta_x}$$

Figure 4 : 2D visualization of real world distance of the bag

Owner Detection & Tracking

For efficient human detection we use a combination of MobileNet-SSD model and Histograms of Oriented Gradient (HOG) with SVM. SIFT is used to extract features and descriptors of all humans in a frame. FLANN is used to match the stored features in consecutive frames and thus track the person.
Once a static object is detected, we backtrack to the frame where the static object was first detected. The distances of humans from this object is calculated for current frame using this approach. The person closest to this object in the frame where this object was first detected is identified as the owner.



(a) (b) (c)

Figure 5 : (a), (b) and (c) shows the person being tracked in consecutive frames

Result

The model was evaluated on PETS 2006 and AVSS 2007 dataset. Our results with evaluation metrics Precision, Recall and F-Score are shown in Table 2.
Though, it has been evaluated using the conventional F-Score metric (for comparison with the other state-of-the-art models), it does not give a good evaluation of the overall steps involved. Therefore, for a better understanding of our model we define a more holistic Accuracy Metric. The metric A is defined as,

$$A = \frac{[N - (N_0 + N_1 \times 0.50 + N_2 \times 0.50 + N_3 \times 0.25)]}{N}$$

Where,
N= Total number of cases tested
N_0= No abandoned object detected
N_1= Abandoned object is detected but owner is incorrectly identified
N_2= Object labelled correctly but human is detected as object
N_3= Object labelled correctly but false blob is also detected
Here, we penalize those cases more which would have the most negative impact at the setting.
Evaluating our model on this metric we get **A = 0.857**

	PETS 2006	AVSS 2007
Precision	0.83	1.0
Recall	0.89	1.0
F-Score	0.86	1.0

Table 2. Results

Conclusion

We present a temporal consistency model combining a reverse traversal algorithm for abandoned object detection in complex environments.
The enhanced background modelling which uses short-term and long-term background model is more effective than single-image based double background modelling.
In order to reduce false alarms we use PFSM to achieve temporal transition information in sequential pattern.

References

1. Zivkovic, Z., Van Der Heijden, F., 2006. Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern recognition letters*, *27*(7), pp.773-780.
2. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C., 2016, October. Ssd: Single shot multibox detector. In European conference on computer vision (pp. 21-37). Springer, Cham.
3. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proc. CVPR 2005. IEEE Computer Society Conference on, vol. 1, pp. 886-893. IEEE, 2005
4. Baur, R., Efros, A., Hebert, M., 2008. Statistics of 3d object locations in images