Unattended Baggage Detection System For Occlusive Complex Environment Using Collaborative Mask

Pulkit Narwal, Rahul Mishra

Abstract— In today's world, security at public places is main concern for authorities. Automated video surveillance is the answer to such situations where the designed system checks for security based on certain defined conditions. From this domain, unattended baggage is a wide concern for security parameters. A baggage left unattended or abandoned baggage could pose serious threats to the security in public places. To minimize this risk, an advanced technological capability in video surveillance system could prove advantageous. The system detects unattended baggage (that is no longer being attended), in real time by analyzing the videos from surveillance cameras and triggering the alert and alarming the authorities whenever such unattended baggage is encountered and detected by the system. The video is analyzed per frame basis. Initial stage involves background subtraction, hence focusing on foreground objects. Collaborative mask is created using first frame as background and Gaussian method. The foreground regions are then checked to qualify for static foreground region. The static foreground regions contain possible candidates for unattended baggage. Baggage identification is performed to verify the presence of baggage and type of baggage using a CNN trained model. Various thresholding checks are performed to ascertain the abandonment of the baggage. The system works in real time and performs computation and check conditions for unattended baggage frame by frame from a video. It takes one frame at a time and checks certain parameters for corresponding consecutive frames. If the conditions get fulfilled, object is termed as unattended baggage and a bounding box is created around that baggage.

Index Terms— Video Surveillance, abandonment, static foreground, background subtraction, baggage identification, thresholding, bounding box, trigger, unattended baggage.

1 Introduction

In recent years, there has been a great increase in video surveillance at public places to ensure safety and security of people and the place. This surveillance of video requires constant manual monitoring by a team of security personnel to ascertain the security assurance. Hence, to maintain regular monitoring , manpower is needed to monitor the systems and check the security and safety parameters at each point of time, which makes it dependent on the human presence at all times and becomes a constraint by limiting the resource. To achieve high efficiency and reduce human monitoring, there is a dire need of an automated system that no longer requires human monitoring to accomplish security. With world security concerns being at high stakes, automated techniques to detect unattended baggage from CCTV surveillance cameras requires more effective advancements. Unattended baggage at public places could lead to serious security issues & can impose as a threat to security parameters. Detecting unattended baggage by analysing frames from the video in real time reduces manual observation & reduces security risks to its minimum.

The Unattended baggage detection in real time incorporates various algorithms which are run simultaneously with the video subject to unattended conditions.

The system performs various tasks to term an object as unattended baggage. Unattended baggage detection involves two aspects: a) whether the object is unattended, b) whether the object is baggage. Both these aspects are further quantified into five steps: 1) Background subtraction, 2) Static foreground region detection, 3) Baggage identification, 4) Thresholding, 5) Unattended Baggage & Alarm Triggering. Initial stage involves background subtraction i.e. to reduce the number of objects in frame that are a part of object or can't be a candidate of unattended object. Connected component analysis is done on the grayscale image, progressively creating contours in the frame. The contour represent a foreground connected component. For every frame contours are calculated. A contour whose centroid did not change for particular time period employs no other contour has been in contact (if so, the contour recombines to create a new contour and hence new centroid) is assumed to be static foreground. This centroid verification of centroid gives the possible candidates for static foreground region objects. Dividing the frames into background and foreground regions ascertain high accuracy and low chances of false positives. Background subtraction is achieved using two techniques and collaborating them. A mask is created by

- *Pulkit Narwal is currently working Part time in H.J.Bhabha I.T.I., New Delhi, India (D.T.T.E. Dept., Delhi State Govt.), & pursuing Masters of Technology in Computer Science & Engineering from C-DAC, Noida, India. E-mail: pulkitnarwal2@gmail.com*

drawing a comparison in the current frame and background frame (first frame of the video). The comparison mask gives the foreground regions obtained. Gaussian mixture model is used to create a second mask of foreground regions. For each pixel of frame 3 Gaussians are maintained with one storing pixels with persistent intensities, other storing the pixels with variable changes and the last Gaussian storing the pixels with fast changing intensities or the dynamic pixels. These two masks are combined to form a collaborative mask which acts as a basis for detecting static foreground regions. The system localizes the possible candidates for unattended object using static foreground region detection. The localized regions are then forwarded to identify type of object the region contains. A pre-trained model is used which is customized to identify three types of baggage. The results contain the type of object (if baggage) and a probability of certainty that the localized object resembles to actual object. After object identification, the possible candidate is passed to thresholding phase. There are certain conditions with time and distance threshold which if qualified by the object. These conditions are employed on every frame. If the conditions/parameters remain true for a certain time period or a successive number of frames. The object is highlighted with a bounding box along with unattended baggage alert to alarm the authorities.

Fig. 1. Detection of unattended baggage- Own DATASET

2 Previous Work

Previous work done in this problem domain uses different techniques at different stages. Most of the work include background subtraction as basis for approach. Liao et al. [4] uses foreground mask sampling for foreground region detection & background subtraction, which creates a mask of foreground pixels. A foreground mask M is an Intersection over Union of static region mask S_k where $k=1$ to n.

$$S_k = p_1 \cup p_2 \cup \ldots \cup p_i \quad (1)$$

$$M = S_1 \cap S_2 \cap \ldots \cap S_n \quad (2)$$

Tian et al. [3] performs background subtraction using Gaussian mixture model given by Stauffer & Grimson. Complementary background modelling is given by [5], which uses short-term and long-term background models to identify foreground static pixels. A mixture of Gaussians is taken into account by [3] to extract foreground static regions. An effective learning algorithm is proposed by Lee [2] based on Stauffer & Grimson's adaptive background mixture model [1] to get background subtracted frame. The algorithm has improved convergence rate and estimation accuracy over the standard method. 1st Gaussian represent persistent pixels or background region. 2nd Gaussian update pixels with variations at a slow rate or relative stationary region & 3rd Gaussian stores pixels with variations at higher rate. If the value of 2nd Gaussian for a pixel is larger than the threshold, the pixel belongs to the static region. Smeureanu et al. [7] applies Static Object Detection (SOD) pipeline by subtracting motion mask from foreground mask & perform IoU (Intersection over Union) on bounding boxes. A cascade of CNN introduced by Smeureanu et al.[7]. The first CNN is pre-trained with type of objects that will be termed as baggage. The mask is passed through filters, whose intersection gives feature information which is compared from ground truth tables. A bounding box is created around objects tested positive. A second CNN is applied to image samples tested positive by first CNN. The bounding box is scaled with twice the height & thrice the width of previous bounding box obtained after first CNN. The new bounding box checks whether the owner is within bounding area or neighborhood of object. Liao et al. [4] implements Selective tracking & motion prediction is performed on owner to compute probabilistic score. This score is compared with predetermined confidence score to declare object as abandoned with a probability certainty. Conditions for object to be termed as unattended according to the following two rules, which are defined by PETS2006 [8].

1) Temporal Rule: The object is abandoned by its owner and is not attended for time T=30 secs.

2) Spatial Rule: If the owner is not in the neighborhood of the object, the distance for that specifies neighborhood is given by D=3 meters.

PETS2006 considered temporal rule & spatial rule as prime conditions to term an object as unattended. Tian et al. [3] detects static region type by region growing segmentation. Background subtraction performed in [6] compares every frame with background frame & determines object as unattended through time thresholding. A pixel based finite state machine model is presented in [5] to use sequence of state transition pattern as a reference to stationary foreground pixels.

3 Proposed System

The proposed system divides the video into frames & works on individual frames. Fig. 1 depicts the workflow of proposed system. The frames obtained from the video are firstly converted into grayscale images and then are converted into binary images after removal of noises and applying morphological functions to improve the quality of images. The resultant images are then passed through background subtraction phase. The frame is processed for three types of objects: (a) background objects, (b) static foreground objects which were not in earlier frames, but are introduced later & has remained static for a certain time duration and (c) dynamic objects which change their relative location with time. Here, the regions in the image are identified as either background region or foreground region. Background region corresponds to the region which has remained constant and there is no change in that area or the corresponding pixels contained in that region show no change in the color intensity. The foreground region is the opposite, it is the region enclosing the pixels with variable color intensities, hence representing dynamic regions. These are identified by checking continuous series of frames to see whether there has been a change in the pixel intensities from previous frames. The background objects or environment are of no interest to system, hence, background region is discarded and focus is put on foreground regions. The detected foreground regions are now used to further detection process. The foreground region may include: Dynamic objects or objects in motion, Objects that were dynamic for some time and then became static. The objects which show continuous movement are dynamic objects.

Once a baggage owner enters the scene with the baggage, the whole contour is treated as a dynamic object. As the owner abandons the baggage, the baggage becomes static foreground region. The system intends to find static foreground region for discovery of unattended baggage. The static foreground region detection is done and highlights the possible candidates of unattended baggage so far. Static foreground region accounts for region that holds object which were introduced in the frames in motion, but became static after certain time. The candidates of static foreground region are then passed to baggage identification phase. To ascertain that the objects captured by static foreground region detection are baggage or not. A model is used to examine objects falling under static foreground region which classifies baggage along with its confidence of that classification. The model parses the results with include type of baggage (if present) and the probability score or the certainty score which represent the accurateness of the result. The baggage identified in static foreground region is then used for thresholding. Thresholding involves certain conditions that are to be checked to call this baggage as unattended.

Condition 1: Distance Rule- the owner which introduced the baggage in the scene is no longer in the neighborhood of the baggage.
Condition 2: Time Rule- The owner of the baggage is no longer in the neighborhood of the baggage for a certain time. This time threshold is an extension of condition 1
Fulfilling to these two conditions, the objected is termed as unattended baggage and a bounding box is created around the baggage, an alarm is triggered to alert the authorities.

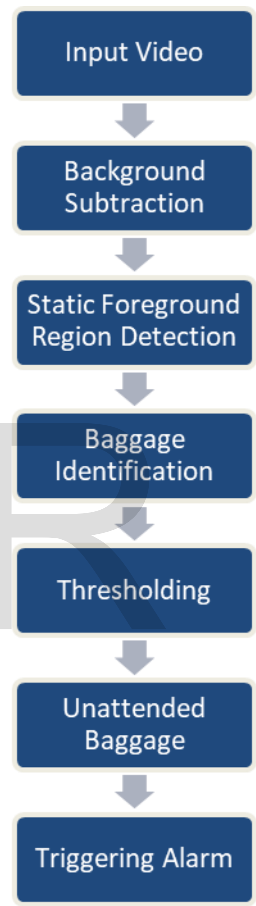


Fig. 2. Proposed System Diagram

3.1 Background Subtraction

The input video is divided into frames. The background subtraction is the step which involves division of regions or areas into background and foreground region and discarding the background. Continuous frames are checked to check the relative motion of each object present in the frame from previous frames. If the object shows a difference in the position, it is considered to be in motion for that sequence of frames during which object's positioning in the frame has changed. The two regions are encountered i.e. background (no motion or remained same during entire video) and foreground (not static). The system discards the background region and focusse3s on foreground region for further processing. To identify a region to be background or foreground, two approaches can be used:

First frame as background. Here the first frame of video is considered to be as background frame, and further frames are compared with this background frame and the difference between the two frames gives the foreground region, hence the common region being the background regions that have not been changed in the frames so far. This approach has good results but has a problem that it considers every object in first frame as background, if there is any moving object in first frame itself, the system fails to detect the background and foreground regions in first frame.

Gaussian Mixture Model. A Gaussian mixture is used. The model uses three Gaussians which store the intensity of each pixel. First Gaussian holds pertinent pixels, or pixels showing no change in intensity. Second Gaussian holds pixels with relative changes. Third Gaussian holds pixels with sudden changes or their intensities change rapidly. The first Gaussian represent a set of pixels representing background region, third Gaussian representing dynamic regions. The second Gaussian holds pixels that achieves a threshold from third Gaussian. The performance lacks as compared to first frame background method.

Analyzing the disadvantages and benefits of each approach. Our approach resolves disadvantages of both. The system creates two separate masks from first frame as background (mask 1) and Gaussian method (mask 2). These two masks are collaborated and bitwise AND operation is performed. The foreground region are the ones which both the masks have determined and are shown positive from both methods. Better object shape and accuracy is given by first mask and checking for objects is performed by Gaussian method. Hence the collaborative mask performs more effectively.

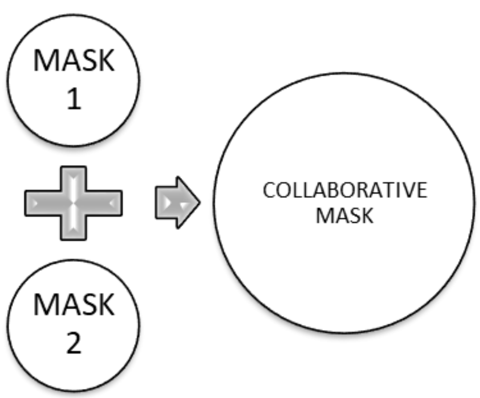


Fig. 3. Background Subtraction Mask (Bitwise- And)

3.2 Static Foreground Region Detection

After stage 1, the background regions from the video has been subtracted. To proceed further, the system divides foreground into two foreground types namely:

Dynamic Foreground. This foreground region consist of the regions containing objects that are in motion or that change their location with respect to several consecutive frames. They are not pertinent, and have dynamic nature. This region encloses the dynamic objects or regions where there is regular change in pixel intensities and are not a part of background region.

Static Foreground. The foreground region or objects that are introduced in the frame afterwards (not a part of background) and become static after certain time. Considering a person walking in the scenario along with a bag, till the time the person and bag remains in motion, it is considered dynamic foreground. As soon as the person leaves baggage and leaves. The region enclosing the bag is now static foreground region, as the region shows no motion & is not dynamic, but was introduced in frame as a foreground object.

The system detects static foreground region and are provided to next steps. Background subtraction and static foreground region detection works in collaboration.

3.3 Baggage Identification

The next step involves the identification of object in static foreground region. The system uses YOLO model [10] proposed by Redmon et al. [10] and configurations to identify the object. The model is customized to search for only human and baggage. The model identifies following baggage:

- Backpack
- Handbag
- Suitcase

The YOLO model [10] works on both images and videos. YOLO is a CNN (Convolutional Neural Network) that uses single neural network for the whole image, it then breaks the image into regions and find objects in those regions creating a bounding box around it along with probability. The identification time in video is not real time i.e. it takes time to analyze video and parse the results. To cover up this problem, the system performs identification on image which is near real-time. The moment a static foreground region is detected, the frame is captured and is passed to the YOLO model [10] for identification. YOLO [10] executes while the next steps for thresholding are done. The system parses the results. The result include type of baggage, if present, with a certainty score that how certain that baggage relates to the images of the baggage that the model is trained with along with the pixel co-ordinates of the box containing the baggage.

3.4 Thresholding

Thresholding is applied to samples tested positive by previous stages. The static foreground region from which baggage id detected is tested with the thresholding parameters of the unattended baggage. The system declares any baggage as unattended if it follows two rules:

Condition 1: Distance Rule- the owner which introduced the baggage in the scene is no longer in the neighborhood of the baggage.

Condition 2: Time Rule- The owner of the baggage is no longer in the neighborhood of the baggage for a certain time. This time threshold is an extension of condition 1

The distance is not constant for distance rule, it depends on the size of the baggage. The blob created for the baggage is

ISSN 2229-5518

scaled to twice its size, this distance around baggage in all direction give the distance threshold within which, if owner not found, distance rule is violated and condition 1 for unattended baggage is true.

The time rule depends on the video frame rate. The system accounts 190 frames and keep tracks. If the owner is no longer in the neighborhood of the baggage for 190 frames or more, the condition 2 for unattended baggage is true. The average video frame rate on tested dataset is 29 frames per second, making the time threshold to be 6.55 sec (approx.) in general.

The system also manages two time threshold parameters considering occlusion scenarios. In certain conditions, occlusion created in crowded scene violates these rules. To remove uncertainties occurred from occlusions in a crowded scene, we use two time threshold parameters given by Tian et al. [3]:

- Time Threshold (t1): Threshold t1 represents the time
- Occlusion time threshold (t2): Maximum Occlusion time specified.

If object is not detected by previous stages for a continuous time period greater than t2, we terminate the process and no alert is triggered. In case the object is detected, we check whether the current time since the region became stationary is greater than t1 in which case we trigger the alert. If object is not detected by previous stages for a continuous time period greater than Occlusion Time Threshold, we terminate the process and no alert is triggered. In case the object is detected, we check whether the current time since the region became stationary is greater than Static Time Threshold in which case we trigger the alert.

3.5 Triggering Alarm

If all the above stages mark object to be positive as a candidate for unattended baggage, the system will create a boundary box around the object and will trigger an alarm with object in the boundary box and the time for which it has remained unattended to alert the authorities. Ascertaining all the steps, the alert trigger event ascertains presence of an unattended baggage in the scene. The system detects and localizes the unattended baggage and creates a bounding box around the baggage with red alert text of unattended baggage.

Fig. 4. Unattended Baggage Detection System on PETS2006 Dataset [8]

4 Result & Discussion

The implementation of the system is done in Python programming language. Various algorithms helped to form the basis of the system including Gaussian Mixture Model [2], & YOLO [10]. The Laptop for implementation has core i7 x64 based processor @ 2.40 GHz, display of 1920 * 1080 & RAM of 16GB.

We identified different strategies used in previous work & their impact on performance measures. The methods used for background subtraction, & determining the object definition of unattended baggage in previous work incorporates various problems which hamper the performance of the system. The proposed system removes their problems that can accumulate more scenarios to detect unattended baggage. Previous works either considered first frame as background for background subtraction or used GMM (Gaussian Mixture Model) for background subtraction. Both these methods had their drawbacks. First frame as background considered every object in the first frame of the video to be a part of background, be it person or any other object. Consecutive comparison of frames gave false positives of static object. The GMM method did not have high accuracy as compared to first frame method. The proposed background subtraction used both these methods and created different masks from individual methods. These masks are then used to create a collaborative mask using BITWISE-AND operation. This removed the inconsistencies of first frame method and gave better results than GMM. The system also detects baggage after object identification which is bespoke to 3 objects: a) Handbag, b) Suitcase, & c) Backpack. The model can be customized to identify more number of objects as per the need of system. The occlusion scenarios which hampered performance in previous work is managed to give more accuracy and handle scenarios with occlusion. The basis parameters for thresholding are distance-rule and time-rule, which are satisfied and held true for a consecutive number of frames to declare an object as unattended baggage. The system parses the results in two phases:

A. Baggage Identification result: The positive candidates of static foreground region are examined for identification. The system uses a model which is customized to detect three type of objects:

ISSN 2229-5518

- Handbag,
- Suitcase &
- Backpack.

B. Unattended baggage detection event: The system after parsing a number of frames checks for the thresholding criteria, if the conditions are true for that object. The result shows the unattended object with a bounding box around it along with alarming text.

We used two datqasets namely: PETS2006 [8] conatining 7 video sequences, own dataset (containing 6 video sequences) & ABODA Dataset [11] (11 video sequences).

TABLE 1. RESULT ANALYSIS

S. No.	Dataset	Detection Score (%)
1.	PETS2006	85.71
2.	OWN DATASET	100
3.	ABODA	63.63

Detection Score consist of videos which showed true positives only. In ABODA dataset [11], some video sequences resulted in true positives and false positives. The detection score has not been considered for such cases having true positives & false positives. The detection score comprises results of those sequences which not only showed true positives but also no false positives.

The system performed well in different environment removing problems identified in previous system hence, incorporating more application areas of the system. We created our own dataset involving different type of scenario.The dataset has 6 video sequences at a frame rate of 29 frames per second. The Gaussian mask is prone is less efficiency if the sudden lightning changes occur. Bad lightning condition and sudden lightning changes may produce false positives and may degrade the system performance. Some scenarios from ABODA Dataset [11] experienced sudden lightning changes and system parsed false results which reduced by the detection score.

```
In [3]: runfile('C:/Users/Lenovo/.spyder-py3/new.py', wdir='C:/Users/Lenovo/.spyder-py3')
backpack  :  55.83977699279785  :  [131, 324, 180, 391]
--------------------------------
backpack  :  55.83977699279785  :  [131, 324, 180, 391]
--------------------------------
```

Fig. 5. Baggage Identification Result

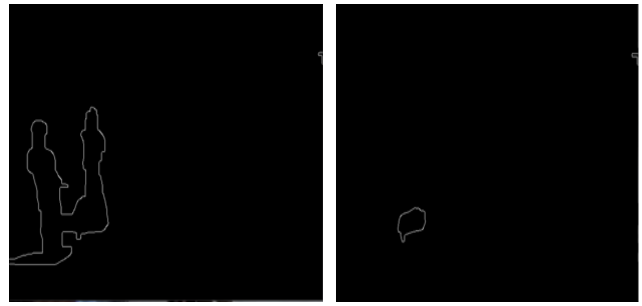
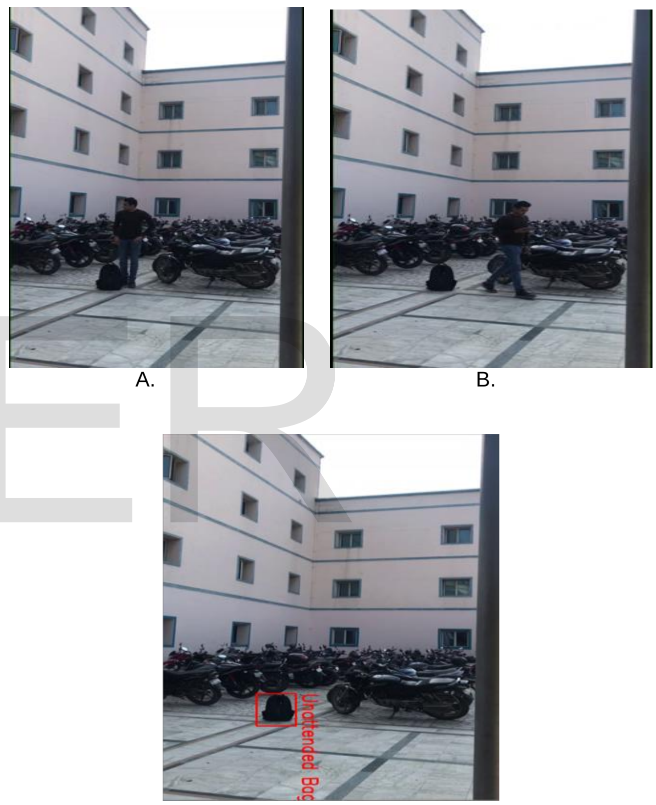


Fig. 6. Background Subtraction Result On ABODA DATASET



A. B. C.

Fig. 7. Unattended Baggage Detection Event –Own Dataset: A. Owner with Baggage, B. Owner going beyond neighbourhood of baggage, C. Unattended baggage alert

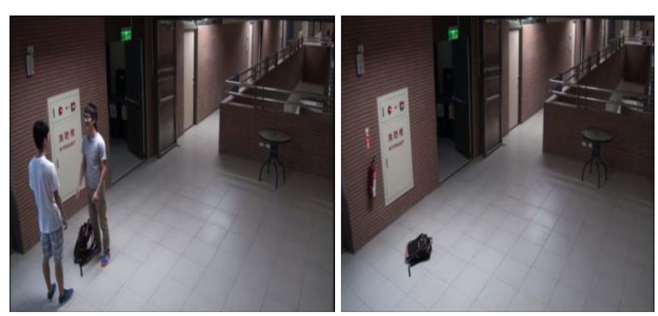


Fig. 6. ABODA S-1 Results

TABLE 2. SUMMARY OF LITERATURE

	Work Details		
	Method Used	**Problems Identified**	**Data Set & Performance**
[7]	Unattended luggage is identified by a cascade of CNN.	To calculate metrics for determining performance, all videos are manually annotated with ground-truth bounding boxes, since the data sets do not provide such annotations.	Data Set: PETS2006, Precision: 95.67%, Recall: 83.74%; Data Set: AVSS2007, Precision: 97.48%, Recall: 66.59%.
[6]	Uses Background subtraction & Static region detection.	-Relationship of object with owner is not taken into consideration. -In presence of occlusions, system does not consider object to be as unattended. -Sudden lightning changes in a challenging issue.	Data Set: PETS2006, Accuracy: 85.71%
[5]	A Finite-state-machine model is introduced to extract stationary foregrounds	-Does not accurately identifies owner of unattended baggage. -Temporary occlusion hinders the performance of system.	Data Set: AVSS2007, Precision: 1.0, Recall: 1.0
[4]	Abandoned baggage are first identified and localized by proposed foreground-mask sampling technique.	Too many abrupt changes in speed and direction of owner makes difficult for the motion prediction algorithm to successfully follow.	Data Set: AVSS2007, Precision: 1.0, Recall: 1.0.
[3]	Uses Gaussian mixture model for static foreground region detection.	-Does not classify between objects. -Sudden lighting changing is an issue. -Detects static person as a possible candidate.	Data Set: Big On-City Test, Detection Rate: 87.5%

5 CONCLUSION

We have proposed a system to detect the unattended baggage by capturing frames from video surveillance cameras in real time. Our system works well under occlusions & removes noisy or irrelevant pixels. We apply CNN procedure to identify those objects that are of interest to us, hence not considering human as a static object. The proposed system removes the drawbacks of previous works given in Table. 2 & we detect the unattended baggage more efficiently with improvised technical method combinations & directives. The futuristic expansion of work will involve more effective identification of owner to increase the accuracy of system. The system can be improved to identify the owner through key-pair with last intersection assumption & perform Selective Tracking. Also, the proposed system can be scaled up for more number of objects being incorporated.

REFERENCES

[1] C. Stauffer and W.E.L. Grimson, "Adaptive Background Mixture Models for Real-Time Tracking," Proc. Conf. Computer Vision and Pattern Recognition, vol. 2, pp. 246-252, June 1999.

[2] Dar-Shyang Lee, "Effective Gaussian Mixture Learning for Video Background Subtraction", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 27, NO. 5, MAY 2005.

[3] Ying-Li Tian, Rogerio Feris, Arun Hampapur, "Real-Time Detection of Abandoned and Removed Objects in Complex Environments", The Eighth International Workshop on Visual Surveillance -VS2008, Marseille, France, Oct 2008 inria-00325775.

[4] Huei-Hung Liao, Jing-Ying Chang, and Liang-Gee Chen, " A Localized Approach to Abandoned Luggage Detection with Foreground-Mask Sampling", IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance, Santa Fe, USA , 2008 pp. 132-139.

[5] Kevin Lin, Shen-Chi Chen, Chu-Song Chen, Daw-Tung Lin and Yi-Ping Hung, "Left-Luggage Detection from Finite-State-Machine Analysis in Static-Camera Videos", 22nd International Conference on Pattern Recognition, 2014, pp. 4600-4605.

[6] Pankit Arora, Akshath Dhar, Tushar Singh, Anju Mishra, "Abandoned Object Identification and Detection System for Railways in India", International Conference on Emerging Trends in Computing and Communication Technologies , Dehradun, India, 2017, pp. 1-5.

[7] Sorina Smeureanu, Radu Tudor Ionescu, "Real-Time Deep Learning Method for Abandoned Luggage Detection in Video", 26th European Signal Processing Conference (EUSIPCO), Rome, Italy, 2018, pp. 1775-1779.

[8] PETS 2006 Benchmark Data, http://www.cvg.rdg.ac.uk/PETS2006/data.html

[9] i-LIDS Dataset for AVSS 2007, ftp://motinas.elec.qmul.ac.uk/pub/iLids.

[10] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 2016, pp. 779-788.

[11] K. Lin, S. Chen, C. Chen, D. Lin and Y. Hung, "Abandoned Object Detection via Temporal Consistency Modeling and Back-Tracing Verification for Visual Surveillance," in *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 7, pp. 1359-1370, July 2015.