A MAJOR PROJECT REPORT

ON

DETECTION OF ABANDONED OBJECTS, IDENTIFICATION AND TRACKING OF THEIR OWNERS FOR EARLY WARNING OF TERROR ATTACKS

Submitted in partial fulfillment of the requirement for the degree of

BACHELOR OF TECHNOLOGY

IN

ELECTRONICS AND COMMUNICATION ENGINEERING

Submitted By:

ALINA HOTA (9918102023)

RADHIKA BERRY (9918102035)

SARAH KHAN (9918102124)

Under the Guidance Of:

DR. KAPIL DEV TYAGI

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY, NOIDA (U.P.)

May, 2022

CERTIFICATE

This is to certify that the major project report entitled, "**Detection of Abandoned Objects, Identification and Tracking of their Owners for Early Warning of Terror Attacks**" submitted by **Alina Hota(9918102023), Radhika Berry(9918102035) and Sarah Khan(9918102124)** in partial fulfillment of the requirements for the award of Bachelor of Technology Degree in **Electronics and Communication Engineering** of the Jaypee Institute of Information Technology, Noida is an authentic work carried out by them under my supervision and guidance. The matter embodied in this report is original and has not been submitted for the award of any other degree.

Signature: 

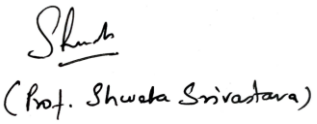
Name of Supervisor:

Dr. Kapil Dev Tyagi

ECE Department,

JIIT, Sec-128,

Noida-201304

Signature: 

Name of HoD:

Prof. Shweta Srivastava

Head of Department ECE,

JIIT, Sec-128,

Noida-201304

Dated: 14/05/2022

DECLARATION

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Place: Noida

Date: 14/05/2022

Name: Alina Hota

Enrollment: 9918102023

Name: Radhika Berry

Enrollment: 9918102035

Name: Sarah Khan

Enrollment: 9918102124

ABSTRACT

Global terrorism has become one of the biggest challenges of our time. The most common pattern of terror attacks is to hide explosives in ordinary-looking luggage/objects in heavy traffic public areas like railway stations, malls, metro stations, etc.

Identifying and monitoring public areas before a threat even poses itself, would protect more lives. As such, prevention is the central focus of counterterrorism strategies. We present a state-of-the-art surveillance system that can detect abandoned objects along with identification and tracking of their owners using video surveillance feeds in real- time, for expeditious alerts to notify officials.

Our model is capable of assigning different threat levels and once the object is flagged if the object was unattended or abandoned, an alarm can be raised thereby notifying concerned authorities immediately.

The algorithm is also immune to illumination changes and performs well with low quality video feeds. This makes our proposed method robust in both day and night where sufficient lighting is not possible.

ACKNOWLEDGEMENTS

The completion of any inter-disciplinary project depends upon cooperation, coordination, and combined efforts of several sources of knowledge. We are grateful to **Dr. Kapil Dev Tyagi** for his willingness to give us valuable advice and direction whenever we approached him with any problem. We are thankful to him for providing us with immense guidance for this project. We would also like to thank our college authorities for giving us the opportunity to pursue our project in this field.

We are also grateful to all our friends, who have graciously helped us with moral support and valuable suggestions. Finally, we would like to extract our gratitude to all those people who have directly or indirectly helped us in process and contributed toward this work.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

Global rates of terrorism have skyrocketed since 9/11 and is one of the major threats the world is facing today. From terror incidents in public areas as the November 26th attacks in Mumbai to 2008 Delhi market bombings, terrorism is now perceived as a major threat to India. In the previous years, India has been a victim of more than 1500 attacks (highest number of blasts in the world in the years of 2016 & 2017), according to the National Bomb Data Centre.

TABLE I

YEAR WISE BLASTS IN INDIA

Year	Blasts	Killed	Injured	Total Casualties
2016	337	112	479	591
2015	268	117	457	574
2014	190	75	295	370
2013	283	130	466	596
2012	365	113	419	532

Even if the aftermath of an attack is productively managed, it is universally more desirable to prevent such an attack entirely. Identifying and monitoring public areas before a threat even poses itself, would protect more lives. As such, prevention is the central focus of counter-terrorism strategies. Terrorists often resort to using explosive devices hidden in ordinary looking objects in high traffic public places like railway stations, metro stations, shopping

malls and other and public places. It is difficult to distinguish a bomber or a suspicious object in a crowded social setting.

The use of CCTV surveillance for combating terrorism is a rapidly growing phenomenon. There is a rise in the number of cases where officials and the communities they represent are assigned with viewing large stores of video footage in an effort to track down culprits of attacks or other threats to public safety. However, the conventional way of predicting potential danger does not guarantee accuracy and suspicious objects may go unnoticed. The lack of significant amount of manpower can also pose a greater risk.

In this paper, a surveillance algorithm is proposed to automatically detect suspicious objects, identify and track its' real owners by analysing the captured video frames. Our model is capable of assigning different threat levels and once the object is flagged if the object was unattended or abandoned, an alarm can be raised thereby notifying concerned authorities immediately. Apart from this, the algorithm is also immune to changes in light intensity and performs well with low quality video feeds. This makes our proposed method robust in both day and night where sufficient lighting is not possible.

This intelligent surveillance model is highly optimized for use in low energy and low power embedded systems and can be integrated with existing CCTV systems, with the help of Single Board Computers. A lot of experiments, although very accurate in detecting these abandoned objects, require high powered systems to run. This is extremely difficult since every public place requires multiple cameras to cover the entire place. The real-time solution proposed in this paper can easily be uploaded on an embedded chip and attached to existing CCTV cameras without incurring huge costs in upgrading these systems.

Our proposed model can be divided into a pipeline which is further explained in Section 3. The pipeline also allows the frames to be continuously processed instead of waiting for the whole process to be done on a single frame and then moving on to the next. This further helps in increasing the efficiency of our algorithm to be deployed in low powered single board computers.

Pipeline of the Proposed Solution:

Stage 1: Background is removed to ignore the static objects that are present in the view by default.

Stage 2: Classifying an object as stationary or moving.

Stage 3: Classifying of a static object as human or luggage and if it is luggage, then classify it as not unattended, unattended or abandoned.

Stage 4: Backtrack to identify the owner and then track him in the following frames to find her/his current location or direction in which she/he left the view for the authorities to track.

CHAPTER 2
LITERATURE SURVEY

2.1 Robust Abandoned Object Detection Using Dual Foregrounds, 2007.

F. Porikli, Y. Ivanov and T. Haga

As an alternative to the tracking-based approaches that heavily depend on accurate detection of moving objects, which often fail for crowded scenarios, a pixelwise method that employs dual foregrounds to extract temporally static image regions is presented. Depending on the application, these regions indicate objects that do not constitute the original background but were brought into the scene at a subsequent time, such as abandoned and removed items, illegally parked vehicles. Separate long- and short-term backgrounds that are implemented as pixelwise multivariate Gaussian models were constructed. Background parameters are adapted online using a Bayesian update mechanism imposed at different learning rates. By comparing each frame with these models, two foregrounds are estimated. Evidence scores at each pixel by applying a set of hypotheses on the foreground responses were inferred, and then aggregate the evidence in time to provide temporal consistency. Unlike optical flow-based approaches that smear boundaries, this method can accurately segment out objects even if they are fully occluded.

2.2 An edge-based method for effective abandoned luggage detection in complex surveillance videos, 2017.

Dahi, M. Chikr El Mezouar, N. Taleb and M. Elbahri

Recent terrorist attacks in major cities around the world have brought many casualties among innocent citizens. One potential threat is represented by abandoned luggage items (that could contain bombs or biological warfare) in public areas. In this paper, an approach for real-time automatic detection of abandoned luggage in video captured by surveillance cameras is described. The approach is comprised of two stages: (i) static object detection based on

background subtraction and motion estimation and (ii) abandoned luggage recognition based on a cascade of convolutional neural networks (CNN). To train the neural networks, two types of examples were provided: images collected from the Internet and realistic examples generated by imposing various suitcases and bags over the scene's background. Empirical results demonstrating that our approach yields better performance than a strong CNN baseline method were presented.

2.3 Multiple Object Tracking with Motion and Appearance, 2019

Cues Weiqiang Li, Jiatong Mu, Guizhong Liu

Due to better video quality and higher frame rate, the performance of multiple objects tracking issues has been greatly improved in recent years. However, in real application scenarios, camera motion and noisy per frame detection results degrade the performance of trackers significantly. High-speed and high-quality multiple object trackers are still in urgent demand. In this paper, a new multiple object tracker following the popular tracking-by-detection scheme is proposed. The camera motion problem is tackled with an optical flow network and utilize an auxiliary tracker to deal with the missing detection problem. Both the appearance and motion information to improve the matching quality is used. The experimental results on the VisDrone-MOT dataset show that this approach can improve the performance of multiple objects tracking significantly while achieving a high efficiency.

2.4 An Abandoned Object Detection System Based on Dual Background Segmentation, 2009

A. Singh; S. Sawan; M. Hanmandlu; V.K. Madasu; B.C. Lovell

An abandoned object detection system is presented and evaluated using benchmark datasets. The detection is based on a simple mathematical model and works efficiently at QVGA resolution at which most CCTV cameras operate. The pre-processing involves a dual-time

background subtraction algorithm which dynamically updates two sets of background, one after a very short interval (less than half a second) and the other after a relatively longer duration. The framework of the proposed algorithm is based on the approximate median model. An algorithm for tracking of abandoned objects even under occlusion is also proposed. Results show that the system is robust to variations in lighting conditions and the number of people in the scene. In addition, the system is simple and computationally less intensive as it avoids the use of expensive filters while achieving better detection results.

2.5 Abandoned object detection in complicated environments, 2013

Kahlil Muchtar, Chih-Yang Lin, Li-Wei Kang, Chia-Hung Yeh

In video surveillance, tracking-based approaches are very popular especially for detecting abandoned objects in public areas. Once the object has been tracked, the object status can be further classified as removed or abandoned. However, some shortcomings were found on tracking-based approaches, e.g., illumination changes and occlusion. Therefore, in this paper, an alternative approach to detect abandoned objects is proposed by incorporating background modeling and Markov model. In addition, the shadow removal is employed to rectify detected objects and obtain more accurate results. The experimental results show that the proposed scheme is better than other methods in terms of accuracy and correctness.

CHAPTER 3

PROPOSED METHODOLOGY

3.1 OUR APPROACH

Our framework consists of the following pipeline, each of the step utilizes the results produced by the previous one which requires low-level processing of image frames.

A) Preprocessing
B) Foreground Extraction
C) Static Object Detection
D) Static Object Classification
E) Identification of the owner
F) Tracking owner in the current frame
G) Classifying alert level

Flow chart in Fig.3.1 shows the proposed algorithm. The system receives a stream of images at its input, performs an analysis of every frame, and outputs the results in the form of rapid alerts.

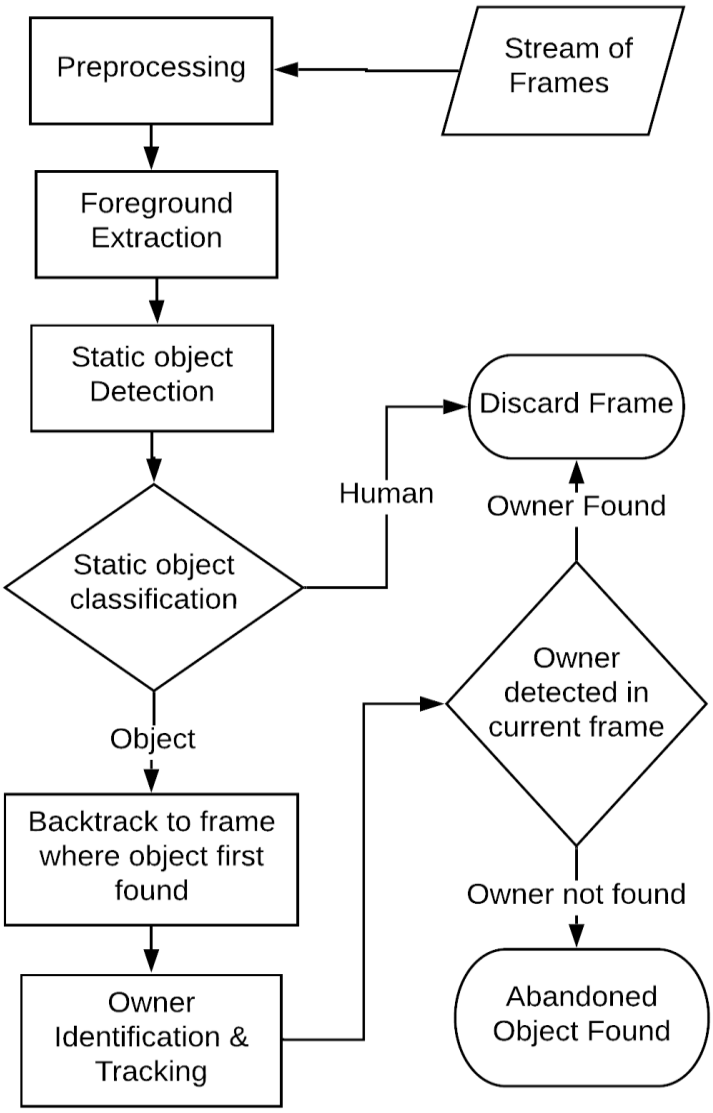


Fig. 3.1 Proposed Methodology

3.1.1 PREPROCESSING

Pixel noise is introduced due to changes in illumination or changes in the positioning of objects or humans. The ongoing video footage from our surveillance system is converted into frames which are then sent for preprocessing in order to remove noise. Frames are converted into grayscale in order to process in a single channel. Processing on a single channel

preserves the data we require and reduces the time complexity required for all the steps.

Blurring operation is then performed using **Gaussian Blur and Gaussian filtering** is done by convolving each point in the input array with a Gaussian kernel and then summing them all to produce the output array as shown in Fig. 3.4. The formula of a Gaussian function in one dimension is:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

where x is the distance from the origin in the horizontal axis, y is the distance from the origin in the vertical axis, and σ is the standard deviation of the Gaussian distribution.

Fig. 3.2 Background Frame *F'*

Fig. 3.3 Sample input to preprocess stage (Section 3.1.1)

Fig. 3.4 Output of the preprocess stage (Section 3.1.1)

3.1.2 FOREGROUND EXTRACTION

Foreground objects refer to moving objects or newly introduced objects in the view of the video which may or may not become static after a predefined time interval. These include both people and luggage. This step imitates the human ability to concentrate exclusively on objects of interest. Frame differencing between frames is done to detect changes in the sequence of frames and to extract foreground objects.

For this, we take a frame F*'* with no one present as our background model as shown in Fig. 3.2, and perform differencing of each incoming frame *F with* the background frame *F'*. This difference gives us a response on each pixel position where there is a new object in the view. The response is shown in Fig.3.5. The resulting response is still susceptible to light changes, shadows and noise. To counter this, we perform **Canny edge detection** [12] on the resulting response. This allows us to ignore light changes and shadows.

The Canny edge detection algorithm is composed of 5 steps:

1. Noise reduction (using Gaussian Blur)
2. Gradient calculation:

The Gradient calculation step detects the edge intensity and direction by calculating the gradient of the image using edge detection operators. It can be implemented by convolving *I* (Intensity) with Sobel kernels *Kx* and *Ky*, respectively:

$$K_x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, K_y = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix}.$$

Sobel filters for both direction (horizontal and vertical)

$$|G| = \sqrt{I_x^2 + I_y^2},$$
$$\theta(x, y) = arctan\left(\frac{I_y}{I_x}\right)$$

Gradient intensity and Edge direction

3. Non-maximum suppression: Ideally, the final image should have thin edges. Thus, we must perform non-maximum suppression to thin out the edges. The algorithm goes through

all the points on the gradient intensity matrix and finds the pixels with the maximum value in the edge directions.

4. Double Threshold: The double threshold step aims at identifying 3 kinds of pixels: strong, weak, and non-relevant. Strong pixels are pixels that have an intensity so high that we are sure they contribute to the final edge.

Weak pixels are pixels that have an intensity value that is not enough to be considered as strong ones, but yet not small enough to be considered as non-relevant for the edge detection.

Other pixels are considered as non-relevant for the edge.

5. Edge tracking by Hysteresis: Based on the threshold results, the hysteresis consists of transforming weak pixels into strong ones, if and only if at least one of the pixels around the one being processed is a strong one.

We could use any other edge detection algorithm but since that is not the aim of the paper and Canny edge detection does not become a bottleneck in our pipeline, we used that in our case. Then, we perform morphological operations in order to remove noise from detected edges.

False positives and small gaps are excluded with the help of Gaussian smoothing and Morphological Closing. The result has blobs signifying all the foreground object in a frame.

Fig. 3.5 Output of frame differencing of frame from Fig.3.2
with Fig.3.3 described in section 3.1.2

Fig.3.6 Output of section 3.1.2 after Canny edge detection

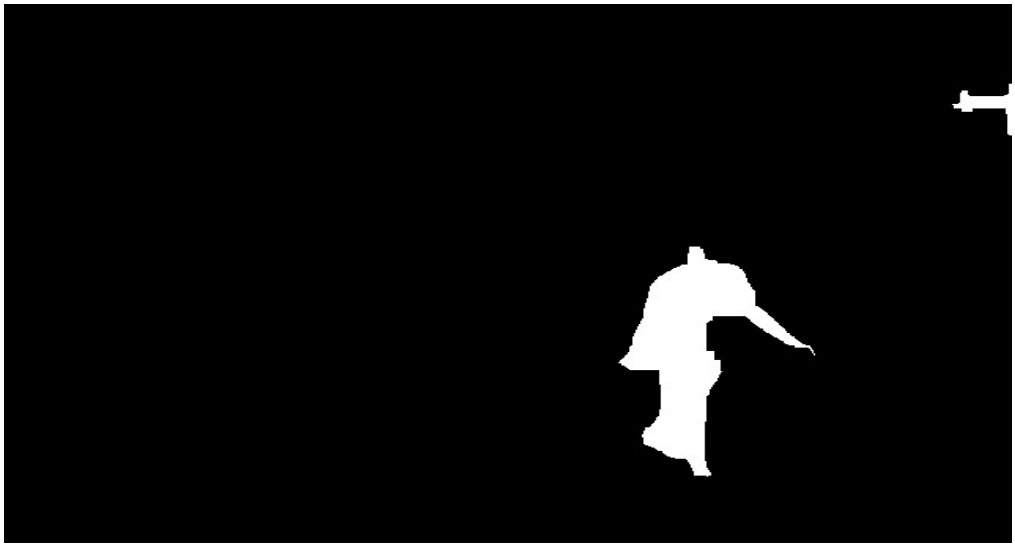


Fig.3.7 Output of section 3.1.2 after Morphological operations on Fig3.6
The shadows of the human in the frame that were visible in the frame difference response in Fig.3.5 have been removed

3.1.3 STATIC OBJECT DETECTION

Contours of all moving objects in the current frame are extracted from the previous steps. These contours might have small changes due to light, shadows, and other noise in the view.

For this we employ an IOU (intersection over union) based mechanism to determine if an object is still stationary or it has moved. For two consecutive frames F_1 and F_2, we construct bounding boxes on all the extracted foreground objects. Let's say we have one object with its bounding box in F_1 as B_1 and corresponding B_2 in F_2. There are two cases possible, either this object is stationary or moving.

$$IOU = \frac{Area\ of\ Intersection\ of\ two\ boxes}{Area\ of\ Union\ of\ two\ boxes}$$

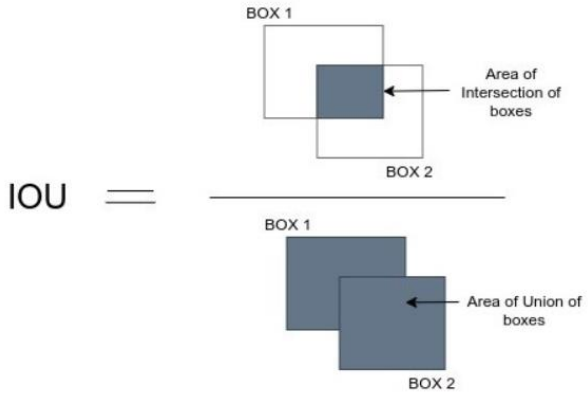


Fig.3.8 Diagrammatic representation of the formula to calculate IOU

Case 1 (Stationary):

If the object is stationary, the IOU of B_1 and B_2 should be 1, but because of noise and other morphological operations that we perform, some changes are introduced in the contours which cause slight changes in the bounding box and the resulting IOU although not 1 but is very high. We use a threshold T of 0.8. If the IOU of box B1 and box B2 in frames F_1 and F_2 respectively is greater than the threshold T, then we count this as the same object.

Case 2 (Moving):

If the object is in motion, the IOU of B_1 and B_2 should be very small, and tending towards 0. Based on the speed of the movement in consecutive frames, this IOU can be varied significantly. The high threshold T helps here too since all values of IOU between 0 and T are considered as objects in motion.

Definition 1: An object is classified as stationary object if for each consecutive frame F_1 to F_n bounding box of the foreground object has an IOU greater than T with the bounding box B_0 in frame F_0, where F_0 is the frame where that object was first encountered at that location. Here n is the minimum number of frames, in which the object must remain stationary to be classified as stationary object.

We define stationary object for our model in 1. The value of n can be chosen according to individual requirements, and it constitutes the minimum time required for an object to be stationary for it to be classified as stationary.

Below is the depiction of 2 cases for static object detection.

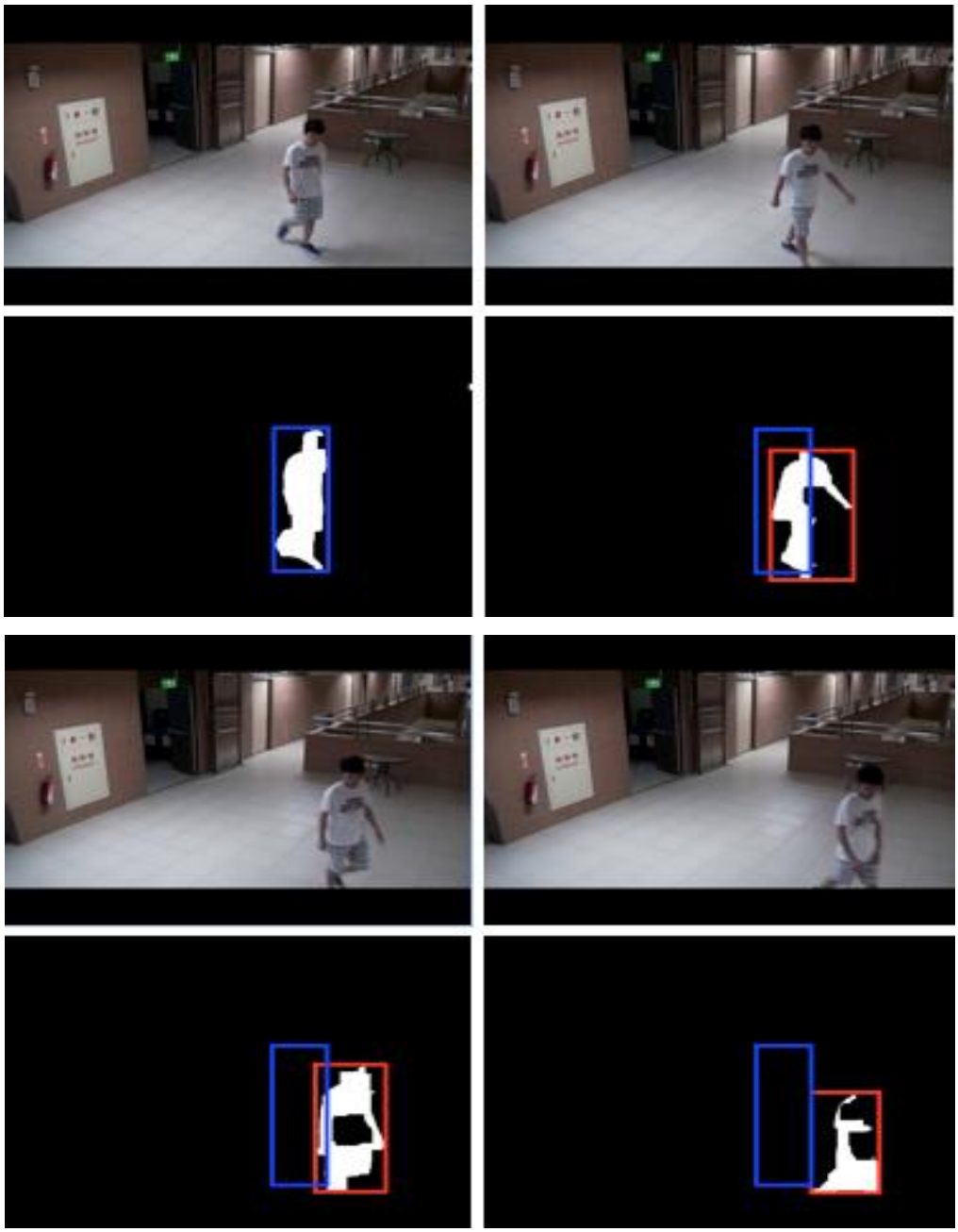


Fig.3.9 When an object is moving, the IOU with the first detection (First image in top row) reduces slowly to 0

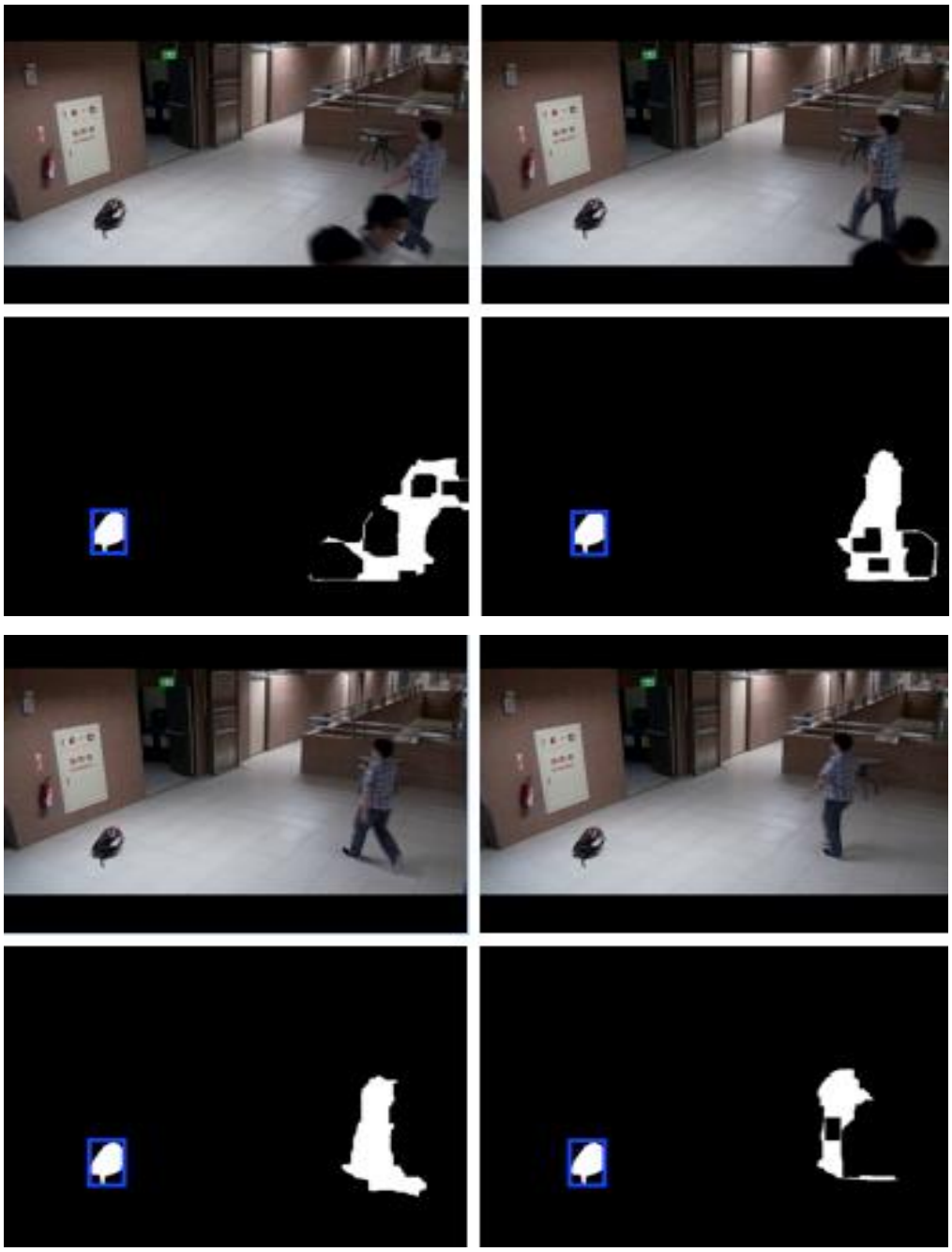


Fig.3.10 When the object is stationary, the IOU remains close to or equal to 1 with the first detection of an object (first image in top row)

Handling occlusions: It is pretty common in public places that the object may get occluded for a short period of time by other moving people in the view. This will cause the object to vanish for a few frames and then reappear at the same location. To handle this, we use buffer frames *b*. If an object that we are tracking is not seen in consecutive frames for more than *b*

frames, we assume that the object has moved. If it reappears before *b* frames, then we start counting the frames again. This works exceptionally well in crowded places where the view is blocked multiple times, but the object has not moved.

Fig. 3.11 Case of Occlusion by moving people in view of the object

3.1.4 STATIC OBJECT CLASSIFICATION

After we have localized a static object from the previous steps, a classification approach is used to differentiate between humans and non-human static objects. A **MobileNet network** from [13] is used. It performs exceptionally well with real- time predictions in limited hardware conditions. The number of parameters is reduced with this network while maintaining great accuracy. Previous studies have shown that MobileNet only needs 1/33 of the parameters of VGG-16 (another popular alternative from [14]) to achieve the same classification accuracy as in ImageNet-1000 classification tasks.

MobileNet is a streamlined architecture that constructs lightweight deep convolutional neural networks using depth wise separable convolutions and provides an efficient model for mobile

and embedded vision applications. MobileNet's structure is based on depth wise separable filters.

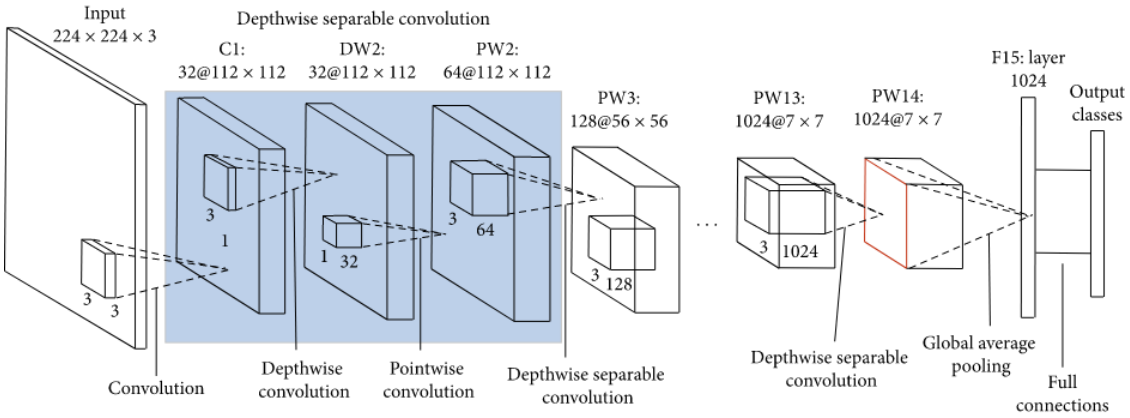


Fig. 3.12 Architecture of MobileNet

In our experiments, with sufficient training, we were able to achieve same levels of accuracy in Mobilenet classifier to the VGG-16 network, with substantial improvements on speed. This was possible since our classes our extremely simplified to human and non-human. For a large number of classes, the accuracy trade off might become large, but it is not relevant to this paper. On Nvidia Geforce 1060 GPU, we were able to obtain 140 fps speed compared to 22 fps of VGG-16.

Another reason for use of a small network like Mobilenet is that it is only 17 MB in size compared to around 500 MB for the VGG-16 based classifier. This is especially useful when using single board computers since they have very limited RAM and storage capacity, often of the order of 1 GB. A small sized network like Mobilenet shines exceptionally well in these limited hardware systems.

Fig. 3.13 Object flagged as Static

3.1.5 IDENTIFICATION OF OWNER

For this, we need to go back to the frame where we first encountered this static object. Fig 3.14 shows the abandoned object, and shows the first frame where it was detected as static object but it is not the frame where the static object was detected. So, we go back frame by frame and get IOU of objects in the frame F_{t-1}, where F_t is the first frame where we encountered this object with the static object in frame F_t. The IOU will slowly decrease from 1 to 0 as we move back. We take frame F_{t-t}*' where* the IOU is 0.6. The frame F_{t-t}*' will* be the one where the object has started entering but has not yet been left by the owner and would thus be closest to the actual owner, which is shown in Fig 3.15. We then perform human detection using **YOLO version** 3 (tiny) proposed in [16]. We use this since it is extremely fast (over 200 fps) and very compact (¡35 MB). Now that we have the bounding boxes of all the humans in frame F_{t-t}*',* we just need to find the closest human to the object. Fig 3.15 is the frame F_{t-t}*'* for object detected in Fig 3.14. With the two humans present, we get two bounding boxes and the closest to the stationary object is the owner.

Identification of owner once a static object has been detected is depicted below.

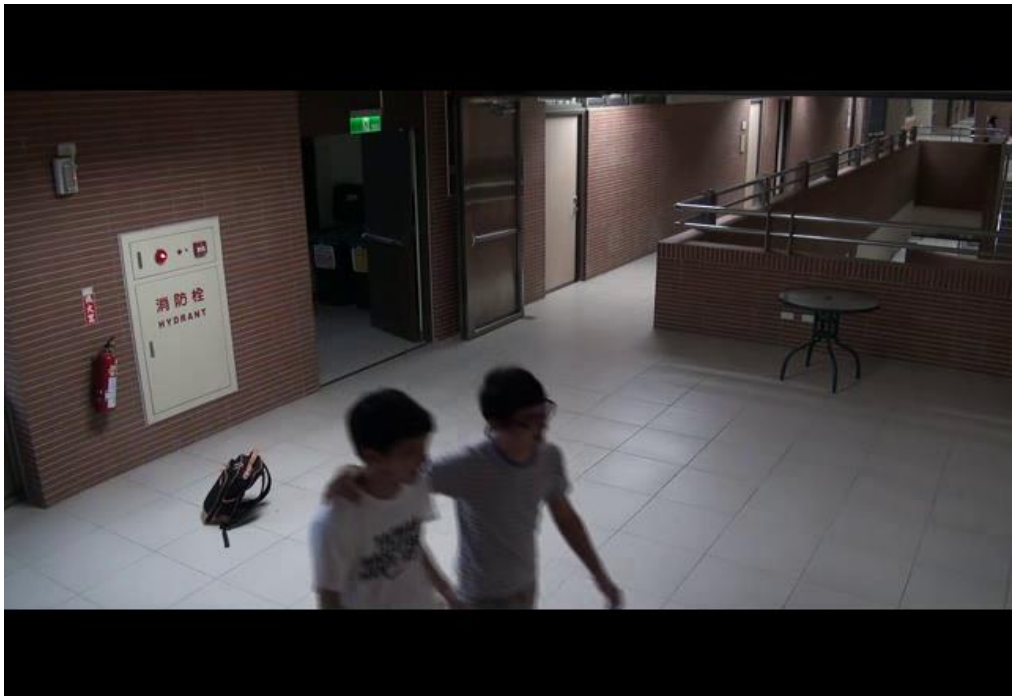


Fig. 3.14 Frame where the object is first detected (Frame F_t) as static

Fig 3.15 Frame $F_{t-t'}$ where the object was entering its current position with **IOU 0.6** and two humans in are detected in the frame

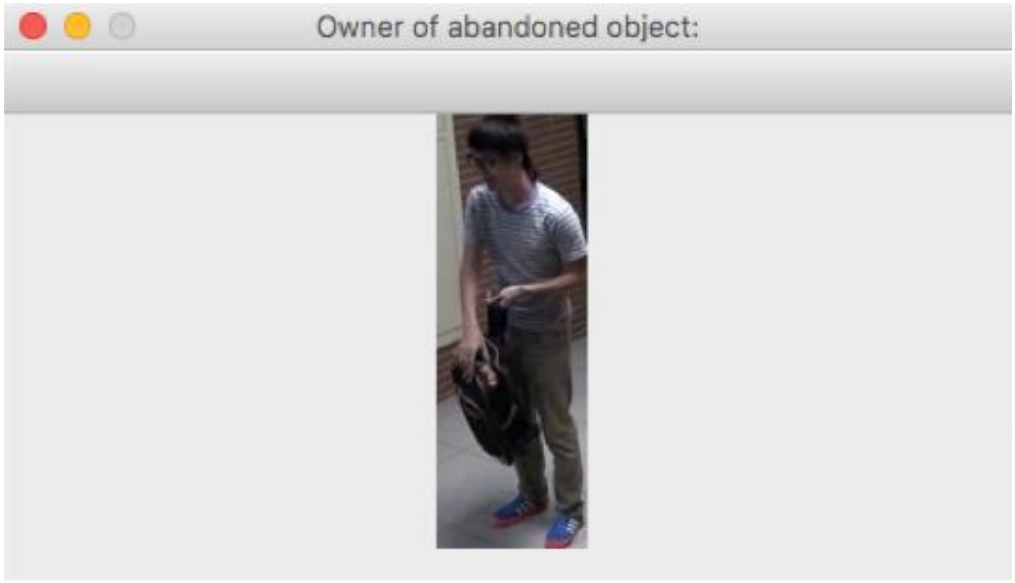


Fig 3.16 Two humans were detected and the closer one is flagged as owner

The YOLO algorithm divides the image into N grids, each of which has an equal dimensional region of SxS. Each of these N grids is in charge of detecting and localizing the object it contains.

These grids, in turn, predict B bounding box coordinates relative to their cell coordinates, as well as the object label and probability of the object being present in the cell.

This method significantly reduces computation because both detection and recognition are handled by cells from the image, but—It generates many duplicate predictions as a result of multiple cells predicting the same object with different bounding box predictions.

Non Maximal Suppression is used to deal with this issue. YOLO suppresses all bounding boxes with lower probability scores in Non Maximal Suppression. YOLO's architecture has a total of 24 convolutional layers with 2 fully connected layers at the end. YOLO accomplishes this by first examining the probability scores associated with each decision and selecting the one with the highest probability. The bounding boxes with the greatest Intersection over Union with the current high probability bounding box are then suppressed. This process is repeated until the desired bounding boxes are obtained.

3.1.6 TRACKING OWNER IN CURRENT FRAME

We found a static object *O* which was first encountered in frame F_t and the owner in frame $F_{t-t'}$, but we would also like to know where the owner went or if he/she is still in the frame to classify the static object as abandoned or not. For this we employ a tracker using Kalman Filter from [15]. Kalman Filter from [15] has a unique ability to not only improve the accuracy of the existing detection but also predict the future position of the object based on its past movement. We start from frame $F_{t-t'}$ where we had identified the humans, and move forward.

For each frame F_n, we perform human detection described in previous Sub-section 3.1.5. For each frame we will have bounding boxes for all the humans. We can now track it using IOU and Kalman Filter. Since each consecutive frame is only a fraction of a second ahead of previous frame, the bounding boxes of same human will have very high IOU. This fact can be used to very efficiently track the human from frame $F_{t-t'}$ to the current frame. This method is really good but fails if the human detection in even a single frame fails or of for some reason the human is occluded for a short period of time, say passing behind a pillar. The Kalman Filter helps here by predicting the future position in the missing frames. It accounts for all the missing boxes for a given human and helps in tracking even in cases of occlusions. This is especially helpful when two humans cross each other in a crowded situation. In that case also, since when they pass, a simple IOU based tracker might get confused as to which human was where, but since Kalman filter tracks using the previous motion, it helps to assign correct labels once the humans have crossed each other since their trajectory of movement will not change. Using these, we can track the owner of the object in each subsequent frame till he/she is in the view of the camera.

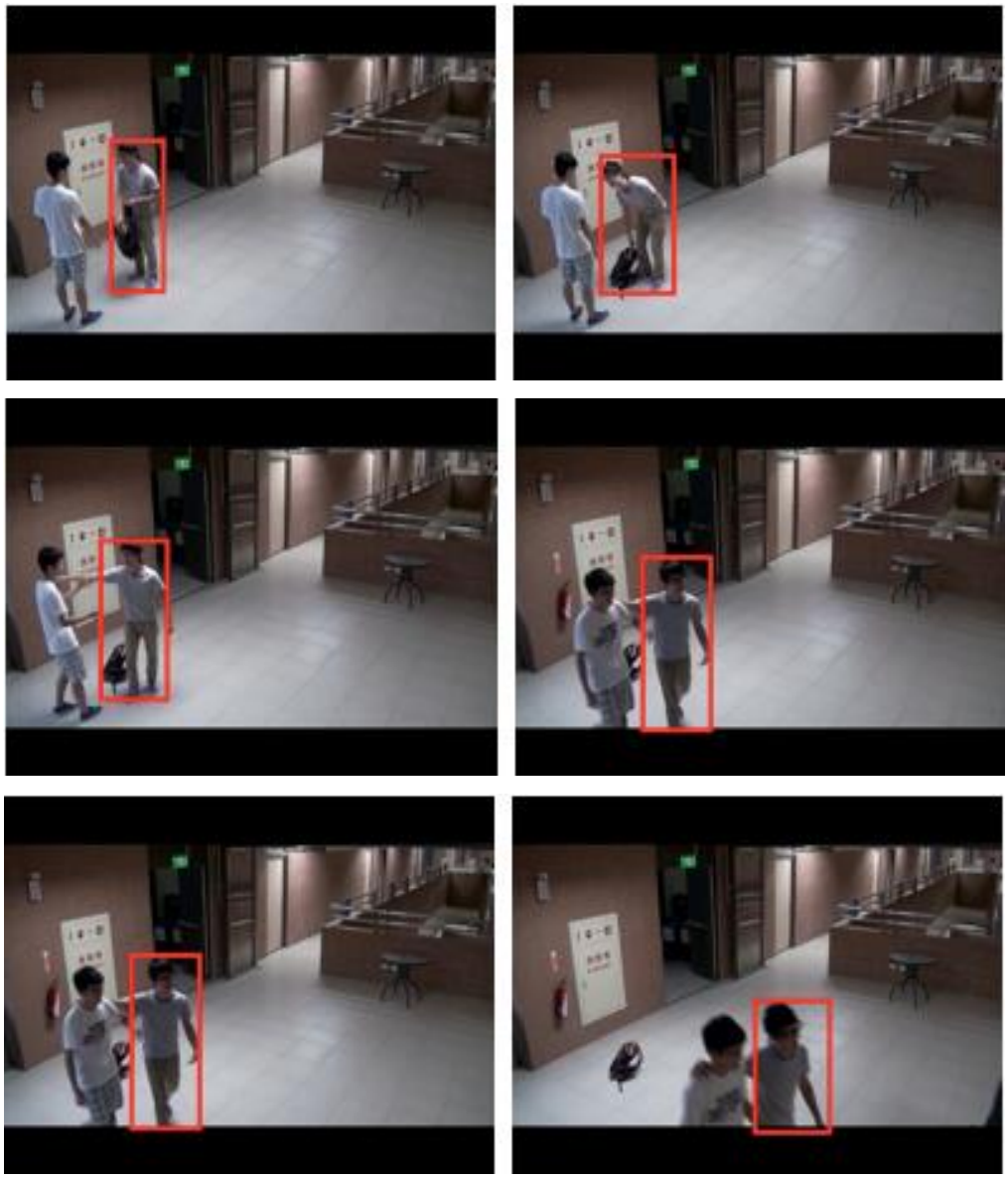


Fig. 3.17 Tracking of owner in frames after $F_{t-t'}$ to current frame

3.1.7 CLASSIFIYING OF ALERT LEVEL

There are three cases for the owner once an object has been detected as static:

Case 1: The owner is right next to the static object and has not left it unattended or abandoned.

Case 2: The owner is still in the view of the frame but is not near the static object, signaling that the object has been left unattended.

Case 3: The owner has left the view of the camera and has abandoned the object in which case the algorithm can raise the alarm and signal the security of possible threat for inspection.

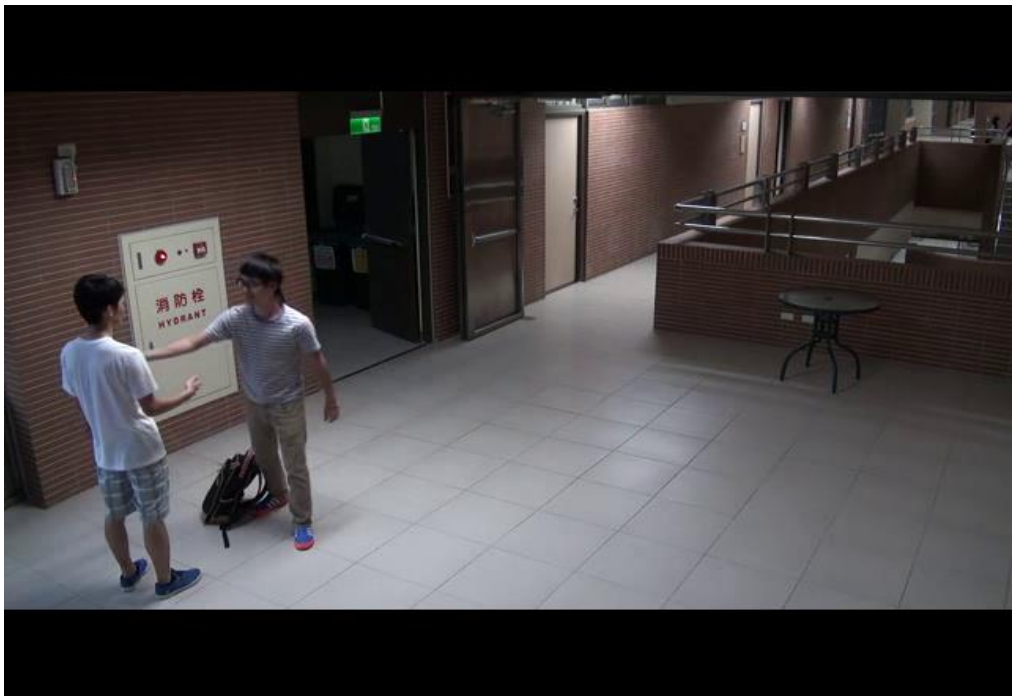


Fig 3.18 Case 1 (Object is attended), No alarm raised.

Fig. 3.19 Case 2, Owner is still in frame (not near the object)

Fig. 3.20 Case 3 encountered; owner has left the view.

Object flagged as abandoned

3.2 TECHNOLOGY STACK

3.2.1 PYTHON (VERSION 3)

Python is a high-level programing language (an open source software). It is an object-oriented language created by Guido van Rossum. It provides us with constructs and libraries that help in programming on both small and big scales, which help programmers to write readable codes.

Some advantages of using Python are as follows:

1. Support of Libraries: Python helps in limiting the length of codes that are to be written with the help of large standard libraries, which are scripted into it. This makes the development process efficient and less time consuming.
2. Productivity and Speed: Python's unit testing framework and its strong integration features enhance its productivity and speed.

3. Open Source and Community: Python is developed under an open-sourced license, it is free to use for both educational and commercial purposes. It also empowers community development.
4. Data structures (user-friendly): Python provides us with built in dictionary and list data structures. These can be used to create faster runtime structures. These also help in reducing the lines of code and increase time efficiency.
5. Dynamic: Python is a dynamic language and assigns data types to variables based on assignment of values. It also performs an automated memory management.
6. Easy to understand: Python is easy to learn and a huge set of resources are available for the same.

3.2.2 PYTHON PACKAGES

The following python packages are used for the implementation of this project.

TABLE II

PYTHON PACKAGES

Packages	Description
1. OPENCV	Open Source Computer Vision Library is a library developed by the Intel Corporation in 1999. It consists of functions customized for real time computer vision. It gives the developers an access to pre-defined functions that can be easily incorporated in their codes. The OpenCV has been developed using the language C++ but it supports a wide variety of languages that include C++, Python, MATLAB and Java. It is a collection of computer vision and machine learning algorithms that are used in wide variety of applications like facial detection, pedestrian density, traffic lights, object detection, human pose estimation etc. The comprehensive step for the successful incorporation of the library function in the code and running it, the following steps must be followed: loading of pretrained models, pre-processing of the inputs and handling the network outputs.
2. NUMPY	NumPy basically started off as a community project. But the major

	contribution credits can be given to Jim Hugunin. It is more of a mathematical solutions (numerical computing) library used by the programs to solve complex multi-dimension arrays. It is just like MATLAB. But NumPy happens to be a more modern version of MATLAB integrated in python. Some of the application fields of NumPy can be signal processing, image processing, 3-D visualization, cognitive psychology, geographic processing and interactive computing. It has the capability to transform the power of languages like C and Fortran into Python.
3. MATH	Apart from the numerical computing there are various mathematical computations that may be needed in programming especially in machine learning. For example, computations of trigonometric functions, power and logarithmic functions, angular conversion, hyperbolic functions, truncation, square root etc. This library gives access to all these integrated predefined functions for the above applications. The math package in Python is basically the same as the C library module but with thin wrappers around it.

4. OS	It is the collection of functions that allows the developer to interface with the operating system on which the Python is being run. The functions can be used to enable and manipulate the paths of the execution log. The arguments can be of byte or string type so as the function can return the same data type of the argument passed. The functions are bound to give OS error if the path is not recognized by the OS even if the arguments passed are of correct syntax. Basically, the functionality can of the package can be said to be of reading, writing and improvising of paths in the operating system.
5. IMUTILIS	Package that supports basic yet crucial image processing tasks. Some of the tasks include conversion of RGB conversions, skeletonization, resizing, rotation, translation etc. This is used in the second step of code implementation that is pre-processing of the input. Every Model has convention set for accepting the input in a certain way, this package supports interconversions of these basic attributes so as to ensure smooth functioning of the code.

CHAPTER 4

EXPERIMENTS & RESULTS

4.1 RESULTS AND VALIDATION

Experiments were conducted to evaluate the performance of the proposed algorithm in detection of abandoned objects, thereby validating the robustness and correctness of the system. The algorithm was implemented in Python 3, programmed using a single thread on two different system configurations.

1. 3.5 GHz Intel Core-i7 processor, 32 GB RAM and Nvidia 1060 GPU.
2. 1.3 GHz Intel Core-i5 processor computer with 4 GB without GPU.

We evaluated our algorithm on ABODA[8], AVSS 2007[9] and PETS 2006[10]. These datasets include indoor and outdoor environments, variations in natural light, crowded surroundings and changes in illumination. Some of the metrics used for quantitative evaluation are as follows:

1. Recall (percentage of detected events) : $\frac{TP}{TP+FN}$

2. Precision (percentage of true alarms) : $\frac{TP}{TP+FP}$

3. F- Score : $\frac{2 * Recall * Precision}{Recall + Precision}$

4. Processing Speed in Frames per second (fps): Number of consecutive images that can be processed each second.

A true positive (TP) represents correct detection of an abandoned object; a false negative (FN) refers to the missed detection; a false positive (FP) refers to the classification of a non-abandoned object as abandoned and a true negative (TN) represents an abandoned object mistakenly classified as non- abandoned.

Results of our method compared to Lin et al.[6], Filonenko et al.[7], Ilias et al.[4], Shyam et al.[11] and Agarwal et al.[5] using ABODA dataset are depicted in Table III. Method [6] gave a false positive for the night vision in Video 5.

Method [6] detects false positives in Video 11 due to partial occlusion of small objects. Method [7] uses a reference background in their method and their framework is unable to adjust to the illumination change, thereby neglecting detection of abandoned objects in Videos 5, 6 and 7.

The comparison of our model using AVSS 2007 and PETS 2006 datasets with several recently published works (Porikli et al.[17], Liao et al.[18], Li et al.[19], Tian et al.[20]) are presented in Table IV.

3.2 COMPARISON WITH EXISTING WORK

The bottle neck in our approach are the Mobilenet [13] classifier and YOLOv3 (tiny) [16] from section 3.1.4 and section 3.1.5 respectively. On our system 1 (specified above), although our pipeline does not require these to run on all frames unless a static object has been detected. But, even on frames where these two steps are run, we get an effective speed of 177 fps.

TABLE III

COMPARISON RESULTS ON ABODA DATASET [8]

Video Sequence	Environment Setup	Ground Truth	Proposed		Lin et al.[6]		Filonenko et al.[7]		Ilias et al.[4]		Shyam et al.[11]		Agarwal et al.[5]	
			TP	*FP*	*TP*	*FP*	*TP*	*FP*	*TP*	*FP*	*TP*	*FP*	*TP*	*FP*
Video 1	Outdoor	1	1	0	1	0	1	0	1	0	1	0	1	0
Video 2	Outdoor	1	1	0	1	0	1	0	1	0	1	0	1	0
Video 3	Outdoor	1	1	0	1	0	1	0	1	0	1	0	1	0
Video 4	Outdoor	1	1	0	1	0	1	0	1	0	1	0	1	0
Video 5	Night	1	1	0	1	1	1	0	1	0	1	0	1	0
Video 6	Illumination Changes	2	2	0	2	0	0	0	2	0	2	0	2	0
Video 7	Illumination Changes	1	1	0	1	1	0	0	1	2	1	0	1	0
Video 8	Illumination Changes	1	1	0	1	1	0	0	1	2	1	0	1	0
Video 9	Indoor	1	1	0	1	0	1	0	1	0	1	0	1	0
Video 10	Indoor	1	1	0	1	0	1	0	1	0	1	0	1	0
Video 11	Crowded Surrounding	1	1	2	1	3	0	0	0	1	1	3	1	3

TP, FP denote True Positive and False Positive respectively

On system 2, which represents most of the general-purpose laptops and personal computers, we get an effective speed of 140 fps. It is interesting to note here that even with limited RAM in system 2 and without the presence of a GPU, we get substantially good results because of the extremely small size of the two models, allows them to load on RAM together.

Our model outperforms the existing models. In Table IV, a comparison of the other models on the PETS dataset is shown. We can see that Porikli et al.[17], Liao et al.[18] and Li et al.[19] get a lower score compared to this model while [18] has a slightly better F - score. On the other contrary, we can see that [18] has a lower F - score compared to [17], [18] and our model.

TABLE IV

COMPARISON ON AVSS 2007[9] AND PETS 2006[10] DATASETS

Algorithm	**AVSS 2007**			**PETS 2006**		
	P	***R***	***F***	***P***	***R***	***F***
Proposed	1.0	1.0	1.0	0.83	0.89	0.86
Porikli et al.[17]	0.05	1.0	0.09	0.03	1.0	0.05
Liao et al.[18]	1.0	1.0	1.0	0.75	1.0	0.86
Li et al.[19]	1.0	1.0	1.0	1.0	0.71	0.83
Tian et al.[20]	0.35	1.0	0.52	0.85	1.0	0.92

P, R, F denote Precision, Recall and F-Score respectively.

As part of our assessment, processing speed of our system was compared with other state of the art models as shown in Table V. For a 320X240 resolution, the processing speed of our model was 140 fps which was greater than other related works.

Also, since the input to Mobilenet in section 3.1.4 is 300X300 px and input to YOLOv3 tiny is 416X416 px, the change in size of input image does not affect our algorithm. We get similar speeds in case of image of size 320X240 or 720X576. But in case of existing state of the art approaches, the input resolution of the video feed greatly affects the speed at which they are able to perform the task.

TABLE V

PROCESSING SPEED ANALYSIS

Algorithm	**Resolution**	
	320X240	***720X576***
Proposed	140 fps	140 fps
Ilias et al.[4]	108 fps	18 fps
Szwoch et al.[1]	49 fps	-
Lin et al.[6]	29 fps	-

CHAPTER 5
CONCLUSION & FUTURE SCOPE

5.1 CONCLUSION

We proposed a method for detection of stationary objects in public spaces using video surveillance, classifying them as attended, unattended or abandoned.

We employed the use of very power efficient and small models described which allow us to perform our pipeline very efficiently. The low memory footprint of our approach allows us to deploy our model in very low powered single board computers and still get real-time detections.

The proposed model can be installed very easily and in a very cost-effective manner to existing video surveillance architecture and aid the security forces in detecting abandoned and potentially dangerous articles in real-time giving the security forces ample time to take action.

We tested and compared our solution on three different datasets which together constitute of wide range of real-world situations of different locations, view angles, crowd, natural and artificial lighting. Our approach was at par and even better than some of the existing state of the art models. We also tested the time complexity of our algorithm on multiple image sizes and different system configurations and outperformed the existing algorithms in real-time inference by a huge margin.

5.2 FUTURE SCOPE

Now a day, security cameras have been installed at a high rate in places where there are extensive grounds and large crowd is there. Surveillance system has become an important aspect in security and a necessity to keep proper check. There was a time when video surveillance was mostly used by the government and big companies. But now the scenario has changed. Presently the use of surveillance camera has increased, and it is increasing more.

The proposed method detects abandoned objects and classifies it with level of danger it may possess. Our project scope includes identification of owner and tracking of owner of the abandoned object in real time. The number of cameras installed for security has been increasing year by year. As the world becomes more and more vulnerable, the need of surveillance also increases with time. Security has become a major issue and will always stay in place no matter what demographic we live in. The main reason is security enhancement including the prevention of incidences of terrorism. The proposed solution proposed in this paper can easily be uploaded on an embedded chip and attached to existing CCTV cameras without incurring huge costs in upgrading these systems.

REFERENCES

[1] Szwoch, G. (2016). "Extraction of stable foreground image regions for unattended luggage detection*". Multimedia Tools and Applications, 75(2), 761-786.*

[2] Li, W., Mu, J., & Liu, G. (2019). "Multiple Object Tracking with Motion and Appearance Cues" *In Proceedings of the IEEE International Conference on Computer Vision Workshops (pp. 0-0).*

[3] Six Terror Attacks that shook India Available: https://economictimes.indiatimes.com/news/defence/six-terror-attacks-that-shook-india/1993-bombay blasts /slideshow/74146291.cms

[4] Ilias, D. A. H. I., El Mezouar, M. C., Taleb, N., & Elbahri, M. (2017). "An edge-based method for effective abandoned luggage detection in complex surveillance videos" *Computer Vision and Image Understanding, 158, 141-151.*

[5] Agarwal, H., Singh, G., & Siddiqui, M. A. (2020). "Classification of Abandoned and Unattended Objects, Identification of Their Owner with Threat Assessment for Visual Surveillance" *In Proceedings of 3rd International Conference on Computer Vision and Image Processing (pp. 221-232). Springer, Singapore.*

[6] Lin, K., Chen, S. C., Chen, C. S., Lin, D. T., & Hung, Y. P. (2015). "Abandoned object detection via temporal consistency modeling and back-tracing verification for visual surveillance" *IEEE Transactions on Information Forensics and Security, 10(7), 1359-1370.*

[7] Filonenko, A., & Jo, K. H. (2016). "Unattended object identification for intelligent surveillance systems using sequence of dual background difference" *IEEE Transactions on Industrial Informatics, 12(6), 2247- 2255.*

[8] Abandoned object dataset (ABODA)," http://imp. iis.sinica.edu. tw/ABODA/index.html

[9] Advanced video and signal based surveillance (AVSS) 2007 dataset http://www.eecs.qmul.ac.uk/andrea/avss2007d.html

[10] Performance evaluation of tracking and surveillance (PETS) 2006 dataset http://www.cvg.reading. ac.uk/PETS2006/data.html

[11] Shyam, D., Kot, A., & Athalye, C. (2018, July) "Abandoned Object Detection Using Pixel-Based Finite State Machine and Single Shot Multibox Detector*" In 2018 IEEE International Conference on Multi- media and Expo (ICME) (pp. 1-6). IEEE.*

[12] Canny, J. (1986) "A computational approach to edge detection" *IEEE Transactions on pattern analysis and machine intelligence, (6), 679-698.*

[13] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T. Adam, H. (2017). "Mobilenets: Efficient convolutional neural networks for mobile vision applications" *arXiv preprint arXiv:1704.04861.*

[14] Simonyan, K., & Zisserman, A. (2014). "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556.*

[15] Porikli, F., Ivanov, Y., & Haga, T. (2007). "Robust abandoned object detection using dual foregrounds" *EURASIP Journal on Advances in Signal Processing, 2008(1), 197875.*

[16] Liao, H. H., Chang, J. Y., & Chen, L. G. (2008, September). "A localized approach to abandoned luggage detection with foreground-mask sampling" *In 2008 IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance (pp. 132-139). IEEE.*

[17] Li, L., Luo, R., Ma, R., Huang, W., & Leman, K. (2006, June). "Evaluation of an ivs system for abandoned object detection on pets 2006 datasets*" In Proc. IEEE Workshop PETS (pp. 91-98).*

[18] Tian, Y., Senior, A., & Lu, M. (2012). "Robust and efficient foreground analysis in complex surveillance videos" *Machine vision and applications, 23(5), 967-983.*

[19] Gurung, S., 2018, March. India witness highest bomb blasts in world in past two years. https://economictimes.indiatimes.com/news/defence/india-witnessed-highest-number-of-bomb-blasts-in-world-in-past-two-years/articleshow/57082541.cms.

[20] 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.*

[21] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C., 2016, October. "Single shot multibox detector". *In European conference on computer vision (pp. 21-37). Springer, Cham.*

[22] Evangelio, R.H., Senst, T., Sikora, T., 2011, January. "Detection of static objects for the task of video surveillance*". In Applications of Computer Vision (WACV), 2011 IEEE Workshop on (pp. 534-540). IEEE.*

[23] Bishop, G., Welch, G., 2001. "An introduction to the Kalman filter" *Proc of SIGGRAPH, Course, 8(27599-3175), p.59.*

[24] Lowe, D.G., 2004. "Distinctive image features from scale-invariant key points" *International journal of computer vision, 60(2), pp.91-110.*

[25] Muja, M., Lowe, D.G. “Fast approximate nearest neighbors with automatic algorithm configuration*”. VISAPP (1), 2(331-340), p.2.*

[26] Pan, J., Fan, Q., Pankanti, S. “September. Robust abandoned object detection using region-level analysis” *2011, In Image Processing (ICIP), 2011 18th IEEE International Conference on (pp. 3597-3600). IEEE.*

[27] 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.*

[28] A. Singh, S. Sawan, M. Hanmandlu, V. K. Madasu and B. C. Lovell, "An Abandoned Object Detection System Based on Dual Background Segmentation," *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance, 2009, pp. 352-357, doi: 10.1109/AVSS.2009.74.*

[29] K. Muchtar, C. Lin, L. Kang and C. Yeh, *"*Abandoned object detection in complicated environments*" 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, 2013, pp. 1-6, doi: 10.1109/APSIPA.2013.6694206.*