Suspicious Activity Detection in Surveillance Footage

Sathyajit Loganathan
Dept. of Information Technology
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
sathyajit.loganathan@my.sliit.lk

Gayashan Kariyawasam
Dept. of Information Technology
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
it16014350@my.sliit.lk

Prasanna Sumathipala
Dept. of Information Technology
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
prasanna@sliit.lk

***Abstract*— Suspicious activities are of a problem when it comes to the potential risk it brings to humans. With the increase in criminal activities in urban and suburban areas, it is necessary to detect them to be able to minimize such events. Early days surveillance was done manually by humans and were a tiring task as suspicious activities were uncommon compared to the usual activities. With the arrival of intelligent surveillance systems, various approaches were introduced in surveillance. We focus on analyzing two cases, those if ignored could lead to high risk of human lives, which are detecting potential gun-based crimes and detecting abandoned luggage on frames of surveillance footage. We present a deep neural network model that can detect handguns in images and a machine learning and computer vision pipeline that detects abandoned luggage so that we could identify potential gun-based crime and abandoned luggage situations in surveillance footage.**

***Keywords*—Gun detection, Abandoned luggage detection, Computer Vision, Surveillance.**

I. INTRODUCTION

With increasing crime rates it becomes a problem if they are not identified in time and necessary precautionary actions taken. Most urban and metropolitan areas have surveillance systems installed which constantly accumulates data. With the vast accumulation of surveillance data there are higher chances of suspicious activities to occur. But these tasks require human supervision to detect such activities as they are too complicated for artificial intelligence to handle and require high resources. Breaking down complicated tasks and detecting sub tasks which lead to potential crimes are one way to simplify an activity to be automated. We focus on two main potential leads to crimes which we attempt to detect through our models.

A. Gun based Crime Detection

Detecting crime which involve the use of guns are of high priority as the tool is of high threat to a human life. To be able to detect such crimes on time could remove the threats that arise to lives. One of our tasks is to identify guns in the frames of the surveillance footage so that potential threats involving guns could be detected.

The main challenge in this was that there was limited datasets publicly available on gun occurrence in surveillance footage. We had to collect our own dataset which had a distribution equivalent to common surveillance footage. Of various techniques object detection is one of the more popular ones to detect objects in image analysis. Utilizing models that were pretrained on large dataset which enables them to extract features from images they were further trained for our specific task to specialize to the task we are focusing on so that the detection of guns perform well with even a small dataset to train on as it has already generalized with the dataset it was trained on.

B. Abandoned Luggage Detection

Visual surveillance is one of the most active computer vision research topics at the moment. Due to increasing concern about terrorist threats and public safety, the amount of monitoring cameras has increased significantly. For any surveillance system, a security guard would monitor the environment continuously and act immediately to deter any crime. Still committed to the viewing of hours of videos and the risk of missing and overlooking critical information. Because of this, automatic methods are necessary for a surveillance system.

Abandoned luggage is the most overlooked security risk due to its difficulty to identify such items in a crowded environment. Because of this, it is clear that there is a need to develop automated methods to detect abandoned luggage.

II. RELATED WORK

Many activities appear suspicious under activities in surveillance footage and some of them even pose a threat to lives. We specifically focus on two cases and identify solutions to detect them.

Detecting objects was a challenge and before arriving at an end to end solution different indirect approaches and pipelines were used. Conventional methods used in indirect approaches included handcrafted features and shallow neural network architectures to pertain to the resource limitations and effective pipelines to detect objects.

A. Gun Detection Research

Some of the early literature involving detection of weapons focused on analyzing x-ray images [6]and infrared images [7] to detect concealed weapons. These systems had machines scanning through individuals and belongings that go through them and the images obtained from those were analyzed using different approaches to detect weapons.

One research [6] utilized the color-based segmentation to distinguish objects and used Harris interest point detector and FREAK to detect guns in the segmented images.

Image processing and video analysis was area of interest to be able to effectively identify objects. In object detection the main tasks [2] are to be able to localize to the object of interest and correctly surround the object with a boundary which illustrates the detection of the object.

We reviewed some research related to weapon detection in images. Some of them involved detection of violent scenes in movie data [8]and handgun detection in videos [9].

Some techniques utilize transfer learning [10] to avoid retraining a whole complex network and also since it requires a much lesser dataset.

B. Abandoned Luggage Detection Research

The majority of works for the identification of abandoned objects utilize background subtraction as the first step. But some works attempt to decrease false-positive detections by means of using a Combination of the Blob Tracker and the Human Tracker [11].

In most cases, the blob tracker tracks objects and locates luggage in the right direction. However, when multiple people appear to be combined at first, it cannot be easily separated by a tracker and the tracker sometimes misidentifies the sort of tracked objects only depending on object mobility. [11]

In some solutions, abandoned luggage is detected via using finite state automata. But Their suggested technique for detection of abandoned luggage missed two incidents, one triggered by a shadow and the other was due to an object monitoring failure. [12]

Guler and Farrow uses moving object tracking to detect abandoned objects with a new stationary algorithm. Their technique comprises of two main parts: a tracker for detecting and tracking movements and a stationary object detector that allows the objects abandoned to be detected and persistently detected rapidly. And Their method uses a background subtraction-based tracker for identify stationary objects. [13]

Some systems have been created for detection of abandoned objects consisting of four primary parts: foreground segmentation, moving object tracking, stationary object monitoring and event detection. [14]

A popular technique for detecting moving objects in a scene is to use a background subtraction in which moving objects are perceived as foreground and non-moving objects are taken as the background. Such methods give good accurate results and computationally less expensive. However, Lim and Keles proposed a triplet CNN model, named as FgSegNet [15] is introduced to further improve accuracy by using deep learning techniques.

III. Methodology

A. Gun Based Crime Detection

1) Data Collection, Annotation and Preprocessing

There are limited publicly available datasets on hand guns so we had or accumulate our own dataset on handguns. Since we are training a deep neural network it requires a sufficiently large amount of dataset to specify to the task we are focusing on. Obtaining a dataset for a specific task is a challenge given the variety of images consisting guns in different environments to be included in the training. By solely relying on searching for training images through a publicly available search engine could only result in a few set of images but for our case we needed more so we relied on collecting data from publicly available datasets on our specific scenario. To build our gun detector we pulled frames consisting of guns from multiple sources to obtain a dataset on guns. We obtained images from the Internet Movie Firearm Database [3], Soft Computing and Intelligent Information Systems [4] , from YouTube videos and through Google Images. A total of 3908 images were collected for training the model.

To be able to test our model performance we had to collect images from our scenario we would apply this detector on, which would be surveillance footage video frames. For this we collected 100 images of guns in surveillance footages. We also collected a test dataset from the Soft Computing and Intelligent Information Systems which was for a classification task to observe the false positives and false negatives of our detector.

As an initial step we needed to annotate these data that we obtained. For annotation we spent a total of 10 hours for all the image examples we obtained. This annotation was the bounding box of the guns available in the image. A result of this annotation would be an xml file which contains the x, y coordinates of the center of the bounding box and width and height of the bounding box. As for our case since we required to identify a firearm we had one category which was our label for the class of the detected object gun.

Before the data could be fed into the model we needed to convert the annotations of the image to a format interpretable by the model. We used python scripts to do this preprocessing step and obtained the format consumable by the model.

2) Detection Model

The detector we used for this case is based on a TensorFlow implementation of Faster R-CNN [1] and uses the Inception v2 [5] network for feature extraction. This model is pretrained on the MS-COCO dataset which is a large general dataset where the detector can learn to generally detect objects. Of the various models this performs comparatively fast inference which a reasonable accuracy which is why this was chosen.

To fine tune this model to our specific task it was trained on our accumulated dataset of guns. Fig. 1 depicts the general workflow involved in building a specific detector. Including the time of trial and error process for finding the optimal hyperparameters the total training time summed up to 20 hours on GPU.

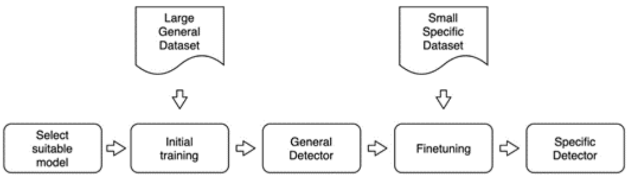

Fig. 1: General workflow of developing a specific object detector.

3) Results

This gun detector's performance was evaluated using the test data we stated before. Our models training accuracy was 91.3% and testing accuracy was 89.4% as per the split of data used to train the model. Since the previous mentioned accuracy includes the localization of the object too which differs from our purpose, detecting a gun in a frame of a surveillance footage is sufficient, we focus on the classification part of the object detector rather than the localization of the detected object (bounding box). The Table 1 is a confusion matrix of the performance of the detector on the test data we aggregated and Fig. 2 shows some detections performed by the detector.

TABLE I. CONFUSION MATRIX OF THE GUN DETECTOR

	Class 1 - Gun (Predicted)	*Class 0 - Other (Predicted)*
Class 1 - Gun (Actual)	302	2
Class 0 - Other (Actual)	33	271

Fig. 2: Sample predictions from the model

B. Abandoned Luggage Detector

1) Static Object Detection

Stationary object detection is a challenging task in this research field. It is hard to achieve a high Recall while keeping a high precision. We describe a stationary object as an object left on the scene and not in motion but wasn't at first. We use a double background subtraction technique, which can detect such items. Not every object in this background is considered as a stationary object, objects which were in the scene from the beginning is considered as a part of the background.

2) Validation

Stationary object detector detects every object that has just entered the scene and has become stationary in the scene including people who arrived at the scene then become still. This results to increase in false positives. Therefore, we suggest the validation step to verify whether a stationary object is inanimate in order to reduce the number of false positives.

For the validation step we use a ResNet 101 model, which is pretrained on MS-COCO dataset. While there are more sophisticated models are available, this is selected due to its comparatively high accuracy and fast inference speeds. Having a high accurate classification model helps to improve precision. However, using a model like MobileNetV2 will improve the inference speed significantly. The recall will not be affected by using this model. However it will increase number of false positives.

3) Results

Our model has been tested with a number of PETS 2006 sequences, including various circumstances, numbers of individuals and baggage kinds. Fig. 3 shows sample detections done by the pipeline. Every scenario is filmed from several cameras with several actors involved. For testing purposes, we used the captured image sequence that is captured from the closest to the abandoned luggage.

The subjective difficulty of detecting the abandoned luggage in each sequence is defined using stars in PETS 2006 Benchmark Data website. One star being the easiest and five stars being the most difficult to identify.

Computation time is a key information in video surveillance model. Our model calculates computation time by taking the amount of time that has elapsed to inference for the entire video and dividing it by the number of frames rendered from the dataset.

TABLE II. PREDICTION SUMMARY OF ABANDONED LUGGAGE

Sequence	Abandoned Objects	Subjective Difficulty	TP	FP	Computation Time per frame
1	1	*	1	0	0.2s
2	1	***	1	0	0.2s
3	1	*	1	0	0.2s
4	1	****	1	2	0.2s
5	1	**	1	0	0.2s
6	1	***	0	2	0.2s
7	1	*****	1	0	0.2s

Fig. 3: Detected Abandoned Luggage

IV. Conclusion

We have proposed techniques to analyze surveillance footages considering two specific cases which are detecting potential gun-based crime and detecting abandoned luggage. To detect guns in surveillance footage we present a deep neural network which can identify guns in images. This is particularly important given the mitigation of risk to human lives if the models were integrated into existing surveillance systems.

The method we propose for detecting abandoned baggage is computationally efficient and our findings indicate that, while achieving an extremely low false alarm rate, we detected most of the abandoned items effectively. We managed to solve shortcomings like having a long-standing individual being identified as a left behind object by adding one extra step to validate our results from the stationary Object Detector. Because of the considerably smaller computational time for each frame, this technique can be used for any implementation in real-time.

V. Future Work

As future works for this project our research suggests trying out with different architectures and compare them to optimize faster predictions for guns. As per concerns of limited time and resource we were able to only bring up the research to the extent mentioned in this paper so further research can be done on how to improve the detection of guns in real time. Incorporating other features rather than just surveillance videos to enhance real-time detections could also be a suggested approach to research on.

The method proposed in this study for detecting abandoned luggage does not address issues like identification of objects in sudden changes of illumination therefore, more studies on this line can be carried out for the development of the subject area.

References

[1] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” in NIPS 2015 - Advances in Neural Information Processing Systems 28. Neural Information Processing Systems Foundation, 2015

[2] Pierre Sermanet David Eigen Xiang Zhang Michael Mathieu Rob Fergus Yann LeCun “OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks” Courant Institute of Mathematical Sciences, New York University. Available: https://arxiv.org/pdf/1312.6229.pdf.

[3] The IMFDB Internet Movie Firearms Database, [Online] Available: http://www.imfdb.org/wiki/Main_Page [Accessed Mar 20, 2019]

[4] Soft Computing and Intelligence Information Systems, [Online] Available: https://sci2s.ugr.es/weapons-detection [Accessed Mar 27, 2019]

[5] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” in arXiv, vol. abs/1512.00567, 2015

[6] Rohit Kumar Tiwari and Gyanendra K. Verma, “A Computer Vision based Framework for Visual Gun Detection using Harris Interest Point Detector”, Procedia Computer Science, vol 54, p. 703 - 712, 2015.

[7] Samir K. Bandyopadhyay, Biswajita Datta, and Sudipta Roy Identifications of concealed weapon in a Human Body Department of Computer Science and Engineer, University of Calcutta, 2012

[8] Claire-Hlne Demarty, et. al The MediaEval 2012 affect task: violent scenes detection Working Notes Proceedings of the MediaEval 2012 Workshop. 2012.

[9] Roberto Olmos, Siham Tabik, and Francisco Herrera Automatic Handgun Detection Alarm in Videos Using Deep Learning Soft Computing and Intelligent Information Systems research group, Department of Computer Science and Artificial Intelligence, University of Granada, 2017

[10] Lisa Torrey and Jude Shavlik, “Transfer Learning”, University of Wisconsin, Madison WI, USA, Available: http://ftp.cs.wisc.edu/machine-learning/shavlik-group/torrey.handbook09.pdf [Accessed Mar 15, 2019]

[11] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” in arXiv, vol. abs/1512.00567, 2015

[12] S. Kwak, G. Bae and H. Byun, "Abandoned luggage detection using a finite state," Optical Engineering, vol. 49, 2010.

[13] S. Guler and M. K. Farrow, "Abandoned Object Detection in Crowded Places," 9th IEEE International Workshop on PETS, 2006.

[14] L. Li, R. Luo, R. Ma, W. Huang and K. Leman, "Evaluation of An IVS System for Abandoned Object Detection," 9th IEEE International Workshop on PETS, 2006.

[15] L. A. Lim and H. Y. Keles, "Foreground Segmentation Using a Triplet ConvolutionalNeural Network for Multiscale Feature Encoding".