AUTOMATIC WEAPON DETECTION IN VIDEOS

BY

NG JOO SIANG

A REPORT

SUBMITTED TO

UNIVERSITY TUNKU ABDUL RAHMAN

IN PARTIAL FULLFILLMENT OF THE REQUIREMENT

FOR THE DEGREE OF

BACHELOR OF COMPUTER SCIENCE (HONS)

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

(PERAK CAMPUS)

JAN 2017

UNIVERSITI TUNKU ABDUL RAHMAN

REPORT STATUS DECLARATION FORM

Title: **AUTOMATED WEAPON DETECTION IN VIDEOS**

Academic Session: January 2017

I **NG JOO SIANG**

(CAPITAL LETTER)

declare that I allow this Final Year Project Report to be kept in

Universiti Tunku Abdul Rahman Library subject to the regulations as follows:

1. The dissertation is a property of the Library.
2. The Library is allowed to make copies of this dissertation for academic purposes.

Verified by,

_____ _____

(Author's signature) (Supervisor's signature)

Address:

B-1-04, Taman Warisan

Jalan Warisan

Jalan Junid 84000

Muar Johor

_____

Supervisor's name

Date: 10 April 2017 **Date**:

DECLARATION OF ORIGINALITY

I declare that this report entitled "**AUTOMATIC WEAPON DETECTION IN VIDEOS**" is my own work except as cited in the references. The report has not been accepted for any degree and is not being submitted concurrently in candidature for any degree or other award.

Signature : _____

Name : _____

Date : _____

AUTOMATIC WEAPON DETECTION IN VIDEOS

BY

NG JOO SIANG

A REPORT

SUBMITTED TO

UNIVERSITY TUNKU ABDUL RAHMAN

IN PARTIAL FULLFILLMENT OF THE REQUIREMENT

FOR THE DEGREE OF

BACHELOR OF COMPUTER SCIENCE (HONS)

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

(PERAK CAMPUS)

JAN 2017

ACKNOWLEDGEMENTS

First of all, I would like to express my sincere gratitude and appreciation to my supervisor, Dr Tan Hung Khoon for his guidance and advice throughout this project. I felt so lucky that Dr Tan can be my supervisor as he always shares his knowledge and techniques in the area of computer vision to me. Without his guidance, this project will never come into existence.

Besides, I would also like to thanks to my friends that willing to give recommendation and criticism to this project. Their opinions allow me to have more ideas which can improve this project.

Last but not least, thanks to my lovely family for being there by my side at my hard time so that I will never feel lonely to face the journey that full of difficulties.

ABSTRACT

Human action recognition is important for wide range application like video surveillance, video indexing and monitoring system. However, human action recognition and analyses is still an open problem in computer vision owing to the variety of human poses and appearances.

In our work, we are interested in tackling the specific issue of dangerous event detection. We are not only interested in detecting the gun in the video scene, but also interested in classifying which person is holding a gun. However, this is a difficult task as the gun is a small object and is easily missed by the current object detection algorithm. Hence, we introduce an approach that can explicitly model the human-object interaction by extracting the interaction feature of the object with respect to the human. In order to extract the interaction features, we employ the state of art technique to localize the human and gun in action. Most importantly, we apply the tracking algorithm to link the object and human detection over time as we noticed the object detection algorithm are far from ideal. The interaction features and 3DHOG feature of the human and object are concatenated into single fixed dimension descriptor and used for training an action classifier.

Our experiment results showed that our approach manages to classify which person is holding a gun in the video scene and from that we can classify the event into dangerous and non-dangerous in section 4.

TABLE CONTENT

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Introduction

Crime is one of the big problems in the world and worrying aspects in any society since crime is increasing at an alarming rate. According to Numbeo (2016), the crime index in Malaysia is 67.43 and rank top 3 among all the Asian countries. Besides, during these recent years, the increase armed crime has clearly worried the public. For example, the incident that happened at Columbine High School, USA which left 15 dead and 24 injured (NY Daily News, 2016a) and also the Nordway attacks by Andreas Breivik which killed 92 victims (Beaumont, 2011). Due to these few factors, several automated methods for video surveillance has been proposed.

One of the strategies to tackle the issue is to install a network of circuit television systems (CCTV) in most public spaces, housing area and office. As a result, CCTV has become ubiquitous. In U.K., 1.85 to 4.2 million CCTV cameras are currently in operation (Security News Desk, 2013). However, the effectiveness of CCTV operators has put into question as there are too many numbers of cameras to monitor. Furthermore, the CCTV cameras are only playing a passive role in the CCTV system which unable to detect crimes. For example, on March 6, a rape incident was captured on a CCTV for nearly 30 minutes, but CCTV system unable to detect the crime (NY Daily News, 2016b). This highlights the limitation of current CCTV system.

In this project, we plan to apply the automated weapon detection algorithm to the CCTV video and realistic video to illustrate on how the automated weapon detection system can aid the CCTV operator or police to classify an action.

However, the developing of the system that employed the algorithm above is not a simple task for the following reasons:

- The system need to cope well with any poor-quality input video
- The system should always provide a reliable result and a low number of false alarms because the user will ignore the result that produced by the system if the system generates too many inaccurate results.
- The weapon is visually a small object and is easily missed by the current object detection algorithm.

For example, the GunDetect smart camera that developed by nanoWatt Design uses computer image processing to automatically detect if a firearm is in the room can achieve 90 percent of accuracy in detecting the firearm if a firearm is clearly visible, but on the other hand, if a firearm is not clearly visible, the firearm may not be detected (Tech Times, 2015).

Hence, we are going to enhance the existing weapon detection technique to achieve a satisfaction and more accurate result by applying the human tracking and object tracking in our proposed solution.

1.2 Project Objective & Scope

This project aims to design a system that is capable to detect weapons automatically. Thus, in order to achieve the goal, the project involves the following objectives and scope:

Main Objectives

- We tackle realistic videos which include CCTV videos and movie videos. In addition, waist shot of video will be used as the input of our proposed algorithm as if the human and object are hardly to be seen in the video, the system may provide a less accurate result.

Sub Objectives

- To develop a human and object tracking mechanism and thus the system able to identify the weapons such as firearms in the video clip.

- To design an algorithm that able to extract the feature of the object that holds by the human and classifies the object into weapon or non-weapon by using the action classifier.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

Several of weapon detection algorithms have been developed and studied over years. Each of them consists of different type of methods and techniques and has its own pro and con. However, all of them have the same goal which is trying to enhance the existing method or create a new method to achieve a better result.

Literature reviews on the Human-Object interaction model (ZhaoZhuo et al., 2015), Automated Weapon Detection in CCTV Image (Grega et al., 2016), and Explicit Modelling of Human-Object Interactions in Realistic Videos (Prest et al., 2013) will be done in section 2.2, section 2.3 and section 2.4 respectively. Lastly, a conclusion will be made in section 2.5

2.2 Human-Object Interaction Model (HOI)

In the system that developed by ZhaoZhuo et al.(2015), they establish an algorithm by constructing the HOI model and hence, the HOI model is used to determine the object in predicting bound drawn by certain direction and distance based on the location of hips. They determined the object based on the location of hips because they realised the other parts of human body are difficult to locate. The object detected is classified by using the Support Vector Machine (SVM) into dangerous and non-dangerous. The overview of the system is shown in Figure 2.1.

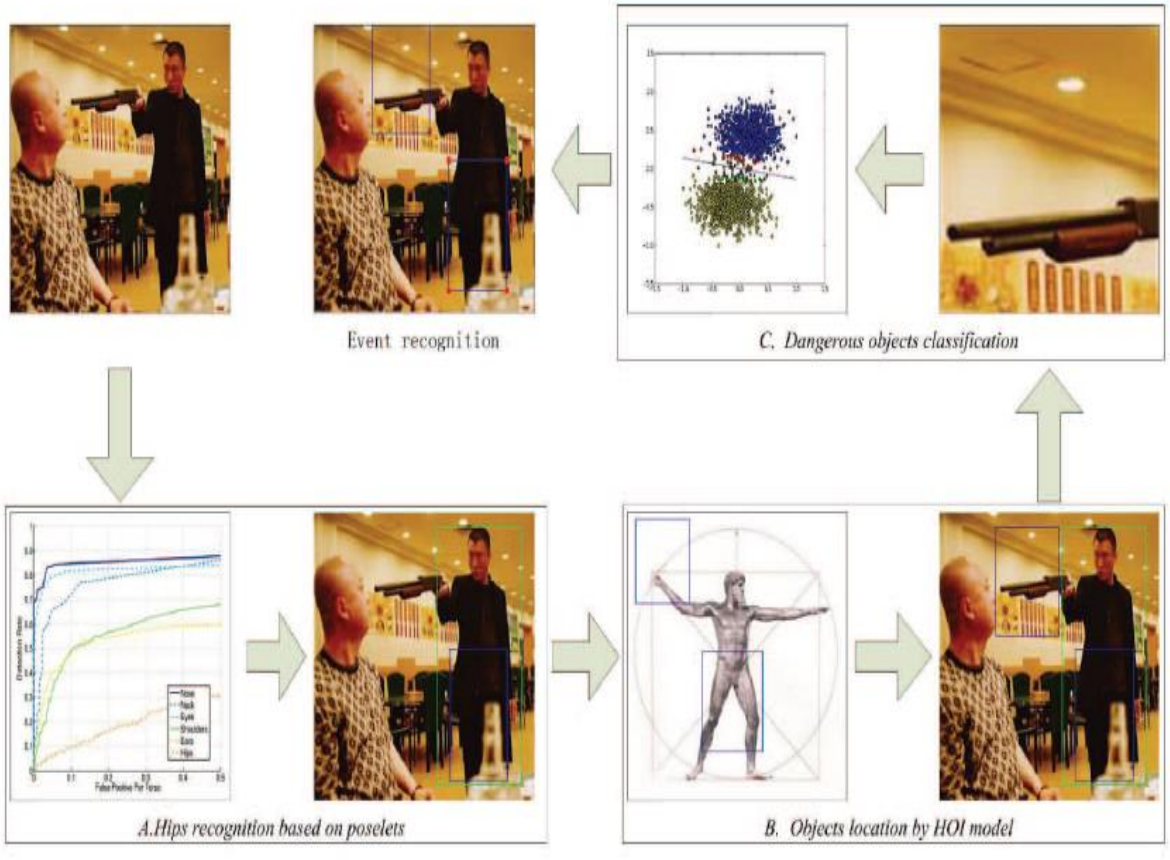


Figure 2.1 Overview of our detecting systems (Zhaozhuo et al., 2015)

Hips recognition based on poselets: First step of the system is to locate the hips location using poselets. To increase the speed of hips recognition, the human bodies are segmented into several parts and Histograms of oriented gradients (HOG) is used to obtain the feature from the part that needed. However, the detection of other body parts is also included as they help in hip recognition.

Objects location by HOI model: Once the location of hips is detected in images, the area in the oblique upward direction of hips on both sides can be drawn based on the HOI model. The speed of searching for an object is clearly increased as the system only need to search the object in the bounds that drawn in the image.

Dangerous object classification: The SVM is adopted for classification in this system for object classification as it provides high detecting rate and the HOG feature of the handheld objects are extracted to act as the input to the SVM classifier.

Discussion: The system that proposed in this literature clearly has its strength, which is the speed of detecting handheld object is increased because the searching of an object is within the bounds that drawn based on the HOI model. However, this system is too rigid from a certain view of the human position and hence may produce an inaccurate result if the view of human position is not similar to the input testing image.

2.3 Automated Weapon Detection in CCTV image

In the system that developed by Grega et al. (2016), they proposed two different algorithms that aim to detect the knife and firearm in an image and alert the human operator before any dangerous event happened. Several related works have further motivated them to solve the problem of weapon detection in the camera video. For example, the Yong et al (2008) have shown that by using microwave swept-frequency radar to detect the metal object such guns and knives. However, this practicality of approach is limited to the financial and health concerns. Figure 2.2 shows the flow of the algorithm for knife detection.

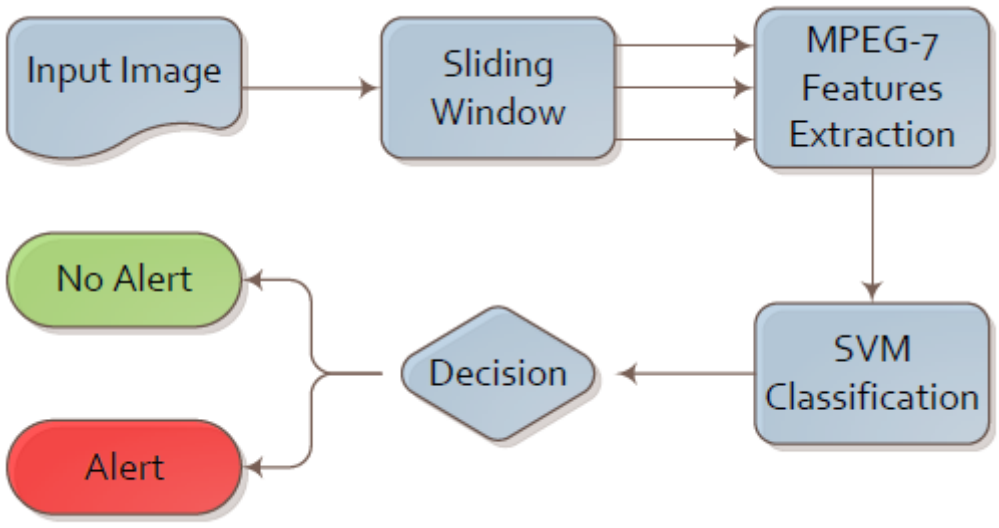


Figure 2.2 Algorithm for knife detection (Grega et al., 2016)

Knife Detection: Firstly, a sliding window technique is used to choose image patches from the input image. The image patches are represented using edge histogram and homogeneous texture descriptor to capture the characteristic features of knives because the edge histogram descriptor contains information of different types of edges

in the image and the homogeneous texture describes the image patterns. Colour-based descriptors are not used as they are not able to deal with different colour balances and light reflections. Furthermore, key point based descriptors are not used because knives do not contain many characteristic features. Then, Support Vector Machine (SVM) is used to act as a classifier based on the extracted feature vector. Figure 2.3 shows the flow of firearm detection algorithm.

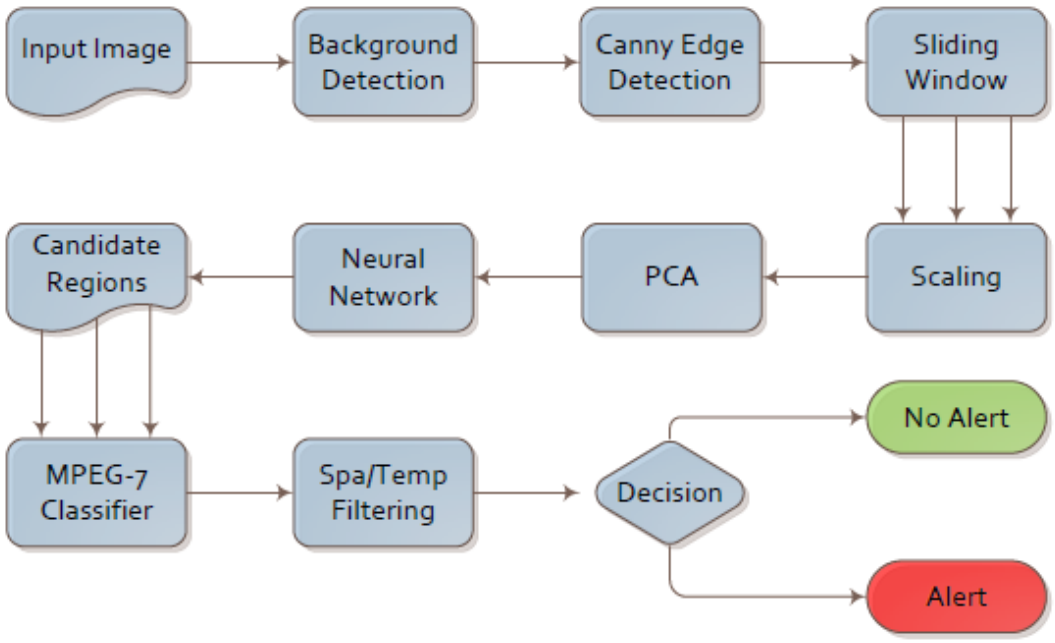


Figure 2.3 Algorithm for firearm detection (Grega et al., 2016)

Firearm Detection: Firstly, a simple background subtraction was applied based on image differences between consecutive frames. Next, to convert the image into a set of edges, the canny edge detection algorithm was employed. The input for the canny edge detection algorithm is the foreground region detected in background subtraction so that the computational power can be conserved. Then, sliding window technique is used to decompose the larger image into small regions and repeatedly scanning in different scale of the same image. Next, the scaled samples are fed into the PCA

(Pearson, K.,2009) to change the dimensionality of input vector to 560 values and this 560 value vector is fed into a three layer neural network (NN) because this NN able to gives low specificity (low number of false alarms) and high sensitivity (low missing of event) (Grega et al.,2016). After that, the MPEG-7 descriptor is used to make a comparison with shape found in candidate region that selected by NN. Lastly, temporal and spatial filtering is applied and an alarm will raise if a firearm is detected.

Discussion: The strength of the solution proposed in this literature can be seen obviously which is the knife algorithm is able to tackle the poor quality and low resolution images due to the edge histogram and homogeneous texture are used instead of colour and key point based descriptors. For firearm detection algorithm, it obtains high specificity (low number of false alarm). However, there is a limitation as the dataset used for the input of the algorithm is gathered in a controlled environment, thus if the both algorithms execute in real environment, the result that produced by the algorithm is still unknown.

2.4 Explicit modelling of Human-Object Interactions in Realistic Videos

In the system that developed by Prest et al. (2013), they proposed an algorithm for model the human actions by using the interaction feature between objects and persons in realistic videos.

Existing work like image gradient or optical flow use low-level features to represents actions. The system for this literature review basically tracks the human and objects over time and represents an action as the trajectory of an object with respect to the human position. Besides, the amount of control needed to train an appearance model of the action object and interaction model can be reduced by using the state-of-art approach.

Human Detection: A generic part-based human detector that consists of four part detectors is employed in the human detection (Prest et al., 2013).

Object Detection: Small objects are harder detected compare to detect human because these objects have different pose and appearance. In the object detection, the system used the detection algorithm of (Felzenszwalb et al., 2010) which produces good results on PASCAL VOC object detection challenge (Everingham et al., 2009)

Tracking: At various stages, tracking is needed as the results of object detection algorithm tend to be sparse. A tracking algorithm that consists of traditional tracking of a target and tracking-by-detection scenario is used to track multiple targets. The

tracking algorithm takes any number of detection windows of the target as input and propagates forth and back in time based on dense point-track.

Discussion: The strength of the solution proposed in this literature review is that the system employs state-of-art object detection technique (Felzenszwalb et al., 2010), which robustly links detection over time and allow the missing of the object in many frames. Next, the methods that employed in the system make the detection does not affected by the background motion.

2.5 Conclusion

One of the weaknesses of the model by ZhaoZhuo et al. (2015) is that it can only handle scenes from a particular viewpoint. As a human may hold the weapon in different positions, the HOI model is impractical for real scenarios.

For the review by Grega et al. (2016), the weakness is that the knife and firearm detection used the dataset that is gathered in a controlled environment to verify the accuracy of the algorithm. If the both algorithms execute in the real environment, the result produced may not be accurate.

In addition, both of the models in these two reviews do not consider the interaction information between the human and object involve in the action. Interaction information can be in terms of the relative motion, relative location and relative area between the object and human (Prest et al., 2013). Interaction features are useful for classifying types of actions that cannot be tackled by Grega et al. (2016) and ZhaoZhuo et al. (2015). A human holding a knife to cut a banana, the HOI model will tend to raise alarm and classify the action as dangerous since there is a knife involved, hence in our project we will employ a similar framework as the work of Prest et al. (2013) that used the interaction information for action classifying.

CHAPTER 3: SYSTEM OVERVIEW

Figure 3.1 shows the overview of our proposed system for gun action recognition using human-object interaction feature. Our system technically employs the well-established object detection (Felzenswalb et at., 2010), human detection (Felzenswalb et at., 2010), and tracking techniques (Sundaram et al., 2010).

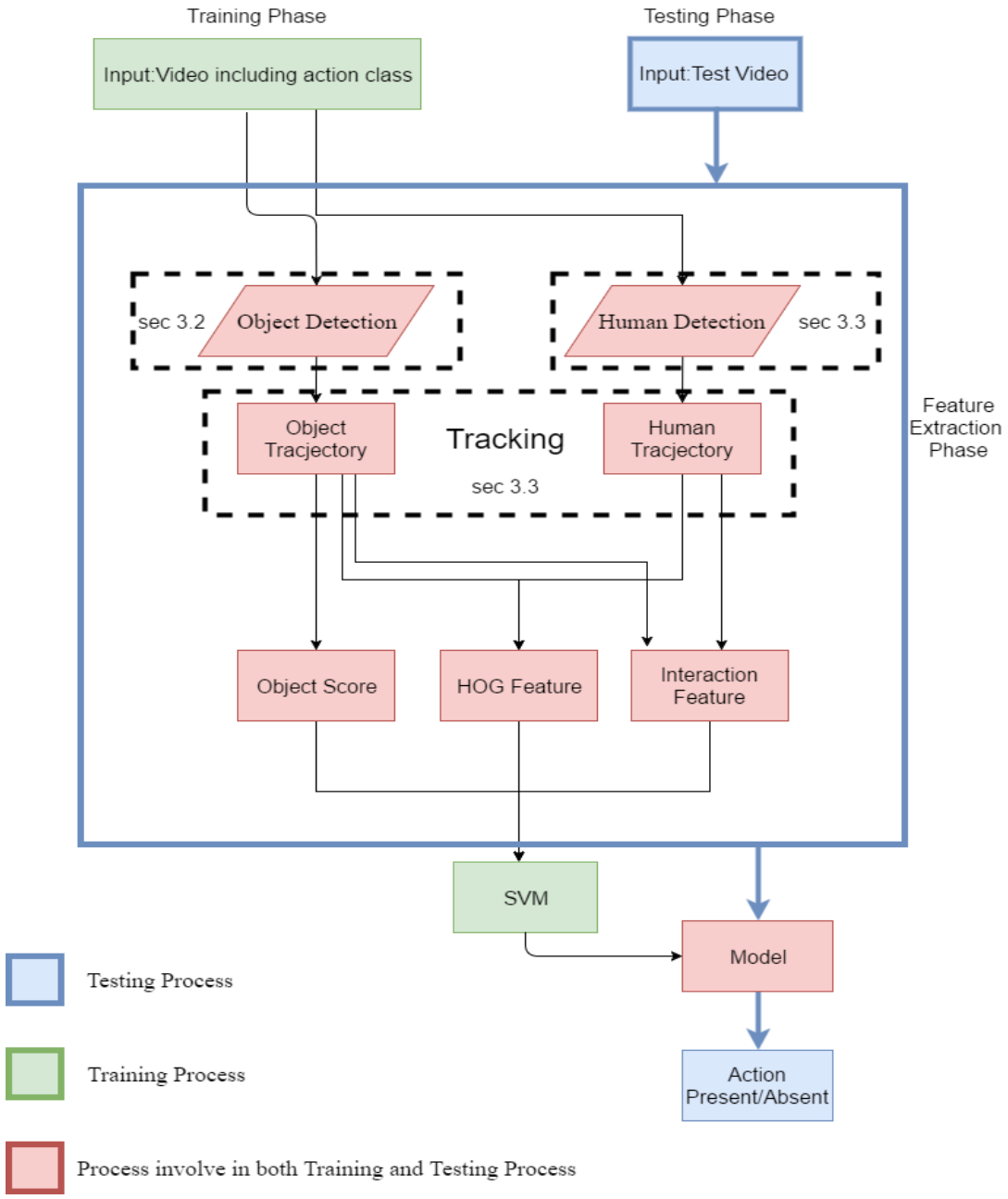


Figure 3.1 System overview

Step 1: Input

This system takes several short videos containing instances of targeted actions as input. For training samples, the spatio-temporal cuboids annotation is provided in terms of the location and time of the targeted action class in the videos.

Step 2: Human Detection

A human detection algorithm (c.f. section 3.1) is used to localize the humans in the video.

Step 3-Object Detection

To learn the interaction features between the human and the object, we need to build objector detector (c.f. section 3.2) to localize the objects of interest in the video. In this project, the objects related to us are guns.

Step 4-Trajectory Extraction

The result of the object and human detection is expected to be sparse as human detection and object detection results are not ideal. To overcome this, a tracking algorithm (c.f. section 3.3) is applied to link the detection over time.

Step 5-Human-Object Pair Extraction

Next, we associate the positive human track and positive object tracks to form the positive human-object pair. For forming the negative human object pair, we repeat the step from step 2 to step 4 with the part that not overlaps with any spatio temporal cuboid.

Step 6-Feature Extraction

Following (Prest et al., 2013), we extract the interaction feature and HOG feature of the object and human as these features are important for training an action classifier. (c.f. section 3.4)

Step 7-Train Action Classifier

At last, we train an action classifier that can classify the person action into gun action or non-gun action based on the features that extracted in step 6 by using non- linear Support Vector Machine (SVM) with RBF kernel.

In the testing process, the step 1 to step 6 are repeated however in the testing we do not know which track is positive or negative, so we associate all the human track with the object track to form human-object pairs and keep all the human objects pairs. Next, we used the action classifier that trained in step 7 to classify the human object pairs into positive or negative. Positive indicate the human is holding a gun, negative indicate the person is not holding a gun.

3.1 Human Detection

Over the past decade, major progress has been made in the area of pedestrian detection in static images as well as human detection in videos. Detecting humans are particularly difficult due to photometric variation, viewpoint variation and intra-class variability. In our system, we adopt the discriminatively trained part based models (Felzenswalb et at., 2010) which achieve state of art results on the PASCAL and INRIA person datasets.

Deformable part model (DPM) is a model that represents an object using mixture of multiscale part models. Each part models captures the local appearance properties of an object. Figure 3.2 shows the human model with different part models.

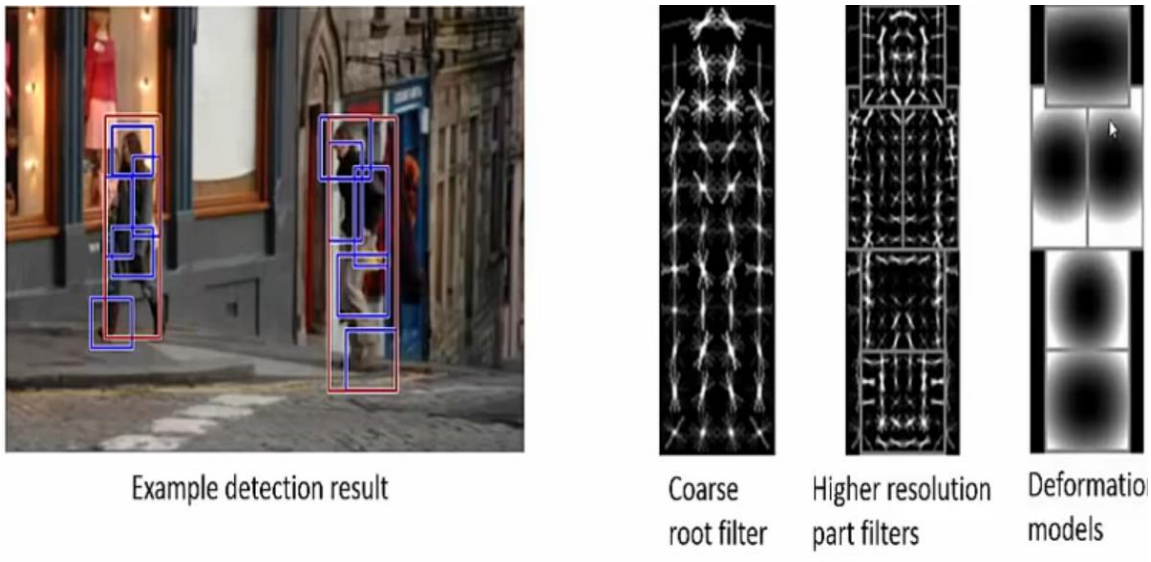


Figure 3.2 Human model (Felzenswalb et at., 2010)

The Dalal-Triggs detector (Dalal et al., 2005) act as the root filter of the DPM. The Dalal-Triggs detector used a filter on the histogram of oriented gradient (HOG) to represent an object category. Sliding window approach is used in the detector and it is

a technique that scan smaller regions of a larger image and then recursively scanning on multi-scale of the same image to identify the exact position and scales of the object. Since Dalal-Triggs detector is a filter, we can compute the score by using the formula:

$$\beta \cdot \varphi(x) \quad (1)$$

where β is the filter, x is the image at specific position and scale and $\varphi(x)$ is a feature vector.

In DPM, the root filter is enriched into a set of part filters. The part filters are the filters that divide the object into several parts and it captures features at twice the spatial resolution relative to the features captured by the root filter. Each of the part filters has a deformation cost which is described in the deformation models. The deformation cost is the penalty given when the part is far away from the location where it is supposed to be. In another word, the higher the deformable cost of certain part, the further the distance from the actual location of the part.

Once the root filter, part filter and deformation model have been trained, the DPM can use to estimate the object location according to the best possible placement of the parts. Figure 3.3 shows the object detection process by using the deformable model.

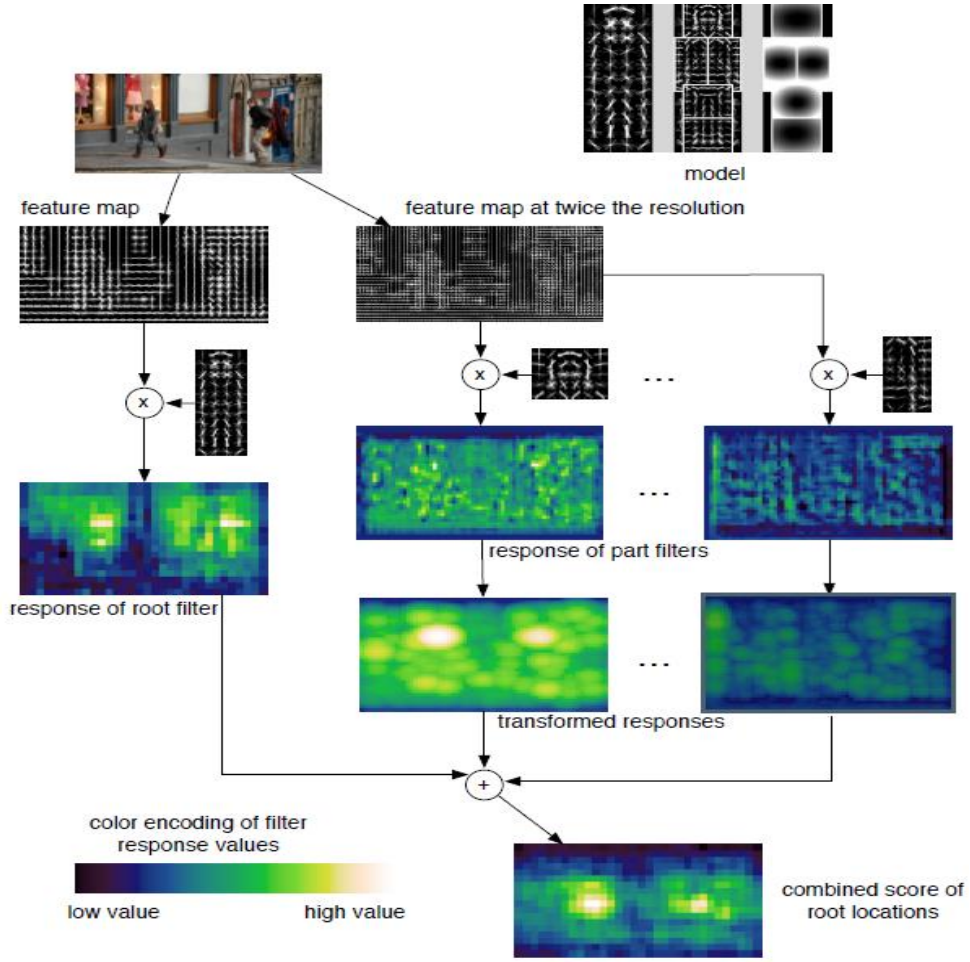


Figure 3.3 Object detection process by using the deformable model (Felzenswalb et at., 2010)

By giving an image, the response of the root and part filters are computed at a different resolution of the feature pyramid. The associate deformation cost for each part filters has been calculated by applying the deformation model. Next, distance transform will be done on the response from the part filters and the deformation model to compute the best location for the part. At last, by combining all the response from the root filter, part filter and deformation model, the DPM can calculate the score of the human detection at the particular point. High scoring indicates true detections.

3.2 Object Detection

Detecting small objects is a difficult task as these objects are rather smaller than the human. In this section, we will implement a gun detector that trained by using the discriminatively trained part based models (Felzenswalb et at., 2010)

As our object of interest is gun, we will collect different positive samples from different type of gun related videos. To minimize the amount of supervision, we provide the bounding box of the gun in the first frame where the gun appears and apply the tracking method that we proposed in section 3.3 to let it track for the gun. We will then use the result from the tracker to annotate the gun location in each image. For the negative samples, we will use all the annotation from the Pascal Visual Object Classes Challenge(VOC) (Everingham et al., 2009) as all the annotations are without any gun object.

Lastly, all the annotation results of the positive sample and negative samples are used as the input for the DPM to train a gun model.

3.3 Tracking

Tracking is required in different stages of our system because the object detector tends to produce sparse detection as the object of interest is too small. Figure 3.4 shows the example of sparse detection result that produced by our gun detector.

Frame 34 Frame 38 Frame 44

Figure 3.4 Sparse detection. Our gun detector manages to detect the gun in frame 34, however, there will be missed detection for 9 frames between the frame 34 and frame 44.

Based on our observation, our gun detector may have a gap of 10 to 20 frames between the positive detections of the corresponding object. Hence, we propose a tracking algorithm which able to track multiple objects simultaneously in every frame and link the detections over time.

Tracker Overview

By given an object detection, the tracker allows us to update and monitor the position of the object at the frame where the detection is being missed. In order to do this, we perform forward propagation and also backward propagation from the frame where the object is being detected. Hence, our tracker is not for the real-time purposes but only suitable in the video that pre-recorded. To give better understanding on the purpose of the tracker, a scenario with explanation will be shown in Figure 3.5.

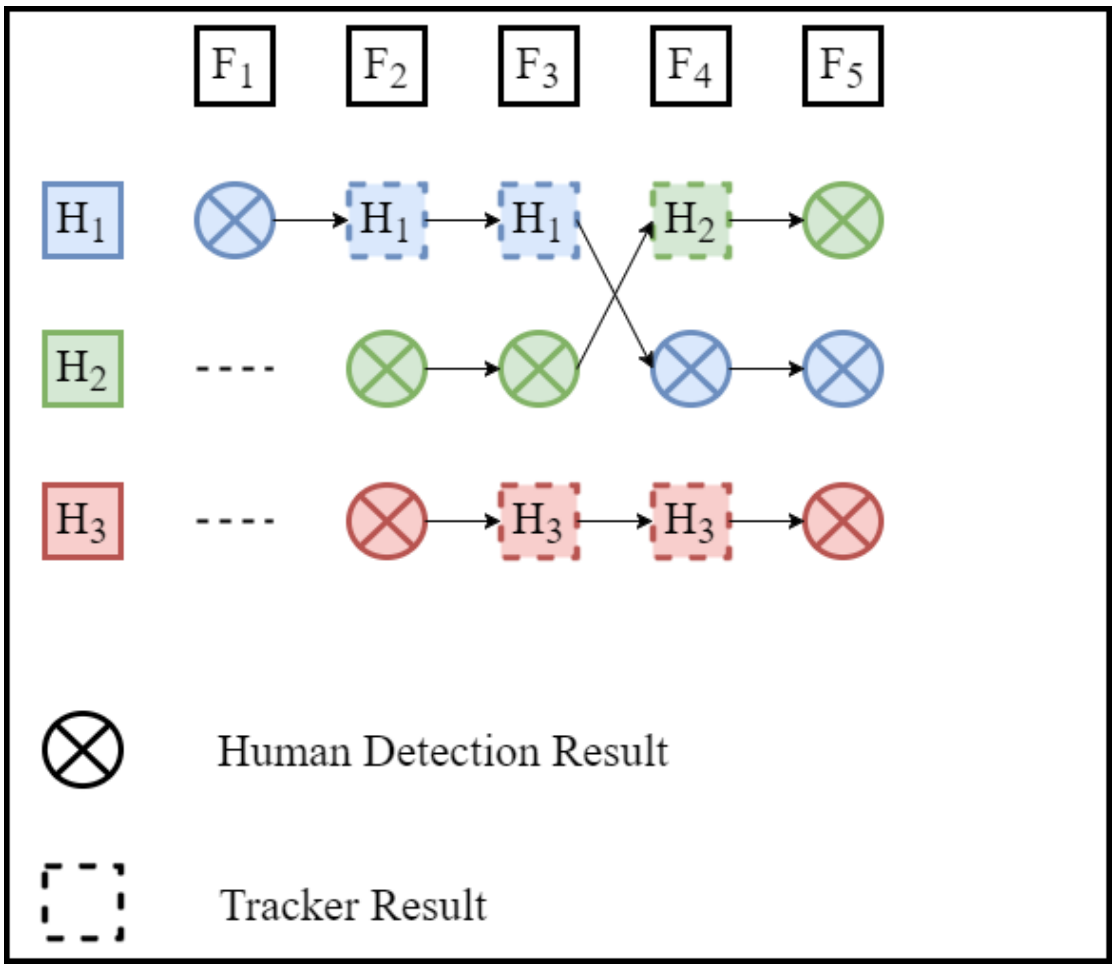


Figure 3.5 Scenario example

Based on the Figure 3.1, there are 3 human detection (H_1, H_2, H_3) that being found across 5 frame (F_1, F_2, F_3, F_4, F_5). In F_1, H_1 is the only detection that being found. However, in F_2, human 1 is not detected and our tracker come into play to track the H_1 in previous frame and link the detection in the subsequent frame. This shows that even there is sparse detection, our system is still able to monitor the location of the object in every frame. Our tracking algorithm works as follow:

Input:

Given a set of detection window D^i for each frame $i \in \{s, \dots, e\}$, we start our tracker at the first frame with at least one detection. If more detections are provided, our tracker will try to link the detection over time. Our tracker manages to work in both situation where one single detection is found across the list of frames or sparse detection produced by the object detector.

Step 1: Initialization

Let frame *m* as the first frame where detection is available. For each detection $D_j^f \in D^f$, we create one track T_j and initialize into the overall tracks T. Each object is one track and more tracks as we go along. This allows us to achieve multiple object tracking.

Step 2: Forward Tracking

Forward tracking is a process that used to estimate and update the location of D^i in the subsequent frame. For each $T_j \in T$, our tracker executes forward over frames *i* from frame *f* to frame *e*.

i. **Dense Optical Flow:** In order to estimate the motion of T_j^i of track T_j in frame $i+1$, dense optical flow is performed. Optical flow is the pattern of motion of local patch between two consecutive frames caused by the movement of object or camera. Each vector in 2D vector field represents the displacement vector of movement of points from the first frame to the second frame. In our system, we used Gunner Farneback algorithm[1] (Gunnar Farneback, G., 2003) to obtain the dense optical flow of each point. Figure 3.6 shows the optical flow of the frame i to frame $i+1$.

ii. **Update Location:** We apply the median filtering on the optical flow of the bounded detection object to compute the median value of the object displacement next frame. From that, we can update the track location in frame $i+1$ by adding the median value with the track location in the frame i.

iii. **Link Detection:** If there is detection D_k^{i+1} in frame $i+1$ which overlaps with the T_j^{i+1}, the detection D_k^{i+1} will be removed from D^{i+1} and the detection D_k^{i+1} will be assign to T_j^{i+1}. We determine D_k^{i+1} in frame $i+1$ is overlap with the T_j^{i+1} by comparing the area of the detection in T_j^{i+1} with the area of D_k^{i+1}.

[1]Algorithm can be found at the OpenCV Websites

Figure 3.6 Optical flow on the gun movement

For the detection D_k^{i+1} that does not include in the overall track T, a new track will be created and initialize into the overall track T.

Step 3: Backward Tracking

Backward tracking is a technique that used in the situation where the first frame that has at least one detection is found in the middle of a list of frames. Step 2 is repeated from the frame *f* to *s* to obtain the full motion path of the object. Figure 3.7 provides the clearer picture on the backward tracking process and forward tracking process.

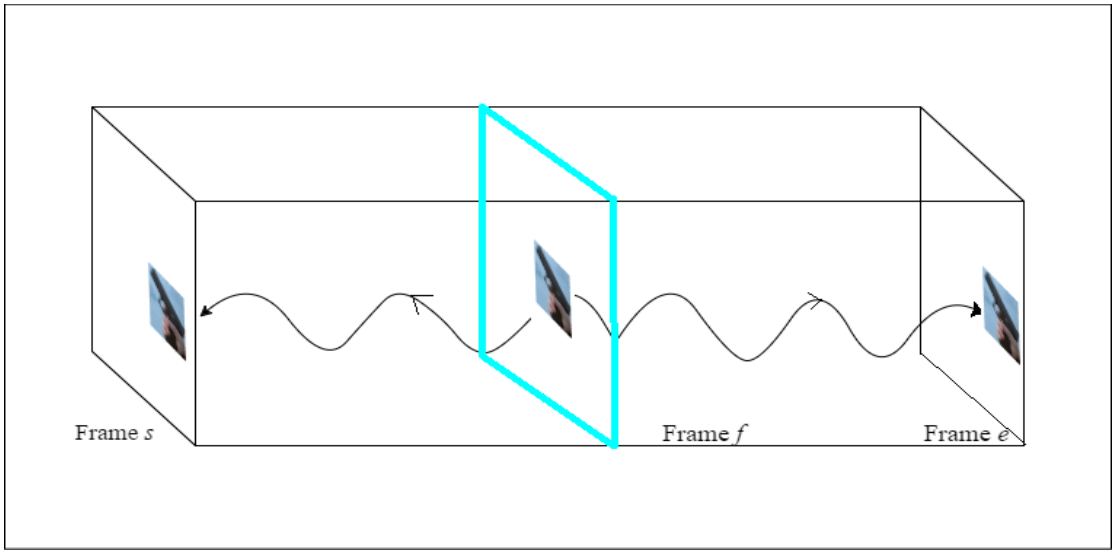


Figure 3.7 Backward tracking from frame f to frame s and forward tracking from frame f to frame e

Step 4: Concatenate Backward Track and Forward Track

We assemble the final track by concatenating the backward track and forward track.

Output:

Figure 3.8 shows how our tracker manages to link the gun detection over time and provide the accurate estimation of object based on the gun detection in the first frame.

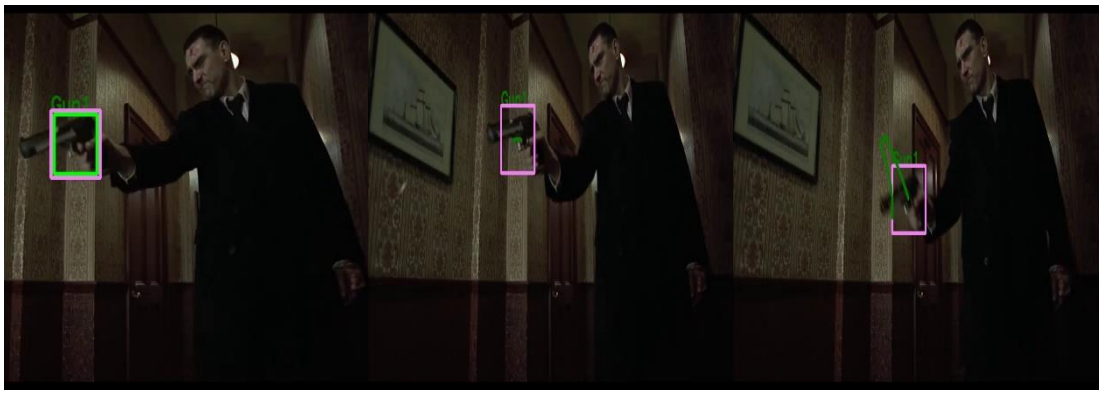


Figure 3.8 Tracker result. In the first frame, the gun is automatically detected by our gun detector (green bounding box). As there is one gun detection in the first frame, our tracker starts to track the object in the subsequent frame (pink bounding box). Green curve line is the full motion path of the gun detected.

3.4 Feature Extraction

We believed that the interaction features of the object track with respect to human track can clearly differentiate and describe each type of action such as gun pointing. Given a list of frames, we compute the interaction feature in every frame where they both exists as our tracker does not skip any single frames.

Interaction Feature:

1. **Relative location** (Prest et al., 2013). The relative location $l(H^t, O^{t)}$ of the object window O^t wrt to human window H^t in frame t

$$l(H^t, O^{t)} = (\frac{O_x^t - H_x^t}{H_W^t}, \frac{O_y^t - H_y^t}{H_H^t}) \quad (2)$$

   where subscripts indicate a window's centre x,y, width W and height H

2. **Relative area** (Prest et al., 2013). The area of O^t relative to H^t

$$a(H^t, O^t) = area(H^t)/area(O^t) \quad (3)$$

3. **Relative motion** (Prest et al., 2013). The relative motion of object with respect to person that define as 2D vector. This 2D vector is the difference $l(H^t, O^t)$ and $l(H^{t-1}, O^{t-1})$ in t frame and it is represent the magnitude and direction

$$m(H^t, O^t) = l(H^t, O^t) - l(H^{t-1}, O^{t-1}) \quad (4)$$

Besides, 3DHOG feature of the object and human are also important for training an action classifier as it captures motion information and low level appearance. Furthermore, we also record the maximum score of the object detection in the object track and also the maximum human detection score in the human track.

At last, we accumulate the value of each interaction feature into a histogram. The relative area is quantized into 4-bins histogram. The 2D relative motion and relative location cues are quantized into 16 bins histogram. Each of the histograms is independently L1 normalize and concatenate with the object maximum score, human maximum score and 3DHOG feature into one single fixed dimensionality descriptor.

CHAPTER 4: EXPERIMENT RESULT

4.1 Human Detector

In the human detection algorithm that we implemented in our system, we adopted the Inria human detector that is pre-computed by Felzenswalb et al. This human detector has achieved excellent result in the Inria testing dataset. Figure 4.1 shows the average precision graph of this human detector.

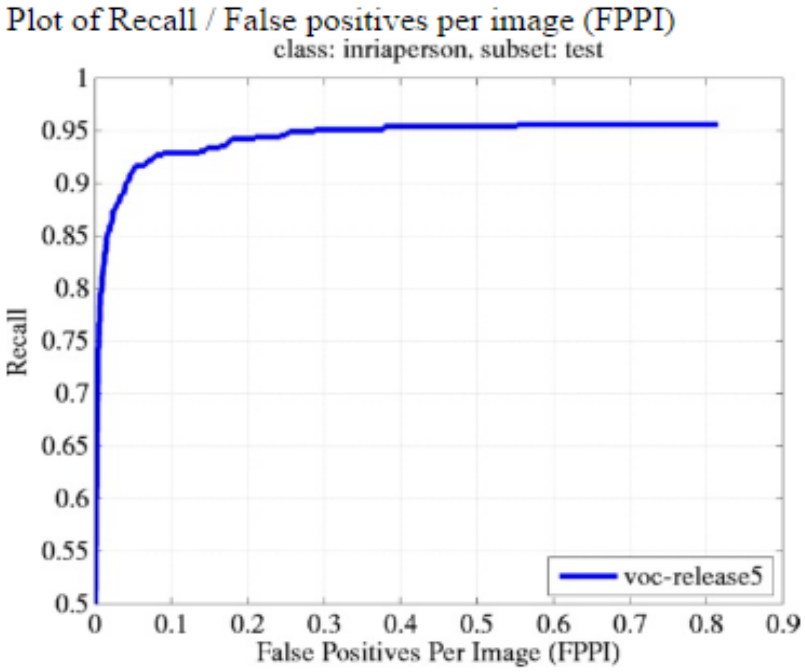


Figure 4.1 Average precision of Inria person detector (People.cs.uchicago.edu, 2017)

Based on Figure 4.1, the average precision for the Inria person detector is 0.88 and the recall of this detector is around 0.8. However, with this precision, the human detector is still far from ideal as it will produce some false detection and sparse detection, hence our tracking method is needed so that the precision of the detection can be increased.

4.2 Evaluation on Gun Detector

In this section, we evaluate the gun detector that trained by using the approach of Felzenswalb et at. The training dataset for training the gun detector has 255 positive images and 4756 negative images while the testing dataset for the gun detector has 200 positive images and 4702 negative images. The positive images for training and the annotations are taken by the method that described in section 3.2. The positive images for the testing are manually selected from the different type of gun related videos and we manually annotated each of the images. There is not bias on the position of the gun for training and testing. The positions of the gun object that included in the training and testing are frontal, left side, and right side. Figure 4.3 shows the gun model with 8 components.

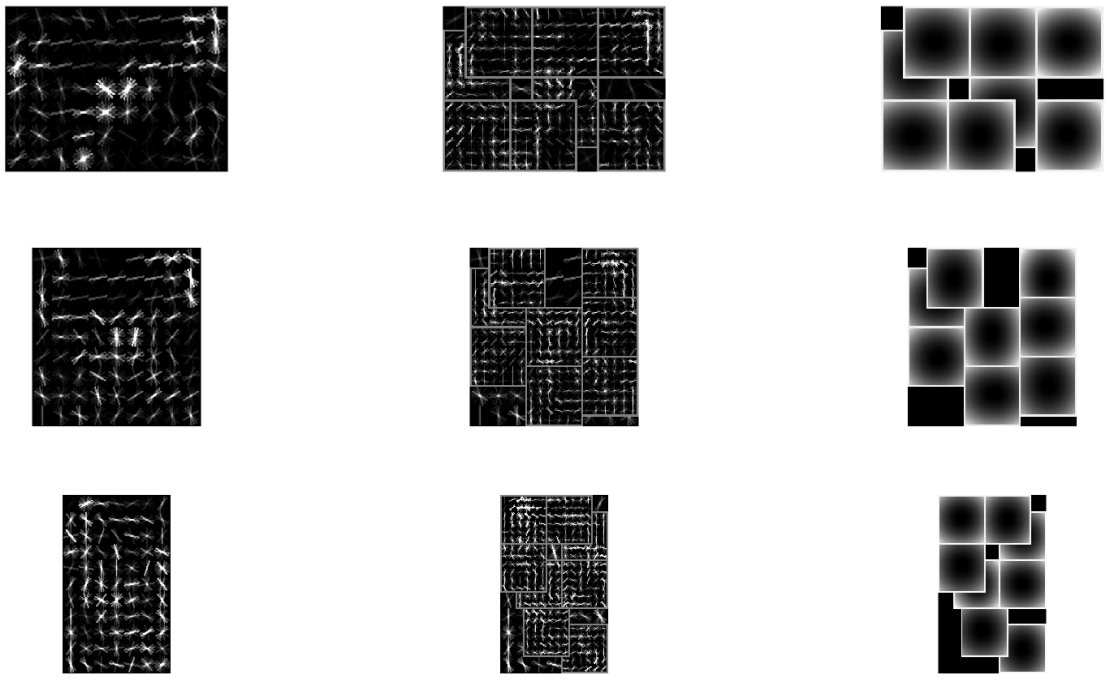


Figure 4.3 Gun model. Top left image show the left side of the gun, bottom left shows the frontal position of the gun and middle left is the right side of the gun.

Figure 4.4 shows the average precision of our gun detector towards the testing dataset.

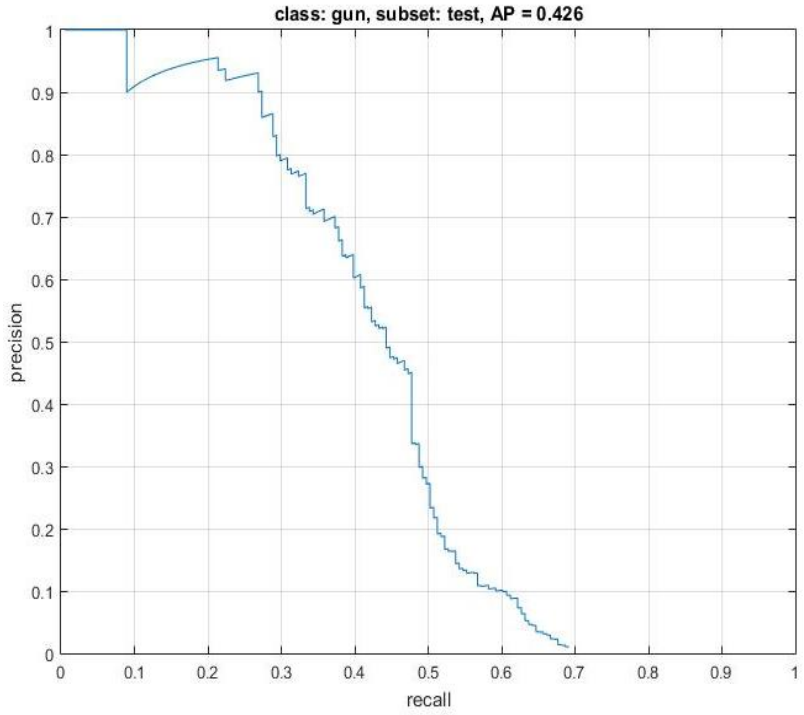


Figure 4.4 Average precision of gun detector

Based on Figure 4.4, the average precision for the gun detector is only 0.426 and the recall rate for the gun detector is approximately 0.6 because the current object detection algorithm is far from ideal and the gun object is relatively small. From that, we know that our gun detector will produce sparse detection. To deal with this issue, our tracking method proposed in section 3.3 has to work with the gun detector.

4.3 Evaluation on Tracker

In this section, we evaluate our tracker that presented in section 3.3. We noticed that tracking of small objects such as guns are rather challenging than tracking the human because the small object easily affected by the lighting condition. Hence, we set up the experiment on tracking of the gun objects instead of the human to evaluate our tracker performance.

Several steps have to be done in order to start the evaluation:

1. We manually annotate the gun objects in each frame of the 5 training videos that used for training gun actions. These 5 training videos consist of total 273 frames. The annotations of objects for these 5 training videos are used to evaluate the tracker only and not for training process in gun action detection.
2. We provide the object location in the key frame of each training video to our tracker and let our tracker track the object location from the keyframe to end frame and keyframe to the first frame.
3. We record the output of our tracker in each frame as correct detection if it overlaps with the ground truth object more than 50% while the other output will record as false-positive.
4. We measure the recall R by using the formula:

$$R = \frac{number\ of\ correct\ detection}{number\ of\ frames\ where\ object\ is\ visible} \quad (5)$$

5. We also record the precision, *P* of our tracker as the percentage of the correct detection.

6. We calculate the F-measure by using the formula:

$$F = \frac{2PR}{P+R} \quad (6)$$

Table 4.1 shows the recall, precision and f-measure of our tracker.

		Tracker
Gun Action	Recall	$\frac{211}{246} = 0.858$
	Precision	$\frac{211}{273} = 0.773$
	F-measure	$\frac{2(0.773)(0.858)}{0.858 + 0.773} = 0.813$

Table 4.1 Tracker Performance

Based on Table 4.1, the recall rate for our tracker is 0.858 and the precision is 0.773 which is high enough to tackle the sparse detection produced by our gun detector. Besides, our experiment shows that our tracker manages to track the object when only single detection result provided. If there are more detections provided, the precision of our tracker will be improved as our tracker will link the detection over time.

Furthermore, another advantage for our tracker is the computational time for computing all the point track is only 1-2 seconds per frame which allows us to apply our tracking algorithm in real time.

4.4 Evaluation on Gun Action Detection

In order to train a gun action classifier which can detect the person that is holding the gun, we manually crop out 25 short videos that contain the gun action from different movies. Each of the videos contains 30 to 100 frames. These 25 short videos contain 147 positive samples and 332 negative samples.

For testing, we crop out 10 short videos from the movies that do not overlap with the training video. Each of the testing videos contains 30 to 100 frames. Figure 4.4 shows the performance of our action classifier.

n=number of human object pairs (270)	Predicted: NO	Predicted: YES:
Actual: NO	TN = 191	FP = 20
Actual: YES	FN = 17	TP = 42

Table 4.2 Confusion matrices on the testing dataset

Based on Table 4.2, we can calculate the precision, recall rate, and misclassification rate of our gun action classifier by using the formulas:

$$Precision = \frac{TP}{Predicted\ Yes} = \frac{42}{62} = 0.68 \quad (7)$$

$$Recall = \frac{TP}{Actual\ Yes} = \frac{42}{59} = 0.71 \quad (8)$$

$$Misclassfication\ Rate = \frac{FP+FN}{Total} = \frac{20+17}{270} = 0.14 \quad (9)$$

The result shows that our gun classifier that trained by using our approach can achieve high precision which is 0.68 and high recall rate which is 0.71. Besides, the misclassification rate of our gun classifier is relatively low. Hence, we can conclude that the interaction feature of the object respect to the human is important to describe an action, for example, gun action.

CHAPTER 5 FUTURE WORK

In order to integrate our system into the CCTV system, our system should be work in real time. According to our observation in implementing the system, the average time for the human detection and object detection in a single frame is 20 seconds so our system is not able to work in real time. Instead of focusing the speed, our currently proposed system is focusing the accuracy of detecting gun in action. Hence, our future work will concentrate on improving the object detection and human detection speed and also the accuracy of the algorithm so that our system can work in real time to aid in monitoring the CCTV.

Besides, our current gun detector is trained by using 250 images only, so in future, we will concentrate on increasing the number of images for training the gun detector to improve the accuracy of the gun detector. Rich object models that trained using a large number of images can overcome the photometric variation and viewpoint variation problem of the objects.

Lastly, our proposed system currently only applicable for detecting gun in the action, so in future, we will train an object detector which can detect the knife in action.

CHAPTER 6: CONCLUSION

In conclusion, we introduce an approach for detecting the gun action by learning the human-object interactions in videos. We explicitly track both object and human and then represent the gun action using the interaction feature such as relative position, relative area and relative motion of the object with respect to the human.

Our experimental result shows that the interaction feature of an object with respect to the human is significant to describe a gun action. However, there is still a roomfor improvement as the training samples for the detecting the gun action is still not enough. Besides, the precision of the gun detector and human detector need to improve by enriching the model.

In term of the contributions of our system, our system can overcome the sparse detection produced by the object detector through tracking the object in every frame. Therefore, the average precision of the object detection in a video can be improved. Besides, our proposed method can be extended and used in another area such as detecting the robbery action.

Lastly, we hope that our approach can successfully implement in real time situation in the future so that the crime rate can be reduced.

BIBLIOGRAPHY

Beaumont, P. (2011). Norway attacks: at least 92 killed in Oslo and Utøya island. [online] the Guardian. Available at: https://www.theguardian.com/world/2011/jul/23/norway-attacks [Accessed 22 Aug. 2016].

Dalal, N. and Triggs, B.(2005). Histograms of oriented gradients for human detection. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) (Vol. 1, pp. 886-893). IEEE.

Everingham, M., Van Gool, L., Williams, C., Winn, J. and Zisserman, A. (2009). The Pascal Visual Object Classes (VOC) Challenge. International Journal of Computer Vision, 88(2), pp.303-338.

Farnebäck, G., (2003). Two-frame motion estimation based on polynomial expansion. In Scandinavian conference on Image analysis (pp. 363-370). Springer Berlin Heidelberg.

Felzenszwalb, P., Girshick, R., McAllester, D. and Ramanan, D. (2010). Object Detection with Discriminatively Trained Part-Based Models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(9), pp.1627-1645.

Grega, M., Matiolański, A., Guzik, P. and Leszczuk, M., (2016). Automated detection of firearms and knives in a CCTV image. Sensors, 16(1), p.47.

Malagón-Borja, L. and Fuentes, O. (2009). Object detection using image reconstruction with PCA. Image and Vision Computing, 27(1-2), pp.2-9.

Numbeo.com. (2016). Asia: Crime Index by Country 2016 Mid-Year. [Online] Available at: http://www.numbeo.com/crime/rankings_by_country.jsp?title=2016-mid&region=142 [Accessed 22 Aug. 2016].

NY Daily News. (2016). Columbine shootings leave 39 dead or injured in 1999. [online] Available at: http://www.nydailynews.com/news/national/high-school-bloodbathgun-toting-teens-kill-25-article-1.822951 [Accessed 22 Aug. 2016].

NY Daily News. (2016). 200 cameras at complex fail to stop rape suspect. [online] Available at: http://www.nydailynews.com/new-york/brooklyn/200-surveillance-cameras-van-dyke-houses-fail-stop-rape-suspect-article-1.289186 [Accessed 22 Aug. 2016].

Pascal.inrialpes.fr. (2017). INRIA Person dataset. [online] Available at: http://pascal.inrialpes.fr/data/human/ [Accessed 8 Apr. 2017].

People.cs.uchicago.edu. (2017). Discriminatively Trained Deformable Part Models (Release 5). [online] Available at: http://people.cs.uchicago.edu/~rbg/latent-release5/ [Accessed 8 Apr. 2017].

Prest, A., Ferrari, V. and Schmid, C. (2013). Explicit Modeling of Human-Object Interactions in Realistic Videos. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(4), pp.835-848.

Security News Desk. (2013). BSIA attempts to clarify question of how many CCTV cameras there are in the UK - Security News Desk. [online] Available at: http://www.securitynewsdesk.com/bsia-attempts-to-clarify-question-of-how-many-cctv-cameras-in-the-uk/ [Accessed 22 Aug. 2016].

Sundaram, N., Brox, T. and Keutzer, K. (2010). Dense point trajectories by GPU-accelerated large displacement optical flow. In European conference on computer vision (pp. 438-451). Springer Berlin Heidelberg.

Tech Times. (2015). This Smart Camera Detects And Alerts You When There Is A Gun In The House. [online] Available at: http://www.techtimes.com/articles/69550/20150716/smart-camera-detects-alerts-when-gun-house.htm [Accessed 22 Aug. 2016].

Xu, Z., Tian, Y., Hu, X. and Pu, F. (2015). Dangerous human event understanding using human-object interaction model. Signal Processing, Communications and Computing (ICSPCC), 2015 IEEE International Conference on, [online] pp.1-5. Available at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7338786 [Accessed 22 Aug. 2016].

UNIVERSITY TUNKU ABDUL RAHMAN

TITLE :AUTOMATIC WEAPON DETECTION IN REALISTIC VIDEO

NAME:NG JOO SIANG | YEAR:YEAR 3 SEM 3 | FACULTY:FICT | COURSE: BSC (HONs) COMPUTER SCIENCE | SUPERVISOR: DR TAN HUNG KHOON

Introduction

- Crime is one of the big problems in the world and worrying aspects in any society since crime is increasing at an alarming rate.
- The increase of the number of incidents with the use of dangerous tools in public area has clearly worried the public
- We are going to enhance the existing weapon detection technique to achieve a satisfaction and more accurate result by applying the human tracking and object tracking in our proposed solution.

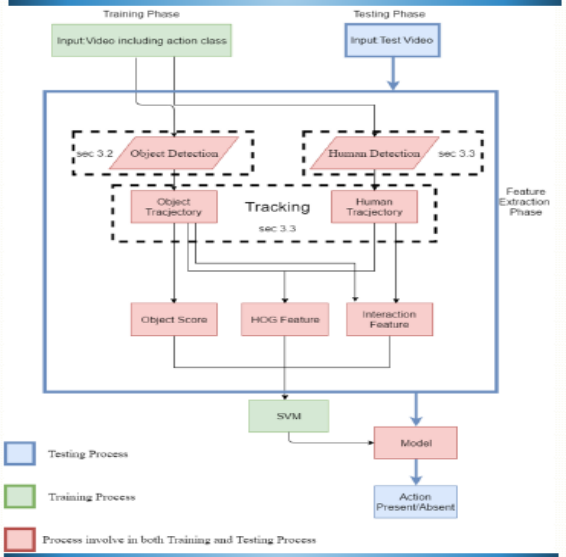
Project Objective & Scope

- We tackle videos which include CCTV videos and movie videos.
- Waist shot of video will be used as the input of our proposed algorithm
- To develop a human and object tracking mechanism
- To design an algorithm that able to extract the feature of the object that holds by the human and classifies the object into weapon or non-weapon by using several types of descriptors.

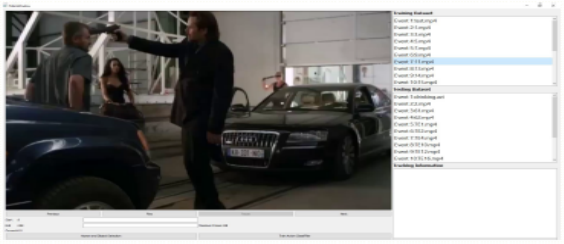
Problem Statement

- The effectiveness of CCTV operators has put into question as there are too many numbers of cameras to monitor
- The CCTV cameras are only play as passive role in the CCTV system which unable to detect crimes.

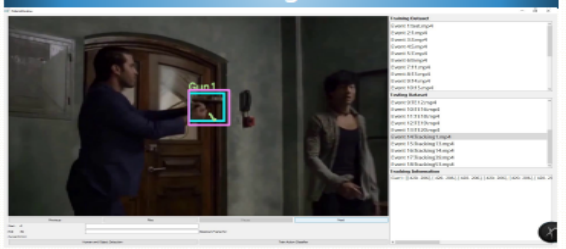
Methodology



Graphic User Interface (GUI)



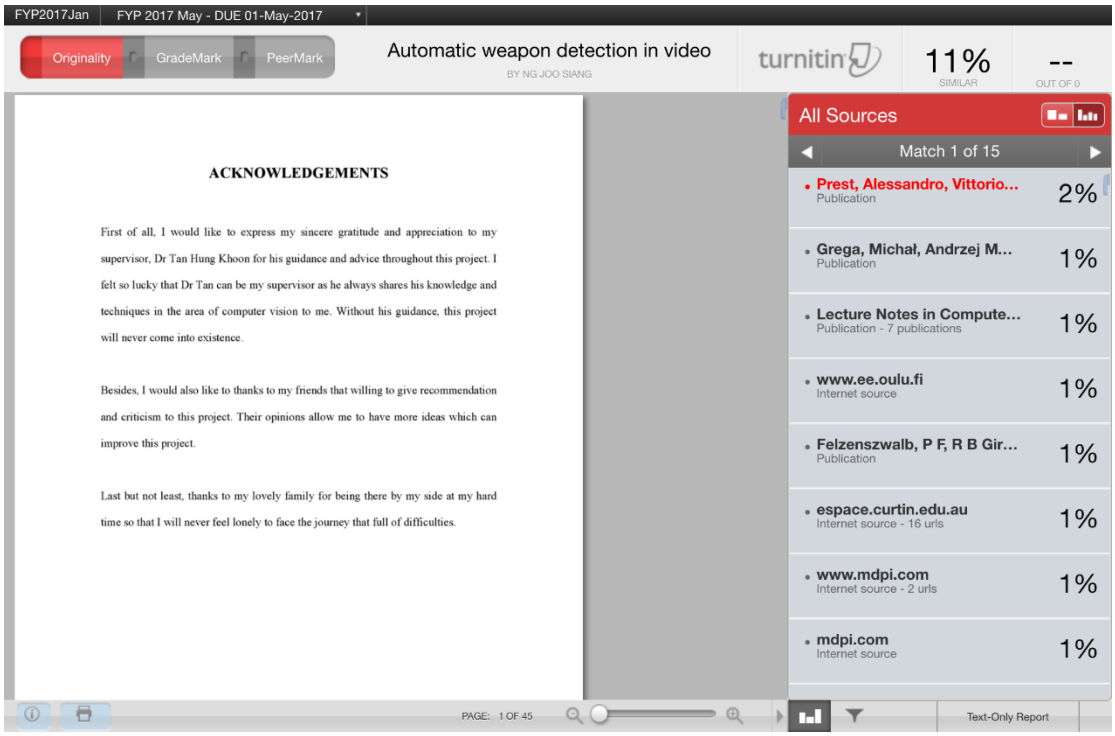
Tracking Result

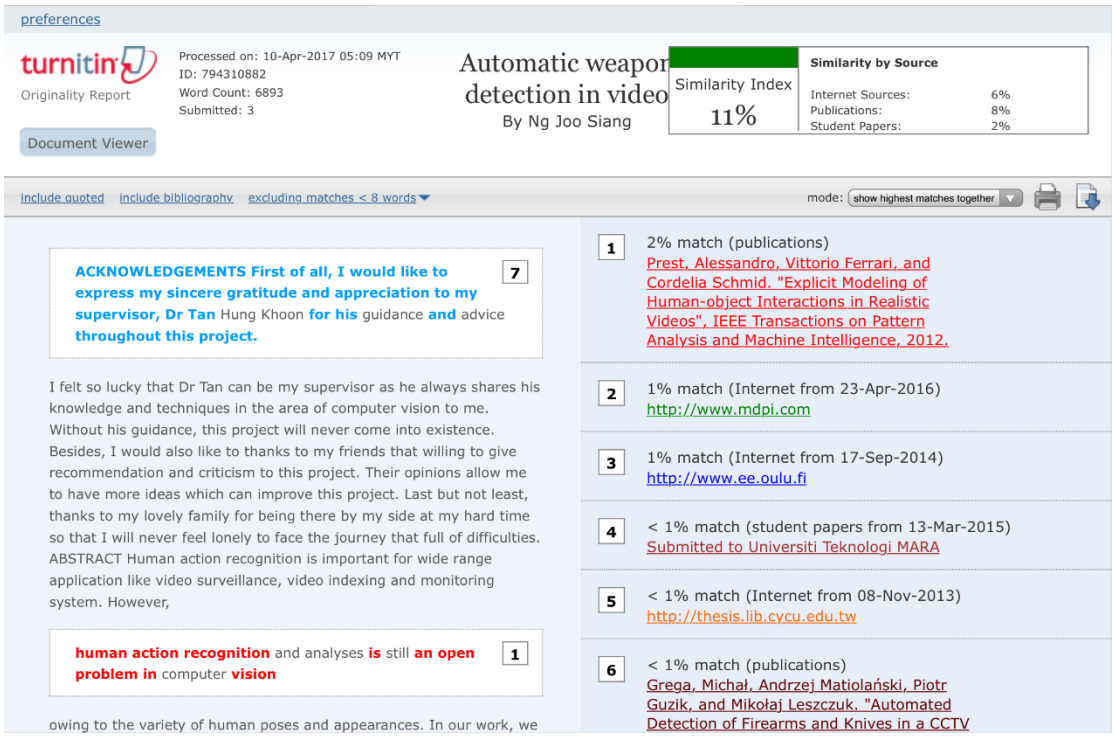


Human Detection & Gun Detection Result

Conclusion

- In term of contribution, our system can overcome the sparse detection produced by the object detector through tracking the object in every frame
- Average precision of the object detection in video can be improved
- Our proposed method can be extended and used in another area such as robbery action.
- Lastly, we hope that our approach can successfully implemented in real time situation to reduce the crime rate.


FYP2017Jan
FYP 2017 May - DUE 01-May-2017
Originality
GradeMark
PeerMark
Automatic weapon detection in video
BY NG JOO SIANG
turnitin
11%
SIMILAR
--
OUT OF 0
ACKNOWLEDGEMENTS
First of all, I would like to express my sincere gratitude and appreciation to my supervisor, Dr Tan Hung Khoon for his guidance and advice throughout this project. I felt so lucky that Dr Tan can be my supervisor as he always shares his knowledge and techniques in the area of computer vision to me. Without his guidance, this project will never come into existence.
Besides, I would also like to thanks to my friends that willing to give recommendation and criticism to this project. Their opinions allow me to have more ideas which can improve this project.
Last but not least, thanks to my lovely family for being there by my side at my hard time so that I will never feel lonely to face the journey that full of difficulties.
All Sources
Match 1 of 15
Prest, Alessandro, Vittorio... Publication 2%
Grega, Michał, Andrzej M... Publication 1%
Lecture Notes in Compute... Publication - 7 publications 1%
www.ee.oulu.fi Internet source 1%
Felzenszwalb, P F, R B Gir... Publication 1%
espace.curtin.edu.au Internet source - 16 urls 1%
www.mdpi.com Internet source - 2 urls 1%
mdpi.com Internet source 1%
PAGE: 1 OF 45
Text-Only Report


preferences
turnitin
Originality Report
Processed on: 10-Apr-2017 05:09 MYT
ID: 794310882
Word Count: 6893
Submitted: 3
Document Viewer
Automatic weapon detection in video
By Ng Joo Siang
Similarity Index 11%
Similarity by Source
Internet Sources: 6%
Publications: 8%
Student Papers: 2%
include quoted
include bibliography
excluding matches < 8 words
mode: show highest matches together
ACKNOWLEDGEMENTS First of all, I would like to express my sincere gratitude and appreciation to my supervisor, Dr Tan Hung Khoon for his guidance and advice throughout this project. 7
I felt so lucky that Dr Tan can be my supervisor as he always shares his knowledge and techniques in the area of computer vision to me. Without his guidance, this project will never come into existence. Besides, I would also like to thanks to my friends that willing to give recommendation and criticism to this project. Their opinions allow me to have more ideas which can improve this project. Last but not least, thanks to my lovely family for being there by my side at my hard time so that I will never feel lonely to face the journey that full of difficulties. ABSTRACT Human action recognition is important for wide range application like video surveillance, video indexing and monitoring system. However,
human action recognition and analyses is still an open problem in computer vision 1
owing to the variety of human poses and appearances. In our work, we
1 2% match (publications) Prest, Alessandro, Vittorio Ferrari, and Cordelia Schmid. "Explicit Modeling of Human-object Interactions in Realistic Videos", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012.
2 1% match (Internet from 23-Apr-2016) http://www.mdpi.com
3 1% match (Internet from 17-Sep-2014) http://www.ee.oulu.fi
4 < 1% match (student papers from 13-Mar-2015) Submitted to Universiti Teknologi MARA
5 < 1% match (Internet from 08-Nov-2013) http://thesis.lib.cycu.edu.tw
6 < 1% match (publications) Grega, Michał, Andrzej Matiolański, Piotr Guzik, and Mikołaj Leszczuk. "Automated Detection of Firearms and Knives in a CCTV

Universiti Tunku Abdul Rahman			
Form Title : Supervisor's Comments on Originality Report Generated by Turnitin for Submission of Final Year Project Report (for Undergraduate Programmes)			
Form Number: FM-IAD-005	Rev No.: 0	Effective Date: 01/10/2013	Page No.: 1of 1

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

Full Name(s) of Candidate(s)	
ID Number(s)	
Programme / Course	
Title of Final Year Project	

Similarity	**Supervisor's Comments (Compulsory if parameters of originality exceeds the limits approved by UTAR)**
Overall similarity index: _____ % **Similarity by source** Internet Sources: _____ % Publications: _____ % Student Papers: _____%	
Number of individual sources listed of more than 3% similarity: _____	
Parameters of originality required and limits approved by UTAR are as Follows: **(i) Overall similarity index is 20% and below, and** **(ii) Matching of individual sources listed must be less than 3% each, and** **(iii) Matching texts in continuous block must not exceed 8 words** ***Note: Parameters (i) – (ii) shall exclude quotes, bibliography and text matches which are less than 8 words.***	

Note Supervisor/Candidate(s) is/are required to provide softcopy of full set of the originality report to Faculty/Institute

Based on the above results, I hereby declare that I am satisfied with the originality of the Final Year Project Report submitted by my student(s) as named above.

_____ _____

Signature of Supervisor Signature of Co-Supervisor

Name: _____ Name: _____

Date: _____ Date: _____