Execution of Coordinate Based Classifier System to Predict Specific Criminal Behavior using Regional Multi Person Pose Estimator

by

Md. Farhan Zaman

17101137

Md. Iftekhar Alam Tousif

17101337

Maliha Monami

17101020

Hazrat Sauda Hossain

17101222

Sanjida Hossain

17101356

A thesis submitted to the Department of Computer Science and Engineering in partial fulfillment of the requirements for the degree of

B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering

BRAC University

September 2021

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at BRAC University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Farhan Zaman	Maliha Monami	Md Iftekhar Alam
Md. Farhan Zaman	**Maliha Monami**	**Md. Iftekhar Alam**
17101137	17101020	17101337

Sauda	Sanjida Hossain
Hazrat Sauda Hossain	**Sanjida Hossain**
17101222	17101356

Approval

The thesis titled “Execution of Coordinate Based Classifier System to Predict Specific Criminal Behavior using Regional Multi Person Pose Estimator” submitted by,

1. Md. Farhan Zaman (17101137)
2. Md. Iftekhar Alam Tousif (17101337)
3. Sanjida Hossain (17101356)
4. Maliha Monami (17101020)
5. Hazrat Sauda Hossain (17101222)

For the summer of 2021, the prerequisite for the degree of B.Sc. in computer science and engineering on September 26, 2021 was accepted as sufficient in part.

Examining Committee:

Supervisor:

(Member)

Md. Khalilur Rhaman

Dr. Md. Khalilur Rhaman, PhD
Associate Professor
Department of Computer Science and Engineering
BRAC University

Program Coordinator:

(Member)

Md. Golam Rabiul Alam, PhD
Associate Professor
Department of Computer Science and Engineering
BRAC University

Head of Department:

(Chair)

Sadia Hamid Kazi
Chairperson and Associate Professor
Department of Computer Science and Engineering
BRAC University

Abstract

There are numerous numbers of issues in society, one of which is crime. While crime refers to a wide range of deliberate, unlawful behaviors, the most archetypal ones involve murder, threatening and violent activities. Its expenditures and consequences affect almost everything but to a certain extent. If we want to prevent crime, we must first identify criminal activity. It is hard to locate unlawful behavior without a lot of effort. With crime surging at an alarming rate, several methods have been developed in the past to predict and prevent criminal activities. However, the methods available currently are not efficient enough to predict the extensive variety of criminal activities that occurs in modern days. To do so, we need to make greater use of technological advancements in order to forecast crime. This paper presents an approach that will be able to detect and predict crime by combining machine learning with a coordinate-based approach. The proposed apparatus integrates existing video footage to detect and analyze human behavior. The system distinguishes between human stances present in the scene in order to detect criminal behavior and subsequently predict crime. Using video processing, the methodology compares human stances with a trained dataset and detects those body positions that may indicate criminal activity.

Keywords:

CBHAC, Convex Hull, Graham’s Scan, Alpha Pose, Ray Casting, Crime Detection, Human Posture, Pose classification.

Dedication

This dissertation is dedicated towards all the front-line workers who are risking their life for the security and betterment of the society.

Acknowledgement

First and foremost, we express our gratitude to Almighty Allah for allowing us to complete our thesis.

Secondly, we would like to sincerely thank our supervisor, Md. Khalilur Rhaman for his support, guidance and constant monitoring, without which completing the thesis was not possible.

Lastly, thanks to our family members specially our parents who supported us constantly throughout the journey and has motivated us to complete this report.

Table of Contents

Declaration i

Approval ii

Abstract iii

Keywords: iii

Dedication iv

Acknowledgement v

Table of Contents vi

List of Figures viii

List of Tables ix

Nomenclature x

Chapter 1: Introduction 1

1.1 Overview 1

1.2 Motivation 1

1.3 Problem Statement 2

1.4 Research Objectives 3

Chapter 2: Background 5

2.1 Literature Review 5

2.2 Algorithm and Framework 11

2.2.1 Alpha Pose 11

2.2.2 Graham Scan (Convex Hull) 14

2.2.3 Ray-Casting Algorithm 16

Chapter 3: Methodology 18

3.1 Data Preprocessing 19

3.2 Data Categorization 20

3.3 Model Description 21

3.3.1 Human Skeleton Key-point Detection 21

3.3.2 Key-point Analysis for Specific Actions 22

3.3.3 Pose Classification from Key-Points..25

Chapter 4: Implementation and result analysis..26

4.1 Extraction of key points...26

4.2 Coordinate Based Human Action Classifier (CBHAC)...27

4.2.1 Pattern Familiarization (PF): 27

4.2.2 Pattern Estimation (PE):..29

4.3 Result analysis: 31

Chapter 5: Conclusion 33

5.1 Limitations...34

5.2 Future Work:...35

Bibliography:...36

List of Figures

2.1 Key Point Representation 12

2.2 Overview of a Convex Hull 14

2.3 Finding the next nearest point from the pivot (red point) 15

2.4 Casting ray from a given point to infinity in a fixed direction 16

2.5 Point inside of polygon 17

2.6 Point outside of polygon 17

3.1 Work Flow Diagram 18

3.2 Data Process Diagram 20

3.3 Data Categorization Diagram 20

3.4 AlphaPose Node Extraction 22

3.5 Scenario 1 23

3.6 Scenario 2 23

3.7 Scenario 3 24

3.8 Scenario 4 24

3.9 Scenario 5 24

3.10 Scenario 6 25

4.1 Line segment that intersects a common point of two hands 30

4.2 Drawing Reverse Line Segment 30

4.3 Sample virtualization for determining D_H and D_O 31

List of Tables

2.1 Results on COCO test-dev 2015 13

2.2 Results on MPII full test set 13

2.3 Multi-Person Pose Estimation (mAP) 13

4.1 Experimentation for value of sigma 32

4.2 Few samples of output 33

Nomenclature

Name	Abbreviation
CCTV	Closed-circuit television
US	United States
FBI	Federal Bureau of Investigation
UCR	Uniform Crime Reports
ATM	Automated Teller Machine
CNN	Convolutional Neural Network
IoT	Internet of Things
PIR	Passive Infra-Red
DSP	Digital Signal Processing
CLAHE	Contrast Limited Adaptive Histogram Equalization
AHE	Adaptive Histogram Equalization
iOS	iPhone Operating System
NJ	New Jersey
GIS	Geographic Information System
PSM	Propensity Score Matching
OR	Odds Ratio
ATT	Average Care on the Treated
LBPH	Local Binary Patterns Histograms
OpenCV	Open Source Computer Vision Library
MSP	Multi-person Stance Predictor
MPII	Max Planck Institute Informatik
COCO	Common Objects in Context
CMU	Carnegie Mellon University
RCNN	Region-Based Convolutional Neural Network
CBHAC	Coordinate Based Human Action Classifier
STN	Spatial Transformer Network
SPPE	Single Person Pose Estimation
SDTN	Spatial De-Transformer Network
NMS	Non-Maximum Suppression
PF	Pattern Familiarization
RMPE	Regional Multi-Person Pose Estimation
PE	Pose Estimation
MSP	Managed Service Provider

Chapter 1: Introduction

1.1 Overview

Crimes all over world has become a daily and common phenomena and people are being murdered, kidnapped in the daytime. When the police comes, the activity of crime has already been done. The privilege of criminal justice and law enforcement experts has been historically resolving crimes. By expanding the use of electronic systems to track crimes, computer data scientists started to assist law enforcement authorities and investigators in speeding up the crime resolution process. Here we adopt an integrative approach to the development of a paradigm combining computer science and criminal law, which might help solve crime more quickly. We will utilize crime activity detection methods in particular to assist identify the patterns of criminal activity through video footage. A system which detects and emphasizes the behavior of aggressive data in videos can only improve monitoring. The system is to collect frame from video footage and detect whether the activity of the person is suspicious enough to be labeled as a crime. This approach has been established on the base of convex hull and ray-casting. The approach involves identifying the points by monitoring the coefficients of wavelets of various scales and calculating the interest range of convex points for extracting the picture region of interest. Also, mentionable that this system has been implemented through manipulation of coordinate base approach which means to determine the convex hull from nodes, and then detecting the pattern from hull.

1.2 Motivation

The deliberate death of one person by another is the gravest conclusion to the range of violent crime. Recent emphasis has been given to the topic of armed violence and the increasing significance of homicide as an indicator. Most young people lose their careers very early in their life when they participate in illegal activities and the rate of crime rises every day. Crime is growing every day in many places of our globe and it is becoming harder. Our world is quite densely inhabited and some nations would punish and jail him for a long period if someone performed a little offense. There are numerous excellent sources on what has happened with violence and crime this year, from criminologists, economists and other data researchers. Jeff Asher, a criminal analyst, provides the latest evidence, examining changes in crime in 51 US towns by 2020 in relation to 2019. He discovered assassinations are 36% higher [1]. Moreover, people are losing their kith and kin due to severely happened crimes. The approach might be helpful to determine the action and take pre caution for the crime. After knowing that somebody is having some weapons or any object that is harmful for lives, people will be cautious. If any criminal activity is detected by the system, crime rate would be decreased. The point of implementing this system is to reduce the crime, save both the victim and the attacker as victim is the sufferer and the attacker is ruining his or her career.

1.3 Problem Statement

When crime can occur is simply hard to tell. In most of the cases by the time law enforcement arrives at the scene, criminals succeed in fleeing. We can all notice if it's pouring, snowing or soaked but crime isn't like the weather. Normally only the offender and the victim and, occasionally, a witness, know this when a crime takes place. Despite the fact that we have an imprecise picture of the crime issue, we are completely conscious of how much crime occurs, who would most likely commit it, and who would be affected by it thanks to a range of resources. According to the FBI, there have been 1,246,248 violent crimes and 9,082,887 property crimes in 2010, for a total of roughly 10.3 million [2]. This is the official criminal record of the nation and is a lot of crime by any standard. However, this statistic is significantly lower than it seems, because more than 50% of any and all victims do not report their crimes to the police, and so the officers are unaware of their offenses. This failure to record crime constitutes a serious validity concern for the UCR (Uniform Crime Reports). There are a number of other issues. First, white-collar crimes are excluded from the UCR and attention is distracted from their impact. Secondly, the number of offenses listed in the UCR are affected by police practice [2]. The police, however, do not treat every citizen complaint as criminal offense. They don't have time to do it occasionally, and the citizen doesn't think sometimes. The FBI does not consider the report to be a crime if it does not record it. Even if the real amount of crime does not vary, the reported crime rate will increase or drop if more complaints are registered by the authorities or if fewer complaints are received by the authorities. In the past twenty years, this has led to criminal reporting scandals since the police departments in numerous large cities have neglected to register many crimes or apparently degraded others so that there seems to be a reduction in the crime rate. With the third issue, the official homicide rate will fluctuate again, despite the fact that the actual crime does not. A third issue arises if perpetrators become more or less reluctant to report their crimes to the authorities. In short, the system could provide a remedy by forecasting the state of illegal activity. Summing the problem, solution could be the system by predicting the posture of criminal activity.

1.4 Research Objectives

We hope to achieve the research goals and those are: 1) Categorize weapon related poses, 2) Determine a confidence level between human posture and criminal activity.

1) Categorize weapon related poses: A framework based on RMPE to collect and monitor data which are related to crime or people holding weapons and any other objects that are responsible for intended crimes. We are to develop a system which uses RMPE as a framework and based on this, we will implement convex hull of nodes and detect the hull. First, we will collect data from video footages. Then the data will give us the detailed postures and we will categorize the exact posture through our system. The details of the proposed model is discussed in the Methodology section. This framework will facilitate collecting data from the videos which are related to crime or planned crimes.

2) Determine a confidence level between human posture and criminal activity: Human posture is not the only thing that will define if a person is committing any crime or murder or

intension of any criminal activities. A person may hold anything but not intending to do any crimes. So, in our system we are determining the confidence level of human posture and criminal activity which indicates the exact posture of pretending the crimes. This will be done by the help of simultaneous analysis through Convex Hull If we could find ways to foresee crime in detail before it occurs, or if we could develop technologies to assist police people, it would relieve the burden on the police and help prevent crime.

Chapter 2: Background

2.1 Literature Review

Machine Learning being one of the most influential sectors of technology, there has been a plethora of research conducted in this area, many of which focuses on detecting and predicting crimes. Some of the notable works done previously regarding this sector uses motion analysis, image processing, gesture detection etc. The system developed by Dorogyy, Kolisnichenko and Levchenko [3] has a distinctive approach towards criminal activity prediction. The system mostly focuses on suspicious activity prediction through data analysis. The most intriguing part of their system is that the system utilizes the common methods used by criminal experts to detect violent crimes. The module requires both primary and supplementary information in order to operate. Primary information refers to the data collected directly from potential crime scenes, which incorporates both visual and auditory information. The secondary information is extracted from different databases. These include police or hospital databases, social networking sites, statistics etc. The system takes in data from both sources and analyses them to categorize an event. In doing so, the system interprets auditory data and recognizes different noises from their trained dataset, which may include loud speech, shouting, gunshot etc. The system also uses a similar method to detect crime from visual data. Through image processing the system recognizes different objects. By combining both of the primary information with secondary data, it classifies the incident. While the system is unique as it incorporates both image and sound processing, it has some limitations as well. Firstly, the system requires high level auditory data to recognize crime, which affects its efficiency. Aside from that, the system may not be applicable for public places as detecting unusual noise in a noisy crowded place can be a challenge. The system also uses face detection and from secondary data generates crime history, which can be ineffective in case of people with no criminal records.

A research by Sikandar, Ghazali and Rabbi [4] focuses on detecting potential criminal activity in ATM booths using image processing. In their research paper they analyze previous image recognition techniques and categorize them in accordance to their weakness and strength. The paper demonstrates a detailed research on conventional image processing methods and provides future research direction. While their research is quite informative as it has extensive analytical research on existing methods, they do not suggest any new method in case of crime detection. On another paper by Ghazali and others [5], a method to detect covered faces was developed. By comparing records of ethnic background, face structure, and facial ratio, their technique can recognize entirely covered, partly obscured, and bare faces. This method is based on the statistics that in most cases of ATM theft, the criminal has either fully covered or partially covered face. By detecting covered or partially covered faces from ATM surveillance cameras, the system can classify the scene. While their approach is somewhat effective in high-security closed indoor areas such as ATM booths, its effectiveness alters in outdoor areas. Besides, its functionality drops significantly in case of people covering faces for religious or health related issues, which can lead to false alerts.

In another paper, Goya and others [6] have devised a system of crime detection using motion analysis. Their system analyzes human behavior through the segmentation of kinetic objects. Their paper detects objects through CCTV camera footage and categorizes objects and extracts features. Using the trained models, the incidents are grouped and based on that classification, authorities are alerted if needed. The system detects moving humans and objects by comparing their position. If sudden movement is detected or human chasing is detected through motion analysis, then it categorizes the incident as criminal behavior. Just like the previously mentioned articles, this system is not efficient in public places, though it functions effectively in secluded areas. Also, the possibility of false alert is quite high as it detects moving objects and humans to determine criminal activity. Again, in another article by 'Feba Thankachan George' and others, describes that along with our intellectual judgement, some common indicators of non-verbal movements may help to successfully recognize a subject with suspicious intentions [7]. In order to detect suspicious non-verbal movements associated with limbs or torso, the device uses Jetson TX1 development kit from NVIDIA ® and several body cameras. This system focuses on the gestures on humans based on their hand placement. However, one of the biggest drawback of the system is that it is quite costly, which makes it difficult to implement on a wider range. Mohammad Nakib and others [8] showed a method for recognizing hazardous objects in images and generating a prediction of whether or not a crime has occurred. They used CNN (Convolutional Neural Network) to detect knives, blood and guns from an image. However, the accuracy of the system is questionable as simply object detection cannot be a proper indication of criminal activities as carrying weapon is legal in many countries [9] and common objects can also be used as a weapon [10].

A system developed by Saranu and others [11] aims specifically on preventing theft inside a house and is based on IoT structure of management. The system uses PIR sensor to detect sudden changes in infra-red energy and camera to detect movement of any object, temperature sensor for detecting changes in temperature. The system uses a Raspberry-pi to analyze the readings sent by the sensors and the camera in order to detect the presence of an intruder in the house and alert the owner on the basis of the result. The system also sends the video footage of the supposed intruder to the owner to observe the scenario and on the basis of that activate the defense system, which is basically a solenoid valve which will emit Chloroform gas to make the intruder unconscious. Owner can then check through the camera whether the thief has been successfully neutralized or not. The system is pretty useful for households, but it does not provide any aid in terms of crimes that are done openly. Moreover, the system itself cannot differentiate whether it is a thief or someone from the house. It has to send video to the owner to observe which can be time consuming. The chloroform gas that it sprays has to be in a limited quantity. Too much spray may cause fatal consequences for the thief as Chloroform can be fatal and possible carcinogenic, according to Health Protection Agency [12].

Patiana Intani and others [13] have developed a system using a more traditional method. They are using CCTV footage and signal processing to determine intruders in an office environment. The system divides the CCTV footage into two comparable portions. One is the background frame where there is no sign of movements and no presence of humans in a room. Another is the sample frame which determines the arrival of a person or a movement. If the background frame changes and any person arrives in the room, the difference in change in frames from the previous still state to the new one gets processed using the digital signal processing (DSP) method which then confirms that there actually is someone in the room. Repeated change in frame can also determine whether there is a fight going on inside the office or not. After figuring out an intruder, alarm sets off and the security guards get notified. The major drawback of this system is that it uses pixel processing to determine result. It cannot be used outside as there are frequent changes of frames and pixels. So, there might be a high possibility of getting false alarms.

Jianyu Xiao and others [14] have developed a system which basically aims at establishing new methods for forensic video analysis to help criminal investigation. First and foremost, they recommended a forensic data analysis system that uses an effective video / picture improvement technique to improve low video performance accuracy. To improve the efficiency of CCTV footage for the use of remote forensic analysis, an adaptive video enhancement algorithm based on CLAHE is implemented. A deep-learning object recognition and tracking algorithm is recommended to help the video-based forensic investigation that can detect and classify possible offenders, instruments, etc. from footage. Specifically, a video-based digital testing platform is presented that addresses difficulties such as low-quality footage, the construction of linkages between items and accessible digital data, video footage identification procedures, and smart techniques that may be employed in modern digital forensics. A full quality enhancement technique is provided for low-quality video, including adaptive histogram equalization (AHE), contrast focused AHE (CLAHE), and others, as well as a comparative scenario to evaluate the various algorithms. In video, a deep learning-based object recognition system is described that may be used to construct relationships between objects, individuals, and their behaviors using accessible footage. The primary goal of forensic video processing is to classify powerful pieces of evidence at multiple stages. In this article, from the perspective of forensics, they concentrate on the contents of the film and establish appropriate video processing techniques. But the mission of this project did not serve properly as the crime was not detected instantly through video as they focused on CLAHE.

Worawut Yimyam and others [15] have developed a system that will detect faces of criminals through CCTV cameras. This paper proposes facial recognition technology for offenders through CCTV surveillance. They used fifteen pictures of the experiments. The findings revealed that both single face detection and group face detection were observed five times in the experiment. As a consequence of the increasing crime and offenders who have left, it is commonly used, many of them are still alive to come to the production of the software through the picture from the CCTV to examine face recognition with picture technology processing. This will encourage police officers and prosecutors to learn that the discipline is easier and more effective. No interruptions will be built via the CCTV face recognition and tracking device. To optimize the image processing, a CCTV camera was added. In the method, there are three sections to use: facial detection and face recognition, and the section on warning. Using facial and empirical identification Human recognition diagnosis and usage. It is determined by the face of each person. They would be able to discern who the user is when using the device. It will record the images found and can be seen on Android, iOS, etc. Next, for indoor use or by venue, this project is necessary. The monitoring system can be uniquely defined, and CCTV can be mounted. But the problem is created in the system when a subtly bent face, bent face, low light, or night spot causes the defective image. The identification is not accurate with an average of 35-55 percent of the face inclined beyond 90 degrees.

Eric L. Piza [16] has developed a project to assess the effects of CCTV in Newark, NJ, across three distinct types of crime: car fraud, car robbery, and violent crime. CCTV view sheds were research units, denoting camera line-of - sight. Through the digitization of live CCTV footage inside a geographic information system (GIS), view sheds for treatment units were developed. Monitor view-sheds were built from Google maps using GIS software and aerial imagery. PSM was used to match test and control instances. The effect on the figures treated was calculated by means of odds ratios and average therapy. Findings lend CCTV limited support as a barrier against car theft while having little impact on other forms of crime. These findings indicate that for municipalities attempting to target automotive crime, CCTV tends to be a feasible choice. The CCTV effect was calculated by odds ratio (OR) and average care on the treated (ATT) statistics in Piza's project. As a barrier to car theft, the results include moderate evidence for CCTV. Against other crime forms, CCTV was inefficient. Through use of PSM to evaluate the circumstances of a quasi - experiment was the paper's most significant contribution, though preserving group equality improves the accuracy of experimentation, the direct CCTV effect of a randomly generated experiment assessment will be extremely beneficial to the discipline. Although considering the difficulty of carrying out a true CCTV experiment, random assignment can be possible in some situations. Installment of Surveillance cameras undertaken in Newark on four distinct occasions over a three-year long time period to enable the creation of a wireless network to display camera video. In his research, since he just concentrated on car theft by GIS, where practically all sorts of crimes will not be decided by his research, it would not be possible to apprehend any weapons or murders.

Matthew P. J. Ashby [17] based his study on significant research into the effectiveness of closed-circuit television (CCTV) as a crime-prevention tool, but less on its worth as an investigative tool. This study examined 251,195 offenses recorded by British Transport Police on the British railway network between 2011 and 2015 to see how often CCTV offers usable evidence and how this is modified by circumstances. In 45 percent of cases, investigators had access to CCTV, which was deemed useful in 29 percent (65 percent of cases in which it was available). Except for drug/weapons possession and fraud, having useful CCTV was linked to a much higher chance of a crime being solved. Images were more likely to be available for more serious offenses, but were less likely to be available for incidents that occurred at uncertain times or in specific locations. Although this study focused on crimes committed on trains, it indicates that CCTV is an effective investigative tool for a variety of crimes. The utility of CCTV is limited by a variety of variables, the most notable of which is the proportion of public locations that are not covered. In his study, he makes several proposals for improving the utility of CCTV. A source of data on the investigative usefulness of CCTV was necessary to answer the research issues addressed in this study. To account for this shortcoming, future studies on the subject should distinguish between the availability and utility of CCTV. These findings lead to a number of practice suggestions. The most crucial point is that if investigators have access to CCTV, it is likely to be relevant in a significant number of cases. CCTV looks to be a powerful investigative tool, at least in the area of railway crime, especially for more serious offenses. As weapon related crimes are occurring here and there, it is now obvious to think about how to reduce murder or intensity to murder by weapons where it is not practiced in Matthew’s paper.

An automatic present generation for criminal databases was suggested in a study by Piyush Kakkar and his colleagues [18], using a recognized Haar feature-based classifier algorithm. This technology will be capable of detecting and recognizing faces in real time. Accurately locating the face remains a difficult challenge. Researchers have commonly employed the Viola-Jones architecture to recognize the position of faces and items in a picture. Public groups, including such OpenCV, provide facial recognition classifiers. In an OpenCV technique for face identification, they employed Haar functionality cascade classifiers. It's a computer method in which a cascade function is learned using a large lot of bad pictures. It's then applied to additional pictures to detect things. Face recognition was also accomplished using Local Binary Histograms (LBPH). The cascade classifier rejects many samples in the first node classifier with an efficient time. So, it can be challenging to identify in a very efficient way, and the movement cannot be detected thoroughly as crimes of murder, attack, and fighting are outside of this research work.

After analyzing the existing papers, it can be concluded that most of the approaches used common weapon detection to detect malicious activity. While it can be effective in certain cases, in most cases these systems are not applicable as carrying weapon itself cannot be termed as a suspicious activity. Keeping that in mind, in our research we developed a model based on detecting potential threatening poses in order to detect criminal activity. Another common approach in the existing models is the use of motion sensors to detect intruder, which is mostly applicable in indoor areas only. However, the system proposed in this paper does not rely on changes in motion. So, it can be implied on both indoor and outdoor situations. Again, another important aspect of the existing systems is that many of the systems were not cost effective, making the implementation challenging. Taking that into consideration, we have tried to develop a method which is both cost effective and easy to implement.

2.2 Algorithm and Framework

2.2.1 Alpha Pose

Alpha Pose is the first open-source multi-person stance predictor (MSP) to accomplish 72.3 mAP on the COCO dataset and 82.1 mAP on the MPII set of data [19]. Stance prediction is a type of computer imaging that predicts and detects a specific individual or object's presence. This is done by analyzing a human's posture and orientation. The basic mechanism is to start with a human detector, then estimate the components, and then calculate the stance for each human. We are utilizing the alpha pose method for human stance recognition. However, the alpha pose algorithm's general working concept is that it tries to find a set of coordinates for each joint in the human body, referred to as a key point, in order to describe a human stance from an image, and then combines these points to create various ways of modeling the human body. There are several methodologies for modeling the human body, including the skeleton based framework, the contour based framework, and the volume based framework. As well as, to attain stance prediction, top down and bottom down methods are commonly implemented. For the sake of this project, a top-down approach is being used for stance prediction and a skeleton-based framework in order to model the human body. In a top-down method, the human will be detected first from the existing frames. It will then attempt to predict key points for each identified human in that frame before attempting to construct the skeleton. Finally, 17 key points of the human will be examined for assembling the skeleton. These 17 key points are as follows,

- Key Point 1: Nose
- Key Point 2: Left Eye
- Key Point 3: Right Eye
- Key Point 4: Left Ear
- Key Point 5: Right Ear
- Key Point 6: Left Shoulder
- Key Point 7: Right Shoulder
- Key Point 8: Left Elbow
- Key Point 9: Right Elbow
- Key Point 10: Left Wrist
- Key Point 11: Right Wrist
- Key Point 12: Left Hip
- Key Point 13: Right Hip
- Key Point 14: Left Knee
- Key Point 15: Right Knee
- Key Point 16: Left Ankle
- Key Point 17: Right Ankle

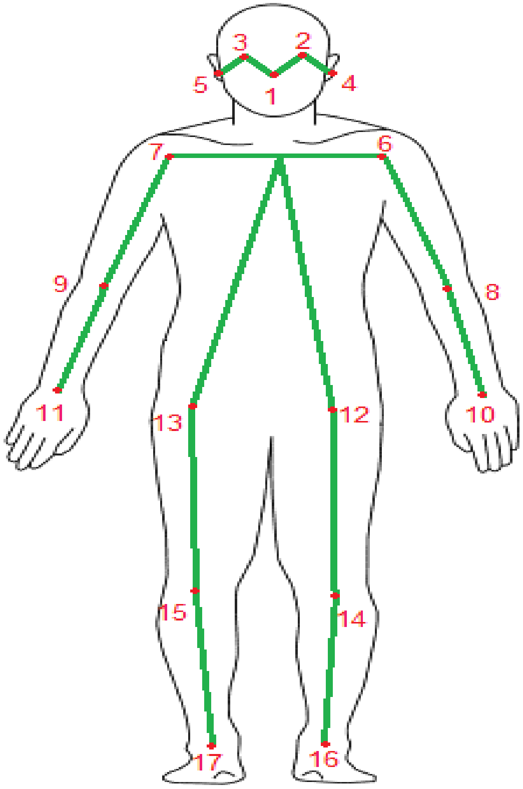


Figure 2.1: Key Point Representation

Each key points of the skeleton has their own confidence score and each skeleton has its cumulative confidence score. So, through detection of human skeletons, Alpha Pose produces results in the following format.

(X coordinate of Key point-1, Y coordinate of Key Point-1, Confidence Score of Key Point-1)

:

:

(X coordinate of Key point-17, Y coordinate of Key Point-17, Confidence Score of Key Point-17)

Cumulative Confidence of key points from the skeleton

Method	**AP @).5:0.95**	**AP @0.5**	**AP @0.75**	**AP medium**	**AP large**
OpenPose (CMU-Pose)	61.8	84.9	67.5	57.1	68.2
Detectron (Mask R-CNN)	67.0	88.0	73.1	62.2	75.6
AlphaPose	72.3	89.2	79.1	69.0	78.6

Table 2.1: Results on COCO test-dev 2015[19]

Method	**Head**	**Shoulder**	**Elbow**	**Wrist**	**Hip**	**Knee**	**Ankle**	**Ave**
OpenPose (CMU-Pose)	91.2	87.6	77.7	66.8	75.4	68.9	61.7	75.6
Newell & Deng	92.1	89.3	78.9	69.8	76.2	71.6	64.7	77.5
AlphaPose	91.3	90.5	84.0	76.4	80.3	79.9	72.4	82.1

Table 2.2: Results on MPII full test set[19]

Method	**Head mAP**	**Shoulder mAP**	**Elbow mAP**	**Wrist mAP**	**Hip mAP**	**Knee mAP**	**Ankle mAP**	**Total mAP**
Detect-and-Track(FAIR)	67.5	70.2	62	51.7	60.7	58.7	49.8	60.6
AlphaPose+PoseFlow	66.7	73.3	68.3	61.1	67.5	67.0	61.3	66.5

Table 2.3: Multi-Person Pose Estimation (mAP)[19]

2.2.2 Graham Scan (Convex Hull)

The convex hull of a form is the shortest convex set. The convex hull is the subdivision of all convex sets of a given domain of Euclidean space, or a collection of all the vertices in the domain. The convex shell may be seen as the form of a rubber strip wrapped around a subset for a limited subset of the plane. In several disciplines, convex hulls have broad applicability. They are part of the two-dimensional image representation of the two data and create randomization rules in the most robust statistics as the outermost outline of Tukey's depth.

A facile polygon is divided between the convex hull and the supplied polygon, one of which being the polygon itself. The remaining parts are termed pockets, which are contained by a polygon chain and a single hull convex edge. The tree of convex disparities, a layered representation of the polygonal curve, is produced repeatedly for each compartment.

There are various types of convex hull algorithm but in our project, we have used only one algorithm which is Graham Scan. Graham's scan is a method for determining the convex hull of a finite number vertices on a plane. This algorithm has a time complexity of O(nlogn). Each point in the sorted array in succession is taken into account in the method. Firstly, the turn from the two vertices immediately prior to the point is detected. Basically, whether the turn from the point to be predicted is leftward or rightward is detected. If the turn on the right is not in the convex hull, and it's in it, the second-to-last point. At this point, the algorithm travels to a number of points in the sorted order subtracting all points found inside the hull, which will then be decided by the set of the last point and the two points immediately prior the point found within the hull, which is replicated till the set of "left turn" is found; the point is not taken into account.

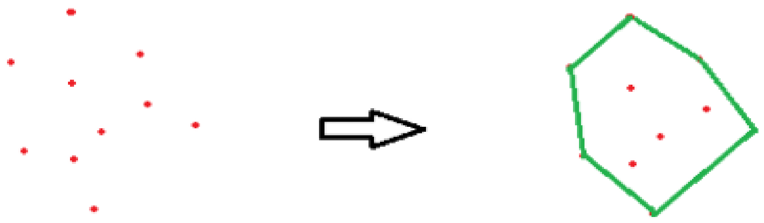


Figure 2.2: Overview of a Convex Hull

This algorithm can be divided into two phases which are phase 1 and phase 2.

Phase 1(Selecting Points): We find the lowest spot first. The objective is to sort the pre-process points to the lowest point. When the points are sorted, a simple, closed route is established.

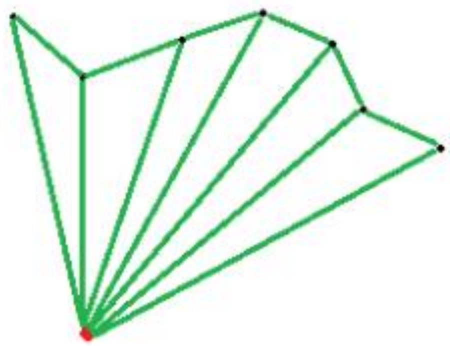


Figure 2.3: Finding the next nearest point from the pivot (red point)

Phase 2(Take or refuse points): Once the track is closed, the following step is to cross the road and eliminate concave spots. There is an issue of how to determine which item to delete and which to maintain. Here again, guidance helps. Convex Hull always comprises 1st and 2nd points in a sorted manner. For other points, we monitor the last three points and calculate their angle. Allow 3 points to be prev(p), curr(c) and following (n). If these points are not directed in counterclockwise, we will discard c, else we will maintain them.

In our project, we used this algorithm for better solution. The nodes which we are getting, needed to be normalized first which means we scaled the nodes and brought it to a limit. After normalizing the nodes, we did augmentation which means we scattered one points to multiple points by noising. So, for one point, we got multiple scattered nodes and for that it needed a boundary. So, this boundary had been created by the algorithm of Graham Scan algorithm of Convex Hull.

2.2.3 Ray-Casting Algorithm

Ray-casting algorithm essentially is used to assess the position of a point in respect to a polygon. This algorithm uses a simple approach to determine whether the position of a point within the area specified by a polygonal curve or on the outside of it. The algorithm detects that by casting a ray from the point to infinity in a fixed direction and counting the number of times the ray intersects with the polygon or the closed curve. If the summation of all the intersecting points are even, the given point is considered to be outside of the given polygon, or else it is considered to be inside the polygon.

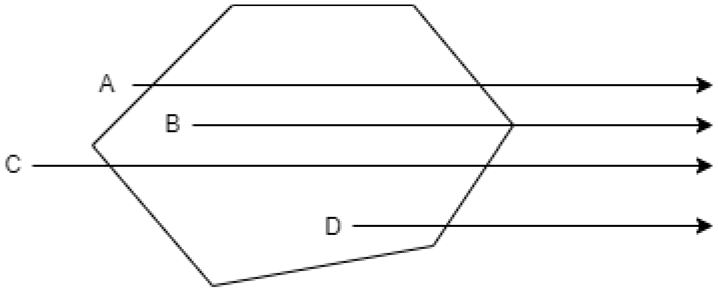


Figure 2.4: Casting ray from a given point to infinity in a fixed direction

This algorithm is based on the Jordan Curve Theorem that states that any simple curve divides the plain in two regions, namely- inside the polygon or outside the polygon. This problem is solved by determining whether the ray is absolutely on top of the ray and only calculating the intersection if the polygonal side's second vertex is well below the ray. On the other hand, the path formed by connecting two points belonging to different classes X and Y must intersect P.

This algorithm first creates a ray from a given point and then sequentially for each side of the polygon checks whether the ray crosses the side. Then it detects whether the total of the intersecting points is even or odd and thus conclude whether the point lies within a polygon.

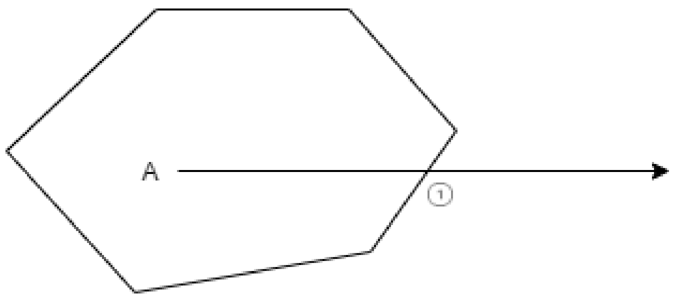


Figure 2.5: Point inside of polygon

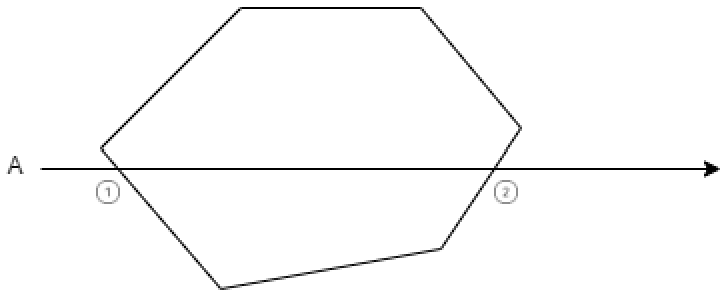


Figure 2.6: Point outside of polygon

However, this algorithm has certain drawback as it cannot be applied for all the points of the plain. For instance, if the point is on the intersection of two sides of the polygon, the algorithm will show incorrect result as in this case, for a point inside the polygon the ray will intersect two sides of a polygon at the same time and thus, conclude that the point is outside the polygon. Another problem is when horizontal rays fall directly on top of a side of a polygon. This problem is solved by determining whether the ray is absolutely on top of the ray and only calculating the intersection if the polygonal side's second vertex is well below the ray. Another issue with the algorithm is that when the point lies too close to the sides, it may show rounding error. However, this error is negligible as the overall speed of this algorithm is way higher than alternative algorithms.

Chapter 3: Methodology

The rapid advancement of technology is no doubt a blessing for the current generation. Security is becoming the demand of the century right now. People want their asset to be safe and if any threat is to occur then rapid action is desired to be taken against in order for the prevention. However, for reasons like keeping hostage or threatening lives, it becomes difficult to reach the security service in certain situations. As a result, a desire for an automated system arises that detects and forwards alarming activities instantly.

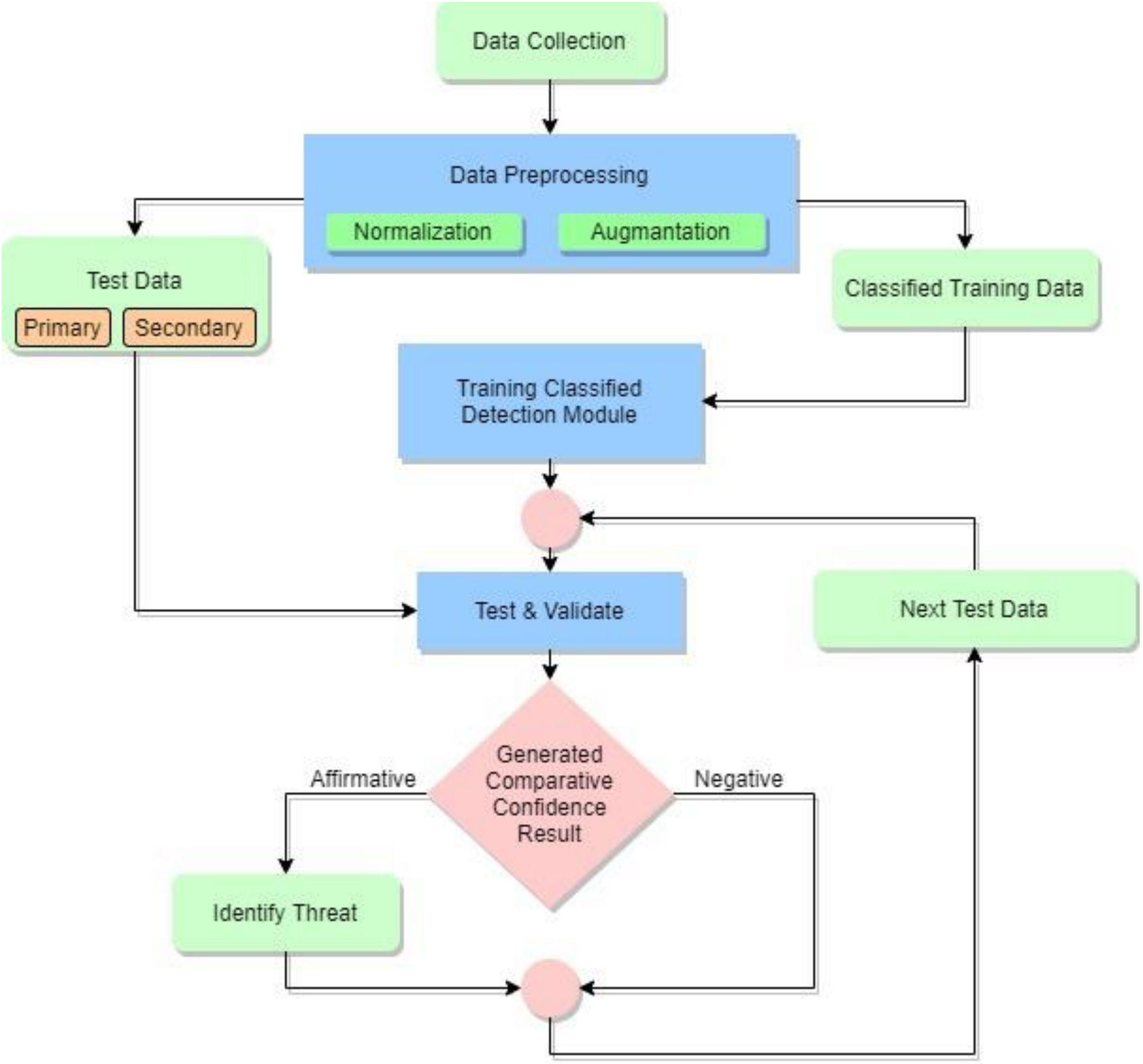


Figure 3.1: Work Flow Diagram

Keeping this principle in mind the entire model is developed. Above is the work flow diagram showing the entire blueprint of the approach is a sky view manner. The first task is to identify and collect relevant data for training the module. After data collection is completed, the dataset will be processed, classified and filtered. After that the data will be divided into two parts, one for testing and the other for training. The training data will then be the input for training the module. After completion of training, the module will be ready for testing. The test data will be given in the module and the module will compare the test data with the classified weights that it has generated. The model will check the data and determine which classification it falls under. Eventually a comparative confidence output will be generated which will determine whether the action is alarming or not. If the result shows affirmation, the action in the data is alarming and will be selected as a threat. Otherwise it will be detected as negative meaning the action does not pose any threats. After making the decision the module will move to check the next test data and similarly determine its threat level. Thus, the module will be running in an iterative manner.

3.1 Data Preprocessing

Preprocessing of data is the process of transforming unprocessed data into an understandable format. Real-world data is usually insufficient, irregular, and/or missing in particular traits or patterns, as well as being riddled with errors. Preprocessing data is a tried and true means of resolving such problems. In order to reach a high confidence level, data preprocessing is a much required approach. Without a proper process of the acquired data, it is very likely to get false outputs. To prevent that from occurring the dataset collected for the module will be thoroughly processed.

The data that is being collected is real world images of human actions that has the potentiality to express feasible criminal activities. The nodes of human joints are being used as coordinates to read a pattern and identify potential criminal actions. However, as the dataset is raw there occurs difficulties in finding a pattern for similar type of actions because of the scaling mismatch of different images. So, in order to generate a similar pattern it is necessary to scale down the similar data nodes in a bounded region. Thus, the concept of normalization is being used. Regardless of difference in the scale of images, normalization narrows down all the nodes in a common region. As a result, patterns can be generated for similar type actions with ease. After that the similar node patterns are being augmented in order to improve the accuracy of the module. The more the data is being augmented, the thicker the region gets which helps in identifying similar but slight off-patterned actions without much difficulties.

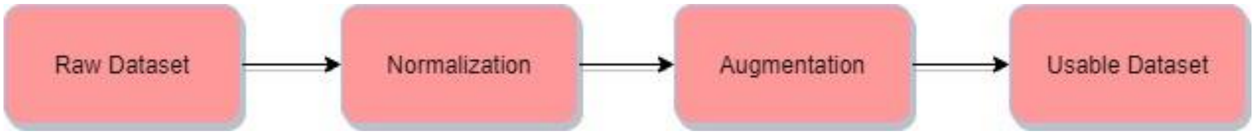


Figure 3.2: Data Process Diagram

3.2 Data Categorization

After processing and extracting out the usable dataset, it will be go through some categorization. The categorization is required in order to know what type of criminal action is being occurred and also invoke a comparative analysis between normal actions and criminal actions. This comparison is beneficial for improving the accuracy of the module even further.

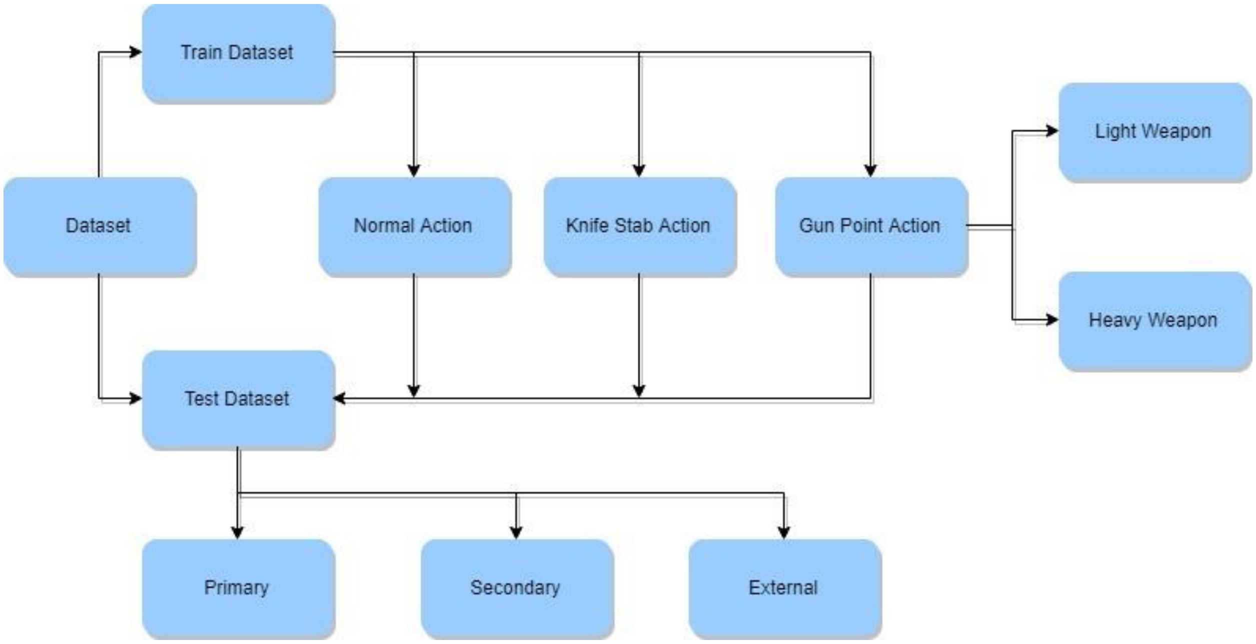


Figure 3.3: Data Categorization Diagram

As shown in the figure above, the collected dataset will be divided into two portions. The division will be in a 70:30 ratio manner, 70% of the dataset will be given for training and 30% will be for testing. The datasets will also be classified into three actions. Having said crime can occur in many forms, two of the standard forms, knife stabbing and gun pointing have been taken for the module training. Moreover, as there are different types of guns like pistol, rifle, shotgun etc., there are various ways to hold and point them. So, in order to get a clear and accurate result, the gun point category has been sub-categorized into light and heavy weapons. In addition to that, actions like standing straight, hands up, walking etc. are taken as normal action in order to compare and make the confidence of the module output even higher.

The testing dataset will also consist the similar categories. Furthermore, the test dataset will consist of three types of data.

- Primary: Data of the Trained 70% of the Dataset
- Secondary: Data of the Remaining 30% of the Dataset
- External: Data collected outside of the Dataset

These three types of test data is essential for checking whether the trained module is working properly or not. Both normal and criminal actions will be checked in the testing phase.

3.3 Model Description

The previous section mentioned how the data is being processed and filtered in an abstract manner. This section will describe the ways of data processing in a deep level. The usage plan of Alpha Pose, Graham Scan and Ray Casting modules will be explained. These modules will aid in extracting the nodes for normalization and augmentation as well as play a part in generating the prediction.

3.3.1 Human Skeleton Key-point Detection

Alpha Pose is used as the RMPE framework module to detect the nodes of human body joints. The module is able to detect multiple human figures in a frame and extract 17 nodes. These nodes are stored in a csv file and each node consists three values, X coordinate Y coordinate and confidence of a particular node. Through brief analysis it has been determined that in order to predict gun pointing and knife stabbing actions, only 8 of the upper body nodes of any human satisfies the requirements. The required nodes are both Shoulders, both Elbows, both Wrists and Left and Right portion of the hip. So, the RMPE framework has been modified in a way that out of the 17 nodes, the required 8 nodes are being extracted.

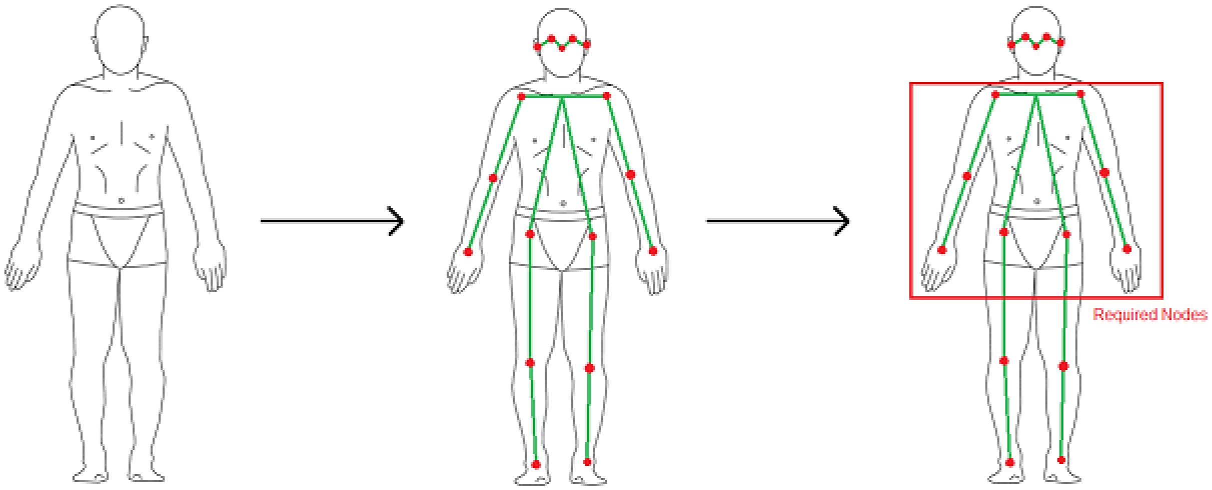


Figure 3.4: Alpha Pose Node Extraction

The above figure shows the process of collecting the required nodes from the RMPE framework. These nodes will be placed in the self-developed Normalization and Augmentation algorithm. Generated output through that algorithm will then be placed in the Graham Scan module for constructing the convex hull. Upon creating the convex hull for each nodes, the weights of training phase will be generated. These weights will be used in the testing phase for predicting each actions.

3.3.2 Key-point Analysis for Specific Actions

Through RMPE and Graham Scan, a particular human will have 8 convex hulls, each indicating the possible area for a node to occur. These set of convex hulls will be called weights. Each weight will express a pattern of an action like gun point or hands up. If a test data matches the pattern of gun pointing better, then it will be considered as a criminal activity. Otherwise it will be treated negatively like a normal action. That being said, there will be certain situations where test data might not have all 8 nodes available but criminal action is occurring. For example, a thug using one hand and shooting while taking cover behind a car. Observing a frame like that will provide only the nodes of shoulders and one hand. Thus only 4 nodes can be obtained in such scenario. Keeping that in mind the testing phase will be developed considering 6 scenarios.

- The first scenario is where all 8 nodes will be compared with all 8 convex hull region.

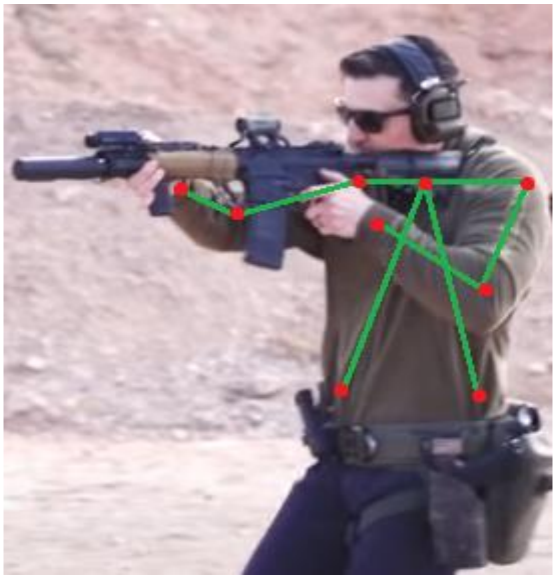


Figure 3.5: Scenario 1

- The second scenario is where 6 of the nodes will be compared with the corresponding convex hull region, leaving both the hip nodes out of consideration.

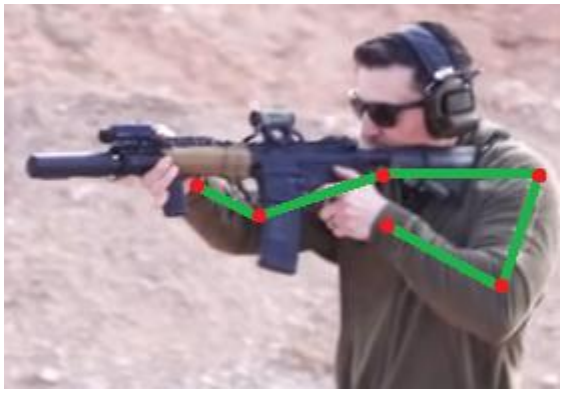


Figure 3.6: Scenario 2

- The third scenario is where 6 of the nodes will be compared with the corresponding convex hull region, leaving the left elbow and wrist out of consideration.

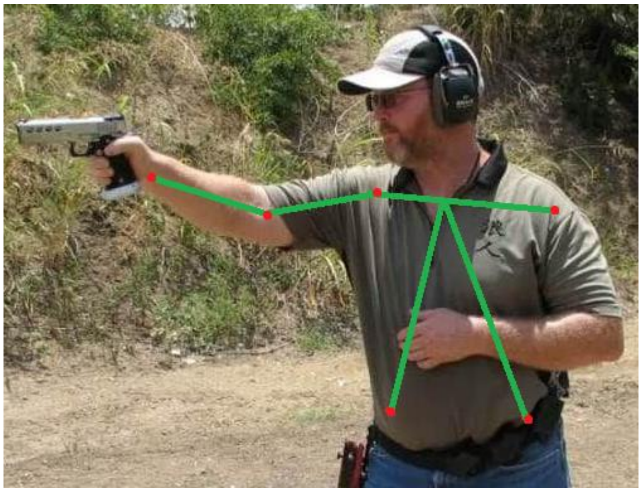


Figure 3.7: Scenario 3

- The forth scenario is where 6 of the nodes will be compared with the corresponding convex hull region, leaving the right elbow and wrist out of consideration.

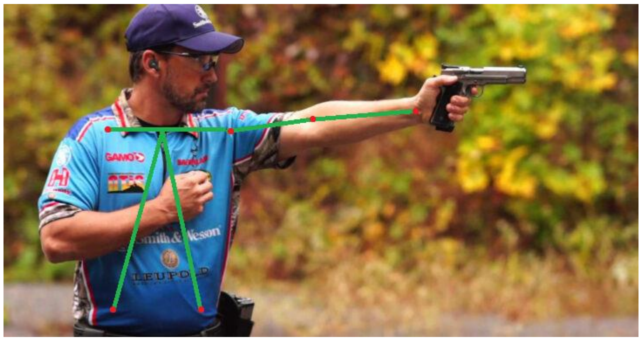


Figure 3.8: Scenario 4

- The fifth scenario is where 4 of the nodes will be compared with the corresponding convex hull region, leaving both the hips, left elbow and wrist out of consideration.

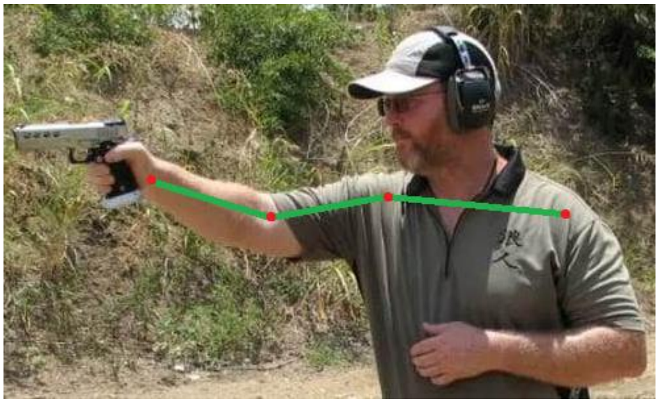


Figure 3.9: Scenario 5

- The sixth scenario is where 4 of the nodes will be compared with the corresponding convex hull region, leaving both the hips, right elbow and wrist out of consideration.

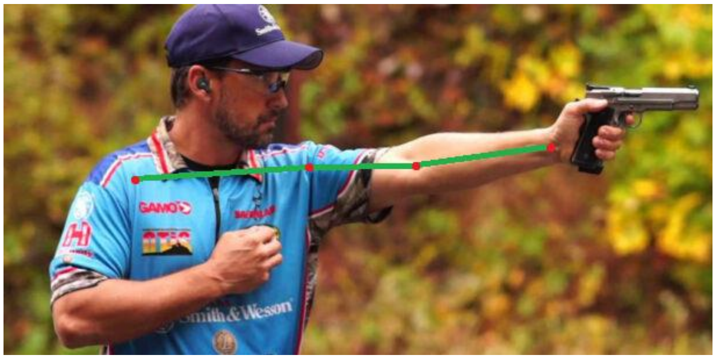


Figure 3.10: Scenario 6

From top to bottom, the test node pattern will be gradually checked in the best fit convex hull will be determined based on each confidence score. As a result, ambiguity due to missing nodes will be resolved.

3.3.3 Pose Classification from Key-Points

This section will discuss the way of predicting whether or not a pattern is indicating criminal action. The Ray Casting module is used here to do so. Given a particular node of an action pattern, the Ray Casting algorithm determines the position of the node with respect to its corresponding convex hull. A node can be either inside or outside of the convex hull. Being inside will indicate that a node is matching with its corresponding hull, thus increasing the confidence level for placement for that particular node. Similarly, all the nodes will be checked with their respective convex hull region and based on being inside or outside, the accuracy for each node will rise or fall. The cumulative confidence score from those nodes of a pattern will determine if the action is actually related to crime or not. So, the classification of the poses based on the key-point nodes will be two types. One being affirmative will be labeled as criminal activity and the other being negative labeled as normal activity.

Chapter 4: Implementation and result analysis

4.1 Extraction of key points

Here Alpha Pose works as a supportive framework in order to get plugin values for coordinate based human action classifier (CBHAC). Alpha Pose works internally going through three processes. Symmetric STN and Parallel SPPE comprise the first component. Because the Single Person Pose Estimation (SPPE) algorithm is trained on a single picture and is susceptible to position mistakes, the human body region frame generated by the target detection method is not well suited for SPPE. Micro-transformation and pruning can significantly enhance the impact of SPPE. Under defective human body area detection findings, symmetric spatial transformer network (SSTN) + Parallel SPPE can significantly increase SPPE's impact. Following spatial transformer network (STN) + SPPE + SDTN (Spatial De-Transformer Network), the imprecise detection of frames estimates the pose and maps the calculated results to the original map, allowing the frame to be modified and be corrected.

Parametric Pose NMS is the second component. Due to the presence of humans, duplicate detection frames and attitude detection are unavoidable. Due to this occurrence, a non-maximum suppression for posture was recommended to minimize redundancy. At first, the posture with the highest level of confidence is chosen as the reference, and the area frame closest to it is removed using the elimination criterion. This procedure is continued until all duplicate recognition boxes have been removed and each recognition box is distinct (no overlap beyond the threshold).

Data Augmentation is the third component. Proper data improvement can enable SSTN + SPPE adapt to defective human body area localization results during Two-Stage pose estimation (first identifying the region, then locating the pose points). Otherwise, while the model is running in the test phase, it may not be effectively suited to unusual human body location data. Using the detected area frame as a training frame is a simple and obvious technique. Target detection, on the other hand, simply generates a person's localization area. A particular impact is created by employing the generated human body localization.

4.2 Coordinate Based Human Action Classifier (CBHAC)

4.2.1 Pattern Familiarization (PF):

Converting points from Cartesian to Polar

The RMPE model's core features of the human body for each pose are fed into the pattern familiarization procedure to train the classifier and tag each pattern appropriately with a label. To reduce computing complexity, points are first transformed from Cartesian to Polar coordinate system. Particularly in data normalization procedure, using Cartesian format can make the system convoluted, as the system needs to consider both X and Y axis for each point for every human gesture and calculate normalized value for each axis for a point, which leads to exacerbation of the computation. This issue can be minimized by using Polar coordinates as in this case the value of one axis (θ) is fixed:

$0 \leq \theta \leq 2\pi$

$$r = \sqrt{x^2 + y^2}$$
$$\theta = tan^{-1}\frac{y}{x}$$

Normalization of the data

The converted key points are then fed into the normalization process implemented using sklearn Preprocessing Library. Normalization refers to the scaling of each point to a standard unit. This is done in order to convert the raw feature vectors into a form that is more suitable for subsequent estimators. This process particularly needed to evaluate the similitude of the key points by measuring them in a common scale, without changing the range difference of the values. sklearn preprocessing library provides a utility class Normalizer, which makes the process easy to execute.

$$\mathrm{mean}(X) = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\sigma^2 = \frac{1}{n}\sum X^{(i)^2}$$

Reconverting points to Cartesian Co-ordinates:

After executing the normalization process, the normalized points are reconverted to Cartesian coordinate system from Polar coordinate system. This is done to represent the actual position of the points and to reduce further computational complexity since the other algorithms used in training process requires the data in Cartesian format.

$$x = r\cos\theta$$
$$y = r\sin\theta$$

Augmentation of the data

Once normalization is done, the normalized data in Cartesian for are then used for data augmentation. Data Augmentation refers to the process of amplifying the amount of data by replicating a marginally reformed value of the existing data or creating a newly formed data based on existing data. Here, data augmentation has been used to increase the accuracy of the system as the nodes of the gestures can vary on the basis of person to person. In fact, even the same gesture can change slightly as based on the movement of the body nodes. So, in order to accurately detect pattern, form the key points in testing phase, data augmentation is needed so that the system that successfully detect a gesture even if the nodes are not exactly the same as the given dataset. In this system, we synthetized the data using a noise amount and a maximum level of data augmentation.

The data augmentation process generates random samples using Normal (Gaussian) Distribution. The probability density of which is given below,

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The distribution of the function depends on standard deviation (σ), which means the function reaches its maximum at $x + \sigma$ and $x - \sigma$.

Convex Hull generation and map labelling

After data augmentation, the new amplified data is then used to create a Convex Hull using the Graham's Scan algorithm. This algorithm first finds the lowest value of y-coordinate among the given value and perform a sorting algorithm. In this model, quicksort has been used to sort the values of each augmented data. If the dataset has multiple value with same y-coordinate, it considers the value of x-coordinate to detect the smallest point of the given dataset. The sorted values are then used sequentially to create a convex hull or the smallest polygon containing all the points of the given data.

In other words, if the dataset is consists of three $P_1 = (x_1, y_1), P_2 = (x_2, y_2)$ and $P_3 = (x_3, y_3)$, with P_1 and P_2 being the previous points of P_3. The system compares the co-ordinates to find whether to travel to previous nodes left or right turn is needed and based on that the convex hull with outer most points are generated.

After the creation of Convex hull for each and every point in key point array obtained from CBHAC, the system them labels the maps in order to categorize the patterns formed by the Convex Hulls. In addition to that, the original point is added with the hull mapping. For each gesture, a different pattern is recognized and familiarized for testing phase.

4.2.2 Pattern Estimation (PE):

Loading weights and a label map are required before the action recognition procedure can begin. Then, for each frame to be processed for pattern coordinates, RMPE is triggered. Similar to the PF method, all the key points get converted into their respective polar coordinate form and the values are now ready for normalization. After normalization is completed, the key points are reverted back to their Cartesian coordinates. As RMPE provides MSP results, these point sets are simultaneously inserted in the next phase of the PE module.

For a particular point $P_{(x, y)}$ and original point O, a self-modified version of the Ray Casting Algorithm is used to determine the following statements,

- Is P inside the convex hull and P and O are the same points.
- Is P inside the convex hull and P and O are not the same points.
- Is P outside the convex hull

The general Ray Casting Algorithm states that a point will be inside a polygon if the line segment from that point is passed through and that line segment intersects the polygon in odd number of times. The point is considered outside the polygon if the number of intersects is even for each hand of the polygon. This principal has some limitations like if the line segments intersect as shown in figure, then the number of intersecting points will be 2 as the point of intersection is a common point of two different hands of the polygon.

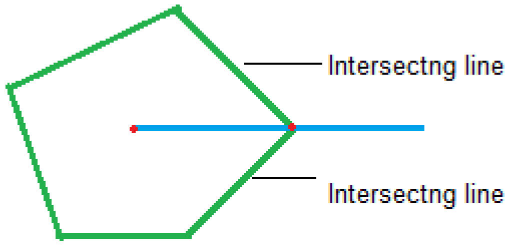


Figure 4.1: Line segment that intersects a common point of two hands

Thus indicating the point is outside the polygon. However, in figure we can see that the point is inside. So, here is where a limitation is being faced. In order to overcome this, a modification has been made where if the intersecting point value is 2, a reverse ray cast is performed as shown in figure and checked whether the reverse line intersects or not. If the line does intersect then it is clear that the point is inside the polygon. Otherwise, the point is outside.

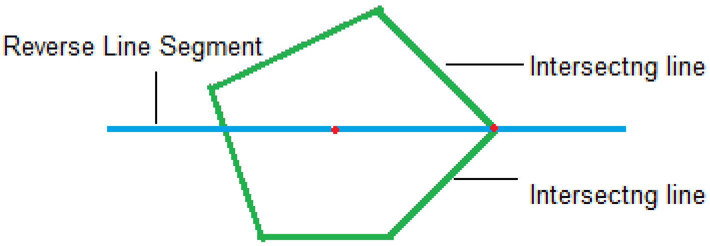


Figure 4.2: Drawing Reverse Line Segment

Now after determining the placement of P, if P is inside the polygon then it is checked whether P and the original point of the hull polygon is same or not. If P is outside the polygon, then following procedures will be calculated,

- Distance, D_H between the original point and the intersection point of P towards convex hull
- Distance, D_O of P with respect to the original key point of P's corresponding convex hull

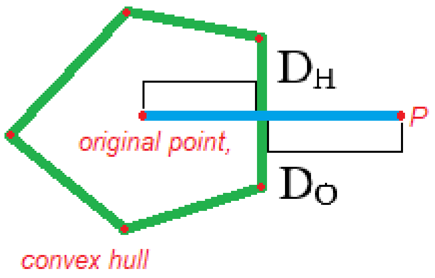


Figure 4.3: Sample virtualization for determining D_H and D_O

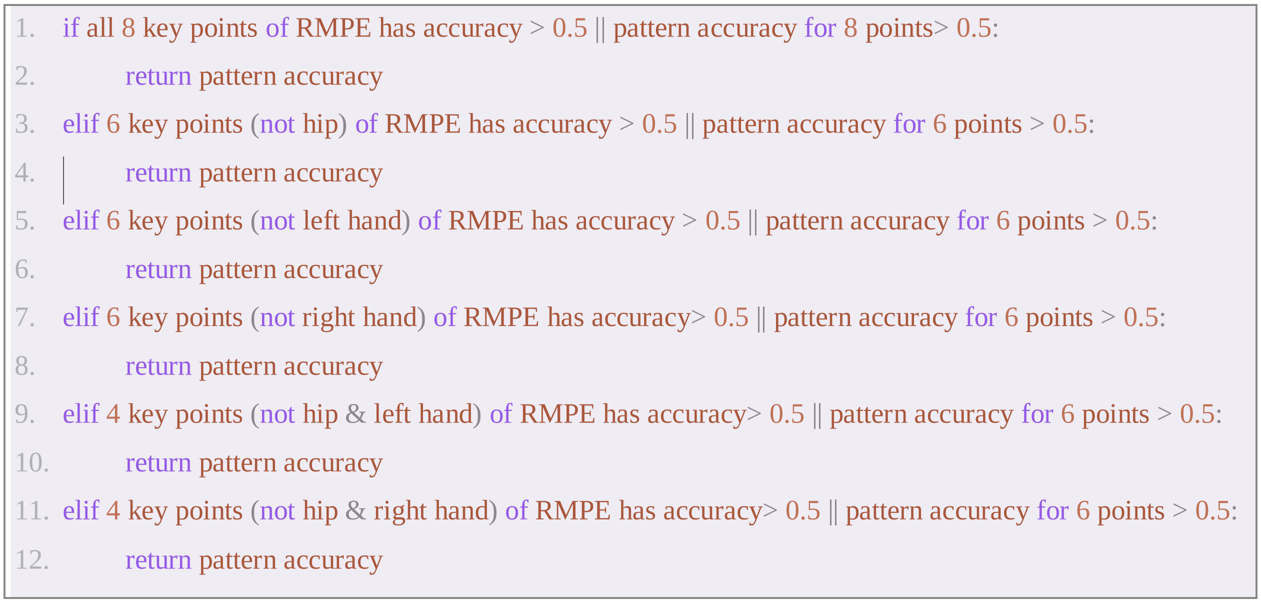
4.3 Result analysis:

For each key point of a single we present an analysis function analysis (Boolean: inside, List: Hull, P, σ).

$$\text{Accuracy}(P) = \begin{cases} \sigma + (1-\sigma) \times \dfrac{D_H}{D_0}; & \text{inside} = \text{True} \\ 0; & \text{inside} \ = \ \text{False} \ \ \&\text{AND} \ \ D_H \geq D_0 \cdot \dfrac{(1-\sigma)}{\sigma} \\ (1-\sigma) \times \dfrac{D_0 \cdot \dfrac{(1-\sigma)}{\sigma} - D_H}{D_0 \cdot \dfrac{(1-\sigma)}{\sigma}}; & \text{inside} \ = \ \text{False} \ \ \text{AND} \ \ D_H < D_0 \cdot \dfrac{(1-\sigma)}{\sigma} \end{cases}$$

Where sigma (σ) is a weight balancing the two distances D_O and D_H to assume a minimum movement in key nodes. This parameter is initially 0.5, following a test-driven manner we can set the value higher than 0.5. For a higher rate of learning data for single class and noise greater than 0.05 while training a higher value of σ performs better.

Arithmetic mean of key points is calculated as pose accuracy and the best fit pose is considered as final result. According to the purpose of this model 6 cases have been defined for generating final result (described in chapter 3.3.2). In order to calculate the accuracy (p), one of the following conditions has to be satisfied with confidence greater than 0.5,



```
if all 8 key points of RMPE has accuracy > 0.5 || pattern accuracy for 8 points> 0.5:
        return pattern accuracy
elif 6 key points (not hip) of RMPE has accuracy > 0.5 || pattern accuracy for 6 points > 0.5:
        return pattern accuracy
elif 6 key points (not left hand) of RMPE has accuracy > 0.5 || pattern accuracy for 6 points > 0.5:
        return pattern accuracy
elif 6 key points (not right hand) of RMPE has accuracy> 0.5 || pattern accuracy for 6 points > 0.5:
        return pattern accuracy
elif 4 key points (not hip & left hand) of RMPE has accuracy> 0.5 || pattern accuracy for 6 points > 0.5:
        return pattern accuracy
elif 4 key points (not hip & right hand) of RMPE has accuracy> 0.5 || pattern accuracy for 6 points > 0.5:
        return pattern accuracy
```

Table 4.1 provides the Classifier results for different values of sigma (**σ**) and noise of data while training. For a greater value of sigma classes may overlap each other which reduces the confidence value as shown. Also, few test samples are shown in Table 4.2 with a sigma 0.6.

Samples	**σ = 0.5,** **Noise = 0.04**	**σ = 0.5,** **Noise = 0.05**	**σ = 0.6,** **Noise = 0.05**	**σ = 0.7,** **Noise = 0.06**
S. 1	0.615	0.651	0.682	0.671
S. 2	0.771	0.774	0.809	0.810
S. 3	0.773	0.801	0.830	0.782
S. 4	0.698	0.703	0.761	0.771

Table 4.1: Experimentation for value of sigma

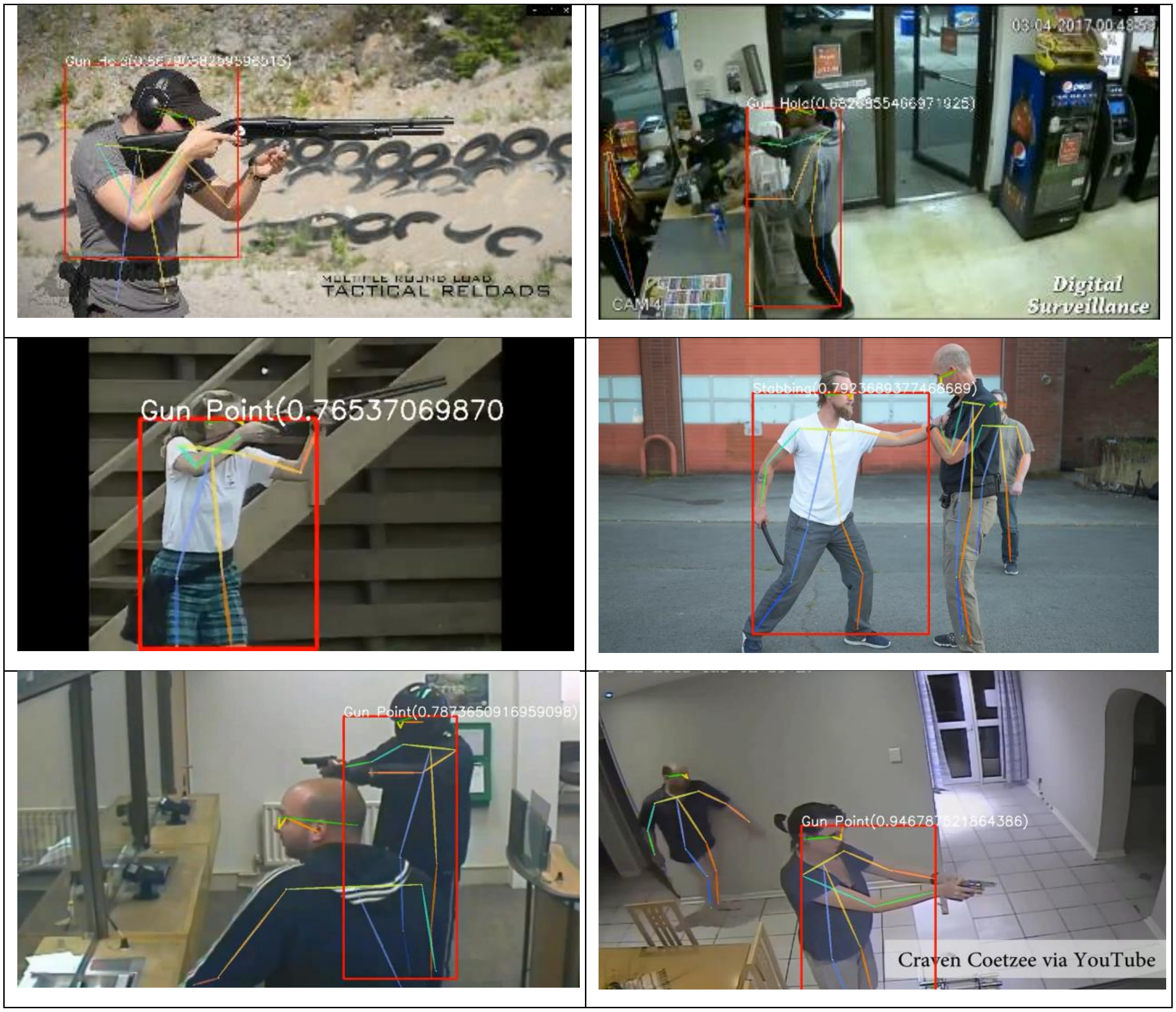

Gun_Hold(0.5679058259598515)
MULTIPLE ROUND LOAD
TACTICAL RELOADS
03-04-2017 00:48:59
Gun_Hold(0.6828955406971926)
CAM4
Digital Surveillance
Gun_Point(0.76537069870
Stabbing(0.7923689377468689)
Gun_Point(0.7873650916959098)
Gun_Point(0.946787521864386)
Craven Coetzee via YouTube

Table 4.2: Few samples of output

Chapter 5: Conclusion

Reports have shown that rate of crime has risen significantly in recent years, possibly due to the effects of pandemic [20][21]. So, a system to detect suspicious activity has become extremely crucial in given situation. With that in mind, we tried to develop a method of detecting common crimes that would be easy to implement and cost-effective. The model discussed in this paper have shown promising result in detecting suspicious or threatening gesture, which can be used to detect and eliminate criminal activities, particularly in case of standard crimes like robbery, theft, murder etc. The system can be implemented using CCTV footages and can be implemented on a wide range, as it is able to detect suspicious activities in both indoors and outdoors.

5.1 Limitations

Despite promising result shown by the model discussed in this paper, our model has some limitations as well, which are discussed briefly below:

- The model is developed with only common criminal activities and gestures in mind, namely, pointing gun, knife attack and so on. As a consequence, this model might not be suitable for detecting every possible kind of criminal activity, particularly in case of detecting non-conventional crimes. Although we tried to incorporate gestures corresponding to different types of guns and knives, there are many weapons can could not be included as they rarely, so collecting dataset with corresponding gestures was challenging.
- This model might not show accurate results while detecting gestures that only needs one hand and shoulder nodes. For example, when a person is pointing his/her hand at something, the system might detect the gesture as gunpoint since it is quite similar to holding revolver with one hand. To eliminate this issue, weapon detection has to be incorporated with the system and the confidence of the system has to be modified in such a way that the system will require detection of both the gun and corresponding gesture to consider a pose as threatening or suspicious.
- The more data we provide through training, the larger the generated weight becomes. Thus it will take longer to load the weight in case of testing. As a result, the total time this model takes for predicting criminal actions will increase.

5.2 Future Work:

Some work has to be carried out to finalize the framework. These are the works we have chosen to conduct in the future:

- **Object detection:** For more perfection in accuracy, we can take real time data and add object detection with posture detection to get maximum accuracy. For instance, if a person is holding a gun with one hand, this will be similar to the fact if a person is pointing finger to someone. But the actions are totally different whereas with the help of object detection we can differentiate if that person is pointing a gun or a finger as here gun and finger are considered to be an object. So in our future work, we can add object detection with posture detection to determine and finalize solid accuracy output.
- **Versatility:** Our model has been implemented in our work for pattern detection and through this we could detect crime pattern through posture. But this model can also be implemented in many other areas like symbol sign detection, alphabet sign recognition etc. After preparing the data for feeding to the model, it can be split into train and test sets, with the images and other basic information standardized. This can be compiled by training; we aim to discover the best weight to decide what to do. The lower function that will be used to evaluate a set of weights, the optimizer that will be used to find various network weights, and any optional metrics that will be gathered and reported during the training process must all be defined.

Bibliography:

[1] Lopez, German. “The Murder Increase in the US, Explained.” Vox, 2 Dec. 2020, www.vox.com/2020/8/3/21334149/murders-crime-shootings-protests-riots-trump-biden.

[2] 8.1 The Problem of Crime – Social Problems. (n.d.). University of Minnesota Libraries Publishing. Retrieved September 26, 2021, from https://open.lib.umn.edu/socialproblems/chapter/8-1-the-problem-of-crime/

[3] Dorogyy, Yaroslaw & Kolisnichenko, Vadym & Levchenko, Kseniia. “Violent Crime Detection System,” 352-355. 10.1109/STC-CSIT.2018.8526596, September 2018. [Online], Available: https://www.researchgate.net/publication/328815099_Violent_Crime_Detection_System

[4] Sikandar, Tasriva & Ghazali, Kamarul & Rabbi, Mohammad. “ATM crime detection using image processing integrated video surveillance: a systematic review”. Multimedia Systems. December 2018. [Online], Available: https://www.researchgate.net/publication/329584911_ATM_crime_detection_using_image_processing_integrated_video_surveillance_a_systematic_review

[5] Sikandar, Tasriva & Samsudin, W. & Rabbi, Mohammad & Ghazali, Kamarul. “An Efficient Method for Detecting Covered Face Scenarios in ATM Surveillance Camera”. SN Computer Science. 1. 10.1007/s42979-020-00163-6. May 2020. [Online], Available: https://www.researchgate.net/publication/341201174_An_Efficient_Method_for_Detecting_Covered_Face_Scenarios_in_ATM_Surveillance_Camera

[6] Goya, K., Zhang, X., Kitayama, K., & Nagayama, I. (2009, September). “A Method for Automatic Detection of Crimes for Public Security by Using Motion Analysis”. 2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing. Available: https://ieeexplore.ieee.org/document/5337276

[7] George, F. T., Patnam, V. S. P., & George, K. (2018, May). “Real-time deep learning based system to detect suspicious non-verbal gestures”. 2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC). Available: https://ieeexplore.ieee.org/document/8409864

[8] Nakib, M., Khan, R. T., Hasan, Md. S., & Uddin, J. (2018, February). “Crime Scene Prediction by Detecting Threatening Objects Using Convolutional Neural Network”. 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2). Available: https://ieeexplore.ieee.org/document/8465583

[9] Gun Ownership by Country 2021. (n.d.). World Population Review. Retrieved September 26, 2021, from https://worldpopulationreview.com/country-rankings/gun-ownership-by-country

[10] Coomaraswamy, K. S. “Predictors and Severity of Injury in Assaults with Barglasses and Bottles.” Injury Prevention, vol. 9, no. 1, BMJ, Mar. 2003, pp. 81–84. Crossref, doi:10.1136/ip.9.1.81

[11] Saranu, P. N., Abirami, G., Sivakumar, S., Ramesh, K. M., Arul, U., & Seetha, J. (2018, February). “Theft Detection System using PIR Sensor”. 2018 4th International Conference on Electrical Energy Systems (ICEES). Available: https://ieeexplore.ieee.org/document/6704220

[12] Agency for Toxic Substances and Disease Registry (ATSDR). Toxicological Profile for Chloroform. 1997, US Department of Health and Human Services: Atlanta, US. Available: https://www.atsdr.cdc.gov/toxprofiles/tp6.pdf

[13] Intani, P., & Orachon, T. (2013, October). “Crime warning system using image and sound processing”. 2013 13th International Conference on Control, Automation and Systems (ICCAS 2013). Available: https://ieeexplore.ieee.org/document/8443215

[14] Xiao, J., Li, S., & Xu, Q. (2019). “Video-Based Evidence Analysis and Extraction in Digital Forensic Investigation”. IEEE Access, 7, 55432–55442. Available: https://ieeexplore.ieee.org/document/8700194

[15] Yimyam, W., Pinthong, T., Chumuang, N., & Ketcham, M. (2018, November). “Face Detection Criminals through CCTV Cameras”. 2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS). Available: https://ieeexplore.ieee.org/document/8705938

[16] Piza, E. L. (2016). “The crime prevention effect of CCTV in public places: a propensity score analysis”. Journal of Crime and Justice, 41(1), 14–30. Available: https://www.researchgate.net/publication/307513460_The_crime_prevention_effect_of_CCTV_in_public_places_a_propensity_score_analysis

[17] Sikandar T., Ghazali K. H., Md. Rabbi (2018). “ATM crime detection using image processing integrated video surveillance: a systematic review”. Available: https://www.researchgate.net/publication/316344070_The_Value_of_CCTV_Surveillance_Cameras_as_an_Investigative_Tool_An_Empirical_Analysis

[18] Kakkar P , Sharma V. (2018). “Criminal Identification System Using Face Detection and Recognition”. International Journal of Advanced Research in Computer and Communication Engineering. Available: https://ijarcce.com/upload/2018/march-18/IJARCCE%2046.pdf

[19] Fang, Hao-Shu, et al. “RMPE: Regional Multi-Person Pose Estimation.” Computer Vision and Pattern Recognition, 2016, https://arxiv.org/abs/1612.00137

[20] Islam, Nazrul. “Five Types of Crime Increase during the Pandemic.” Prothom Alo [Dhaka], 27 July 2021, en.prothomalo.com/bangladesh/crime-and-law/five-types-of-crime-increase-during-the-pandemic

[21] “Overview of Preliminary Uniform Crime Report, January–June, 2020.” Federal Bureau of Investigation, FBI National Press Office, 8 Dec. 2020, www.fbi.gov/news/pressrel/press-releases/overview-of-preliminary-uniform-crime-report-january-june-2020