

# **Machine Learning based Human Observer Analysis of Video Sequences**

by

Seema Faqeer Abdul Rahim Al-Raisi

A Doctoral Thesis

Submitted in partial fulfilment  
of the requirements for the award of

Doctor of Philosophy

of

Loughborough University

2017

© Copyright 2017 Seema Al-Raisi

# Abstract

The research contributes to the field of video analysis by proposing novel approaches to automatically generating human observer performance patterns that can be effectively used in advancing the modern video analytic and forensic algorithms. Eye tracker and eye movement analysis technology are employed in medical research, psychology, cognitive science and advertising. The data collected on human eye movement from the eye tracker can be analyzed using the machine and statistical learning approaches. Therefore, the study attempts to understand the visual attention pattern of people when observing a captured CCTV footage. It intends to prove whether the eye gaze of the observer which determines their behaviour is dependent on the given instructions or the knowledge they learn from the surveillance task. The research attempts to understand whether the attention of the observer on human objects is differently identified and tracked considering the different areas of the body of the tracked object. It attempts to know whether pattern analysis and machine learning can effectively replace the current conceptual and statistical approaches to the analysis of eye-tracking data captured within a CCTV surveillance task.

A pilot study was employed that took around 30 minutes for each participant. It involved observing 13 different pre-recorded CCTV clips of public space. The participants are provided with a clear written description of the targets they should find in each video. The study included a total of 24 participants with varying levels of experience in analyzing CCTV video. A Tobii eye tracking system was employed to record the eye movements of the participants. The data captured by the eye tracking sensor is analyzed using statistical data analysis approaches like SPSS and machine learning algorithms using WEKA.

The research concluded the existence of differences in behavioural patterns which could be used to classify participants of study is appropriate machine learning algorithms are employed. The research conducted on video analytics was perceived to be limited to few

projects where the human object being observed was viewed as one object, and hence the detailed analysis of human observer attention pattern based on human body part articulation has not been investigated. All previous attempts in human observer visual attention pattern analysis on CCTV video analytics and forensics either used conceptual or statistical approaches. These methods were limited with regards to making predictions and the detection of hidden patterns. A novel approach to articulating human objects to be identified and tracked in a visual surveillance task led to constrained results, which demanded the use of advanced machine learning algorithms for classification of participants

The research conducted within the context of this thesis resulted in several practical data collection and analysis challenges during formal CCTV operator based surveillance tasks. These made it difficult to obtain the appropriate cooperation from the expert operators of CCTV for data collection. Therefore, if expert operators were employed in the study rather than novice operator, a more discriminative and accurate classification would have been achieved. Machine learning approaches like ensemble learning and tree based algorithms can be applied in cases where a more detailed analysis of the human behaviour is needed. Traditional machine learning approaches are challenged by recent advances in the field of convolutional neural networks and deep learning. Therefore, future research can replace the traditional machine learning approaches employed in this study, with convolutional neural networks. The current research was limited to 13 different videos with different descriptions given to the participants for identifying and tracking different individuals. The research can be expanded to include any complicated demands with regards to changes in the analysis process.

# **Dedication**

I dedicate this thesis to the two most important men in my life my father and my best friend Khalid (my beloved husband) for their endless support and unconditional love. My father always believed in me even in times when I was full of doubt in myself. My beloved husband sacrificed his own future in order to make my dream come true. He was always there cheering me up and stood by me through the good and bad times.

I hope both of you are proud of me.

# Acknowledgements

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

First, I would like to thank Allah Almighty for giving me such an opportunity & ability to undertake this PhD project and complete it with satisfaction. This achievement wouldn't have been possible without his blessings.

I would also like to thank my supervisor Professor Eran, there aren't enough genuine words in which I can describe an appreciation to my supervisor. One simply could not wish for a better or friendlier supervisor. He has set an example of excellence as an instructor and a mentor. I have been extremely lucky to have a supervisor who cared so much about my work, health and family.

To my dear supervisor, I am grateful to you for giving me the right directions during studies, your understanding, extraordinary patience, lavish assistance and support. I must say that without your advice and trust, I'd never come so far. Thank you very much professor Eran for everything you did to help and support me.

My family deserves a special mention for their endless supports, love and prayers day and night to make me achieve success and honor during my PhD. A special thanks to my father for his moral support and constant encouragement. Many thanks to my mother and sister Samiya for their love, supports and encouragements. I will forever be thankful to them.

My deepest appreciation and love goes to my husband Khalid, who has been with me from beginning of this PhD journey until its end. Thank you for your endless love, your encouragement, your entire confidence in me and for being my true best friend. Thank you for everything you have ever done for me.

My warmly acknowledge to all the colleagues of Computer Science Department at Loughborough University, specially to all Omani colleagues and friends.

Finally, I would like to thank his majesty Sultan Qaboos government the Ministry of Manpower (my sponsor) and Cultural Attaché (Embassy of Oman) for providing the financial assistance and support throughout the research period.

**Seema Al-Raisi**

# Table of Contents

Abstract.....	ii
Dedication.....	iv
Acknowledgements.....	v
List of Figures.....	xii
<b>Chapter 1. Introduction.....</b>	<b>1</b>
1.1. Aim & Objectives .....	4
1.2. Research Contributions .....	5
1.3. Thesis overview.....	8
<b>Chapter 2. Literature Review.....</b>	<b>9</b>
2.1. Introduction.....	9
2.2. Eye tracker research.....	9
2.2.1 Medical Applications .....	10
2.2.2 Applications in Education .....	11
2.2.3 Applications in Aviation.....	12
2.2.4 Applications in Usability Studies .....	12
2.2.5 Applications in Marketing .....	13
2.3. Eye Movement Tracking in CCTV .....	13
2.4. Machine Learning and Surveillance .....	16
2.5. Summary & Conclusion.....	19

<b>Chapter 3. Research Background</b> .....	<b>21</b>
3.1. The Human Visual System.....	21
3.1.1 Human Eye: perception.....	21
3.1.2 Visual Perception and Attention.....	22
3.1.3 Computer vision versus human vision .....	22
3.1.4 Eye Movements & Visual Attention .....	23
3.2. Measurement of eye movements .....	24
3.2.1 Eye tracker.....	24
3.2.2 Tobii studio .....	26
3.3. Visual stimuli: static versus dynamic stimuli. ....	27
3.3.1 Text-based stimuli.....	27
3.3.2 Image-based stimuli .....	27
3.3.3 Dynamic stimuli .....	28
3.4. Area of Interest .....	29
3.5. Data mining tools.....	30
3.5.1 SPSS.....	30
3.5.2 Statistical Method (ANOVA).....	30
3.5.2.1 Interpreting the key results of One-Way ANOVA .....	31
3.5.3 Machine learning.....	33
3.5.3.1 Predictive models.....	35
3.5.4 Machine learning software.....	38
3.6. Statistical Analysis Versus Machine Learning .....	39
3.7. Think Aloud Method(TAM) .....	39
3.8. Person or human re-identification .....	40
3.9. Summary .....	41



<b>Chapter 4. Experimental Design &amp; Data Gathering .....</b>	<b>42</b>
4.1. Research questions .....	42
4.2. Methodology for data collection and analysis .....	45
4.2.1 Experimental design .....	46
4.2.2 Data Collection Phase .....	60
4.2.2.1 The procedure .....	60
4.2.3 Data Analysis.....	65
4.3. Summary .....	68
 <b>Chapter 5. Statistical Analysis of Visual Attentional Patterns in Video Footage.</b>	<b>69</b>
5.1. Data preparation.....	70
5.2. Experimental Results and Discussions.....	71
5.2.1 Research Question 1.....	74
5.2.1.1 Summary Research Question 1.....	77
5.2.2 Research Question 2.....	78
5.2.2.1 Summary Research Question 2.....	84
5.2.3 Research Question 3.....	85
5.2.3.1 Summary Research Question 3.....	88
5.2.4 Research Question 4.....	89
5.2.4.1 Summary Research Question 4.....	96
5.2.5 Research Question 5.....	97
5.2.5.1 Summary Research Question 5.....	98
5.2.6 Research Question 6.....	99
5.2.6.1 Summary Research Question 6.....	100
5.3. Summary and conclusion.....	103
5.4. Human Observer Behaviour Analysis of CCTV Video Surveillance using Linear Regression .....	105

5.4.1	Data capture .....	106
5.4.2	Data pre-processing or preparation .....	107
5.4.3	Removal of Missing Values .....	107
5.4.4	Attribute or Feature Selection .....	108
5.5.	Dataset Representation .....	109
5.6.	Modelling Fixation Duration .....	112
5.7.	Interpretation and Analysis of Regression Models .....	115
5.8.	Summary and Conclusion .....	122
<b>Chapter 6. Use of Machine Learning Algorithms to Classify Different Groups of</b>		
<b>    Participants based on Eye Gaze Patterns .....</b>		<b>123</b>
6.1.	Introduction .....	123
6.2.	Experimental Procedure .....	125
6.3.	Experimental Results .....	126
6.3.1	Results for the classification of experts from novice participants: ...	126
6.3.2	Results for the classification of Female from Male participants: .....	128
6.4.	Analysis of Results .....	130
6.5.	Summary & Conclusions .....	132
<b>Chapter 7. Conclusion and future work .....</b>		<b>133</b>
7.1.	Future Work .....	136
<b>References .....</b>		<b>138</b>
Appendix 1 .....		148
Appendix 2 .....		180
Appendix 3 .....		187
Appendix 4 .....		194

# List of Tables

Table 4.1: The written descriptions of the objects to be identified and tracked in the videos .....	52
Table 4.2: The written descriptions of the objects to be identified and tracked based on level of description from simple to complex.....	55
Table 4.3: DAOI for each video .....	57
Table 4.4: Eye-gaze metrics used in the conducted experiments .....	63
Table 4.5: The two test videos considered for detailed data analysis .....	67
Table 5.1: Mapping between research questions and data metrics .....	72
Table 5.2: Analysis of Order of Fixation for (a) Video 1 and (b) Video 10.....	92
Table 5.3 Attributes and attribute ‘type’ for each DAOI .....	106
Table 5.4 WEKA Feature Selection Methods.....	108
Table 5.5 Selected Attribute when all data are considered.....	109
Table 5.6: Sample data selection for video-10.....	110
Table 5.7: Sample data selection for video-1 .....	111
Table 5.8 The regression models for video-1 .....	114
Table 5.9 The regression models for video-10 .....	114
Table 5.10: Models for video-1 .....	118
Table 5.11: Models for video-10 .....	120
Table 6.1: Groupings of participants .....	124
Table 6.2: Results of classification of Experts vs Novice participants .....	126
Table 6.3: Results of classification of Female vs Male participants.....	128

# List of Figures

Figure 3.1: Tobii Eye tracker Source .....	24
Figure 3.2: Illustration of how eye tracking works .....	25
Figure 3.3: The DAOIs in research conducted .....	29
Figure 3.4 : Reporting one way ANOVA in APA style .....	33
Figure 3.5 : Types of ML .....	34
Figure 3.6 : The general process of machine learning .....	34
Figure 3.7: Linear correlation .....	36
Figure 3.8: Weka Machine learning software .....	38
Figure 3.9: Weka Machine learning software .....	40
Figure 4.1 : Stages followed in a Research Project .....	45
Figure 4.2: Pilot Study Venue (Usability Lab, Haslegrave Building) .....	47
Figure 4.3: Field-of-view of each captured video.....	49
Figure 4.4: Objects that are requested to be identified and tracked by the participants.....	50
Figure 4.5: Objects that are requested to be identified and tracked by the participants.....	56
Figure 4.6: The calibration procedure for a participant .....	59
Figure 4.7: Overview of study setup with regards to data collection and information display .....	64
Figure 5.1: Field-of-view of each captured video.....	73
Figure 5.2: Percentage of respondents that have fixated at least once within the defined DAOIs in (a) video 1 [ <i>Find a women wearing white trousers and carrying a pink shoulder handbag</i> ] and (b) video 10 [ <i>Find a person wearing a blue top and white pants</i> ] .....	75

Figure 5.3: Average fixation duration for each DAOI in (a) video 1 and (b) video 10 .....	80
Figure 5.4: Average visit count for each DAOI in (a) video 1 and (b) video 10 .....	82
Figure 5.5: The average visit duration for each DAOI in (a) video 1 and (b) video 10 .....	83
Figure 5.6: Average fixation time to first fixation for each DAOI in (a) video 1 and (b) video 10 .....	87
Figure 5.7: Novice and Expert Average time to first fixation for each DAOI in (a) video 1 and (b) video 10 .....	95
Figure 5.8: Female and Male Average time to first fixation for each DAOI in (a) video 1 and (b) video 10 .....	96
Figure 5.9: Analysis of Average Time to First Fixation for (a) video 1 and (b) video 10 .....	98
Figure 5.10: Scan paths of all participants (a) before, (b) during and (c) after the target object appears on screen for video-1 Analysis of Average Time to First .....	102
Figure 5.11: A linear relationship .....	113
Figure 5.12: Correlation coefficient obtained for each DAOI of video-1 and video-10 .....	117
Figure 6.1: Results of classification of Experts vs Novice participants (a) Without bagging (b) with bagging .....	127
Figure 6.2: Results of classification of Female vs Male participants (a) Without bagging (b) with bagging .....	129

## List of Abbreviations

<b>ANOVA</b>	<b>A</b> nalysis <b>O</b> f <b>V</b> ariance
<b>AOI</b>	<b>A</b> rea <b>O</b> f <b>I</b> nterest
<b>Bagging</b>	<b>B</b> ootstrap <b>A</b> ggregation
<b>CC</b>	<b>C</b> orrelation <b>C</b> oefficient
<b>CCTV</b>	<b>C</b> losed <b>C</b> ircuit <b>T</b> elevision
<b>DAOI</b>	<b>D</b> ynamic <b>A</b> reas <b>O</b> f <b>I</b> nterest
<b>df</b>	degrees of freedom
<b>FC</b>	<b>F</b> ixation <b>C</b> ount
<b>FD</b>	<b>F</b> ixation <b>D</b> uration
<b>GPU</b>	<b>G</b> raphics <b>P</b> rocessing <b>U</b> nit
<b>HVS</b>	<b>H</b> uman <b>V</b> isual <b>S</b> ystem
<b>LR</b>	<b>L</b> inear <b>R</b> egression
<b>ML</b>	<b>M</b> achine <b>L</b> earning
<b>MLP</b>	<b>M</b> ulti <b>L</b> ayer <b>P</b> erceptron
<b>P value</b>	<b>P</b> robability value
<b>PF</b>	<b>P</b> ercentage <b>F</b> ixated
<b>RepTree</b>	<b>R</b> educed-error pruning
<b>RF</b>	<b>R</b> andom <b>F</b> orest
<b>SPSS</b>	<b>S</b> tatistical <b>P</b> ackage for the <b>S</b> ocial <b>S</b> ciences
<b>TAM</b>	<b>T</b> hink <b>A</b> loud <b>M</b> ethod
<b>TTF</b>	<b>T</b> ime <b>T</b> o <b>F</b> irst <b>F</b> ixation
<b>TTFMC</b>	<b>T</b> ime <b>T</b> o <b>F</b> irst <b>M</b> ouse <b>C</b> lick
<b>VC</b>	<b>V</b> isit <b>C</b> ount
<b>VD</b>	<b>V</b> isit <b>D</b> uration
<b>WEKA</b>	<b>W</b> aikato <b>E</b> nvironment for <b>K</b> nowledge <b>A</b> nalysis

# Chapter 1.

## Introduction

While the novel fundamental principles and latest trends are continuously improving technology, one fact that remains is that the general public security is improving, specifically after the introduction of Closed-Circuit Television commonly known as CCTV cameras. The fundamental purpose of the cameras is to provide on-time and recorded visual analytic and forensic services. The cameras have been widely used in almost every organization as they serve the security of individuals, groups or public with visuals at areas that especially authorities have minimal presence in person. The security authorities and departments of such organisations use such videos to manually or automatically review the recorded videos. The advantages brought by the cameras is the key aspect of CCTV systems praised in most popular literature; however, the capability of added computational intelligence potentially possible behind the cameras have been of great interest to computer scientists as the cameras bring to light the possible application of two artificial intelligence branches - machine learning and pattern recognition algorithms within the area of computer vision.

Most videos recorded using CCTV are usually of low visual quality as the cameras are aimed at recording specific areas that are at a long distance or in a dark environment. Further the cost, hardware and software related constraints of cameras themselves may lead to cheaper technology being utilised in the design and manufacture of the CCTV cameras that brings to light the potential use of machine learning and pattern recognition in the post manufacture quality improvement of captured video footage. Nevertheless, these two-computer science related areas have been importantly used, supported by latest development of eye tracking technology, to provide useful information capture from within the contents of recorded CCTV video footage. In particular machine learning and pattern recognition has been used as an aid to the users and/or operators of CCTV

systems due to the ability they provide to especially understand the data provided, identifying and uncovering hidden patterns in the footage.

An Eye Tracker is a versatile device often applied in many fields of applications to study human behaviour. It has been hailed as, “a great opportunity for fascinating researchers to look into the mind” of the participants by analysing their eye movements. Thus far, this device is considered to be the only powerful and fast-growing method for studying and understanding human behaviour research. It has been used in the different fields of application of camera technologies including video surveillance in order to identify, analyse and deliver valuable insights into the gaze patterns of the video observers [24] [25] [26].

Eye trackers and eye movement analysis technology has been used in psychology[6], medical research[18] [19] [20] [21], advertising[9][10] and cognitive science. Today, using CCTV cameras, human eye gaze (i.e. eye movement pattern) is relatively recorded with ease and with high reliability. Since eye movement is primarily a combination of voluntary and involuntary cognitive processes, eye data analysts need to be careful in interpreting the participant’s appearance as well as movements, as this is what is used to infer the participant’s behaviour and intentions. The eye movement data collected from the eye tracker can be analysed using statistical and machine learning methods.

In this thesis, we provide the results of a rigorous study that was carried out to investigate how humans observe and analyse a CCTV surveillance video, when instructed to look for a person with a given description. The pilot study, which lasted for approximately 30 minutes per participant, consisted of observing 13 different pre-recorded CCTV clips of a public space with a clear given written description of the targets to look for, in each video. The study was conducted with the help of 24 participants having varied levels of experience analysing CCTV video. During the experiments, the participants' eye movements were recorded by a professional Tobii eye tracking system. Based on the data captured by the eye tracking sensor and by making use of statistical data analysis approaches using SPSS and machine learning algorithms using WEKA, the observed behaviour of participants is analysed, given a specific task.



To the authors' knowledge, all existing studies of CCTV observer performance analysis use statistical and/or conceptual data analysis approaches using tools such as ANOVA that does not allow observer behaviour pattern analysis [24] [25] [26] as ANOVA is not designed for time based predictive analysis but rather to provide statistical metrics that subsequently needs manual analysis. Further such studies conducted are limited in their analysis as the investigations are carried out with regards to observing eye tracking data captured in visually identifying particular human objects based on their entire body areas that significantly constrains the potential to carry out detailed behaviour analysis. The purpose, however, of using machine learning techniques is to better analyse, identify, learn, understand, make predictions and uncover hidden patterns in the eye tracking data.

Given the above reasons the research proposed and presented in this paper could immensely benefit the video analytic community in the future by providing the means to use the automatically generated human performance patterns in making modern video analytic and forensic algorithms, smarter and more time-efficient, especially by making to mimic the behaviour of a human.

Motivated by the above reasoning the proposed research will focus on answering the following research questions:

- 1) In carrying out a visual surveillance task using captured CCTV footage, is the behaviour of a human observer measured by his/her eye gaze tracking, dependent on the instructions given or knowledge they process about the surveillance task.
- 2) Is the human observer's visual attention on human objects being identified and tracked, different with regards to the different areas of the tracked object's body and if so how and why.
- 3) Can pattern analysis and machine learning be effectively used to replace existing statistical and conceptual analysis approaches, in analysing eye-tracking data captured within a CCTV surveillance task.

In the research conducted within the research context of this thesis the above research questions are investigated in detail and answered.

## 1.1. Aim & Objectives

The overall aim of this research study is defined as follows:

*“Investigate and understand how a human observes and analyses a CCTV surveillance video specifically when they are asked to look for a target with a given description ”*

The above aim is achieved via making an effort to meet the following thesis objectives:

1. Conduct a comprehensive review of literature to investigate existing work on human observer eye tracking behaviour analysis applied especially to analysing CCTV footage.
2. To investigate in detail, via statistical and conceptual approaches, the eye movement behaviour of a human observer when carrying out an instructed video surveillance task.

In order to achieve this objective five general sub-objectives will be met:

- a) Investigate if the participants ‘fixate their gaze’ on all parts of the target human body in the same way or not.
- b) To investigate which specific parts of the target body captured the most attention, and why.
- c) To investigate what target body attracts attention first when participants have been instructed very specifically about the appearance of a person being searched.
- d) To investigate the order in which the participants paid attention to given body regions, given an instruction and in the presence of potential distractions.
- e) To investigate how long did it take to spot each part of the body.

3. To investigate using both statistical and data mining approaches, if the written description for search of a specific person has any influence on the order in which participants look at target body parts and pay attention to them.
4. To investigate using advanced machine learning algorithms if the data captured via eye tracking can be used for general behaviour analysis and classification of different groups of people such as experts vs. novice observers, male vs female observers, etc.

The research conducted within the context of this thesis has met all above objectives and have led to the following original contributions.

## **1.2. Research Contributions**

The research conducted within the context of this thesis has met all above objectives and have led to the following original contributions.

### **1. Conceptual Contribution**

Previous research with respect to the human observer analysis of CCTV video have not been sufficiently detailed and rigorous enough to analyses observer attention on different parts of the tracked object's body, given different instructions[24] [25] [26]. This lack of detailed investigation leads one to a number of open research questions. Chapter-4 of this thesis defines a novel method to break-down a human body into four areas of significant visual importance and carry out eye tracking data capture and analysis, based on Dynamic Areas Of Interest (DAOIs) using rigorous statistical and conceptual approaches.

## 2. Technical Contribution

- a) Analysing eye tracking data using statistical and conceptual approaches using SPSS and linear regression using a machine learning suite WEKA to provide valuable insight to the participant visual attention features associated with the surveillance task assigned (when carrying out an instructed video surveillance). **Chapter-5** of this thesis proposes the use of linear regression which is a statistical approach by using machine learning suite called WEKA for the detailed analysis of eye tracking data captured when human observers are carrying out pre-instructed visual surveillance tasks on videos. The novel approaches of data analysis supported by linear regression in WEKA proposed in this chapter not only benefits the research community within CCTV video analytics/forensics, but beyond, specifically those who carry out medical, material inspection and environmental monitoring. This contribution has resulted in the following conference paper.

S. F. Al Raisi and E. Edirisinghe, “A Machine Learning Based Approach to Human Observer Behaviour Analysis in CCTV Video Analytics & Forensics,” Proceedings of the 1st International Conference on Internet of Things and Machine Learning. ACM, 2017. [1] (**Appendix 4**)

- b) To the best of our knowledge previous work on classification of human observer groups based on data captured via eye tracking devices has been limited to direct use of derived statistical data and the drawing of graphs for the visualization of differences for the ease of human interpretation. These approaches are not only tedious but will also not be able to identify the presence of fine detailed discriminative features between data captured from different groups. In **Chapter-6** we propose the use of machine learning algorithms for the classification of experts vs. novice participants and male vs female participants, of the experiments carried out within the context of this thesis. The research outcomes prove the capability of machine learning algorithms to carry out such discriminate tasks with a very high degree of accuracy.

### 3. Overall Contribution

In **Chapter-7** we provide an insight into how the outcomes of the research conducted within the context of this thesis was effectively used within another research project to improve the performance of a purely computer vision based human object re-identification algorithm, enabling the algorithms to be re-designed using agent based technology in a manner similar to how a human will carry out the task of visual forensics, showing improvements to the accuracy of the tasks carried out. This work proves the usefulness of the novel research work carried out in this thesis and its potential contributions to CCTV analytics and forensic research.

The research conducted within the remit of this thesis has resulted in the following Secondary contributions:

1. A comprehensive review of literature to investigate existing work on human observer eye tracking behaviour analysis applied especially to analysing CCTV footage (**Chapter-2**).
2. Data collection (dataset): It was noted that a dataset of similar nature in which the analysis of the eye tracking information based on separate parts of the human body has not been conducted prior to this research and hence no public database is available to support the original research presented in this thesis. Hence it was essential to carry out the tasks relevant to capturing this novel dataset. The dataset collected during this research is going publicly available and this will be itself contribution to the future research committee (**Chapter-4**).

### 1.3. Thesis overview

For clarity of presentation the thesis is organised into seven chapters as follows:

**Chapter 2:** Provides a review of literature on the use of Eye Tracking equipment and data thus gathered on studying the visual attention associated features and patterns.

**Chapter 3:** Introduces the reader to the research background and covers details of the human visual system, visual attention, eye tracking equipment used in the experiments and the theoretical background of various statistical and machine learning algorithms used in this thesis for data analysis.

**Chapter 4:** Provides details of the experimental design, the processes adopted in carrying out the subjective experiments, data capture and data preparation for the analysis to be conducted in the contributory chapters that follow.

**Chapter 5:** Proposes the use of statistical and conceptual approaches to data analysis, providing a valuable insight to the participant visual attention features associated with the surveillance task assigned. It's also proposes the novel use of linear regression which is a statistical approach by using matching suit called WEKA for the participant visual attention behaviour analysis when conducting the tasks assigned. use

**Chapter 6:** Proposes the use of machine learning algorithms to classify and distinguish between various groups of participants, e.g. expert vs novice and female vs male.

Finally, **Chapter 7** concludes with an insight to future work, including details of a separate project conducted to prove the concepts that are the outcomes of the research presented in this thesis.

# Chapter 2.

## Literature Review

The aim of this chapter is to provide a thorough and extensive review of the research areas related to the work that are relevant to the research that is being undertaken as a result of this project. To this effect the chapter presents previous research on eye tracking experiments supporting a number of applications with the main aim of reviewing how in these attempts the data obtained from eye tracking was analysed.

### 2.1. Introduction

The eye tracking experiments and related developments have been gaining popularity and has been used to great effect within the field of human computer interaction [2]. By integrating eye tracking into HCI based studies it has helped researchers to explore and understand the human Physico -visual system in more detail. Eye tracking equipment have become valuable tools in helping the researchers to understand human behaviour [3] and investigate the relationship between the design of a system, where people are looking and which point of interest within the scene being viewed has captured their attention when completing a task [4]. It is this relationship that can identify whether a system is deemed usable by developing a deeper understanding of the evaluation and the execution of a task.

### 2.2. Eye tracker research

Eye tracking has been a main tool used in human observer behaviour studies [5]. There are many disciplines that had benefited from eye tracking systems both in their research and applications, such as human psychology [6], marketing [7] [8] [9] [10], aviation [11], education [12] [13] [14] [15] [16], usability [17], medicine [18] [19] [20] [21], marketing [9][10], pilot training assistance [22], entertainment [23], surveillance [24], [25], [26] [27] and so on. In the research presented in this thesis we focus on using eye tracking systems in CCTV operator behaviour analysis and therefore review literature in

this specific discipline in more detail is section 2.3. In the following sub-sections, we present the use of eye tracking systems in other disciplines stated above.

### **2.2.1 Medical Applications**

#### **Detecting Readers with Dyslexia Using Machine Learning with Eye Tracking Measures. [19]**

Rello & Ballesteros in 2015 Applied machine learning and eye tracking methods to automatically classify readers with dyslexia and without dyslexia. The data collected from the output was used in the classification tasks.

#### **Visual search behaviour during laparoscopic cadaveric procedures. [20]**

Dong et al. 2014 presented results of an investigation carried out to discover if there is any difference between surgeons' eye movements during an actual operation and while viewing the same in a pre-recorded operation.

#### **Predicting diagnostic error in radiology via eye tracking and image analytics: Preliminary investigation in mammography. [21]**

The main aim of the research study was using an eye tracker and machine learning to predict any diagnostic error in mammography by using expert observers' (radiologists') gaze behaviour and image content. The second aim was study group based and aimed to create personalized predictive models for radiologists depending on their previous skill and knowledge. The data analysis of this research was carried out using ML.



## **2.2.2 Applications in Education**

### **Eye Tracking and Studying Examples: How Novice and Advanced Learners Study SQL Examples. [13]**

Shareghi Najar, Amir Mitrovic & Kourosh 2015 presented research findings of a project where the main aim was to investigate and analyse how advanced and novice students study SQL examples. The data analysis of this research was carried out using ML.

### **Teachers' gaze and awareness of students' behaviour: using an eye tracker.[14]**

In this work authors conducted research to investigate the gaze behaviour of teachers while they watched pre-recorded video of a student's behaviour in the classroom. The participants' gaze was recorded using a Tobii T60 and T120.

### **Tracking learners' visual attention during a multimedia presentation in a real classroom. [15]**

Yang et al. 2013 investigated students' visual attention while viewing a multimedia presentation (Power Point) in a real classroom. In order to achieve the aim of the research an eye tracking sensor technology was applied to record and analyse eye movement allocation over a multimedia (PPT) presentation.

### **Learner Behaviour Analysis on an Online Learning Platform.[16]**

The first aim of this study was to investigate the learners' behaviour on an e-learning platform in order to produce user profiles that reorganized and regrouped learners depending on their behaviour on the e-learning platform. Secondly, the research resulted in the creation of a system to better understand the requirements of the online learning user in order to improve the learning situation. It was shown that the proposed system can be integrated with the learner agent of an intelligent tutoring system (ITS).

### 2.2.3 Applications in Aviation

#### **Pilot's Attention Allocation During Approach and Landing -Eye- and Head-Tracking Research in an A330 Full Flight Simulator. [11]**

Anders 2001 investigated the use of eye tracking technology within an advanced flight simulator. The professional pilot's head and eye movements were monitored and recorded under real flight conditions in an Airbus A330 full flight simulator (FFS). This recording was used to investigate human computer interaction (HCI) behaviour significantly to study information selection and management and mode awareness in a modern glass cockpit.

### 2.2.4 Applications in Usability Studies

#### **Effects of User Age on Smartphone and Tablet Use, Measured with an Eye tracker via fixation Duration, and Saccades Proportion. [17]**

The aim of this research study was to provide an insight into the effects of user age on interactions with smartphones and tablets applications. An eye tracker (Eyelink-1000 desktop device amounted with IR illuminator) was used to track and record users' eye movements which were analysed to understand the effects of age and screen-size on browsing effectiveness. The study proved that the elderly users faced difficulties when using or interacting with smart phones and tablet devices as compared to other age groups. All other age groups were affected by screen sizes; the small screen size has smaller cascades proportion indicating uneasy interface browsing compared to large screen size. The results have been statistically evaluated using two-way ANOVA.

## 2.2.5 Applications in Marketing

### **Game, Set, Match! Brand eye-tracking on TV sport programmes.**[10]

Roy et al. 2008 conducted research to discover the distribution of attention between football games, and banner advertisements by using a sample of people involved in a football broadcast event.

### **Seeking attention: an eye tracking study of in-store merchandise displays.** [9]

Huddleston et al. 2015 The aim of this research was to understand online shoppers' search behaviours. The eye tracking related work where machine learning has been used for observer behaviour analysis presented above has focused specifically on application areas that are not directly related to the subject area of research presented within this thesis. The review also shows that the main approach that is used to analyse the data that gets captured is using statistical methods. In the few attempts machine learning has been used, the study has been limited to the classification of expert and novice users based on their behaviour that is analysed by the eye tracker/s.

The research of direct relevance to the key area of application this thesis is to focus, is detailed below.

## 2.3. Eye Movement Tracking in CCTV

### **Statistical analysis of visual attentional patterns for video surveillance.** [24]

In 2013, Roffo et al. showed that understanding the way people visually analyse video sequences, beyond the content they observe in a video, is vital for the understanding and the prediction of people's activities. The analysis presented in this study was based on eye tracking data on CCTV video sequences. Because of the extensively investigated higher capabilities of expert operators in predicting violence in surveillance footage, the main goal of the research proposed was to understand how expert CCTV operators analyse such videos, and if there is a difference between expert operators and novice participants. It is noted that all the analysis was carried out using statistical approaches.

## **Suspiciousness perception in dynamic scenes: a comparison of CCTV operators and novices. [25]**

In 2013 Iain Gilchrist et al. Proposed that CCTV operators sometimes may be able to predict trouble, and trouble hotspots, very rapidly and they can take less than two-second delay before making any decision. The study suggests that a waiting strategy permits them to take an action and identify supplementary visual information, and may be a thoughtful approach or method to help the operators make a correct decision and reduce the number of false alarms. In this study, the CCTV operators were requested to view 80 clips of recorded footage (such as a night-time view of a car park, a shopping street underpass, a nightclub entrance and a cash point) for one minute each. They were instructed to observe the clips for behaviour believed to be suspicious enough in order to alert the relevant authorities and used a joystick device to specify the perceived level of suspicious behaviour. The eye tracking sensor data were used to calculate the relationship between these ratings and the CCTV operators' patterns of gaze.

The research results were compared against a group of untrained observers (novices). A statistical data analysis tool was used to analyse the operator's performance. The study established that trained operators spend more time than untrained observers when determining whether a scene is suspicious. The collected data also suggested that trained operators moved their eyes to the significant part of the scene earlier, and they followed very similar viewing patterns. They would normally look at specific areas of focus rather than moving quickly between random locations as was the case with untrained operators and were also much more consistent in spotting suspicious events in ambiguous footage. In addition, the study revealed that CCTV operators decide to undertake additional visual processing to reduce their doubt and help them to make an accurate decision while viewing the footage. While in safety and security situations, where quick and accurate judgements are critical to public safety, this is clearly of significant importance. The data analysis of this research was carried out using statistical approaches.

## **Automatic Human Behaviour Recognition and Explanation for CCTV Video Surveillance. [26]**

Robertson's et al. research found that explanation of global scene activity, mostly where interesting events had happened, can aid human observer analysis of high volumes of captured data. This was mainly achieved by using an extensible, rule-based method that was generated based on past studies of observer behaviour analysis using statistical methods. This research paper was concerned with creating high-level text reports and explanations of people's activity in video from single, static cameras, with the motivation being to allow surveillance analysts to provide situational awareness regardless of the presence of huge data. The paper focused on urban surveillance where the pictured person was shown in low/medium resolution. The final output required was text descriptions that explained the interactions that took place and described what was happening (observed human activity). The research also states that the whole system denotes a general technique for video understanding, which involves a guided training phase via an experienced analyst. It is however noted that all the data analysis conducted within the scope of the research presented in this paper were carried out using statistical approaches.

## **The ( Change ) Blindingly Obvious : Investigating Fixation Behavior and Memory Recall during CCTV Observation. [27]**

Dr. Gemma Graham has integrated eye tracking technology into many research projects. In one of her studies she investigated how people observe the CCTV footage[27], mostly when they are given instructions to focus on given features in the video and seeing different severity of crimes. In this research four experiments were carried out in order to investigate the following research questions: “(a) does instruction and/or event type impact where observers view during CCTV footage observation?; (b) how do task instructions and central and marginal information influence fixation behaviour during CCTV observation?; (c) what is the effect of change detection on memory recall during CCTV observation?; and (d) do verbalization, attentional set and/or repeated viewing improve change detection rates and memory recall for CCTV footage?”.

The study was conducted with a different number of participants for each experiment. During the experiments, the participants' eye movements were recorded by a professional eye tracking system. The observer performance was analysed by using a statistical data analysis tool called ANOVA. This research found that the instructions did not significantly affect gaze behaviour for dynamic scenes and any changes in detectors evoked more accurate detail from the CCTV footage as related to the non-detectors, but only in the case where the seriousness of the crime had increased. However, when the observer was repeatedly shown the CCTV footage the rates of change detection improved greatly, but verbalisation made no difference in terms of change detection and the accuracy of memory recall. The last finding obtained from this research can provide help to inform training courses aimed at instructing users on how to optimally attend to surveillance video.

## **2.4. Machine Learning and Surveillance**

An American pioneer named Arthur Samuel defined machine learning as “a field of study that gives computers the ability to learn without being explicitly programmed”[28]. The researchers conducted many studies on machine learning in order to make more intelligent machines that can replace as many human tasks as possible. Machine learning is employed in many fields such as education [13] [29], surveillance systems, medicine [30] [31] [32] [33], advertising [34], entertainment (Netflix), autonomous vehicles (self-driving) [35], face recognition [36] and many more.

The following section describes numerous applications of machine learning in the field of surveillance.

### **Automated Detection of Firearms and Knives in a CCTV Image. [37]**

In this paper, the researchers proved that it is possible to build a warning system in order to provide advance alert for any dangerous situation, which may help in reducing the number of victims in a very short time. This research paper concentrated on two tasks of automated detection and recognition of any dangerous activities appearing on CCTV systems. The paper proposed an algorithm to notify the CCTV operator when a firearm or knife is visible in the image thus being able to alert the operator to take quick action.

### **Identifying moving bodies from CCTV videos using machine learning techniques. [38]**

This paper proposed face detection from surveillance cameras using SURF feature extraction and for comparing image descriptors. Normally the task of recognition involves comparing each face detected from the video with all the other faces saved in the database. When new faces come into the camera view the database of images will be updated. However, the face labelling can be done by a human or not be done at all. According to the researcher, if the proposed system is used, it does not require a database of images to start with. The system automatically creates its own collection of images, and then searches for the future occurrences of those images.

### **A Machine Learning System For Human-in-the-loop Video Surveillance.[39]**

This paper introduced a novel real-time surveillance monitoring and offline retrieval surveillance system that studies the properties of items that are interesting to a CCTV operator. This proposed system combines methods from different areas such as human-computer interaction, computer vision, and machine learning. In this system, the items that are of interest to the CCTV operator are automatically studied via tracking the operator's eye gaze positions while they monitor the surveillance video in order to find interesting activities and synthesize interesting actions first.

The researchers plan to extend the proposed system to enable it to work with multiple video and CCTV operators where individual operators view the surveillance video with different interests in mind.

**Predictive Policing: Using Machine Learning to Detect Patterns of Crime.[40]**

In this paper, the authors proposed a machine learning pattern detection method named as “Series Finder” that can support the police in determining and gaining a better understanding of patterns of crime. This method produces a pattern of crime, starting from a seed of two or more crimes. Basically, the method was trained to discover patterns in housebreaking (i.e. theft within houses) by learning from historical data collected from the Cambridge Police Department’s crime analysis unit.



## 2.5. Summary & Conclusion

The research work reviewed above reveals that a significant number of application domains benefit from observing human behaviour and intent via the use of eye tracking devices. Each of these studies requires an approach to analyse the eye tracking related data that is captured and the main method adopted has been the use of statistical approaches and functionality within statistical software packages such as ANOVA.

All previous research mentioned in this chapter with respect to the human observer analysis of CCTV video have not been sufficiently detailed and rigorous enough to analyse observer attention on different parts of the tracked object's body, given different instructions [24] [25] [26] [37] [38] [40]. Further these studies adopted statistical and/or conceptual data analysis approaches that are limited in their analysis capability in particular in their ability to learn from past observed behaviour and predict towards future outcomes, based on behaviour. Further the investigations carried out with regards to observing eye tracking data captured via visually identifying human objects based on their entire body areas significantly constrains the potential to carry out detailed behaviour analysis. The research presented in this thesis argues that any analysis that will produce realistic behavioural analysis should analyse a human object being viewed in the manner that human's would naturally do and this needs attention on articulated body parts, separately. For example, humans, when looking at another human would normally either look at the face, the clothes being worn or anything being carried, e.g. colourful handbag etc..This lack of detailed investigation in previous literature leads one to many open research questions that still needs investigation.

Further to the above observation the literature review conducted above clearly demonstrates that most of the work in classification of different human observer groups based on data captured via eye tracking devices has been limited to the direct use of derived statistical and/or conceptual data analysis approaches, data and the drawing of graphs for the visualization of differences for the ease of human interpretation [24] [25] [26] [37] [38] [40]. These approaches are not only tedious but will also not be able to identify the presence of fine detailed discriminative features between data captured from different groups.

The capabilities of machine learning when used in data analytics not only could complement statistical data analytics, but could also enable effective behaviour analysis, pattern recognition and behaviour prediction, all of which are vital aspects in detailed human observer analysis. Although a very few attempts have been made in using machine learning for human observer analysis in particular with regards to CCTV operator analysis, machine learning will be able to deliver significant additional pattern recognition, prediction and behaviour analysis tasks, the key focus of the research conducted in this thesis.

In the proposed research the data mining tool WEKA is used for CCTV observer performance analysis, using a linear regression model and other machine learning algorithms, to analyse, identify, learn, understand, make predictions and uncover hidden patterns in the eye tracking data / behaviour. Use of WEKA prevents this research having to focus on the implementation of standard machine learning algorithms, saving time and effort

The research conducted within the context of this thesis and presented in **chapters 4, 5,6 and 7** contributes towards closing this research gap as briefly highlighted in **Chapter 1**.

# Chapter 3.

## Research Background

The purpose of this chapter is to provide a thorough and extensive information about the conceptual and theoretical background of the subject areas this thesis covers. To this extent the chapter initially presents the Human Visual System (HVS), operational aspects functionality of eye tracking systems, and details about visual stimuli. The chapter also presents information about different data mining tools and methods that have been considered to analyse the data collected via the eye tracking systems in this thesis, the difference between statistical approaches to investigating collected data and the proposed machine learning algorithms based approach and concept behind the thinking aloud method often used to complement the data capture during an eye tracking experiment. The information provided in this chapter is most essential to understand the main body of the thesis, if the readers are interested in more detail regarding this information can refer to the original publications and references where appropriate.

### 3.1. The Human Visual System

#### 3.1.1 Human Eye: perception

The human eye is one of the most vital, important, valuable and very complex human organs. It has a very complicated structure. It uses light and allows humans to see objects and the colourful world around them [41]. When a light beam is reflected off an object it enters the eye through the pupil and passes through the lens. The lens reverses the viewed object's picture upside down and focuses it onto the retina. The retina has two types of light-sensitive cells called rods and cones, which transmit impulses to the brain via the optic nerve. Finally, the optic nerve transfers the readings of the light sensitive cells from the eye to the brain. When the brain obtains this reading, it directly tries to process it by

turning the image back to the right way up and detects the object which has been seen by the eyes [42].

### **3.1.2 Visual Perception and Attention**

One of the ways we experience the world surrounding us is via our vision (eyes). Visual perception can be simply defined as the way humans get to know and understand the world surrounding them through what they see by using two organs, the eyes and brain[43]. Therefore, what is seen through our eyes is interpreted by the brain. The brain is the most significant organ in the human body and enables one to work out what the eyes see [44]. It is also known as sight, vision or eyesight. It is caused when visible light reaches the eye and we have the ability to interpret the information and our surroundings [45]. In psychology, visual perception is the process or ability by which sensory information from our eyes is transformed to produce an experience of shape, size and brightness of objects and we perceive the distance telling us how close or how far away an object is [46]. Moreover, it involves the brain's ability to interpret the surrounding environment by processing information; that is, it processes and integrates incoming functional activity as an outcome. Whenever we look at the world surrounding us, we do not process every single piece of information available. We basically perform a perceptual selection process called attention. In this process, we select and filter out unnecessary or less significant information so that we can deal with the most important elements.

### **3.1.3 Computer vision versus human vision**

Computer vision is a field that includes methods for acquiring, processing, analysing, and understanding images and, in general, high dimensional data from the real world in order to produce numerical or symbolic information, e.g., in the forms of decisions[47]. However, for many complex problems raised in the field of computer vision, human learners are much better than machines in learning and recognizing. The humans visions have highly accurate internal recognition and learning mechanisms that are not yet fully understood by researchers. They often have experienced more extensive training data

during their lifetime with the visual world [48] that a computer vision based system will never be exposed to, during training. The most challenging goal of computer vision at present is to solve visual challenges for which human observers have gained effortless expertise, such as face recognition, object recognition, image segmentation, medical image analysis and more [48]. However, there is still some obvious gaps between human performance and performance of a computer vision application. With the advances made in research and technology development, the researchers are trying their best to improve current methods to not only mimic the human visual system, but also reach its performance levels. However, the changes to date are below the level of human performance.

### **3.1.4 Eye Movements & Visual Attention**

The eye plays an important role in helping humans and other species to easily detect and recognise objects. The human eyes are always scanning and moving until they stop and focus on a point of interest. However, during visual perception the eye has two major functions, fixation and tracking[49][50]. Fixation is to position the target object into the fovea. This function allows our eyes to maximize the focus we can give to the object even when the object is moving. Tracking an object is important because most real world objects move, and without the ability to track, we will have a very difficult time perceiving things around us[50].

The human eye makes different kinds of eye movements, and each type of movement provides a different function in visual perception. There are over ten different types of eye movements, of which the most important ones for us and for the eye tracking perspective are saccades, fixations and smooth pursuit.

A fixation is the sustaining of the eye focus on a single area [51]. When the eye moves very quickly between any two given consecutive fixation points on the viewed scene, these movements are called saccade. The quick eye movements allow the eyes to concentrate on various parts of the visual world to gather as much information as possible and draw a general picture in our brain [50]. However, the main movement of interest for our research is smooth pursuit. Smooth pursuit movements are used to track moving

objects, and as stated previously this is an important function because most objects in the real world move. In our daily life this eye movement is most vital, as without the ability to track moving objects surrounding us we would not be able to walk safely on the streets because we wouldn't be able to perceive cars and bikes etc. [50]

## **3.2. Measurement of eye movements**

In 1897 the first eye tracker was developed[52]. Since many different methods and sensors have been developed for measuring eye movements. The equipment (sensor) we used for measuring eye movements in this thesis is a Tobii eye tracker, which consists of a computer screen with a built-in camera that captures the participants' eye movements.

### **3.2.1 Eye tracker**

In the simplest terms, eye tracking refers to recording eye movements whilst a participant examines a visual stimulus [53]. It basically tracks the eye movements and that tells us where, how and when people look. This gives meaningful insights about behaviour and performance that may be important for any research. Eye tracking technology is increasingly applied to research in several areas, such as design, testing and diagnostics, and entertainment etc. Eye tracking data is collected using either a remote or head-mounted eye tracker connected to a computer.



**Figure 3.1: Tobii Eye tracker Source [54]**

An eye tracker consists of two main parts: sensor (camera & projectors) and algorithms. The projector sends out near infrared light beams toward person's eyes. The eyes receive this light and reflects the light. The reflected light is then captured by the eye tracker's cameras. Finally, the image processing algorithms find specific details in the user's eyes and reflection patterns, and understand the image stream generated by the sensors. They basically calculate the user's eyes and gaze point on a device screen. Based on details gathered, a mathematical algorithm calculate's the eye's positions or in other words eye tracker knows where the human subjects looking. The figure 3.2 illustrates how an eye tracking works.

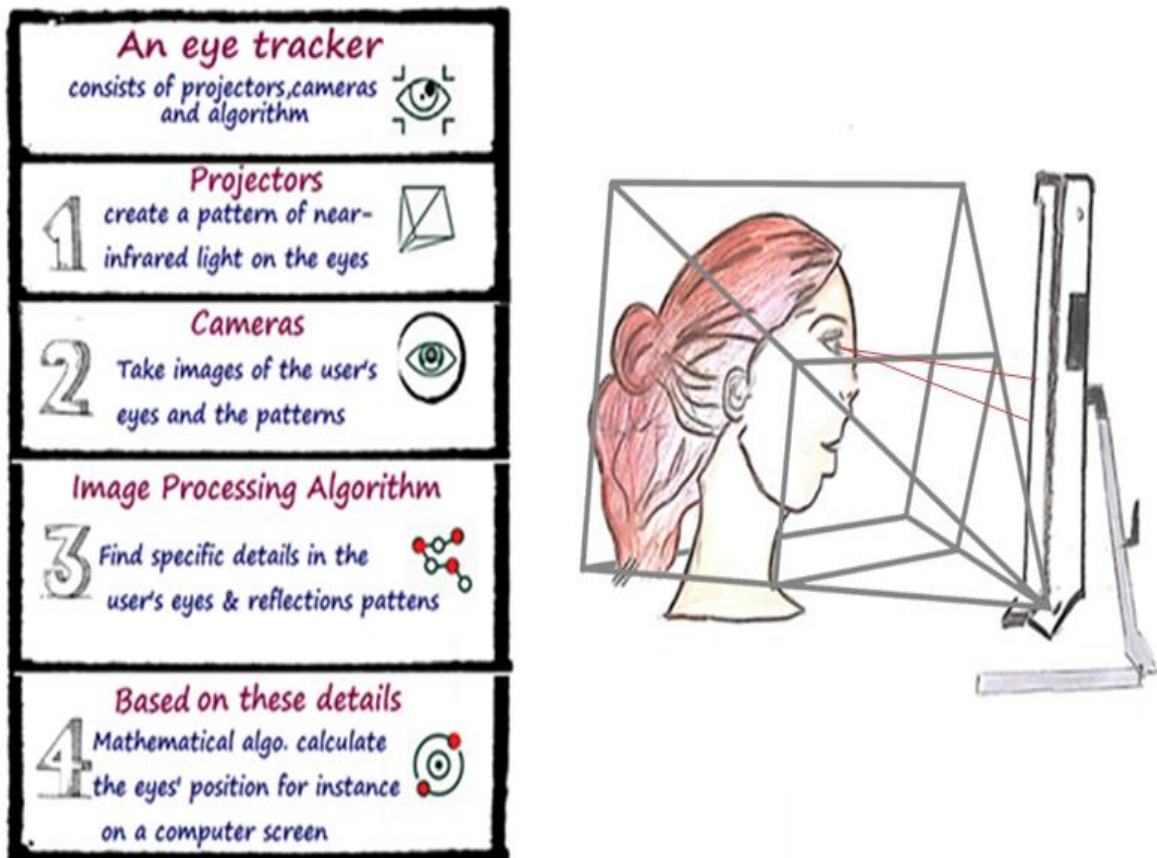


Figure 3.2: Illustration of how eye tracking works[55]

For this study, we collected experimental data using the Tobii T60 eye tracker [54]. This non-invasive device is connected to a computer; it captures and records the position of the user's gaze on the screen using infrared cameras. This particular eye tracker was chosen for a number of reasons:

1. A user friendly and easy to use device.
2. Simple, quick and fully automatic calibration producer: The system displays a number of points to follow in order to help the eye tracker to learn or know the characteristics of participant's eyes. This procedure helps the sensor to work as expected, for eye tracking to work accurately. The calibration is done in less than a minute and it can save and reuse calibrations for future use.
3. Allow the participants to freely move their head.
4. Integrated and comprehensive eye tracking, analysis and visualization software provided with the eye tracker.

### **3.2.2 Tobii studio**

Tobii eye trackers offer a unique platform and an opportunity to work with Tobii Studio; it is considered one of the most advanced tools for visualization and analysis [56]. The eye tracking data is easily processed for useful comparison, interpretation and presentation. This software provides a complete platform for easy designing, presenting testing procedures, recording, observing, visualizing and analysis of eye gaze data by using statistics. Tobii Studio features a new 'Area of Interest' (AOI) tool that allows users to easily create areas of interest on stimulus, both on static (images) and on dynamic areas of interest (video and animations). This allows the eye tracking researcher or analyst to calculate quantitative eye movement measures easily by automatically analysing the AOI [56]. In addition, setting up, participant eye calibration and carrying out an experiment with Tobii Studio software is easy as compared to other competing eye trackers that are available commercially. These amazing built-in features make the software user friendly and allow users to work with the system without extensive training or the need for the participation of experts.



### **3.3. Visual stimuli: static versus dynamic stimuli.**

A static stimulus is a visual scene that does not change in time, such as images, text and advertisements (containing images and text or both).

#### **3.3.1 Text-based stimuli**

When we look at a visual scene or read a text, a series of eye movements occur known as the scan path. The scan path is the sequence of eye fixations and saccades made when someone looks at a visual scene or reads a text [57]. In 2011 Holland et al. [58] proposed a method for biometric identification based on the scan path using a text as a stimulus. In this paper, the authors have reported the eye trajectories of the observers during a short reading task and analysed quantitative properties of the scan path produced while reading; for example, the number of fixations and the average duration of each. During repeated viewings of the same text-based stimulus most of the participants tend to repeat certain scan paths. The results indicate that variations in the scan between observers and the similarity of paths in the same person might make the text-based stimulus hold promise to excite the behaviour of the eyes for an eye movement biometric technique.

#### **3.3.2 Image-based stimuli**

Komogortsev et al. [59] studied the use of static images as a stimulus for eye movements. In this study, the authors used a complex pattern image called Rorschach inkblots and concluded that the inkblots could be used for eye movement biometrics. The inkblots were used in psychological tests where the subjects' perception of the inkblots is recorded and analysed.

### 3.3.3 Dynamic stimuli

Dynamic stimulus is a visual scene that changes in time, such as videos, animated pictures, dynamic visualizations and real-world scenarios.

- **Jumping point stimuli**

The first work, which attempted to study eye movements as a personality trait for biometric identification, was conducted by Kasprowski and Ober[60]. The researchers designed a stimulus consisting of a 3x3 matrix of yellow dots and used this stimulus to excite the movements of the eyes for biometric identification. Each yellow dot flashed one after the other for 550ms to create an animation of a jumping point. The jumping point stimulus has been extensively used by many researchers for eye movement biometrics. Komogortsev et al. [61] studied a similar approach to evaluate the feasibility of using eye movements for biometric identification. The purpose of the study was to identify the movements that could be used for biometric identification by using a simple dynamic stimulus such as a jumping point.

- **Movie stimuli**

Kinnunen et al. [62] used a movie as a stimulus for exciting the eye movements for biometric identification. The reason for using a dynamic stimulus in their research was that watching a movie is not associated with a specific task or instructions to perform, and movies as a stimulus method can be used for stealth identification.

In the study, all the participants were requested to watch 31 movies in 25 minutes presented on an eye tracker screen one after another.

The eye movement signal was segmented into short-term signals to determine how much time was necessary for collecting usable eye movement data for biometric identification. The authors worked out that five minutes of movie watching was sufficient to collect eye movement data for biometric identification.

### 3.4. Area of Interest

The AOI (Area Of Interest ) and DAOI (Dynamic Area Of Interest) is a powerful concept that allows the eye tracking researcher or analyst to easily define areas of interest on the stimulus such as videos and images[63]. This concept is integrated in Tobii Studio and helps us easily and simply create a boundary around a target body. For this study, we used a DAOI because the target is not static; the target is dynamic as it is moving or walking across a scene in CCTV footage. The DAOI calculation tool in Tobii is considered a powerful way to quantify gaze data; rather than knowing where the test participants were looking, we can get information on which object (DAOI) they were looking at by automatically calculating and analysing DAOI. The DAOIs utilised in the research conducted within the scope of this research is illustrated below.



**Figure 3.3: The DAOIs in research conducted**

## **3.5. Data mining tools**

In this section, we present the different data mining tools and methods that were used to analyse the data collected in the research of this thesis.

### **3.5.1 SPSS**

In Chapter-5 we show the use of Statistical Package for the Social Sciences (SPSS) version 23[64] in the analysis of the data collected via the eye tracker sensor used in the proposed experiments. SPSS Statistics is one of the most popular advanced software packages used for statistical analysis. This software is commonly used in the social sciences and many other disciplines for managing data and calculating a wide variety of statistics. The reason for using SPSS in our research is that SPSS includes many features and functions that help us to make the most out of the statistical data we obtained from Tobii studio software. It helped in finding trends in the data to produce a clear and correct outcome and offered the means to draw effective graphs and provided many other benefits.

### **3.5.2 Statistical Method (ANOVA)**

Using data generated from Tobii Studio, a one-way analysis of variance (ANOVA) data analysis method was used in the proposed work to establish if there are differences in the areas of interest for different eye tracking measurement metrics. The ANOVA is appropriate for various reasons. Firstly, each measurement metric focuses on a single variable. Secondly, One-way ANOVA is an inferential statistic that allows generalization of the findings from a sample to the entire population. Lastly, ANOVA allows for comparison of more than two groups. In this case, there are three or more groups that represent each area of interest. In addition to One-way ANOVA, graphs were used to provide visual representation for comparison of the groups.

### 3.5.2.1 Interpreting the key results of One-Way ANOVA

After carrying out statistical analyses of data, it is very important to report findings (what we did and what we found) in a simple and easy to understand language.

The following are the relevant parts of the SPSS output that is required in order to report our findings: the  $p$ -value  $F$ -value and the  $df$  value [65] .

#### a. $p$ -value

$P$  is a short form of probability. It is used in order to determine whether our condition means were statistically significant or not significant (different from one another). The  $p$ -value ranges from 0 to 1, If the  $p$ -value is  $>.05$  this indicates that there is no significant difference between the means conditions. However ,if the  $p$ - value is  $\leq .05$  this indicates that there is a statistically significant difference between means conditions [66] .

#### b. $df$

$df$  stands for degrees of freedom. It is very important in order to calculate statistical significance. It is used to represent the size of the sample, or the number of samples used in the test.

#### c. $F$ value

The  $F$ -value is the value of the test statistic in ANOVA. The  $F$ -test determines whether group means are equal. It is determined by dividing the mean square between groups by the mean square within groups.

Mean squares are used to represent estimates of variance across groups. It is calculated as a Sum of Squares ( $SS$ ) divided by degrees of freedom ( $df$ ). Assume that  $N$  is the total number of samples in a study, and  $K$  the number of groups, then the:

$$\text{Mean squares} = \frac{SS \text{ total}}{N-1}$$

Mean Square between groups compares the means of groups to the grand mean:

$$\frac{SS \text{ between groups}}{k-1}$$

If the means between groups are close together, this number will be small.

However, Mean Square within groups calculates the variance within each individual group:

$$\frac{SS \text{ within groups}}{N-K}$$

Finally, the F-value can be calculated as:

$$\frac{MS \text{ between groups}}{MS \text{ within groups}}$$

In this thesis, the result of one way ANOVA is reported in APA Style [66] as illustrated in figure 3.5. In order to interpret one way ANOVA results, one has to inspect the last column for Sigvalue (p-value in our definitions above) in the ANOVA box figure 3.5. If the Sig value is > .05, we can conclude that there is no statistically significant difference between conditions. However, if the Sig value is <= .05, we can conclude that there is a statistically significant difference between some or all conditions. The result can be reported as shown below and in figure 3.5.

There was a statistically significant difference in the Fixation Duration between the four areas of p-value < 0.05 as determined by one-way between subjects ANOVA [**F (2,57) = 8.207, p = 0.01**] for video-10

"F [df Between Groups, df Within Groups) = F-value, P-value]"

**ANOVA**

Fixation Duration

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4.233	2	2.117	8.207	.001
Within Groups	14.700	57	.258		
Total	18.933	59			

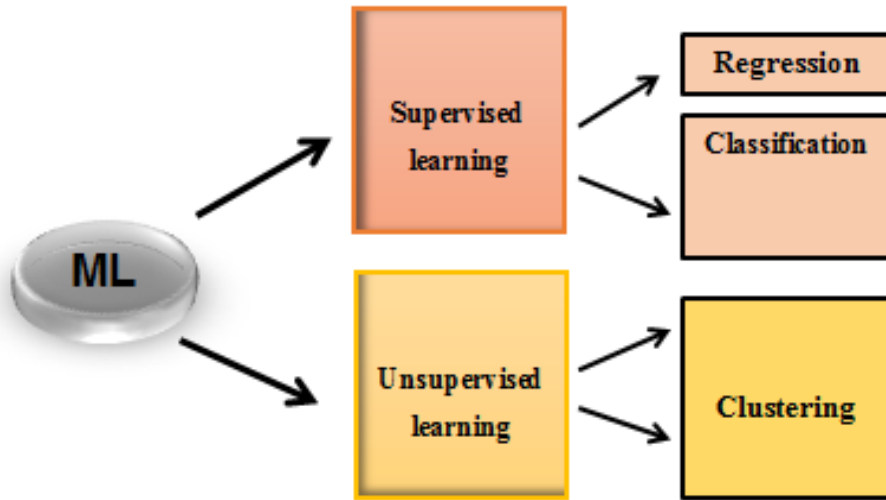
- ① degrees of freedom between groups
- ② degrees of freedom within groups
- ③ f value
- ④ p value

**Figure 3.4 : Reporting one way ANOVA in APA style**

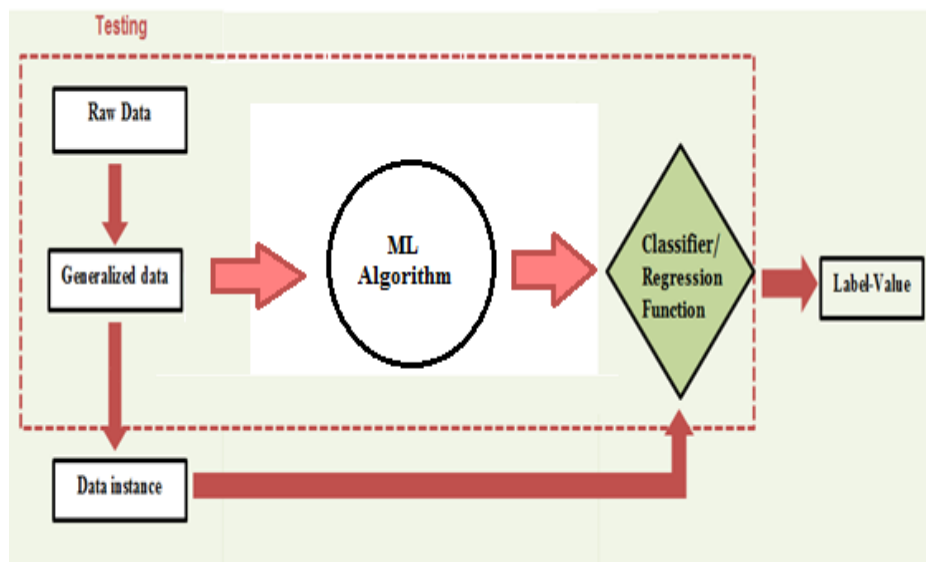
### 3.5.3 Machine learning

Machine learning (ML) is the area of study involving different algorithms and data analysis techniques for recognising or identifying patterns and relationships in given data [67] . Essentially, this field gives computers the ability to learn, make and improve predictions or behaviours on data without being explicitly programmed [68]. Many researchers in various fields such as medical science, pharmacology, agriculture, archaeology, games, business, education and others have been fascinated by machine learning and its use. Therefore, research continues to be conducted in machine learning in order to make more intelligent machines that can replace or relieve human tasks. Machine learning can be classified into two types of learning: supervised learning (regression and classification techniques) and unsupervised learning (clustering techniques) see Figure 3.5. In supervised learning, the data is labelled based on prior knowledge in order to help

the program to make a decision or predictions. In contrast, unsupervised learning does not require any labelling for data as the algorithm itself can learn and will find the relationships between different inputs. In this thesis, supervised learning, also known as the predictive modelling approach is being used. Figure 3.6 illustrates the general process of machine learning (supervised learning) in which the testing processes have been separately illustrated.



**Figure 3.5 : Types of ML**



**Figure 3.6 : The general process of machine learning**



### 3.5.3.1 Predictive models

The predictive model or supervised learning is divided into two types to make a decision or predictions.

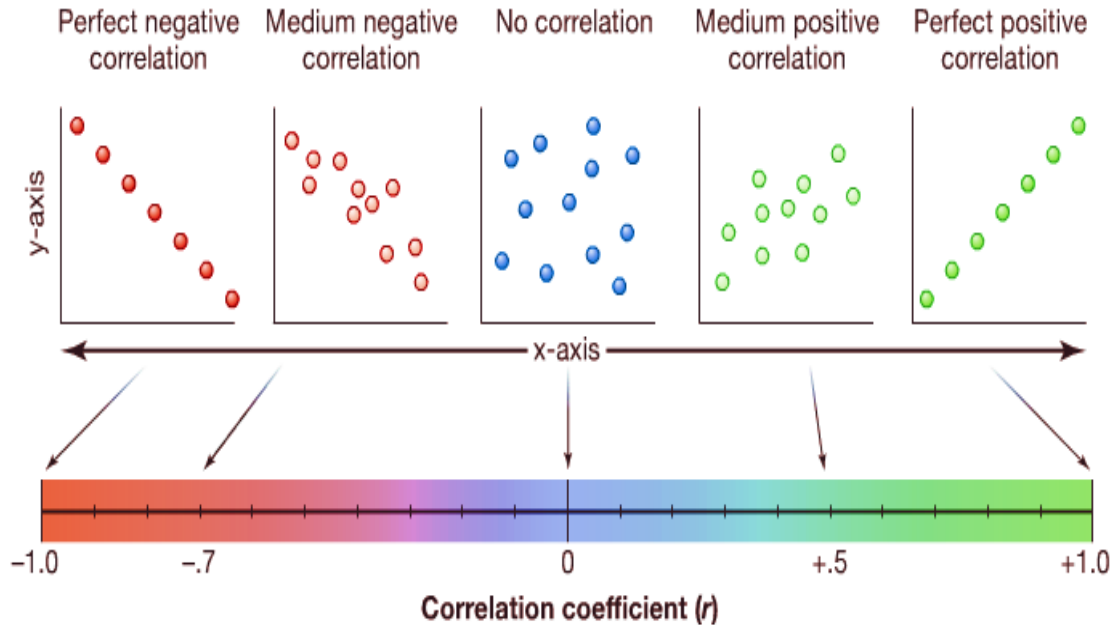
1. Regression Methods (the target class is numeric)
2. Classification Methods (the target class to be predicted is nominal)

#### 1. Machine Learning Regression Methods

Regression analysis is one of the most frequently used tools in predictive modelling. It is used to find the relationships between dependent and independent parameters (variables)[69]. It is also used to find the strength of the impact of multiple independent parameters on a dependent variable. In this thesis, a machine learning method with linear regression was applied to our data and is proven to be effective despite its simplicity as compared to other more complex regression methods. It is noted that other regression methods exist, e.g. polynomial regression and other regression approaches based on specific mathematical functions.

- **Linear regression(LR)**

Linear regression[70][71] is one of the most common modelling techniques used for predictive modelling. It is used to describe the relationship between two variables named the dependent variable and the independent variables in the form of a linear equation of the type:  $Y = a + bX$ , where Y is the dependent parameter (variables) and X is the independent parameter (variables). In order to show the power of a relationship between variables, a correlation coefficient is used. The correlation coefficient describes the magnitude of relationship between two variables. The greater the absolute value of a correlation coefficient the stronger is the linear relationship. The correlation coefficient ranges from 1 for perfectly correlated results, through to 0 when there is no correlation, to -1 when the results are perfectly correlated but negatively, i.e. for example the reduction of the value of the independent variable results in an increase of the value of the dependent variable and vice versa (see Figure 3.7).



**Figure 3.7: Linear correlation**[72]

## 2. Classification Methods

In machine learning, classification involves predicting to which set of categories a new observation belongs[73] , on the basis of a set of training data containing observations whose category is previously known. The following are ML algorithms that have been used in our study presented in Chapter-6.

### a) Multilayer Perceptron (MLP) Algorithm

Multi-layer perceptron (MLP) [28] is a feed forward neural network with one or more layers between input and output layers. Feed forward simply means that data flows in one direction from input to output layer and it never goes backwards. This type of network is learned by the back propagation learning algorithm [28] .It is used to amend the weight of hidden nodes based on their individual contributions to the final data prediction.

## **b) Random Forest (RF)Algorithm**

Random forest [74] is an ensemble learning algorithm which uses multiple learning algorithms to build a single model. It is considered one of the most useful algorithms capable of performing both regression and classification. This algorithm works building on many decision trees at a training time which is based on random selection of given data and a random selection of variants.

## **c) RepTree Algorithm**

Reduced-error pruning or in short Reptree is a fast decision regression tree algorithm that creates a regression tree by using information gained as the splitting criteria. The pruning mechanism reduces the size of decision trees by deleting parts of the tree that reduce the performance to classify instances. To clarify, Reptree considers each attribute as a candidate for pruning. The technique deletes a sub-tree of a node if there are any chances to increase the performance and assigns the class for the node. Otherwise, the model will go onto the next node. This process will be repeated until the leaves are reached [75].

## **d) Bagging Algorithm**

Bagging [74] [76] [70] is a short form of Bootstrap Aggregation; it is an ensemble learning algorithm that can be used for classification or regression in order to improve the performance of the model. In the bagging algorithm, each associate classifier is built from a different training data set; while each training data set is generated by sampling from an original data set with uniformly random replacement. The final model output from the bagging method frequently performs better than a single model that performs input on the whole dataset and it never always gets worse. It is noted that the random forest algorithm is a kind of bagging on its own [77]. It performs better than bagging due to the extra randomness utilised while building the model. However, the random forest act differs from bagging when it splits the node of a tree. Thus, rather than looking for the best point to split the node among the entire set of variables, it randomly chooses sub-features to search for.

### 3.5.4 Machine learning software

A popular free open source machine learning software is WEKA (Waikato Environment for Knowledge Analysis) [78]. WEKA is a collection of state-of-the-art machine learning algorithms and data pre-processing tools. It provides wide support for the whole process of experimental data mining. It can evaluate many learning methods statistically, and visualize the input data and learning results at the same time.

Further, WEKA contains many learning algorithms including a wide range of pre-processing tools for data pre-processing, classification, regression, clustering, association rules and visualization. It is also well suited for developing new machine learning schemes. In addition, we can associate different methods and select the most appropriate one for our problem. In the research conducted we used WEKA both for experimental analysis of data and also visualisation of results.



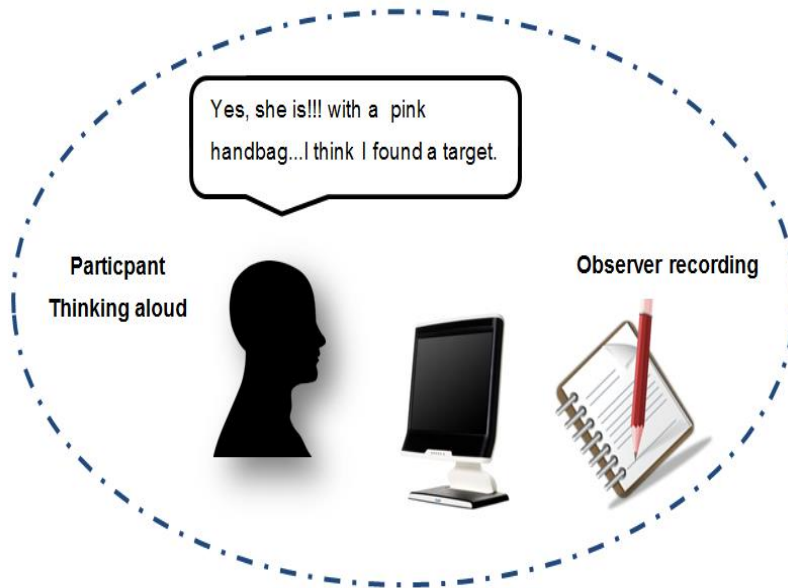
Figure 3.8: Weka Machine learning software

### **3.6. Statistical Analysis Versus Machine Learning**

A statistical model is more about developing a model that explains the data, while machine learning approaches aim is to develop a method to solve a problem[79]. Although they are two different branches, according to Witten et al. [70] they are almost exactly the same. The main concern for statistics is the hypothesis testing and a main concern of machine learning is to develop the process to look for possible hypotheses.

### **3.7. Think Aloud Method(TAM)**

The think aloud technique has been applied and integrated into many research studies that uses eye tracking devices, such as investigating reading strategies [80], decision making [81] and sport performance [82]. Unfortunately, up until today there has been little forensic research that has integrated the think aloud technique in the study[27]. Among the few is Weaver and Carroll's research study [83], in which two groups of people (expert and novice shoplifters) are requested to think aloud regarding how they would plan to shoplift while they are walking inside the shop. In the think, aloud method the participants are requested to think aloud or verbalize their thoughts as they are performing a set of given tasks. In thinking aloud methods, participants are asked to think aloud and say whatever they are looking at, doing and feeling, while the observer is recording their thought process [84]. The main objective of this method is to gain a valuable insight into the thought process behind the user's action. Eye tracking should be combined with another method and data because eye movement cannot always be interpreted correctly without participants providing context to the data [85], so it is better to gain some important data or information from participants through an approach such as using the 'Think Aloud Method'.



**Figure 3.9: Weka Machine learning software**

### **3.8. Person or human re-identification**

Person re-identification is one of the most vital tasks in automated video surveillance and has become one of the most interesting and challenging topics under consideration in the area of intelligent video surveillance research in the past few years.

Re-identification is basically a pipelined process consisting of a series of image processing techniques that finally indicates whether it is the same person who has appeared on different cameras [86]. Thus, the term human or person re-identification can be defined as the task or problem of establishing whether a person currently in the field of view of a CCTV camera has been seen earlier either by that camera or another at an earlier time instance. A typical human object re-identification process starts by giving an image or a video of a person taken from one camera, then finding or identifying the person of interest among pre-recorded images, sequences of photos or video frames that track the individual in a network of cameras in real time [86]. While monitoring the human behaviour or activities (a surveillance scenario) a human object disappearing from a particular camera view needs to be matched with similar human objects appearing in one or more other views obtained at different physical locations, over a period of time, and be differentiated from many other human objects in the same views. In surveillance,

the video monitoring task for observation can be very helpful to establish if a specific person who enters and exits a building is the same person identified within another different building, within a public area, workspace, university campus, school, train station or airport etc. It is noted that in answering the above question the views of surveillance footage may be taken from different angles and distances, backgrounds, lighting conditions and various degrees of occlusions [87].

The research conducted in this thesis proposes how both statistical and machine learning approaches can be used to rigorously study the eye tracking information gathered from a group of participants who have been requested to search for people with given description within CCTV footage that will be presented to them. The manual task that the human participants are carrying out in these experiments is, ‘human object re-identification’ in the terminology of automatic video analytics and forensics. In Chapter-7 we present how the proposed machine learning based eye tracking information gathered could be effectively utilised in the performance improvement of ‘human object re-identification’.

### **3.9. Summary**

This chapter initially presented the Human Visual System (HVS), its operational aspects, functionality of eye tracking systems and details about visual stimuli. The chapter also presented information about different data mining tools and methods that have been considered to analyse the data collected via the eye tracking system in this thesis. The difference between statistical approaches to investigating collected data and the proposed machine learning algorithms based approach and the concept behind the thinking aloud method often used to complement the data capture during an eye tracking experiment, have also been presented. The information provided in this chapter is most essential to understand the main body of the thesis. If the readers are interested in more detail regarding this information, they are referred to the original publications and references, where appropriate.

# Chapter 4.

## Experimental Design & Data Gathering

This chapter introduces the research methodology adopted in the design of the experiments used for data collection needed to support the research proposed in this thesis. Initially, the research questions that are intended to be answered by the data collection exercise, is presented in section 4.1. This is followed by several sections on experimental methodology that includes details about the data collection, experimental design (section 4.2), data analysis (section 4.2.3) and a final subsection concluding the results of the data collection exercise (section 4.3).

### 4.1. Research questions

The aim of the experiments conducted for data collection is to support a study that will be conducted within this thesis to *“investigate and understand how humans look at and analyse a video when they are instructed to search for a certain person with a given description”*.

The given description/instruction to a human observer could be as simple as “Find a person who is wearing a red jacket” or relatively more detailed such as “Find a woman who is wearing a yellow skirt and carrying a child”. However, the given descriptions are limited to a specified human object to be observed in the video (e.g. colour/texture of clothing) and/or is closely associated with such an object, such as the child being carried in the above example or a bag being carried, cycle being pushed etc.

In order to achieve the aim of the study to be conducted, following are the research questions to be answered by the data collection and analysis exercises to be conducted within this thesis:



1. Did the participants fixate their gaze on all or parts of the target human body?  
*(Chapter 5, Section 5.2.1, page 64)*
2. Which specific parts of the target human body captured the most attention?  
*(Chapter 5, Section 5.2.2, page 67)*
3. Which target human body part attracted attention first given the description of the target human object disclosed to the participant? *(Chapter 5, Section 5.2.3 page 71)*
4. If the target human object's body is divided into sub-areas, given the description of the target body, what is the order in which the participants generally looked at the different sub-regions? *(Chapter 5, Section 5.2.4, page 73)*
5. How long did it take to spot for the first time each part of the body? *(Chapter 5, Section 5.2.5, page 76)*
6. Has the written description of the target human object given to the participants any influence on the order in which participants looked at the different target human body parts? *(Chapter 5, Section 5.2.6, page 77)*
7. Is there consistency between what participants are asked to look for, they actually look at and also say is important to them when it comes to searching for the target human object with a given description? *(Chapter 5, Section 5.4)*
8. Is there a distinguishable difference between the observation patterns of more and less experienced participants and between male and female participants in analysing image/video footage? *(Chapter 5, Section 5.4)*

All the above research questions were setup up after we met the police investigators and CCTV operators. We found out from discussions that they want us to investigate and analysis human observer behaviour in a bit more details rather than just looking at the

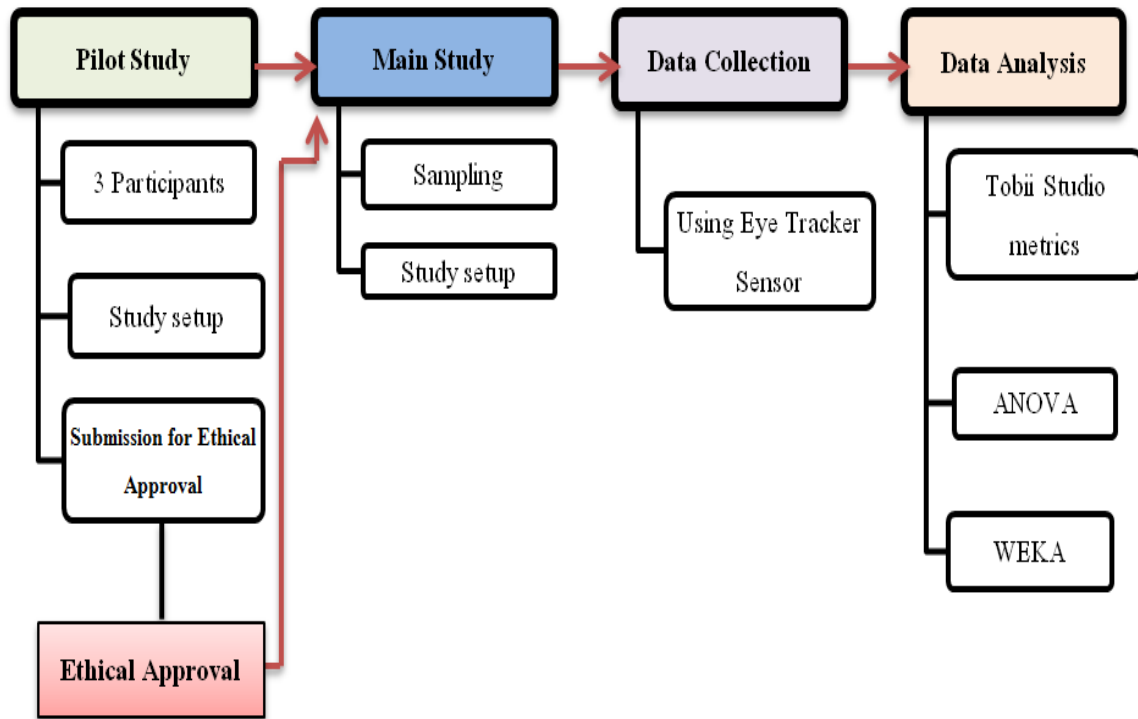
whole body as a single entity or object. This motivated me to divide the human parts into different parts. Then I also conducted a pilot study with the help of three participants (section 4.2), each person had a different level of expertise in visually analysing images and video for content identification. The pilot study was useful it helps me to redesign the experiment before the full study was conducted and also provided help in creating research question from collecting data from eye tracking. The eye tracking software (Tobii studio) was used to check what data were captured by the eye tracker so basically, I checked in each area what parameters picked up and try to understand each parameter (I asked myself a question from this data what questions can I created or answered. Finally, I start creating those questions by going into more detail analysis how human observe and analysis

Whilst some of the above research questions will be answered in the **section 5.2 (research question 1 to 7)** because of the analysis that is carried out in this chapter, more detailed and meaningful analysis of results and answers will be provided in the contributory **section 5.4** of this chapter and **Chapters 6**.

The **sections 4.2** present the experimental procedure adopted to collect data that leads to data analysis in chapters 5 and 6.

## 4.2. Methodology for data collection and analysis

Figure 4.1 illustrates the research methodology followed for data gathering and subsequent analysis.



**Figure 4.1 : Stages followed in a Research Project**

The first stage is to carry out a pilot study with a small number of participants (3 in the research conducted) in order to better understand the challenges, one has to meet in the design of the data collection experiments and while conducting the subjective testing exercises that involve a larger number of participants. The pilot study was useful in fine tuning the experimental designs before the full study was conducted and also provided enough feedback to the ethical clearance process. In the research conducted within this thesis the pilot study was conducted with the use of three participants with relatively low, medium and high level expertise in visually analyzing images and video for content identification.

Ethics are essential when carrying out any type of research and/or experiments where human/animal participants/subjects are involved. Therefore, the proposed research required the approval of the Loughborough University Ethics Committee. The experimental design, conducting the detailed subjective experiments with the full set of participants and data collection was done after successfully obtaining ethical clearance from the Loughborough University Ethics Committee.

### **4.2.1 Experimental design**

This section presents design and set-up information about the subjective experiments conducted to gather sufficient eye-tracking data in order to answer the research questions listed in section 4.1. The main aim was to provide the opportunity for each participant of the experiments to view a number of different videos on a computer screen and in each video to be requested to identify and track a human object with a given appearance related description. Each participant is shown a number of different videos of the same duration and in each video, has to identify and track one specific human object with a given description. While each participant is carrying out the requested task, a professional quality eye tracker is used to monitor eye movements. Each participant is requested to ‘think aloud’ (i.e., express in words what they are interpreting from watching the video) and click on the object once they are sure that they have located it. The specific details from the thinking aloud process is recorded (a voice recorder was used) and the observers eye tracking data was stored for subsequent analysis and processing.

More specific design details of the experiments are provided in the following sub-sections:

#### **a. Study Sampling / Participants**

While conducting any research project where human subjects are involved, special consideration must be made in deciding on the required number of participants and how these participants will be grouped and their contributions effectively used. In this research, the experiments were conducted with 24 participants enrolled for contribution within the Loughborough University premises. The participants were made aware of the aim and objectives of the study, through personal e-mails and via verbal communication

(face-to-face explanation) before the commencement of the experiments. The 24 participants were equally divided gender-wise into 12 females and 12 male participants. The 12 females and 12 male participants were again divided into two further groups, namely; a Novice group (**6 Male + 6 Female**) that included participants who have never had any specific experience in visual image analysis of any kind and an Expert group (**6 Male + 6 Female**) that included participants having substantial exposure to visual image/video analysis/analytics. Most of the expert group included PhD students from within Vision, Imaging and Autonomous Systems research group of the Department of Computer Science and the Novice group consisted mostly the other PhD students with little exposure to image analysis, friends etc. All members of the groups viewed the same pre-recorded videos and were given the same descriptions, per video.

## **b. Experimental Setup**

### **▪ Location**














The lab study was set-up and conducted in the ‘Usability Research Lab’, Haslegrave Building, Department of Computer Science, Loughborough University. The room provided a private and secluded space ideal to carry out the controlled experiments with no external distractions to the participants.



**Figure 4.2: Pilot Study Venue (Usability Lab, Haslegrave Building)**

- **Chosen Stimuli / video**

A total of 13 different pre-recorded video recordings captured by the researcher (videos of random people walking in the Loughborough town centre) were used for this study. The videos thus captured were of HD resolution with wide angle view, mimicking the quality of present day HD CCTV cameras. The length of time each video was presented to each of the 24 participants was one minute. The videos randomly appeared one after another on the screen. Further, before each video clip is presented to a participant, written descriptions regarding target appearance to look for, were shown on the screen. The objects that were requested to be identified in the chosen videos used in the study are outlined in Figure 4.4. Figure 4.3 gives an idea of the field-of-view of each captured video to give the reader an idea about the objects visible within the wider view angle of the captured video. In all test videos the objects of interest (shown in figure 4.4) either appear first far from the camera and then approaches towards the camera, or appears close to the camera first and walks away or sideways. The duration of appearance of each person varied depending on their movement.

<b>Video 1</b>	<b>Video 2</b>	<b>Video 3</b>
		
<b>Video 4</b>	<b>Video 5</b>	<b>Video 6</b>
		
<b>Video 7</b>	<b>Video 8</b>	<b>Video 9</b>
		
<b>Video 10</b>	<b>Video 11</b>	<b>Video 12</b>
		
<b>Video 13</b>		
		

**Figure 4.3: Field-of-view of each captured video**

Video 1	Video 2	Video 3	Video 4	Video 5
				
Video 6	Video 7	Video 8	Video 9	Video 10
				
Video 11	Video 12	Video 13		
				

**Figure 4.4: Objects that are requested to be identified and tracked by the participants**



- **Written descriptions of objects to be identified and tracked**

There has been much research concluding that task instructions influence where an observer's eye is fixated while viewing static scenes or images [88]. In 2011 a research conducted on a dynamic scene showed similar findings [89]. So, in our experiment all the participants are provided with identical written instructions per human object to be identified and tracked in each video, displayed on the screen before the video is displayed. It is noted that in each instruction there are some important keywords participants need to follow in order to find the correct target. For example, in video 1 the following instruction was given to the participants

“Find a **woman** wearing **white trousers** and carrying a **pink shoulder handbag**”.

In this instruction, there are three important keywords “*woman*”, “*white trousers*” and “*pink shoulder handbag*”, the key-information that the participants will extract from the given descriptions that will be used in the search for the object. Since none of the instructions given it was not mentioned whether the person is static (not moving or stationary) or moving (walking), it was likely that all participants would attempt to analyze and look for both types of objects/people. The written descriptions given for each of the 13 videos to the participants involved in the study are tabulated in Table 4.1. It is noted that these questions for each of the videos were designed based on the visible articulation of details of the human objects to be tracked agreed by the research team conducting this research, with a view of mimicking the descriptions used in the search of known criminals in CCTV footage as advised by the police and also with the view of articulating a human object into sufficient detail so that ambiguity related challenges to be met by the computer vision algorithms to follow, will be minimized.

Video	Written Description
1	Find a woman wearing white trouser and carrying a pink shoulder handbag
2	Find a man wearing white shirt carrying black laptop bag
3	Find a man wearing a red half sleeve T-shirt and pants
4	Find a person wearing blue shirt/top and sleeveless cardigan
5	Find a person wearing a pink short and white sleeveless top
6	Find a person wearing a red shirt/top with stripe
7	Find a person wearing red trousers
8	Find a man wearing dark blue sport trousers
9	Find a woman wearing red sleeveless shirt with a bike
10	Find a person wearing a blue top and white pants
11	Find a man wearing a yellow shirt
12	Find a woman wearing a white shirt
13	Find a woman wearing a blue dress

**Table 4.1: The written descriptions of the objects to be identified and tracked in the videos**

Obviously, the people differ from each other, they have different level of learning and different ways to describe people or a situation. Some people are not very good at describing they struggle as explaining things due to a lack of memory required for remembering, recalling or describing situation occurred. In other word, they simply have difficulties in recalling exactly what happened and what they saw to pin down the details. However, some people have good memory for facts and knowledge of the past but they can struggle to remember specific information.

In this research study conducted with help of two groups Expert and Novices Group included participants having substantial exposure to visual image/video analysis/analytics and Novice group consisted mostly the participants with little exposure to image analysis All members of the groups viewed the same pre-recorded videos and were given the same descriptions, per video. The instruction has been consisting of simple and complex instruction regrading person to look for. The purpose of giving different level of instruction because in the respect to CCTV expert and less expert observer differ in perceiving, understanding, monitoring, and processing visual information so due that we split the instruction to simple and complex instruction the possibility of different groups expressing different observing behavior.

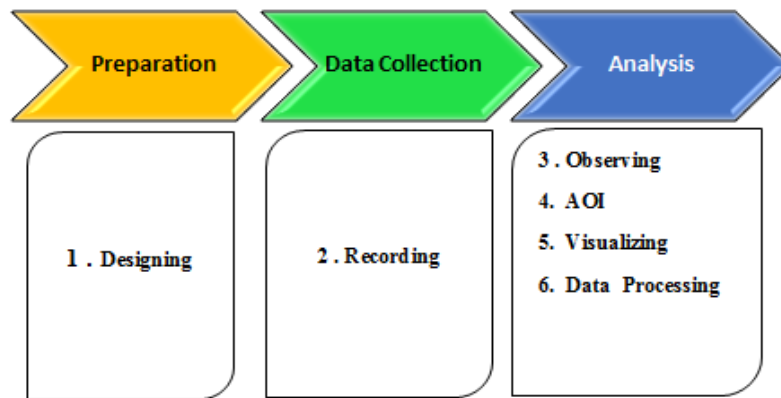
<b>Video</b>	<b>Written Description</b>	<b>No. of keyword</b>	<b>Important Keywords</b>	<b>Level of description</b>
1	Find a woman wearing a white trouser and carrying a pink shoulder handbag	3	Woman white trouser pink shoulder handbag	<b>Full Description</b> (Complex)
2	Find a man wearing a white shirt carrying a black laptop bag	3	man white shirt black laptop bag	<b>Full Description</b> (Complex)
3	Find a man wearing a red half sleeve T-shirt and pants	3	man red half sleeve T-shirt pants	<b>Full Description</b> (Complex)
4	Find a person wearing a blue shirt/top and sleeveless cardigan	2	blue shirt sleeveless cardigan	<b>Short Description</b> (simple)
5	Find a person wearing pink shorts and a white sleeveless top	2	pink short white sleeveless top	<b>Short Description</b> (simple)
6	Find a person wearing a red shirt/top with stripes.	2	red shirt/top stripe	<b>Short Description</b> (simple)
7	Find a person wearing red trousers	1	red trousers	<b>Short Description</b> (simple)
8	Find a man wearing dark blue sport trousers	3	man dark blue sport sport trousers	<b>Full Description</b> (Complex)
9	Find a woman wearing a red sleeveless shirt with a bike	3	woman red sleeveless shirt bike	<b>Full Description</b> (Complex)

<b>Video</b>	<b>Written Description</b>	<b>No. of keyword</b>	<b>Important Keywords</b>	<b>Level of description</b>
10	Find a person wearing a blue top and white pants	2	blue top white pants	<b>Short Description</b> (simple)
11	Find a man wearing yellow shirt	2	man yellow shirt	<b>Short Description</b> (simple)
12	Find a woman wearing a white shirt	2	woman white shirt	<b>Short Description</b> (simple)
13	Find a woman wearing a blue dress	2	woman blue dress	<b>Short Description</b> (simple)

**Table 4.2: The written descriptions of the objects to be identified and tracked based on level of description from simple to complex**

- **Eye tracking sensor/technology, hardware & software environments**

The required eye tracking of participants was carried out using a Tobii Eye Tracker[54]. This specific eye tracker was chosen for data capture due to a number of reasons that included, availability, accuracy and ease of integration into lab studies. The presentation of the stimuli was designed, presented and controlled by using the associated Tobii Studio Software[90]. The Tobii studio project consisted of 13 written descriptions and pre-recorded videos. These descriptions and videos were displayed one after another on a screen. Figure 4.5 illustrates the testing environment set-up procedure that was adopted with the Tobii eye tracker.



**Figure 4.5: Objects that are requested to be identified and tracked by the participants**

- **Dynamic Areas of Interest (DAOI) design**

In order to investigate the human observer’s attention given to different parts of the human body, namely the head area, torso area, legs and any object being carried/pushed/pulled etc., (i.e. an area connected to the human body area that contains an object external to the human body), the DAOI tool provided by the Tobii Studio software was used for the creation of dynamic areas of interest within and around a target body. The tool provides the user the ability to define a rectangular area of interest in an image being viewed that the software tool will then track to the subsequent frames of the video.

Due to the human objects to be detected and monitored being of a dynamic/moving nature and very different to each other between the videos the design of the DAOIs in the thirteen different videos were different and hence the definition of the DAOIs was a tedious and time-consuming task. However, it is an essential task that ensured the uniqueness of this study as compared to the closest research presented in existing literature, related to video surveillance. Table 4.2 illustrates the definitions of the DAOIs for the objects intended to be detected and tracked in each of the videos to be monitored.







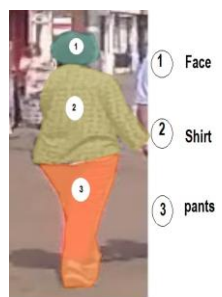
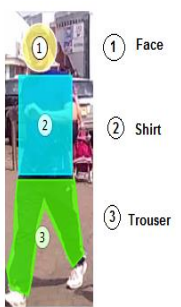





Video 1	Video 2	Video 3	Video 4	Video 5
 <p>① Face ② Shirt ③ Handbag ④ Trousers</p>	 <p>① Face ② Shirt ③ Laptop bag ④ Pants</p>	 <p>① Face ② Shirt ③ Plastic bag ④ Pants</p>	 <p>① Face ② Shirt ③ pants</p>	 <p>① Face ② Shirt ③ pants</p>
Video 6	Video 7	Video 8	Video 9	Video 10
 <p>① Face ② Shirt ③ pants</p>	 <p>① Face ② Shirt ③ pants</p>	 <p>① Face ② Shirt ③ Trouser</p>	 <p>① Face ② Shirt ③ Bike</p>	 <p>① Face ② Shirt ③ Handbag ④ Trouser</p>
Video 11	Video 12	Video 13		
 <p>① Face ② Shirt ③ pants</p>	 <p>① Face ② Shirt ③ pants</p>	 <p>① Face ② Shirt ③ pants</p>		

Table 4.3: DAOI for each video

- **Eye tracker calibration Procedure**

Calibration is the process that trains the eye tracker sensor on the characteristics of a specific participant's eyes. A simple calibration procedure for each participant was conducted before starting the experiment [91] for accurate estimation of a participant's gaze point [92]. The Tobii eye tracker sensor used in the studies conducted within this thesis offers a quick and automatic calibration procedure. During eye calibration procedure, the participant is requested to look at specific points, coloured in 'red', that are displayed on the screen (calibration dots or circles); during which time several images of the eyes are gathered and analysed. In the research conducted a specific default calibration that consists of the presentation of five red calibration dots on a white background was used. The five dots are presented consecutively, one after another and the participant was required to look (or gaze) at the centre of the dots. If the calibration was unsuccessful (not enough calibration data collected), a re-calibration procedure is carried out until sufficiently good calibration data is collected. The occurrence of a weak calibration is common for eye-tracking devices with regards to certain participants. This can be due to many reasons such as the shape and structure of the participant's facial features such as eye, the colour of their eyes, too much eye blinking during viewing, or the need for glasses (prescription glasses) to be worn. Therefore, a re-calibration procedure is performed in order to obtain accurate calibration data. If the calibration remains weak after several attempts of calibration the participant was excluded from the experiment. There were a total of 24 participants in this study, four subjects had unmeasured eye tracking metrics due to weak calibration (missing data) and were thus excluded from data analysis.



- **The test set-up**

The total duration of each participant's subjective experiment session associated with the study carried out was approximately 30 minutes. The following procedure was followed during each session.

1. The session commenced with the participant being explained the purpose of the research study and given an overview of what was expected from the participant.
2. The monitor and chair were appropriately adjusted to ensure that the participant will remain feeling comfortable during the study.
3. The participant is placed and sets up in front of the eye tracker sensor.
4. The eye tracking camera is calibrated for the participant concerned (see figure 4.6). The participant was asked to look at the calibration points on the screen. If the calibration is noticed to be weak, re-calibration is carried out as required.
5. Once the system setup procedure is completed, the subjective study for the participant commences while the eye tracker is recording the participant's visual attention and making the Tobii studio software generate the related data.



**Figure 4.6: The calibration procedure for a participant**

## **4.2.2 Data Collection Phase**

Within any research project, the data collection phase is very important, as it helps to gather, study and record information, and make/improve decisions about important problems. The most important part of data collection is being sure to record it carefully, so that the output data is rich, accurate and complete. This step helps to increase the chances of success and the achievement of the primary aim of the research as a whole.

### **4.2.2.1 The procedure**

During the study, each participant was set up in front of the eye tracker sensor as explained above. This sensor will measure each participants' eye positions and eye movement information. Once the system was set up, the participant was requested to relax and perform the task. The task was to read a short written descriptions of human object to appear on a video and view 13 different videos (videos of a public scene where people are walking in the Loughborough town centre) individually on the screen. All the 13 videos are displayed in a random order, one after the other on the screen. This random order of displaying videos to a participant is important to try and minimize the impact that participants will have in gaining experience in carrying out the tasks assigned to them and hence all videos tested getting a fair response from the group of participants.

The following are the steps followed during the study:

- 1.** The experiment is started with a fixation cross sign displayed at the centre of the screen for re-calibration of the device. This sign appears at the beginning of the experiment and after each piece of video footage. When the sign appears, the participant is requested to look at it.
- 2.** When the written description to find a target (person, man or woman) appears on the screen the participant is required to read it carefully and then press the space bar on the keyboard to allow a video to start.

3. When the video starts, the participant is given 1 minute to look at the video and find a person who matches the description.
4. As soon as the participant finds the person, the participant is requested to use the mouse to point at the target and click.
5. This cycle is repeated 13 times for the different videos (The videos randomly appeared one after another on the screen); during this time, the sensor records eye tracking data with respect to the participant's natural reactions to the content of the video given the specific instructions.
6. From the beginning of the experiment until the end, participants are requested to verbalise, i.e. think-aloud, their thoughts during the process of carrying out the tasks.
7. From the start of the experiment until the end the participant's reactions are observed by the Tobii eye tracking system and comments are recorded via a voice recorder, in order to support potential subsequent detailed analysis of eye tracking data.

As a participant reads the descriptions and watches the videos, the eye-tracking device focuses on the pupil of the participant's eye and determines the direction and concentration of their gaze. The software generates data about these actions in the form of heat maps and saccade pathways (see Chapter-3). The following data is collected: Where participants are looking, how long they are looking for, how their focus moves from item to item on the screen, what parts of the interface they miss and how the size and placement of items on the screen affects attention.

- *Eye-movement Metrics*

During the experiment, the eye tracker software calculates many different statistics regarding the participant's visual behaviour and attention to the different DAOIs. These metrics can easily be extracted by using Tobii Pro Studio. Tobii Studio provides various eye-movement metrics. The most important and most frequently reported eye movement metrics are fixation duration, fixation count and time to first fixation [93]. These fixation related metrics have helped many researchers in the past to learn what grabbed the attention of the participant first, how interested the participant was in a certain part or section, or if a task was difficult to achieve. Table 4.3 illustrates relevant metrics depending on the research questions that have chosen to analyse our research objectives and questions.

<b>NO.</b>	<b>Tobii Metrics</b>	<b>Abbr.</b>	<b>Description</b>
1	Times To First Fixation(sec)	<b>TTF</b>	The time from the start of stimulus display until the test participant fixates on the AOI or DAOI.
2	Fixation Duration	<b>FD</b>	Is the total amount of the time that gaze was fixated within the target area.
3	Visit Duration	<b>VD</b>	Duration of each individual fixation within an AOI or a DAOI.
4	Visit Count	<b>VC</b>	Number of visits within an AOI or a DAOI.
5	Percentage Fixated %	<b>PF</b>	Percentage of participants that fixated at least once within an AOI or a DAOI.
6	Fixation Count	<b>FC</b>	The number of fixations in particular AOI or DAOI.
7	Time To First Mouse Click	<b>TTFMC</b>	The time taken before the test participant clicks on an AOI or a DAOI for the first time.

**Table 4.4: Eye-gaze metrics used in the conducted experiments**

Figure 4.6 illustrates an overview of the experimental procedure adopted in collating eye tracking data as described above. The completion of this process completes data collection. The data will now be ready for analysis using Tobii studio software associated with the eye tracker or via other statistical and machine learning procedures.

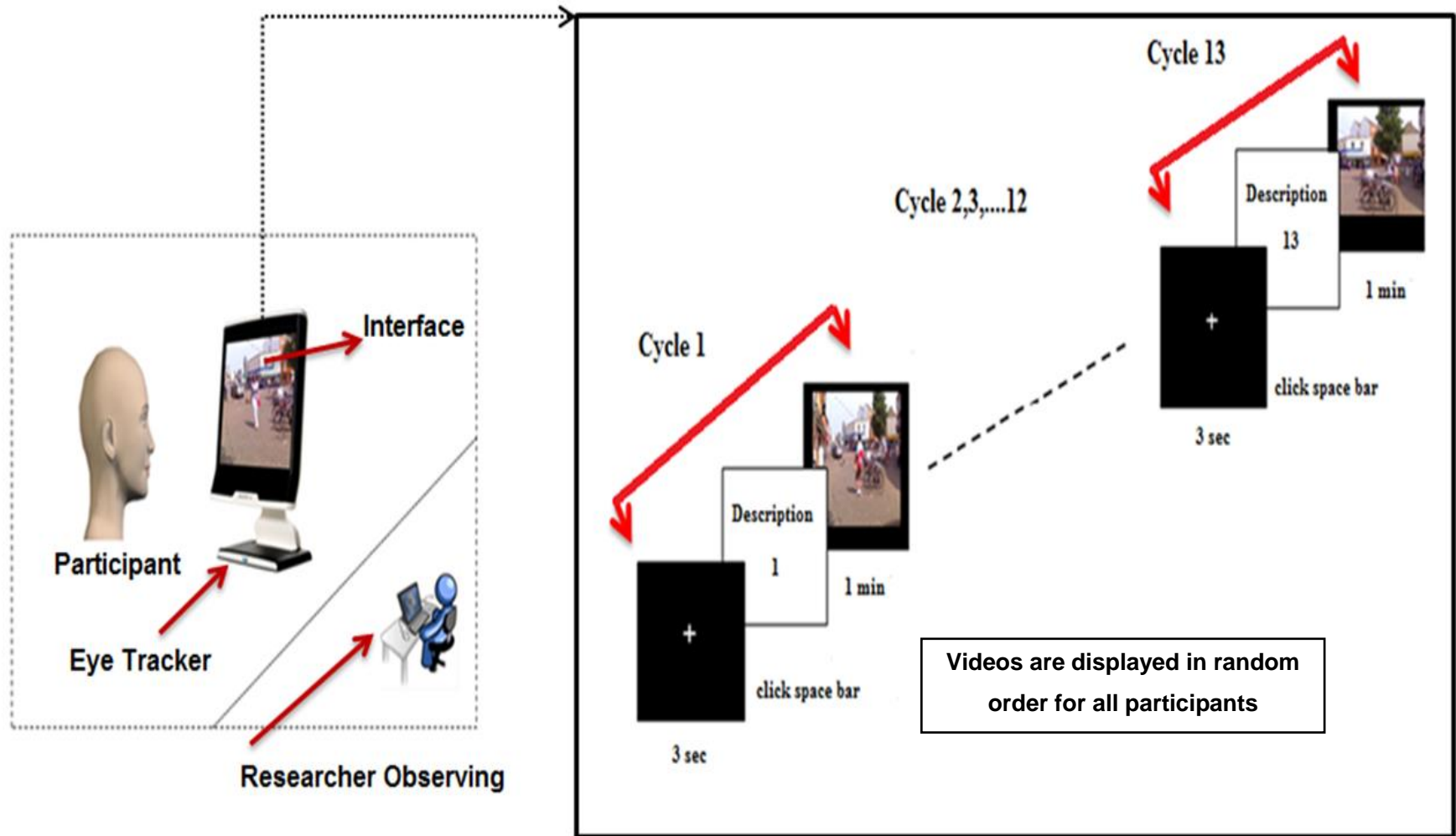


Figure 4.7: Overview of study setup with regards to data collection and information display

### 4.2.3 Data Analysis

Once the required data was collected from the eye tracker sensor during the data collection phase explained above, the next step was to analyse the data by converting each data stream into meaningful information. During this stage, the collected data is analysed by a number of data analysis methods that include conceptual, statistical and machine learning based approaches to convert the collected data from each video into meaningful information with regards to the human observer's behaviour in identifying and tracking the object concerned. It is noted that the novelty of the work presented in this thesis is based on the detailed DAOI based detailed analysis conducted based on the conceptual and statistical approaches and this being the first attempt in using machine learning to analyse human observer eye tracking data using machine learning approaches.


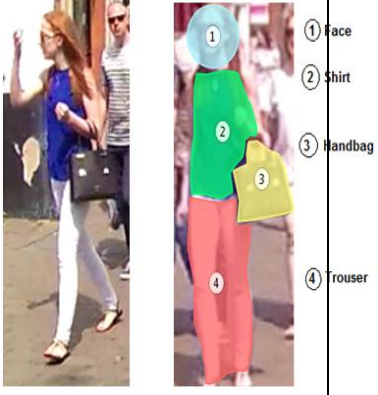
As mentioned before the study implemented consisted of participants being observed attempting to identify and track human objects of a given description in 13 different videos. In this thesis, due to limitation of space and purpose of clarity only two representative videos are used for the detailed presentation of results and analysis even though the remaining videos were also tested in the research conducted and confirmed to behave in a similar manner. Where ever any ambiguity of analysis of results are indicated with regards to any of the remaining videos, this is stated and is explained in detail in the relevant sections. The two videos considered (see table 4.3) are quite different to each other in terms of the descriptions provided to the human observers and the objects being described for identification and tracking.

In **Chapter 5**, the above data obtained from the eye tracker is analysed using Tobii Studio and a statistical data analysis approach called ANOVA and linear regression using WEKA machine learning suite. In addition, based on the data captured a conceptual analysis of the human behaviour is also carried out and provided as a means for the explanation of the results obtained and the conclusions that can be drawn upon based on the statistical data analysis approaches.

In **Chapters 6** however, the same data is analysed using machine learning approaches, confirming the results of the statistical analysis and generating novel knowledge from the information gathered, such as demonstrating the ability to make predictions of future observer behaviour etc., a capability that the statistical data analysis approaches do not provide.

We use the results of two of the thirteen videos tested for the conceptual and statistical analysis, i.e. Video-1 and Video-10 because each video differs from each other as described in table 4.4. However, the rest results described in **appendix 1**.



	Video 1	Video 10
<b>Target</b>		
<b>Instruction</b>	Find a woman wearing white trouser and carrying a pink shoulder handbag	Find a person wearing a blue top and white pants.
<b>Instruction type</b>	Long description	Short description.
<b>Important keywords</b>	Woman, white trousers, pink and shoulder handbag	Blue top and white pants.
<b>Target Gender</b>	Given	Not given
<b>Time at which the object first appears in the video</b>	31.5 sec	Start, i.e. 0 sec
<b>Total Duration of appearance</b>	Appearing for a long time	Appearing for a short time
<b>Pattern of movement</b>	Person coming from far to the nearest camera then disappearing	The person faces the camera and then disappears
<b>Hand carried item</b>	Carrying a bright handbag (pink)	Carrying a dark handbag (black) .

**Table 4.5: The two test videos considered for detailed data analysis**

### **4.3. Summary**

This chapter provided the methodology adopted for data collection and capture. The experiment design details, information on the conduct of the subjective tests, data collection and an overview to the data analysis approaches to be adopted in chapters **5,6 and 7** were presented.

It was noted that a dataset of similar nature in which the analysis of the eye tracking information based on separate parts of the human body has not been conducted prior to this research and hence no public database is available to support the original research presented in this thesis. Hence it was essential to carry out the tasks relevant to capturing this novel dataset.

# Chapter 5.

## Statistical Analysis of Visual Attentional Patterns in Video Footage

In this chapter, we provide a detailed statistical analysis of the human observer visual attention patterns when carrying out task specific observations in analyzing captured video footage. Where ever appropriate additional analysis and justifications are provided via a conceptual analysis of the captured data.

In order to achieve the aim of the study to be conducted, the research questions to be answered by the data collection and analysis exercises to be conducted within this thesis were presented in section 4.1. In this chapter, these research questions will be answered based on the statistical data analysis to be conducted. For ease of reference with regards to the analysis of the results the research questions are re-listed below.

1. Did the participants fixate their gaze on all or parts of the target human body?
2. Which specific parts of the target human body captured the most attention?
3. Which target human body part attracted attention first given the description of the target human object disclosed to the participant?
4. If the target human object's body is divided into sub-areas, given the description of the target body, what is the order in which the participants generally looked at the different sub-regions?
5. How long did it take to spot for the first time each part of the body?

6. Has the written description of the target human object given to the participants any influence on the order in which participants looked at the different target human body parts?
7. Is there consistency between what participants are asked to look for, they actually look at and also say is important to them when it comes to searching for the target human object with a given description?
8. Is there a distinguishable difference between the observation patterns of more and less experienced participants in analysing image/video footage? Is there a significant difference between the observation patterns between male and female participants?

## **5.1. Data preparation**

Prior to carrying out detailed analysis of the data it is essential that the collected data is checked for missing data, inaccurate data, any outliers and discrepancies. It is noted that the above could significantly impact on the overall accuracy of data analysis and the subsequent conclusions made. Thus, it is essential for such data to be either corrected or removed.

In the original data gathering exercise carried out, a total of 24 participants were involved. Unfortunately, the data gathered by observing four subjects had to be excluded from the gathered dataset as our close investigations revealed that the device calibration has been weak (eye tracker was not able correctly recorded participant eye movement) during data gathering and hence the captured data was unreliable. The occurrence of a weak calibration is common in using eye-tracking devices, which can be due to many reasons such as, the shape and structure of the participant's facial features such as eyes, the colour of their eyes, too much eye blinking during viewing, or the need of spectacles (prescription glasses) to be worn.

As the result of data preparation carried out above, only data captured on behalf of 20 observers are used for the analysis presented below. The data analysis is conducted both by using statistical data analysis software SPSS and Tobii Studio software. It is noted here that SPSS is used as the Tobii Studio Software does not provide sufficient means for detailed statistical data analysis. One-way ANOVA analysis was used for further investigation if there were differences in observing the three or four areas of interest based on different eye-tracking measurement metrics obtained using IBM SPSS Statistics Version 23.

## **5.2. Experimental Results and Discussions**

Table 4.4 listed seven eye-gaze metrics that the Tobii eye tracker captures during subjective experiments for each defined Dynamic Area of Interest (DAOI) per given video illustrated in Table 4.3. In Chapitre-4 it was also mentioned that depending on the objects to be searched for, in each of the test videos, the human body area of the object to be identified and tracked by the human observers, are divided into 3 or 4 DAOIs by defining the rectangular DAOIs encompassing head(face), top body, bottom body and/or carried object regions. In the experiments conducted the human observers were asked to click anywhere on the screen when they are satisfied that they have found the person with the given description. The time to this click is one of the seven metrics recorded. Due to the nature of instruction given to the human observers as described above, per human object being tracked for all 3/4 DAOIs, only one value is recorded Therefore, for a video with an object consisting of 3 DAOIs a total of  $(6 \times 3 + 1) = 19$  metrics are recorded and for a video with an object consisting of 4 DAOIs, 25 metrics are recorded.

The relevant metrics measured and recorded by the Tobii eye tracker to answer each of the eight research questions are tabulated in Table 5.1. Following sub-sections explain how each of the metrics can be used conceptually to answer, in particular the research questions 1-6 (inclusive) raised.

Research Objectives							
1	2	3	4	5	6	7	8
Percentage Fixated %	Percentage Fixated %	Fixation Duration	Time to First Fixation	Time to First Fixation	Time to First Fixation	Time to First Fixation	All metrics
	Visit Duration Visit Count						All metrics

**Table 5.1: Mapping between research questions (section 4.1) and data metrics**

this chapter, we use the results of two of the thirteen videos tested for the conceptual and statistical analysis, i.e. Video-1 and Video-10 (see Figure 5.1). In Video-1 the instruction was to ‘Find a woman wearing a white trouser and a pink shoulder handbag’. In Video-10 the instruction was to ‘Find a person wearing a blue top and white pants’.

The metrics produced by the Tobii eye tracker is used for conceptual analysis and formulating the answers for question 1-6 supported by ANOVA analysis using SPSS where ever needed. It is noted that the answers to questions 7 and 8 are investigated in Chapters 6 and 7 respectively as the conceptual / statistical analysis alone cannot provide sufficiently justified answers.



(a) Video-1



(b) Video-10

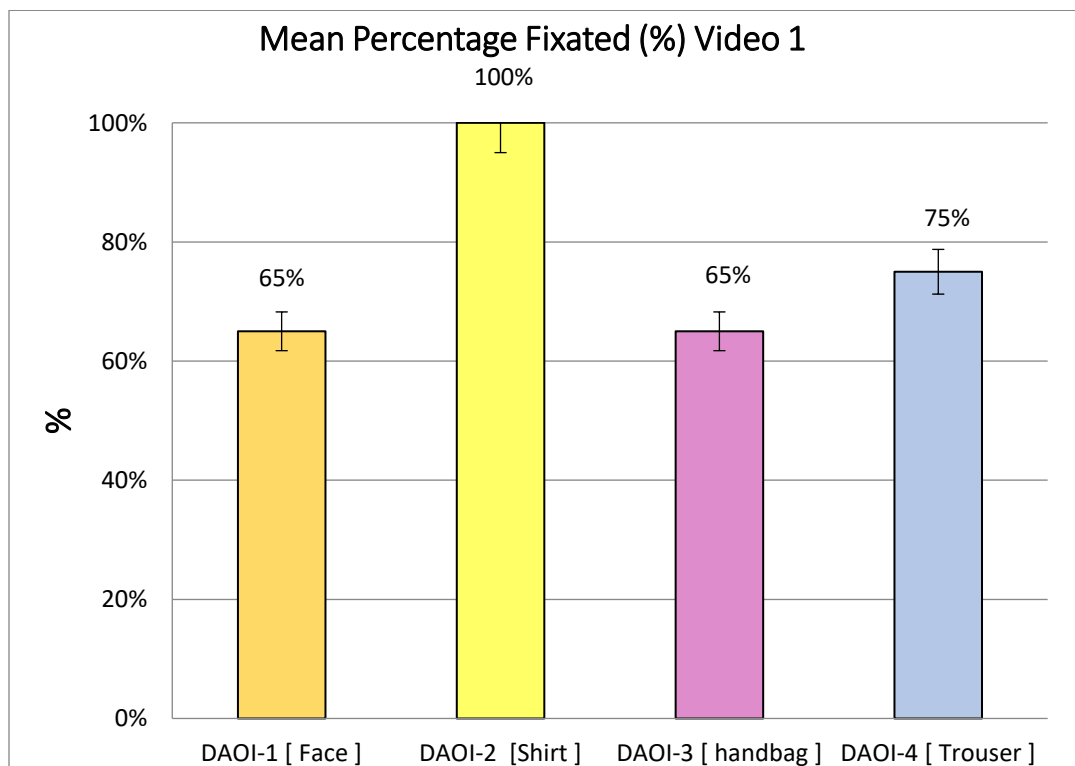
**Figure 5.1: Field-of-view of each captured video**

The following upcoming sections analyses the results obtained to formulate the answers to each of the research questions:

## 5.2.1 Research Question 1

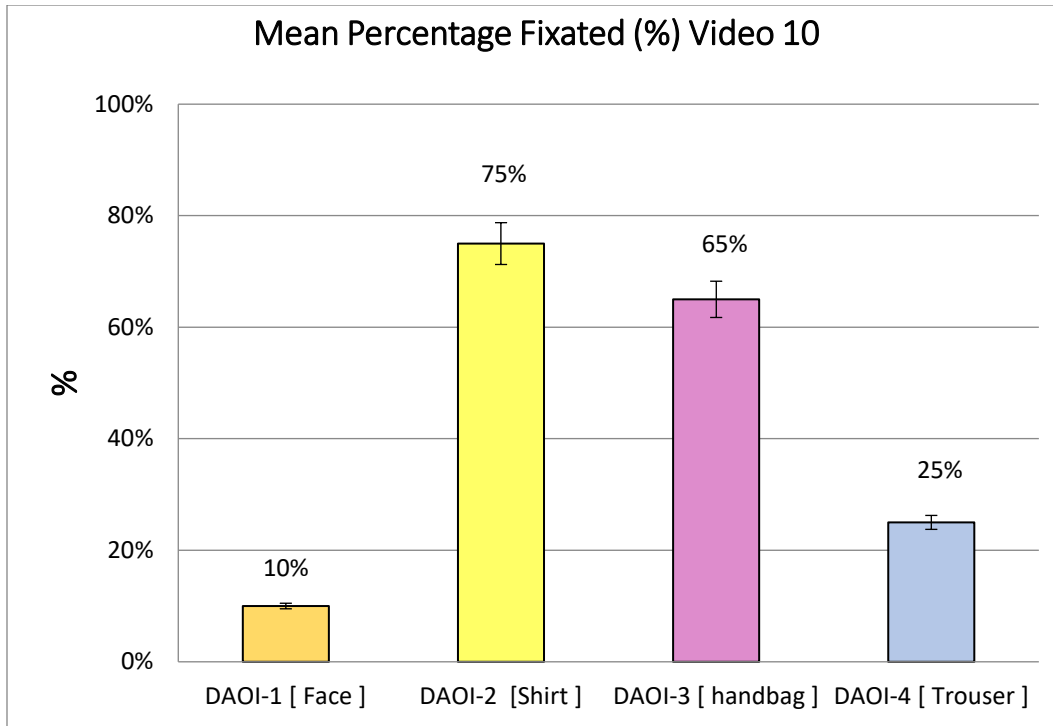
**Did the participants fixate their gaze on all or parts of the target human body?**

The metric, ‘Percentage Fixated (%)’ determines the percentage of participants that have fixated at least once within a given DAOI. Therefore, this metric provides a means to analyse the different level of attention each DAOI has received from the set of participants and hence answer the question. Figure 5.2 illustrates bar charts indicating the ‘Mean Percentage Fixated’, i.e. the mean of ‘Percentage Fixated’ over the set of 20 participants, for each DAOI for both videos 1 and video 10.



(a)





(b)

**Figure 5.2: Percentage of respondents that have fixated at least once within the defined DAOIs in (a) video 1 [ *Find a women wearing white trousers and carrying a pink shoulder handbag*] and (b) video 10 [ *Find a person wearing a blue top and white pants*]**

It is clear that in both videos all participants fixated their views on each of the four DAOIs. It is also clear that there is a significant relative difference between the % of participants who have fixated their views on the different parts of the body. [DAOIs 1-4 refers respectively to the ‘head’, ‘shirt/top’, ‘bag being carried or pulled’ and ‘trouser/legs’]. In comparing and detailing the results obtained for Video-1 with that of Video-10, it is indicated based on the instructions given to the participants, different levels of attention have been received by the different DAOIs.

In both videos, the DAOI-1[Face] has received minimal attention. Human participants will consider looking at the head[Face] area of a human object to determine whether the

person they are looking at is a woman or a man. In Video-1 the instruction mentions to look for a woman but in Video-10 the sex of the person is not specified. This could explain why the Mean Percentage Fixated on DAOI-1 [Face] for Video-10 was 10% as against 65% for Video-1.

In both videos the DAOI-2 [Shirt] has received significant attention. However, in the case of video-1, DAOI-2 [Shirt] has received 100% attention, i.e. attention from all participants as against 75% for video10. This could be due to the fact that in making a decision whether the person is male or female this part of the body can play a significant role, especially when the woman in video-1 has short hair that makes it difficult to make that judgement based only on DAOI-1 [Face]. Thus, the DAOI-2 [Shirt] of video-1 received attention from all participants even though the colour of the top was not a part that needed identification. Hence a unique feature that was important in making a decision has made this region more significant in video-1.

In both videos, the DAOI-3 [handbag] received similar attention. It is recalled that in video-1 the details of the handbag was specified and in video-10 the handbag was not mentioned. i.e. even though this difference of description existed the area received attention from the same, relatively high percentage of participants. This can only be justified by the fact that at a distance like what the objects were visible from the camera, the video-10's handbag attracted attention due to its nature and angle of appearance.

The DAOI-4[Trouser] of video-1 has received significant more attention than the DAOI-4[Trouser] of video-10 despite the fact that in both descriptions, the colour of the trouser was specified. It is seen that in video-1 the colour of the trouser is less clear to be 'white' than in video-10. Hence this lack of clarity would have attracted more individuals to 'fixate' their view until a decision could be made. In video-10 it's very clear that the colour is 'white' and hence about 75% of the participants did not even require to 'fixate' their view in order to make a judgement that the trouser is white.

### 5.2.1.1 Summary Research Question 1

Almost all participants fixate their gaze on task relevant aspects of the footage but the attention given to the different parts of the body differs depending on the instructions given to them. It can be concluded that given instructions have been in sufficient detail to influence fixation behaviour of all participants. When a task specific decision was easy and quick to make, the relevant areas did not receive 'fixations of gaze' from most participants.

## 5.2.2 Research Question 2

**Which specific parts of the target human body captured the most attention?**

In order to answer this research, question the four metrics, Percentage Fixated, Fixation Duration, Visit Count and Visit Duration are used as a quantification of ‘attention’ could be derived from one or more of these metrics.

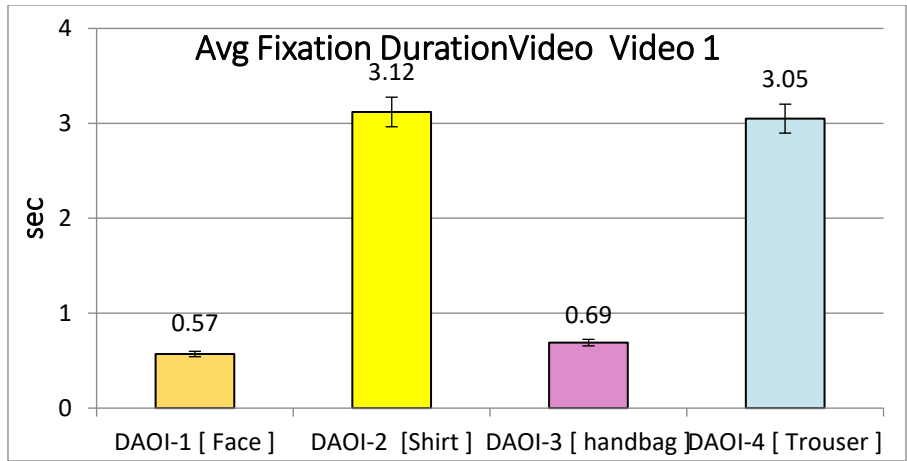
In section 5.2.1 the ‘Percentage Fixated’ metric was used for the analysis of the results which revealed that the level of attention from the human observed received by each DAOI will depend on the description given about the target body. Specifically, it showed that the DAOI-2 [Shirt] was receiving more attention and this was in the case of video-1 to support the identification of the human object to be ‘female’ and in the case of video-10 due to the given description of the colour of the ‘top/shirt’, worn by the person.

In order to analysis the statistical significance of the above results a one-way between subjects ANOVA analysis (see Chapter-3) was conducted using the output of SPSS[64]. There was a statistically significant difference on the percentage fixated between the four areas at  $p\text{-value} < 0.05$  as determined by one-way between subjects ANOVA ( $F(3,75) = 3.35, p = 0.024$ ) for video-1 and ( $F(3,76) = 10.7, p = 0.00$ ) for video-10, respectively.

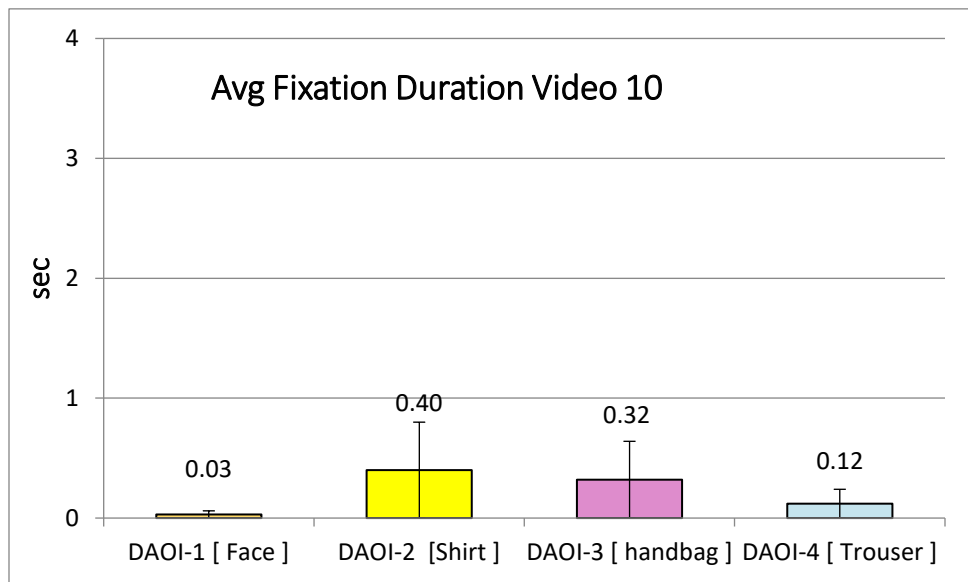
The metric ‘Fixation Duration’ determines the duration (in seconds) of each individual’s fixation of gaze within a DAOI. This metric allows the identification of DAOIs that receive participant attention in the form of the participants spending more time fixating gaze, i.e. receiving more participant attention. A DAOI with a longer duration fixation can indicate possibilities of either higher levels of attention or interest and sometimes even creating participant confusion [94][95][96].

Our result showed, that the participants had a significantly longer fixation duration on the DAOI-2 that relates to the shirt/top area of the human body and shorter fixation on the DAOI-1 [ Face] that relates to the head. In both videos, however the fixation times recorded were different for the two videos for each of these DAOIs for reasons explained in section 5.2.1. It is shown that the average fixation duration on DAOI-1 [ Face] of video 1, 0.57s, is significantly higher than the significantly shorter average fixation duration on DAOI-1 [ Face] of video 10, 0.03s. This can be attributed to the fact that in video-1 the written description requires the user to locate a woman with a given description on clothing and items carried and the video-10's written description do not specify the sex of the person being looked for. For the same reason, we believe that DAOI-2[Shirt] of video-1 receives 3.12s of fixation duration as against 0.4s of fixation duration for the DAOI-2[Shirt] of video-10, despite the fact that for video-10 the color of the 'top' has been specified but not for video-1. Further, it is observed that the fixation duration for DAOI-4[Trouser] of video-1 is significantly higher as compared to that DAOI-4[Trouser] of video-10 despite the fact that in both description the colour of the trouser was mentioned. This can directly be attributed towards the fact that in video-1 the colour of the trouser is not easily be differentiable as 'white' and this may have required the users to fixate their gaze for a longer period, before a decision could be made that it is indeed a 'white' coloured trouser.

It is noted that this video does not contain a second person wearing a white trouser. In summary, there was a statistically significant difference of fixation duration for the four areas at  $p$ -value  $< 0.05$  as determined by one-way, between-subjects ANOVA analysis ( $F(3, 57) = 9.57, p = 0.0$ ) for video-1 and [ $F(3, 76) = 7.863, P = 0$ ] video-10 respectively.



(a)



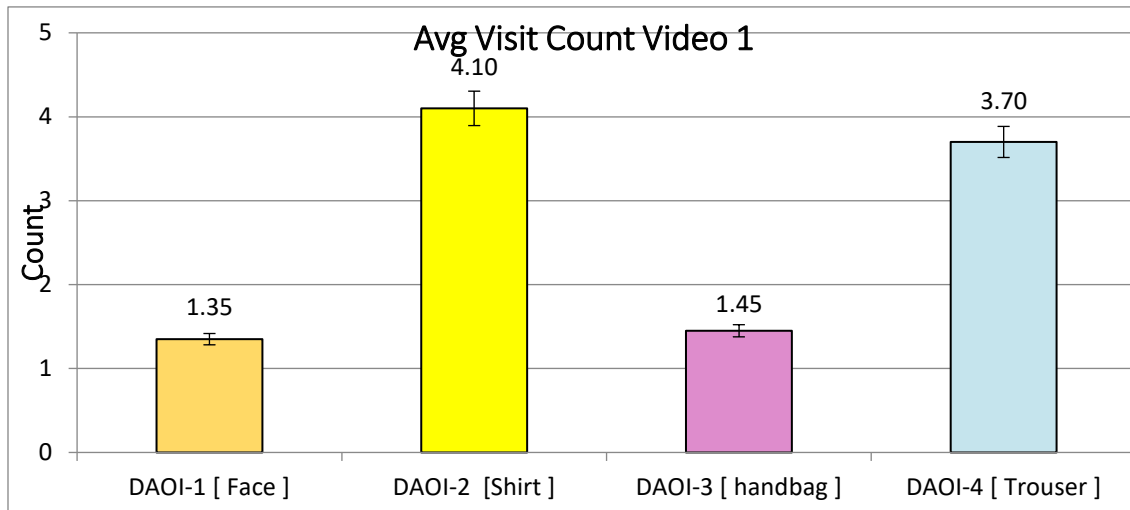
(b)

**Figure 5.3: Average fixation duration for each DAOI in (a) video 1 and (b) video 10**

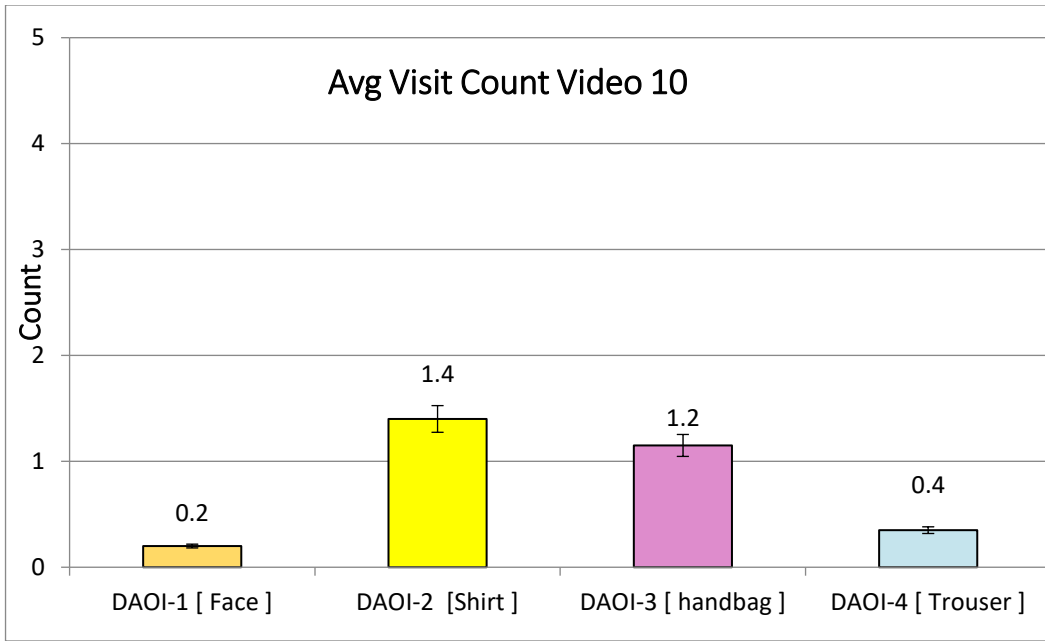
Visit duration and visit count are two further metrics that can quantify human participant attention to different DAOIs. For DAOI-2[Shirt], in video-1 the average visit count is high as 4.1 visits per participant and the visit duration is as high as 3.22s. For the same DAOI, in video-10 the visit count of 1.4 visits is relatively low as compared to that of video-1, and the visit duration is also low 0.52s. This for video-10 relatively fewer visits

of shorter duration is received by DAOI-2[Shirt] (although as compared to the other DAOIs of the same video, the figures are relatively high). This indicates that in video-1 DAOI-2[Shirt] generates more interest and hence attention, and in video-10 the DAOI-2[Shirt] is visited more often than other DAOIs of the same video, but very quickly, perhaps to check whether the ‘top’ is of ‘blue’ colour. Making the decision appears to be easier due to the very low visit duration.

For DAOI-3[handbag] the Visit Count is slightly higher for video-1 as compared to that of video-10. This may be because the description of the handbag was given in video 1, but not in video 10. This is further justified as the average Visit Duration of the area of the handbag in video-1 is close to double of that video-10. For DAOI-4[Trouser], both the average Visit Count and average Visit Duration are significantly higher in video-1 than in video-10 despite the fact that both descriptions specified the colour of the trouser. It is very clear that the participants found it harder to conclude that in video-1 the colour of the trouser is white as it rather appears to be ‘beige’ in colour, i.e. ‘off-white’ in colour.

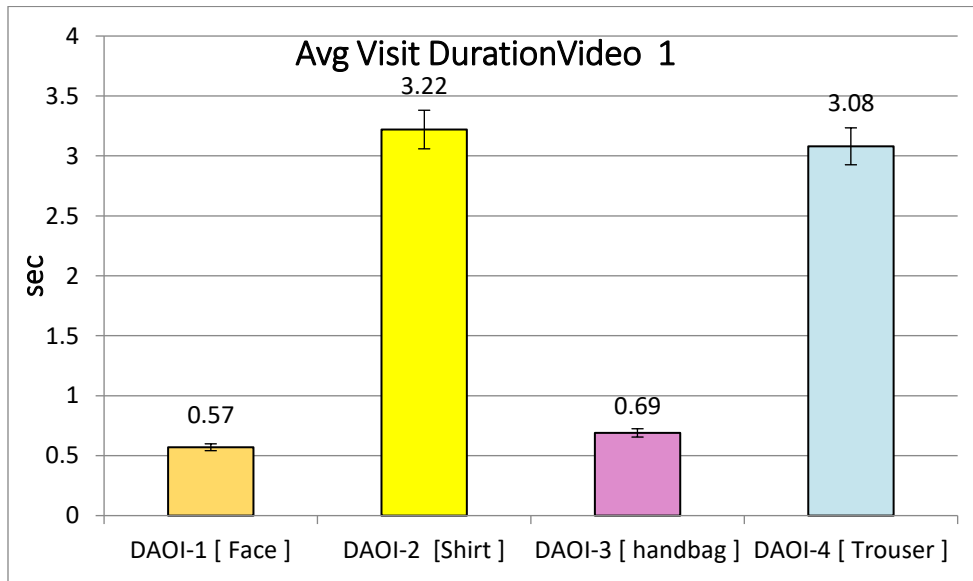


(a)



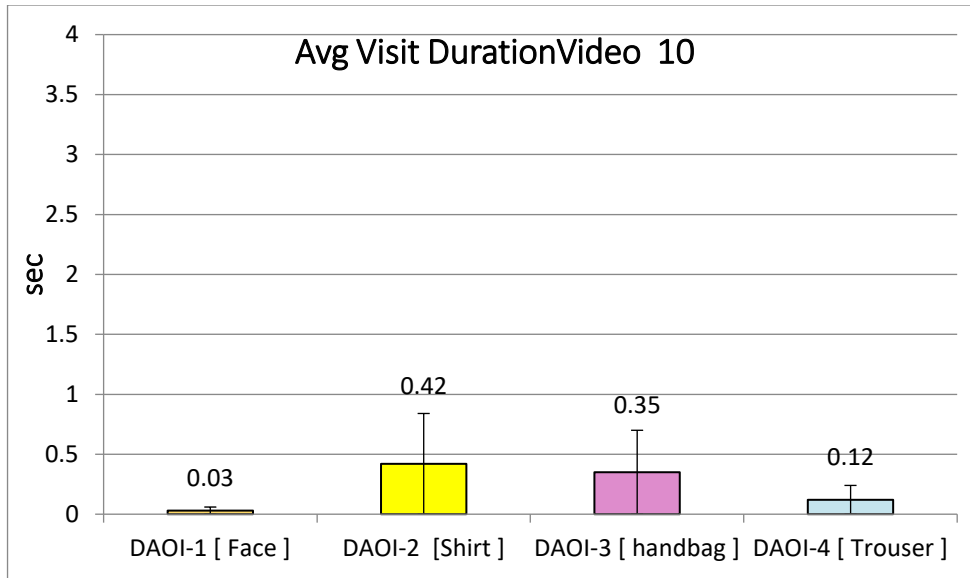
(b)

Figure 5.4: Average visit count for each DAOI in (a) video 1 and (b) video 10



(a)





(b)

**Figure 5.5: The average visit duration for each DAOI in (a) video 1 and (b) video10**

### 5.2.2.1 Summary Research Question 2

To answer this research question four attention metrics were used, Percentage Fixated, Fixation Duration, Visit Count and Visit Duration.

A DAOI with a longer duration fixation, more Percentage Fixated and more visit count can indicate possibilities of either higher levels of attention or interest and sometimes even creating participant confusion.

According, to experiential results we observed that most of the participants fixated more and for longer time on DAOI 2. The majority of the participants first gave attention to the DAOI as the instruction given predominately included a description of the trouser or handbag, i.e. it being 'white' in colour. Most of participants gave attention to the relevant DAOI areas which was specified in the instruction or capture their attention These results show that the region of interest that managed to attract a human observer's attention were the middle and the lower parts of the body despite no direct information was given to describe that part in the description given.

In summary when observing the visual attention of human observers, it is seen from this experimental analysis that different parts of the human body plays a different level of significance in the recognition and tracking (re-identification). This different in attention could be due to the specific instructions given to the observer or due to attractive/distractive other objects, features etc. of a region.

### 5.2.3 Research Question 3

**Which target human body part attracted attention first given the description of the target human object disclosed to the participant?**

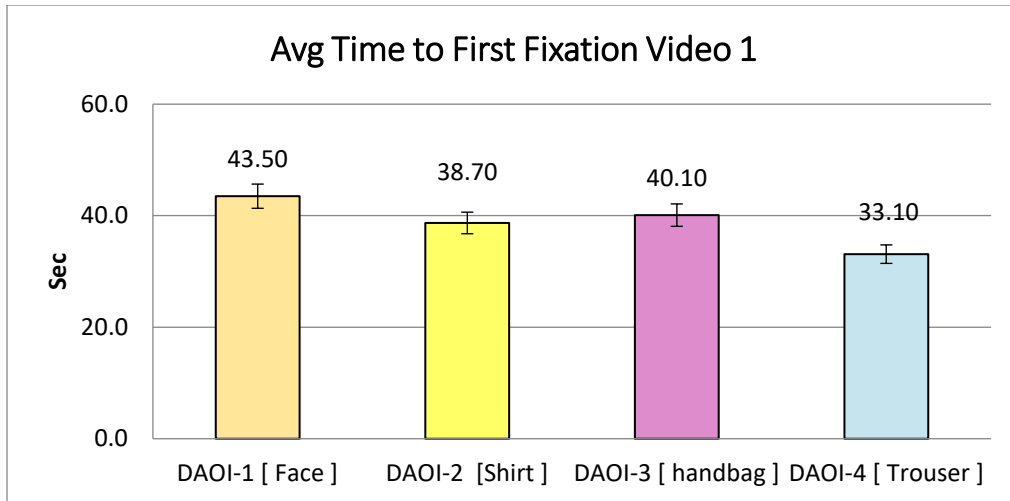
The metric, 'Average time to first fixation' produced by the Tobii eye tracker measures how long it takes before a test participant fixates on a named active DAOI for the first time. A shorter average time to first fixation indicates that the specific region has received very quick attention.

For video-1, on average, the participants of the study took the longest time for the first fixation on DAOI-1[Face] (43.5 Sec, head region) and the shortest time for the first fixation on the DAOI-4[Trouser] (33.1 Sec, leg region) (see figure 5.6). Since the majority of participant (70%, Mean percentage fixated) had seen leg area, i.e. DAOI-4[Trouser] first (had lowest time to first fixation), DAOI-4[Trouser] ranks as number one region attracting the quick attention of participants. The region receiving the second ranked attention is the shirt area, i.e. DAOI-2[shirt].

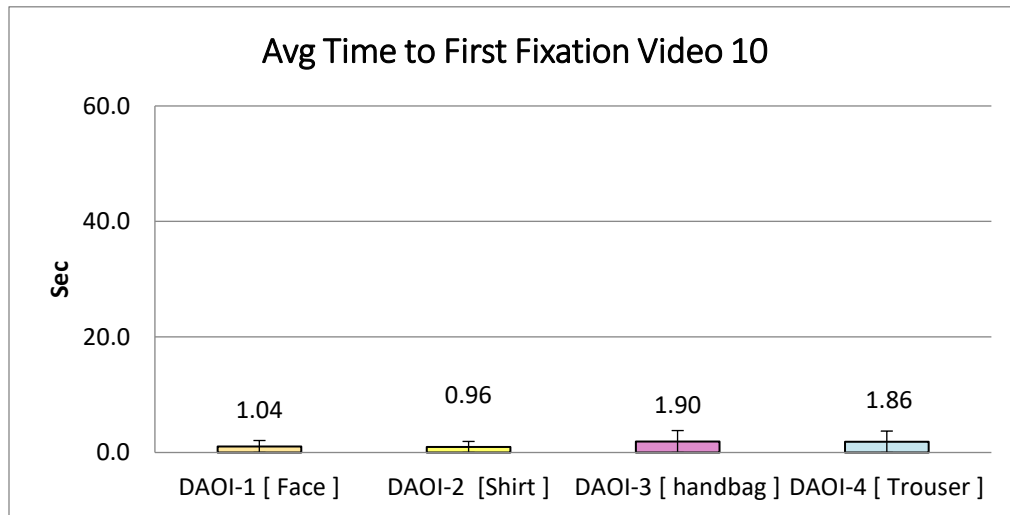
By using the think aloud method during the experiment, helped the researcher to gain a valuable insight into the thought process behind the user's action. Therefore, the reasoning behind the above highest ranking of DAOI-4[Trouser] in terms of the average time to first fixation can be attributed to the fact that a description about the trouser is given for the participants to target their search and it is also not easy to observe that the trouser to be 'white' in colour, thus naturally receives observer attention more as it is a rather difficult question to answer. However, the head area, referred to be DAOI-1[Face] receives the highest metric value as observers appear to be using DAOI-2[Shirt] as a means for confirming that the person is a 'woman' rather than using DAOI-1[Face] perhaps due to the relatively smaller size of face that makes it difficult to visually analyse (hence more time to make a decision), short hair and masculine nature of the person's face that makes it difficult to look at the face and determine the sex of the person being

searched for. There was a statistically significant difference in the Time to first Fixation of the four areas at p-value  $< 0.05$  as determined by one-way ANOVA [ $F(3, 63) = 11.87$ ,  $p = 0.00$ ] for video-1. Therefore, there is sufficient empirical evidence that the DAOI-4[Trouser] has the lowest time to first fixation while the DAOI-1 [Face] has the highest time to first fixation for video-1.

In observing video-10, on average, the participants took the longest time for the first fixation on the DAOI-4[Trouser] (1.9 Sec), i.e. the region of the leg area and the short time for the first fixation on the DAOI-2[Shirt] (0.96 Sec), i.e. the region of the top of the human body. Since the majority of participants (70%, mean percentage fixated) had seen DAOI-2[Shirt] first (had lowest time to first fixation), DAOI-2[shirt] ranks as number one region of the human figure attracting the participants' attention. The above ranking of attention can be explained by the fact that the given description specifies the colour of the 'top' the person is wearing to be 'blue' and it's much more eye-catching than the colour of the trouser, being 'white' (thus attention at the end). There was a statistically significant difference in the Time to first Fixation of the four areas at p-value  $< 0.05$  as determined by one-way ANOVA [ $F(3,33) = 7.38$ ,  $P=0$ ]. There is sufficient empirical evidence that the DAOI-2[shirt] has the lowest time to first fixation while the DAOI-4[Trouser] has the longest time to first fixation for video-10. However, in section 5.2.6 we explain that this aspect should be carefully considered in the detailed analysis of human behaviour due to the nature of the output that is produced by the Tobii eye tracker.



(a)



(b)

**Figure 5.6: Average fixation time to first fixation for each DAOI in (a) video 1 and (b) video 10**

### 5.2.3.1 Summary Research Question 3

The time to first fixation indicates the amount of time takes for a participant to first fixate on a specific area of interest for the first time. Taken on its own, this metric doesn't tell much. However, when compared to other areas of interest, time to first fixation can show you across the board which sections or parts of the body are catch a user's attention in the context of the task they are asked to perform.

As we said previously DAOI with a longer duration fixation can indicate possibilities of either higher levels of attention or interest and sometimes even creating participant confusion

Pervious eye tracking research and in our research, we observed that if the description provided to participants is not enough, if the part is eye catching and if the object parts is not clear (i.e. colour not clear, size small etc.), thus naturally receive observer attention more. Sometimes due to the smaller size of person face, short hair and masculine nature of the person face that makes it difficult to look at face to determine the gender of the person being searched for so the judgement can be taken by using middle part(shirt). If the part is eye catching that part often get attention first although if it was not mentioned in the description.

Finally, instructing participants to observe a video and detect a described person has significant effect on eye movement behavior.

## 5.2.4 Research Question 4

**What is the order in which the participants generally looked at the different sub-regions?**

The ‘average time to first fixation’ metric also helps to find out the order in which participants viewed each DAOI. This metric does not directly provide the order in which each DAOI was seen by each participant. The simplest way to determine the order in which target body parts/DAOIs were looked at is to extract and organize the ‘average time to first fixation’ values of DAOIs of each video, for each participant and perform the analysis as shown in table 5.2 below. Table 5.2 (a) and (b) respectively provides the results obtained from and prepared for the videos 1 and 10. It is noted that for detailed analysis purposes the results for each participant is provided with an indication of whether the individual is a novice (G1/Group1) or an expert (G2/Group2) and whether the individual is female (F) or male (M).

A close analysis of the results tabulated in the tables 5.2 (a) of video-1 reveals that the overall view order is DAOI-4[Trouser], DAOI-2[Shirt], DAOI-3[handbag] and DAOI-1[Face]. This order is clearly justified as the instruction given requests to search for a woman, wearing white pants and carrying a pink shoulder handbag. In section 5.2.1 it was explained that even though the colour of the ‘top’ of the woman is not being specified, the participants appear to be using the DAOI-2[shirt] area to determine whether it’s a man or a woman as the DAOI-1[Face] that consists of the face does not easily reveal this information.

The ‘white trouser’ is the object that catches most attention in terms of % fixations with 70% of participants fixating their view on this part first. This can be attributed to the fact that the colour of the trouser is not clear to be ‘white’ and requires more attention from the observer before a decision could be made. An interesting observation is the fact that even though the colour of the handbag is specified to be ‘pink’ and is bright and eye-catching in nature, 40% of participants have not fixated their view on DAOI-3[handbag],

ever. These participants may have very quickly and via peripheral vision seen that the handbag is pink and made the decision quickly rather than actually fixating their gaze for long enough for the eye-tracker to notice this.

A close analysis of the results tabulated in the tables 5.2 (b) of video-10 reveals that the overall view order is DAOI-2[shirt], DAOI-3[handbag], DAOI-1[Face] and DAOI-4[Trouser] based on the Percentage of Participants Fixating their view on the respective regions. This result is justified as the description is to find a person (male or female, thus DAOI-1[Face] is not relatively important) wearing a blue top (DAOI-2[shirt], most important) and white pants (DAOI-3[handbag], an easy decision and thus not necessarily needing fixation). No participant fixated their view for the first time on DAOI-4[Trouser] and this could be due to the fact that despite the description given to find the 'white' trouser it is a less interesting area to view. Large percentages of participants never fixated their view on DAOI-1[Face] (90%) and DAOI-4[Trouser] (75%) as descriptions of these parts were not needed in making the decision or the decision was very easy to make, respectively for the two regions.



Video 1								
Sample	DAOI Time to First Fixation				DAOI view order			
	DAOI 1	DAOI 2	DAOI 3	DAOI 4	1	2	3	4
G1-F01	45	34.7	44.7	33.0	4	2	3	1
G1-F02		31.6	33.8			1	2	
G1-F03	47.7	44.6	34.8	34.2	4	3	2	1
G1-F04	46.1	43.1		37.9	3	2		1
G1-F05	45.7	4.9	43.2	31.6	4	3	2	1
G1-M01		43.2	41.9	37.9		3	2	1
G1-M02	47.5	39.1	39.3	1.6	4	2	3	1
G1-M03	33.8	41.8		32.7	2	3		1
G1-M04	37.7	32.5	42.9		2	1	3	
G1-M05	48.2	36.1		34.8	3	2		1
G2-F01	33.7	31.6	45.4		2	1		3
G2-F02	42	44.6			1	2		
G2-F03		44.8		31.6		2		1
G2-F04		45.9		31.6		2		1
G2-F05	47.5	46.3	31.6	32.2	4	3	1	2
G2-M01	44.7	35.3	41.9	31.6	4	2	3	1
G2-M02		45.6	46.8	31.6		2	3	1
G2-M03		45.2	42	33		3	2	1
G2-M04		43.8		31.8		2		1
G2-M05	45.6	38.4	33.1		3	2	1	

Sample	
<b>G1-F</b>	Novice -Female
<b>G1-M</b>	Novice -Male
<b>G2-F</b>	Expert -Female
<b>G2-M</b>	Expert-Female

Order No	DAOI 1	DAOI 2	DAOI 3	DAOI 4
1st	5%	15%	10%	70%
2nd	15%	55%	25%	5%
3rd	15%	30%	25%	5%
4rd	30%	0%	0%	0%
Not seen	35%	0%	40%	20%

Avg Time To First Fixation			
DAOI 1	DAOI 2	DAOI 3	DAOI 4
43.5	38.7	40.1	33.1

(a)Video 1

video 10								
Sample	DAOI Time to First Fixation				DAOI view order			
	DAOI 1	DAOI 2	DAOI 3	DAOI 4	1	2	3	4
G1-F01		0.21				1		
G1-F02		0.59	2.15			1	2	
G1-F03		0.88	2.29			1	2	
G1-F04		1.09	2.13			1	2	
G1-F05		0.83		1.3		1		2
G1-M01		0.86	1.94			1	2	
G1-M02			2.48				1	
G1-M03			2.38				1	
G1-M04		1.86				1		
G1-M05		1.07	2.42	2.64		1	2	3
G2-F01		0.65	1.84			1	2	
G2-F02	1.39	0.83			2	1		
G2-F03	0.68	2.72			1	2		
G2-F04		0.73	1.43			1	2	
G2-F05		0.87		2.8		1		2
G2-M01		0.48	2.19	1.18		1	3	2
G2-M02			0.77				1	
G2-M03		0.74	2.08	1.22		1	3	2
G2-M04			1.4				1	
G2-M05			1.14	2.02			1	2

Sample	
<b>G1-F</b>	Novice -Female
<b>G1-M</b>	Novice -Male
<b>G2-F</b>	Expert -Female
<b>G2-M</b>	Expert-Female

Order No	DAOI 1	DAOI 2	DAOI 3	DAOI 4
1st	5%	70%	25%	0%
2nd	5%	5%	30%	25%
3rd	0%	0%	0%	5%
4rd	0%	0%	0%	0%
Not seen	90%	25%	35%	75%

Avg Time To First Fixation			
DAOI 1	DAOI 2	DAOI 3	DAOI 4
1.04	0.96	1.9	1.86

**(b)Video 10**

**Table 5.2: Analysis of Order of Fixation for (a) Video 1 and (b) Video 10**

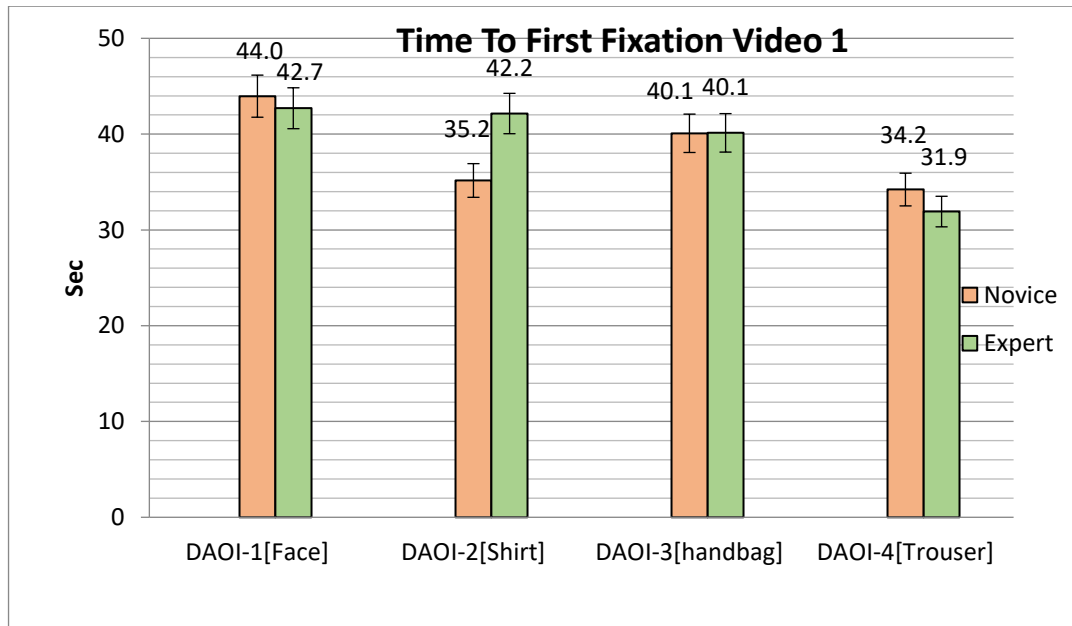
However, if the ‘Average Time to First Fixation’ is considered, the view order is remarkably different, which is DAOI-2[Shirt], DAOI-1[Face], DAOI-4[Trouser] and DAOI-3[handbag]. A closer look at the results tabulated in the tables 5.2 (b) reveals that this is due to the significant percentages of participants not fixating their views on DAOI-1 [Face] and DAOI-4 and in the few cases that some participants fixate their view, that happened very quickly, hence recording low average Time to First Fixations for regions DAOI-2[Shirt] and DAOI-4[Trouser].

Our result for both videos showed, that the Novice and Expert participants had a significantly shorter fixation duration on the relevant aspects of the target body as mentioned in the written description. It is indicated based on the instructions given to the participants, different levels of attention have been received by the different DAOIs.

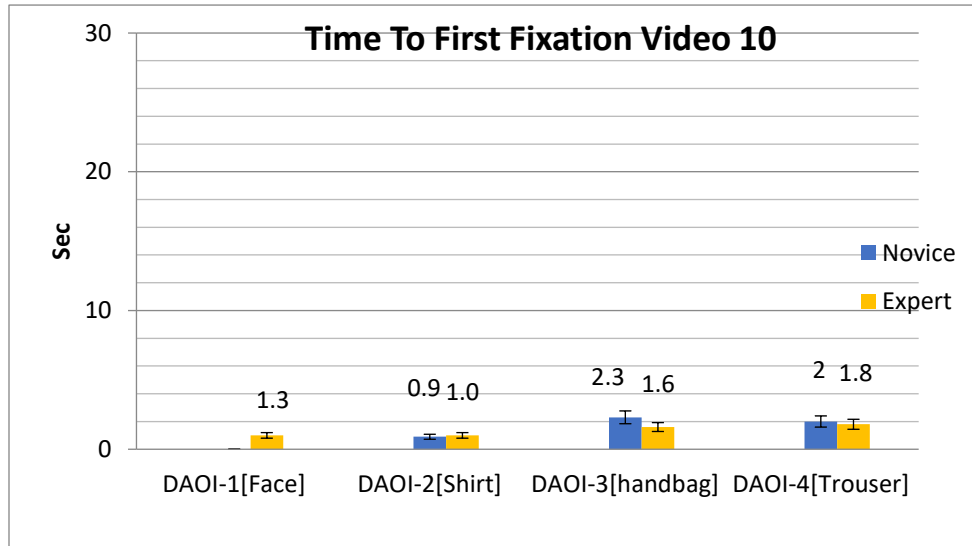
For video-1, on average, the Novice participant took the longest time for the first fixation on DAOI-1[Face] (44.0 Sec, head region) and the shortest time for the first fixation on the DAOI-4[Trouser] (31.9 Sec, leg region) (see figure 5.6). As shown in the figure DAOI-4[Trouser] ranks as number one region attracting the quick attention of both groups (Expert and Novices )participants and the region DAOI-2[shirt] received the second ranked attention .The reasoning behind the above highest ranking of DAOI-4[Trouser] in terms of the average time to first fixation can be attributed to the fact that a description about the trouser is given for the both participants to target their search and it is also not that clear to be ‘white’ in colour, thus naturally receives observer attention more as it is a rather difficult question to answer. However, the head area, referred to be DAOI-1[Face] receives the highest metric value (participants took the longest time for the first fixation) as observers appear to be using DAOI-2[Shirt] as a means for confirming that the person is a ‘woman’ rather than using DAOI-1[Face] perhaps due to the smaller size of face, short hair and masculine nature of the person’s face that makes it difficult to look at the face and determine the sex of the person being searched for.

In observing video-10, on average, Novices participants took the longest time for the first fixation on the DAOI-3[handbag] (2.3 Sec) and the short time for the first fixation on the DAOI-2[Shirt] (0.9 Sec). However, Expert participants took the longest time for the first fixation on the DAOI-4[Trouser] (1.8 Sec), and the short time for the first fixation on the DAOI-2[Shirt] (1.0 Sec).

The DAOI-2[Shirt] had lowest time to first fixation, therefore DAOI-2[shirt] ranks as number one region of the human figure attracting the participants' attention. The above ranking of attention can be explained by the fact that the given description specifies the colour of the target shirt to be 'blue' and it's much more eye-catching than the colour of the trouser, being 'white' (thus attention at the end).

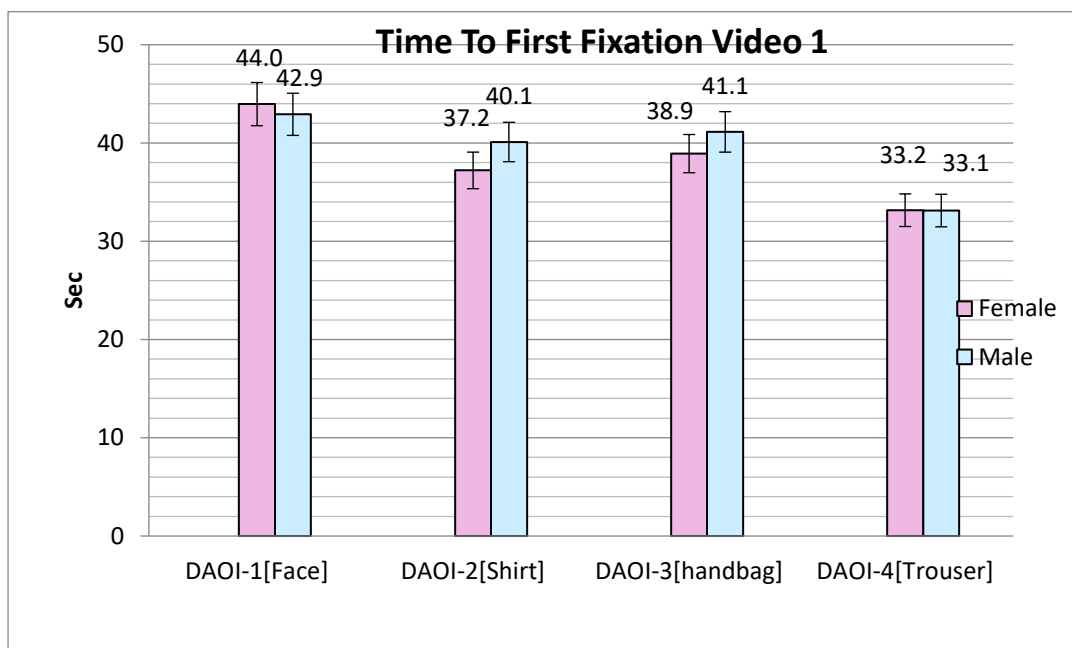


(a) Video 1

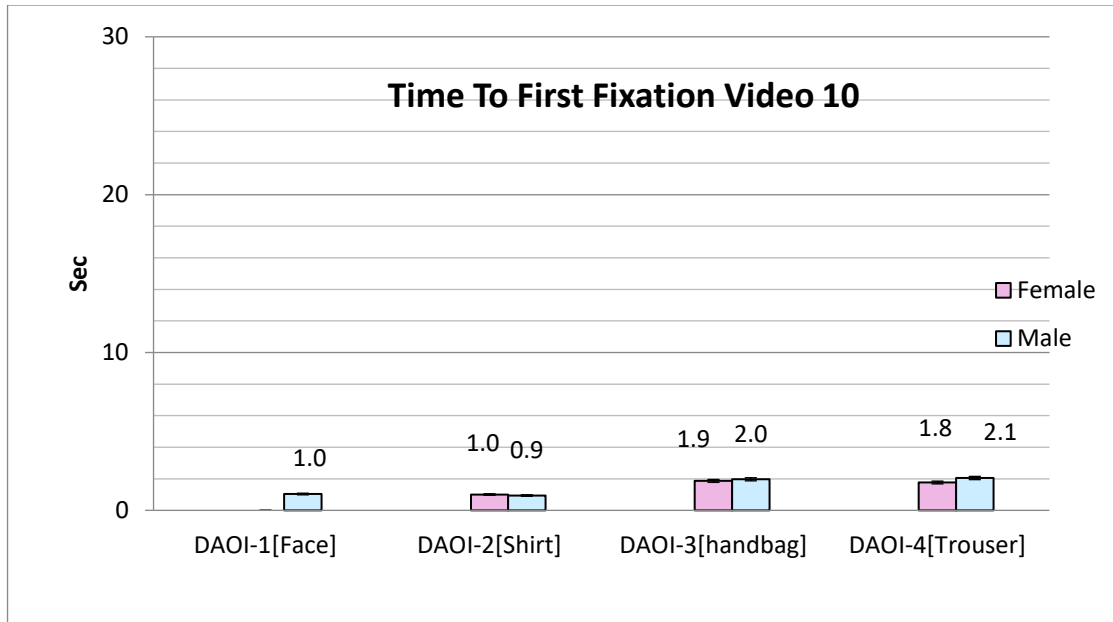


**(b) Video 10**

**Figure 5.7: Novice and Expert Average time to first fixation for each DAOI in (a) video 1 and (b) video 10**



**(a) video 1**



(b) Video 10

Figure 5.8: Female and Male Average time to first fixation for each DAOI in (a)video 1 and (b) video 10

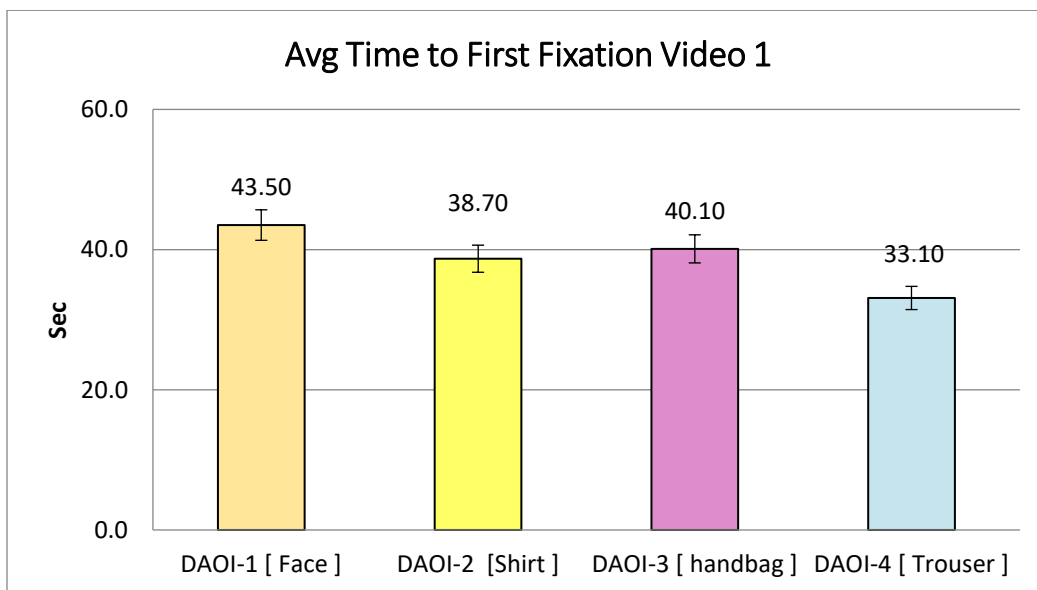
#### 5.2.4.1 Summary Research Question 4

All participants were very much engaged with the search process and were effective at accurately detecting and identifying the described person (target). However, the patterns of search for each participant was somehow different with some clear outliers due to participant nature, ability to concentrate on the task, remember the descriptions and speed of reaction

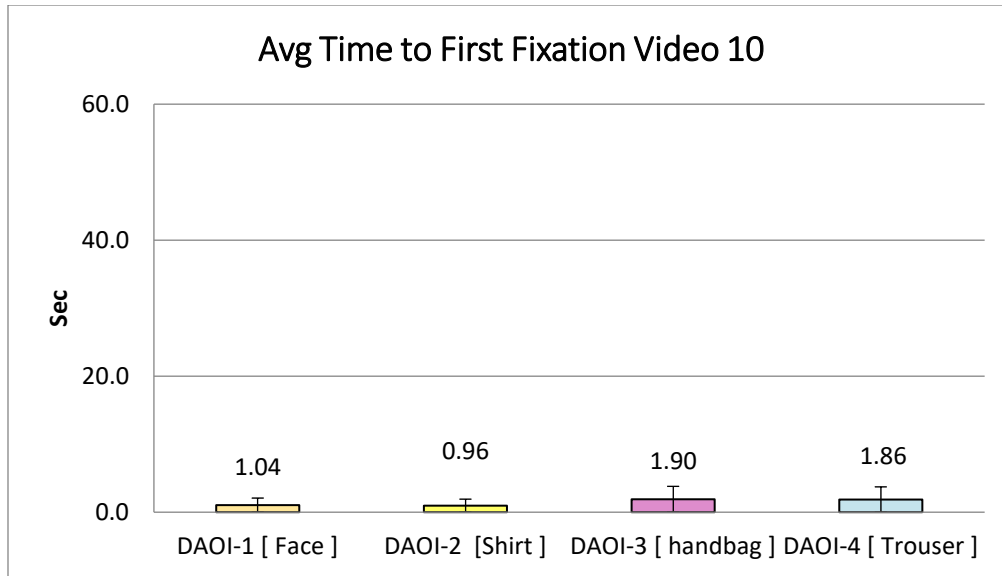
## 5.2.5 Research Question 5

**How long did it take to spot for the first time each part of the body?**

When considering the metrics that Tobii eye tracker captures, from the definitions given for each metric, the first impressions suggest that this question can be answered by analysing the values of the metric, 'Time to First Fixation'. Figure 5.7 illustrates the Average Times to First Fixation for each of the four regions for the two test videos, video-1 and video-10. In video-1 it takes 43.5, 38.7, 40.1 and 33.1 sec to first fixate on DAOI-1[Face], DAOI-2[Shirt], DAOI-3[handbag] and DAOI-4[Trouser] respectively. For video-10 it takes 1.04, 0.96, 1.90 and 1.86 sec to first fixate on DAOI-1[Face], DAOI-2[Shirt], DAOI-3[handbag] and DAOI-4[Trouser] respectively.



**(a) Video 1**



(b) Video 10

**Figure 5.9: Analysis of Average Time to First Fixation for (a)video 1 and (b) video 10**

### 5.2.5.1 Summary Research Question 5

As discussed in section 5.2.4 the metric 'Time to First Fixation' is not suitable to make an accurate and reasonable judgement as it ignores the large percentage of participants who never fixate on certain DAOIs (for example for video-10, DAOI-1[Face] and DAOI-4[Trouser]).

Therefore, this metric should be carefully used to answer this question, considering and separating/removing the DAOIs that receives minimal participant attention first and then ordering the remaining DAOIs in the rank order of this metric to answer the research question.



## 5.2.6 Research Question 6

**Has the written description of the target human object given to the participants has any influence on the order in which participants looked at the different target human body parts?**

There has been much research concluded that task instructions influence where observer's eye fixated while viewing static scenes or images [89]. In 2011 a research project conducted on dynamic scenes has shown a similar finding [88]. However, with regards to analysing CCTV footage for people recognition, re-identification and tracking, research has not been conducted to this effect and has hence been the focus of the research conducted within the research context of this thesis.

The analysis of results presented in section 5.2.1 to 5.2.5 above and the answers thus obtained for the research questions 1-5 above clearly reveal the answer to this research question. The written description of the target human object given to the participants has a direct influence on the order in which the participants looked at the different target human body parts.

### 5.2.6.1 Summary Research Question 6

In order to provide a wider context to answering these questions in figure 5.8 (a), (b) and (c) below we plot the gaze-plots for all participants before the object of interest appears in the scene, during the time the object continues to be present and visible in the scene and after the object of interest has disappeared from the scene for video-1.

The gaze plots indicate that specific focus is given to the object's full Dynamic Region of Interest (includes all DAOIs analysed above) during its appearance and a detailed analysis of what happens during this time was presented in sections 5.2.1 to 5.2.5. When the object is not in the scene, with the given instructions in mind, the participants are scanning the scene in the most probable regions that a human object is likely to appear.

The written description of the target human object given to the participants has a direct influence on the order in which the participants looked at the different target human body parts.

The figure 26 illustrates scan paths of all participants before, during and after the target object appears on screen for video-1 Analysis of Average Time to First.



(a) before target appear on the screen



(b) during the target appear on the screen



(c) after the target object disappears from the screen

**Figure 5.10: Scan paths of all participants (a) before, (b) during and (c) after the target object appears on screen for video-1 Analysis of Average Time to First**

### **5.3. Summary and conclusion**

The analysis of experimental results obtained via the Tobii eye-tracker software in the form of eye-tracking related metrics and the statistical analysis conducted via SPSS for checking the statistical significance of the results that have been obtained led to the following conclusions.

- Before the target appears on the scene, the participants' eye gaze is random but is more focused around the regions where a human object is likely to appear. When the object appears at a distance the eye gaze becomes focused on the human object until details begin to appear that can lead to further detailed visual analysis of the human figure.
- Giving specific instructions to look for specific people, wearing specific clothing or carrying/holding specific objects has a significant impact on the participants' eye movement behaviour after the target appears and before it disappears.
- Almost all participants fixate their gaze on task relevant aspects of the footage. It can be concluded that given instructions have been in sufficient detail to influence fixation behaviour of all participants. When a task specific decision was easy and quick to make, the relevant areas did not receive 'fixations of gaze' from most participants.
- All participants were very much engaged with the search process and were effective at accurately detecting and identifying the described person (target). However, the patterns of search for each participant was somehow different with some clear outliers due to participant nature, ability to concentrate on the task, remember the descriptions and speed of reaction.

It is noted that although the seven-metrics produced by the Tobii eye tracker was able to answer the research questions 1-6, answering the research questions 7 and 8 will require further analysis. In particular, if one is to investigate if there is a difference in the observation patterns of males vs females and novice vs experienced participants, the consideration of the large number of metrics that results from the different DAOIs will be close to impossible, via statistical and/or conceptual means. The novel approaches presented in Chapters 6 address answering this research question by the use of machine learning algorithms.

## **5.4. Human Observer Behaviour Analysis of CCTV Video Surveillance using Linear Regression**

In **section 5.2 of this Chapter(chapter-5)** a detailed human observer performance analysis was conducted based on conceptually and statistically analysing the seven-feature metrics produced by the Tobii eye tracker, for different DAOI. The analysis conducted was focused on answering six of eight stated research questions. In answering each of the six research questions it was sufficient to analyse one metric at a time.

Two further research questions remained un-answered that required a broader analysis where if relationships between the recorded metrics can be analysed they could lead to further information about human behaviour analysis.

The section 5.4 of this Chapter aims to use linear regression which is statistical technique found in the popular Machine Learning Toolbox WEKA in order to carry out data modelling, correlation analysis and predictions based on the seven parameters recorded and additional derived information from the subjective tests carried out.

It is noted that in literature, linear regression with WEKA has not been used in relation to the human observer performance analysis of CCTV video footage. Such analysis could lead to gathering information that was previously considered to be impossible to be gathered via statistical approaches. This is the key contribution of the research conducted in this chapter.

The popular Machine Learning Toolbox WEKA [78] was used in this research to provide the implementations of data pre-processing, preparation, reduction, feature extraction, modelling and classification algorithms that are required to carry out the proposed machine learning based analysis of observer behaviour.

The following sections 5.4.1 to 5.4.4 present the methodology adopted:

### 5.4.1 Data capture

As discussed in section 5.1, during each subjective experiment, i.e. for the experiments of each of the thirteen videos be viewed by each of the 20 participants, seven attributes related to each of the four DAOIs are recorded. Table 6.1 lists the seven attributed indicating the attribute type as ‘numeric’. The attribute ‘type’ plays a key role in the modelling and machine learning experiments to be conducted in the following sections. All attribute values are automatically captured and provided by the Tobii Studio software.

<b>Variables</b>	<b>Parameters</b>
<b>VD</b>	<b>Visit Duration</b>
<b>VC</b>	<b>Visit Count</b>
<b>TTF</b>	<b>Time to First Fixation</b>
<b>FFD</b>	<b>First Fixation Duration</b>
<b>TFD</b>	<b>Total Fixation Duration</b>
<b>FC</b>	<b>Fixation Count</b>
<b>PF</b>	<b>Percentage Fixated</b>

**Table 5.3 Attributes and attribute ‘type’ for each DAOI**



## 5.4.2 Data pre-processing or preparation

The data captured as described in section 6.1 is not fit to be directly used for the purpose of modelling or machine learning. A close investigation of the captured data revealed that some data values are missing. In Chapter-5, it was revealed that missing data is due to issues with regards to some human eye, registration issues with some subjects. When the conceptual and statistical analysis was done in the case of two users whose eyes did not get registered appropriately the entire records of the data were removed and thus excluded from consideration. While this approach is the best when the majority of the data captured for one specific user during a specific experiment is not recorded or lost, if only few values are missing, this approach could lead to substantial data loss. WEKA provides different algorithms to deal with missing data and in the work presented in this chapter, the possibility of using such algorithms were considered.

In addition to the above, not all attributed captured are required for the purposes of modelling and machine learning. Some attributes may not contribute to the models or to the machine learning to be carried out as they are redundant with respect to the creation of any additional knowledge/information. Having such attributes considered in the building of the models can have an adverse impact in terms of reducing the accuracy of the models created. Therefore, in this research the possibility of attribute selection was considered. WEKA has a number of attribute selection algorithms implemented within itself, which were tried and tested within the context of the experiments conducted.

## 5.4.3 Removal of Missing Values

In the experiments conducted the filter named `RemoveWithValues`

listed under the preprocessing algorithms in WEKA (***Filters*** → ***Unsupervised*** → ***Instances*** → ***RemoveWithValues***) was used for the removal of missing values. Where data is values are missing in the records this filter removed the whole record, rather than attempting to fill the missing values, which often lead to inaccuracies in the substitute values generated for the missing values.

#### 5.4.4 Attribute or Feature Selection

There are many features selection options/ filters implemented within WEKA. The most popularly used approaches in literature are listed in Table 5.4. All of these approaches were tested on the data that had been captured for this research.

	<b>Method evaluator</b>	<b>Search Method</b>
<b>1</b>	cfsSubsetEval Evaluator	Best First and Greedy Stepwise Search methods
<b>2</b>	ReliefFAttributeEval	Attribute ranking
<b>3</b>	Principal Components	Attribute ranking
<b>4</b>	WrapperSubsetEval	Best First and Greedy Stepwise Search methods

**Table 5.4 WEKA Feature Selection Methods**

It was found that "cfsSubsetEval" with search method greedy stepwise filters improved the accuracy of data modelling therefore positively impacting on the effective use of data. Unfortunately, the rest of the filters did not improve the accuracy of data modelling. It is noted that in the attribute reduction experiments conducted here data captured from all the videos for all DAOIs and all participants were considered. When considering data related to specific videos only it was observed that the features that remained were a subset of the features presented here. Table 5.5, lists the attributes selected as a result of the above selection procedure.

Dataset Attributes	Feature Selection
Time to First Fixation (Sec)	✓
Fixation Count	✓
Visit Duration (Sec)	✓
Visit Count	✓

**Table 5.5 Selected Attribute when all data are considered**

As indicated by the contents of Table 5.5, insignificant attributes were excluded as expected from use in modelling, thus impacting positively on the accuracy of modelling. One of the key attributes removed was the ‘Percentage Fixated %’ indicating that the percentage of participants fixating their view on a specific DAOI has no impact on the duration spent of inspecting the said DAOI. By the subject for any given video.

## **5.5. Dataset Representation**

Samples of data instances for all the DAOIs, i.e. DAOI 1, DAOI 2, DAOI 3 and DAOI 4, for the video 10 and Video 1 are presented in tables 6.4 and table 6.5 respectively. The data is arranged in .CSV file format and fed into WEKA for the purpose of modelling. It is noted that our intention is to model Fixation Duration (i.e. the dependent variable) based on the remaining four parameters, namely, Fixation Count, Visit Duration, Visit Count and Time to First Fixation (i.e. the independent variables) for each DAOI.

Time to First Fixation	Fixation Count	Visit Duration	Visit Count	Fixation Duration
1.3	1	0.3	1	0.3
2.64	1	0.52	1	0.52
1.18	1	0.69	1	0.69
1.22	1	0.38	1	0.38
2.02	1	0.26	1	0.26
0.21	2	0.31	2	0.31
0.59	3	0.31	3	0.31
0.88	1	0.89	1	0.89
1.09	1	0.42	1	0.42
0.83	2	0.55	2	0.55
0.86	1	0.28	1	0.28
1.86	1	0.34	1	0.34
1.07	1	0.74	1	0.74
0.65	1	0.75	1	0.75
0.83	4	0.39	3	0.26
2.72	1	0.12	1	0.12
0.73	3	0.28	3	0.28
0.48	1	0.7	1	0.7
0.74	3	0.3	3	0.3

DAOI 2

Time to First Fixation	Fixation Count	Visit Duration	Visit Count	Fixation Duration
2.15	2	0.24	2	0.24
2.29	1	0.35	1	0.35
2.13	2	0.28	2	0.28
1.94	1	0.19	1	0.19
2.48	1	0.61	1	0.61
2.38	1	0.54	1	0.54
2.42	1	0.22	1	0.22
1.84	1	0.53	1	0.53
2.19	1	0.01	1	0.01
0.77	2	0.58	2	0.58
2.08	2	0.28	2	0.28
1.4	1	0.39	1	0.39
1.14	2	0.43	2	0.43

DAOI 3

Time to First Fixation	Fixation Count	Visit Duration	Visit Count	Fixation Duration
1.3	1	0.3	1	0.3
2.64	1	0.52	1	0.52
1.18	1	0.69	1	0.69
1.22	1	0.38	1	0.38
2.02	1	0.26	1	0.26

DAOI 4

Table 5.6: Sample data selection for video-10

Time to First Fixation	Fixation Count	Visit Duration	Visit Count	Fixation Duration
43.44	2	0.17	2	0.17
43.36	1	0.48	1	0.48
47.77	1	0.13	1	0.13
46.14	5	0.24	4	0.19
45.69	4	0.3	4	0.3
37.74	6	0.43	5	0.35
48.25	2	0.28	1	0.1
33.69	2	0.5	2	0.5
42.23	4	0.43	4	0.43
47.53	2	0.22	1	0.1
44.68	4	0.34	2	0.14
45.54	4	0.65	2	0.3

DA01 1

Time to First Fixation	Fixation Count	Visit Duration	Visit Count	Fixation Duration
31.6	6	0.22	6	0.22
33.89	3	0.31	3	0.31
34.8	2	0.7	2	0.7
41.91	1	0.11	1	0.11
41.92	2	0.22	2	0.22
42.15	2	0.58	1	0.22
42.09	1	0.01	1	0.01
44.69	1	0.13	1	0.13
41.91	2	0.1	2	0.1
43.09	1	0.25	1	0.25
44.85	1	0.11	1	0.11
41.91	3	0.43	2	0.28
41.91	10	0.44	4	0.12
45.14	2	0.25	2	0.25

DA01 3

Time to First Fixation	Fixation Count	Visit Duration	Visit Count	Fixation Duration
40.01	3	0.21	3	0.21
35.13	17	0.56	9	0.24
31.57	22	0.97	9	0.34
44.64	5	1.02	2	0.4
41.53	6	0.22	5	0.15
40.9	4	0.23	3	0.17
41.53	7	0.28	4	0.12
39.98	10	0.35	5	0.14
33.79	8	0.64	4	0.27
32.94	14	0.81	7	0.39
45.8	5	0.51	2	0.16
31.75	14	1.02	8	0.53
42.03	8	0.62	5	0.35
44.8	3	0.42	2	0.27
45.94	7	0.48	3	0.17
46.14	3	0.68	1	0.22
37.01	11	0.59	6	0.3
41.53	4	0.23	3	0.14
32.55	9	0.46	5	0.23
34.17	10	0.62	6	0.35

DA01 2

Time to First Fixation	Fixation Count	Visit Duration	Visit Count	Fixation Duration
32.11	3	0.35	3	0.35
41.73	1	0.25	1	0.25
34.24	12	0.45	8	0.29
38.15	3	0.2	3	0.2
31.57	3	1.51	3	1.51
37.95	4	0.2	4	0.2
31.75	7	0.11	7	0.11
32.34	10	0.89	6	0.5
34.39	11	0.42	10	0.38
31.57	6	1.2	6	1.2
32.24	13	0.58	9	0.37
31.57	4	1.25	3	0.87
33.03	9	0.92	6	0.58
31.78	7	0.83	5	0.59

DA01 4

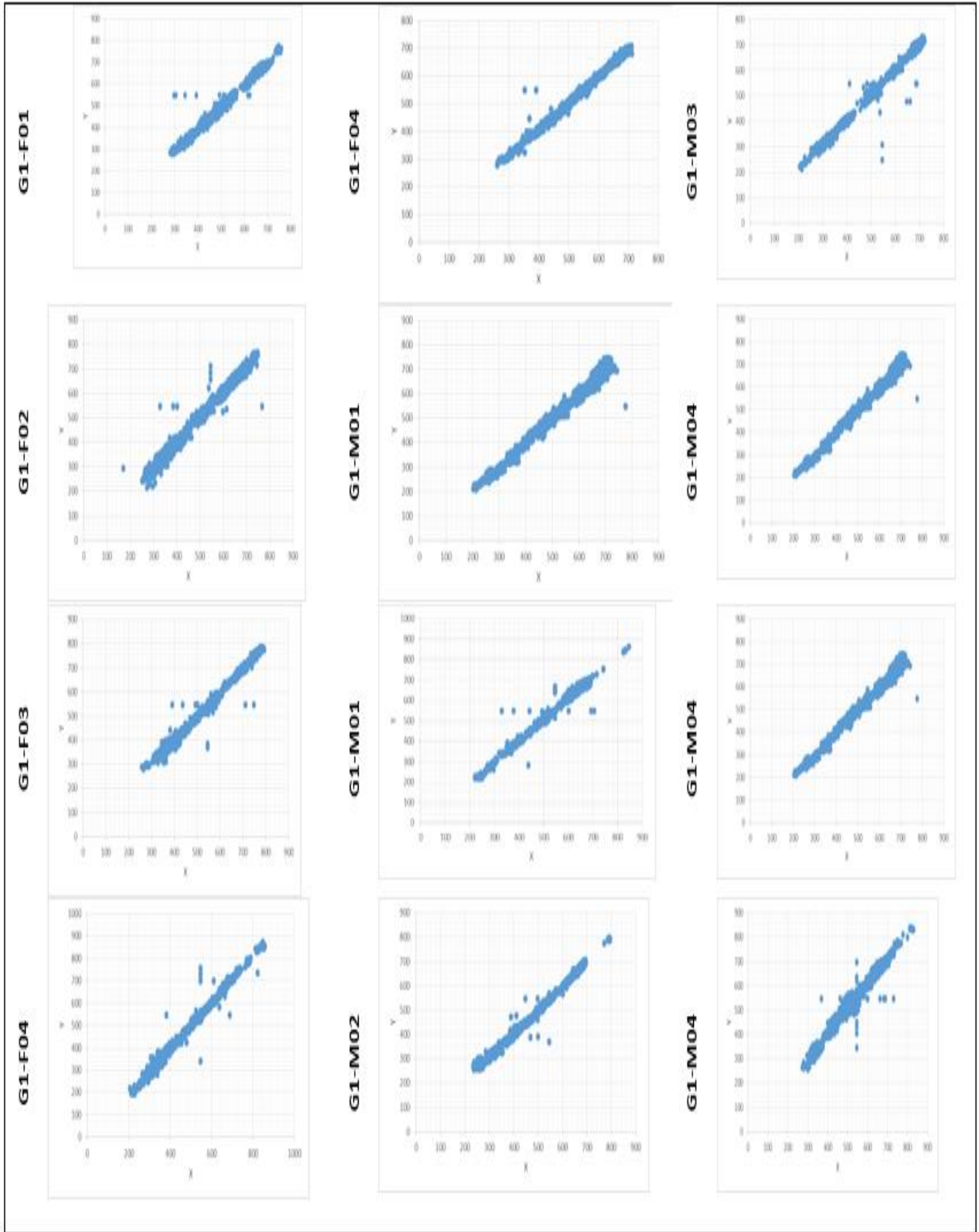
Table 5.7: Sample data selection for video-1

## 5.6. Modelling Fixation Duration

Fixation Duration on a DAOI indicates a measure of visual importance of the DAOI in terms of the search task assigned to the subjects, given the videos and the descriptions given [74][75]. The purpose of this modelling exercise is to model the Fixation Duration, hence the visual significance given the task at hand and video of each DAOI based on the parameters that determine the visual attention pattern, such as the Fixation Count, Visit Count, Visit Duration and Time to First Fixation. Given that each video is different with the object of attention appearing in the scene at different times as compared to the viewing start time and the descriptions given are different, the modelling has to be carried out per video, per description but for all subjects.

Given the high possibility of a linear relationship (see figure 27), the first attempt was to use a Linear Regression Model [70][71] to model visual attention on each DAOI. In this thesis, the modelling results obtained for the video-1 and video-10 are presented and analysed. [It is noted that similar modelling was carried out for all remaining eleven test videos producing similar levels of accuracy of modelling.]

Linear Regression is used to model the visual attention on different parts (DAOI 1, DAOI 2, DAOI 3 and DAOI 4) of the object being identified and tracked on both video 1 and video 10 and to thus identify Important features of the eye-gaze patterns and behaviour of participants when presented with the visual surveillance task. Basically, the data provided in table 5.6 and 5.7 for video-10 and video-1 respectively, for each DAOIs are the inputs to the linear regression modelling process. The data modelling was carried out using linear regression with the Fixation Duration used as the dependent variable and the remaining four attributes used as the independent variables. Ten-fold cross validation was used to optimize prediction accuracy and reduce the bias of selecting a specific dataset for testing.



**Figure 5.11: a linear relationship**

The tables 5.8 and 5.9 respectively lists the obtained models for each DAOI of video 1 and 10 and their correlation coefficient as generated by WEKA linear regression for video-1 and video-10 respectively. Note that FC, VD, VC and TTF are denoted as Fixation Count, Visit Duration, Visit Count and Time to First Fixation respectively

	DAOI			
	DAOI 1	DAOI 2	DAOI 3	DAOI 4
Linear Regression Model	Fixation Duration= -0.1091 * FC+ 0.7547 * VD+ 0.1213 * VC + -0.0058 * TTF+ 0.3021	Fixation Duration= -0.0315 * FC+ 0.4714 * VD+ 0.065 * VC + -0.0309	Fixation Duration= -0.0541 * FC + 0.7686 * VD+ 0.0721 * VC + -0.0017	Fixation Duration= -0.08 * FC+ 0.08 * VD+ 0.0966 * VC + -0.0391
Corr. Coeff.	<b>0.9804</b>	<b>0.9619</b>	<b>0.9079</b>	<b>0.9786</b>

Table 5.8 The regression models for video-1

	DAOI			
	DAOI 1	DAOI 2	DAOI 3	DAOI 4
Linear Regression Model	Only one person seen target face	Fixation Duration = -0.13 * FC + 1 * VD + 0.13 * VC + 0	Fixation Duration = 1 * VD + 0	Fixation Duration = 1 * VD + 0
Corr. Coeff.		<b>1</b>	<b>1</b>	<b>1</b>

Table 5.9 The regression models for video-10



## 5.7. Interpretation and Analysis of Regression Models

In this section, a detailed interpretation and analysis of the regression models are provided to identify the observer behaviour via eye gaze attention patterns.

The resulting linear regression functions for DAOI 1, DAOI 2, DAOI 3 and DAOI 4 are the final modelling outcomes produced by WEKA. Following are the models obtained for video 1, with  $f(1)$  representing the FD of DAOI 1,  $f(2)$  representing FD of DAOI2,  $f(3)$  representing the FD of DAOI 3 and  $f(4)$  representing the FD of DAOI 4. Note also that FC, VD, VC and TTFE have been replaced by the variables  $x1$ ,  $x2$ ,  $x3$  and  $x4$  respectively.

$$f(1)_{DAOI1} = -0.1091 * x1 + 0.7547 * x2 + 0.1213 * x3 - 0.0058 * x4 + 0.3021$$

$$f(2)_{DAOI2} = -0.0315 * x1 + 0.4714 * x2 + 0.065 * x3 - 0.0309$$

$$f(3)_{DAOI3} = -0.0541 * x1 + 0.7686 * x2 + 0.0721 * x3 - 0.0017$$

$$f(4)_{DAOI4} = -0.08 * x1 + 0.08 * x2 + 0.0966 * x3 - 0.0391$$

(1)

Following are the models obtained for video-10 using the same notations for FD, FC, VD, VC and TTFE (if relevant).

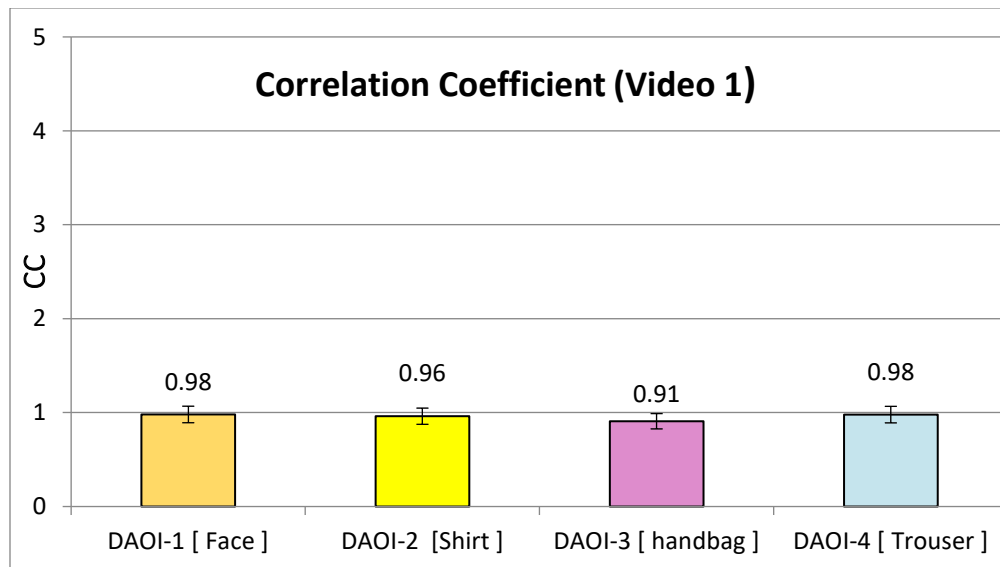
$$f(2)_{DAOI2} = -0.13 * x1 + 1 * x2 + 0.13 * x3 + 0$$

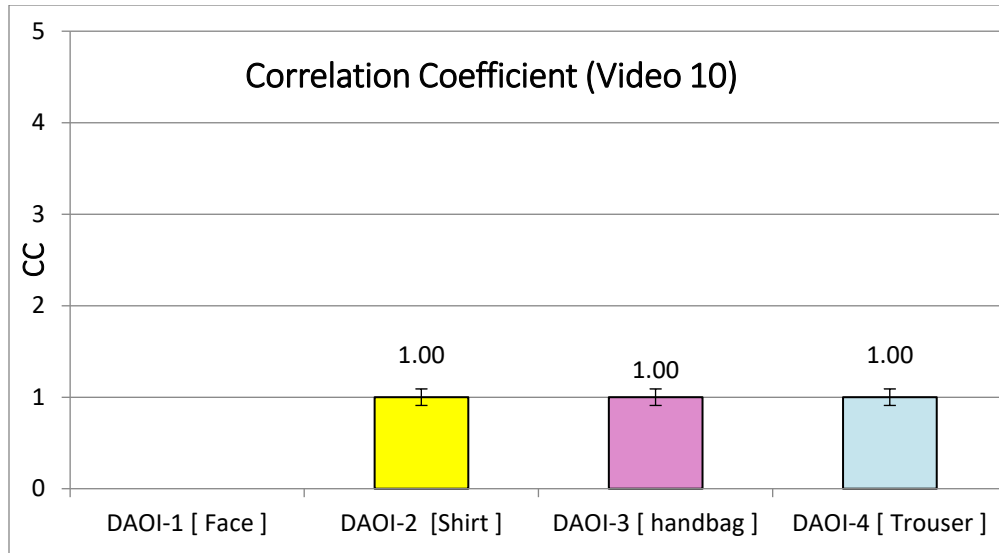
$$f(3)_{DAOI3} = 1 * x2 + 0$$

$$f(4)_{DAOI4} = 1 * x2 + 0$$

(2)

Given the correlation coefficients listed tables 6.6 and 6.7, obtained from the WEKA based modelling process for each DAOI of each of the two videos, it is seen that the Fixation Duration predictions obtained for all DAOIs in both videos were very accurate, in particular for DAOI 1, DAOI 2 and DAOI 4 of video-1 (all having correlation coefficients  $\geq 0.96$ ) and DAOI 2, DAOI 3 and DAOI 4 of video-10 (all having correlation coefficients of precisely 1). The correlation coefficients for the two models are plotted in Figure 5.12 for the purpose of ease of comparison. Overall the very high correlation coefficients obtained for all models justify the proposed use of a Linear Regression model for the prediction of the visual attention to the various parts of the human body.





**Figure 5.12: Correlation coefficient obtained for each DAOI of video-1 and video-10**

A detailed analysis of the models obtained for each of the DAOIs of video1 can be presented as follows, leading to the identification of a number of interesting behavioural patterns of observers when inspecting the video under the given instructions:

It is noted that when inspecting the video-1, the instruction given to all participants is, ***“Find a woman wearing white trouser and carrying a pink shoulder handbag”***. In this specific video the object concerned, first appears in the scene in a far-away location and then comes closer to the camera and disappears at the front of the scene, close to the camera.

For ease of reference for detailed analysis, Table 5.10 tabulates the coefficients obtained for each DAOI, for each independent variable, for video-1.

	$x1 / FC$	$x2 / VD$	$x3 / VC$	$x4 / TTFF$	$C.C$
DAOI-1 (Head area)	- 0.1091	0.7547	0.1213	-0.0058	<b>0.9804</b>
DAOI-2 (Upper body)	-0.0315	0.4714	0.0065	0	<b>0.9619</b>
DAOI-3 (Carried bag)	-0.0541	0.7686	0.0721	0	<b>0.9079</b>
DAOI-4 (Lower body)	-0.0800	0.0800	0.0966	0	<b>0.9786</b>

**Table 5.10: Models for video-1**

Given the instructions to the participants it is noted that there are three keywords or important features to look for. First to identify that it's a *woman*, and then also to identify the *white trouser* and *pink shoulder handbag*. Nothing has been mentioned about what the woman is wearing on the upper body.

The values tabulated in Table 5.10 show that the coefficient for the Fixation Count for all DAOIs is negative. This indicates that higher the Fixation Count, lower will be the Fixation Duration (due to negative correlation). Higher fixation count can indicate an area that is catching the observer's attention, but intermittently. If the area was actually keeping the observer's attention intact for a reason, then the FC should ideally be low as attention will result in focus for a longer period of time. Therefore, the negative correlation indicated in the models between FD and FC is accurate and justified.

It is also observed that the region of which the attention is mostly impacted by the fixation count is the 'head area'. This observation is justified as the given description requests to look for a woman, and the actual person is a woman, but with short hair and facial features of a less feminine nature. The participants may have attempted to check for few times that it is indeed a woman as it is not necessarily clear by looking at the face that it is actually a woman. It is also noted that the least impact of FC on the FD is shown in the upper body area as the coefficient is the least. What this means is that there is less correlation between the FC and FD in the upper body, indicating that some people will visit this part more often than others despite that this changing nature of behaviour has less impact on the FD. At this moment as all groups of people, i.e. expert/novice,

men/women, have all being considered together it is not possible to say whether this different in behaviour viewing the top body part, depends on the type of person observing.

It is seen that for all the DAOIs, the VD is the parameter that mostly impacts the Fixation Duration. This is indicated by the fact that the largest coefficient is for the VD parameter in all equations. The correlation is positive. The Visit Count (VC) is also positively correlated with the Fixation Duration.

When observing, the parameters obtained for Time To First Fixation, it is seen that apart from the head area, TTFF does not impact the Fixation Duration of any of the other three regions. Even for the head area, the coefficient obtained is significantly small and can be considered as thus impacting insignificantly. However, what this shows is that the observers may have first tried to identify that it is a woman before they tried to look for other features. The Fixation Duration on the head area thus depends slightly on the TTFF. Perhaps the individuals who showed a keener interest on the face, looked at it at an early stage and overall spent more time looking at this area, while identifying and tracking the individual.

An interesting observation with regards to the 'carried handbag' area is that it is the region with the least correlation coefficient, or the least accurate model obtained. This is justified as the handbag is attractive and might receive different amounts of interest from the different observers, making the creation of a simple linear model difficult and hence ultimately less accurate. At this moment, the experiments conducted consider all participants together. In Chapter-6 further experiments are conducted to determine whether there is a difference in the way different groups etc., men/women, experienced/novice observe the videos. For the remaining three areas the correlation coefficient of the models obtained are very accurate. This indicates that all observers behaved in a very similar way when inspecting these regions.

It is seen that even though no description was given to the middle part of the human body being looked for participant attention is being received by this part. However, for this

area the magnitude of the coefficients are the smallest, indicating that the Fixation Duration on this area is least impacted by, especially by FC, VD and VC. What this means is that this part of the body, even without being described with a feature is receiving attention from the participants, in a manner not similar to the other three regions. The fact that the lack of clarity whether it is a woman or a man by looking at the face may ensure added attention to this area is a way to justify this observation.

Given the above explanations it is seen that the Linear Regression models obtained do not only illustrate the possibility to model observer behaviour accurately, but also allows a detailed behavioural analysis to be conducted.

A detailed analysis of the models obtained for each of the DAOIs of video-10 can be presented as follows, leading to the identification of a number of interesting behavioural patterns of observers when inspecting the video under the given instructions:

It is noted that when inspecting the video-10, the instruction given to all participants is, ***“Find a person wearing a blue top and white pants”***. In this specific video the object concerned, first appears very closely to the camera and disappears from the scene very quickly (the person stayed for just a few seconds) moving from the right side to the left of the scene.

For ease of reference for detailed analysis, Table 5.11 tabulates the coefficients obtained for each DAOI, for each independent variable, for video-10.

	$x1 / FC$	$x2 / VD$	$x3 / VC$	$x4 / TTF$	$C.C$
DAOI-1 (Head area)	N/A	N/A	N/A	N/A	<b>N/A</b>
DAOI-2 (Upper body)	-0.13	1.0	0.13	0	<b>1.0</b>
DAOI-3 (Carried bag)	0	1.0	0	0	<b>1.0</b>
DAOI-4 (Lower body)	0	1.0	0	0	<b>1.0</b>

**Table 5.11: Models for video-10**

The given description for video-10 requires the observers to only look to recognize the ‘person’ being searched for is wearing a ‘blue’ top and a ‘white trouser’. Although the

said individual is wearing a handbag in her shoulder, it is not being requested to inspect this area at all. In the video, the said individual only appears for a short duration, entering the scene from the right and leaving it on the left. The depth of view remains also constant and short distance from the camera. Therefore, whatever the observers do, an attempt will be made to very quickly identify the individual.

A closer look at the coefficients for each regression equation obtained for the different DAOIs it is observed that it has not been possible to model for the DAOI-1, the 'head' region. A closer investigation revealed that only one of the 20 participants ever decided to look at the face and hence the reason that WEKA has declined to create a model. The object concerned is moving fast and is close to the camera that most participants will be very much focused in looking to see whether the person is wearing a blue top and white trouser. It is not required to look at the head area that much as the search does not specify person being looked for is a man/woman.

It is noted that for video-10 for DAOI-2, DAOI-3 and DAOI-4 the obtained correlation coefficients are 1. What this demonstrates is that all participants behave very similarly when observing each of these areas.

The most complicated model (yet with a correlation coefficient of 1) obtained is for the upper body area where the model indicated that FC, VD and VC all play a role in FD. The models obtained for DAOI-3 and DAOI-4 are very simple with FD being directly defined by VD.

Therefore, the above analysis shows that the use of Linear Regression to model human observer's attention has been very successful, but in addition also provides one the opportunity to study observer behaviour in much detail.

## **5.8. Summary and Conclusion**

The experiments conducted in this chapter proved that linear regression, which is a statistical approach can be used to accurately model and analyse observer's attention while conducting a search for identifying and tracking a human with a given description on a given video. The model parameters allow one to analyse observer behaviour in detail, accurately. The additional behavioural analysis that the conceptual and statistical methods did not enable has been made possible by the use of machine learning.

It is noted that more complicated machine learning models such as tree based and ensemble learning algorithms may have been able model human behaviour more accurately. However, as Linear Regression has been capable of doing the intended behavioural analysis to be conducted in this chapter to a high level of accuracy, no attempt was made to consider such approaches in this chapter. In chapter-7 we use such approaches for further analysis.



# Chapter 6.

## Use of Machine Learning Algorithms to Classify Different Groups of Participants based on Eye Gaze Patterns

### 6.1. Introduction

In chapter-4 which described the experiments conducted and the data gathered, it was mentioned that the 20 participants who were the subjects of the visual inspection experiments belonged to different groups, namely 10 each of male/female and 10 each of novice/expert image/video analysts.

Usually a task similar to what is carried out is conducted in practice by expert CCTV operators. Unfortunately, due to the security situation that prevailed in the UK during the period this research was conducted it was not possible to enlist the service of a large number of expert CCTV operators. Instead, therefore a decision was made to have two groups of individuals, the first group having worked in the area of image analysis and processing for a long period (more than two years) and thus having experience in searching for described details in the video shown, with the aim, commitment and focus and the second group that consisted of individuals who have had little or no exposure to image/video analysis or visual processing. These two groups of individuals were predominantly chosen from within the student community of Loughborough University and their close family members. A Further, 10 of the above two groups of individuals were male and the other 10 were female. Table 6.1 details groups the 20 participants according to their sex and level of expertise in the visual image analysis.

	<b>Experts</b>	<b>Novice</b>
<b>Male</b>	5	5
<b>Female</b>	5	5

**Table 6.1: Groupings of participants**

Given the different groups of individuals who participated in the subjective experiments, it will be interesting to determine whether there is any distinguishable difference between the eye gaze distribution patterns or behaviour of the individuals given the instructions and the contents of the relevant videos inspected. In previous literature a number of attempts have been made based on purely statistical approaches to separate experts from non experts while conducting visual inspections, especially in medical image analysis and interpretation [21][97]. Such studies required a large number of participants from each group to enable accurate classification based on statistical approaches. Further, in the classification process a large number of attributes were measured and used. In some of the approaches, whether or not a certain parameter will have sufficient impact on the decision to be made was not investigated a priori and hence this would have also led to unnecessary data, with no discriminative capability be used in the classification attempt.

Knowing the capabilities of the traditional and the latest developments of machine learning algorithms (see Chapter 3) in this chapter, given the challenge we only have a very small number of participants from each group (i.e. 10 each) to consider, we conduct machine learning based classification of the collected datasets.

For clarity of presentation this chapter is divided into separate sections. Apart from this section which is a general introduction, section 6.2 presents the experimental procedure adopted. Section 6.3 provides the detailed experimental results obtained. Section 6.4 follows provides a comprehensive analysis of the results. Finally, section 6.5 concludes the chapter.

## **6.2. Experimental Procedure**

In the experiments conducted in this chapter, no specific experiments or data capture/representation was required. The same dataset captured in chapter 4 was used. Data captured for all of the 13 videos for each of the 20 participants were considered. This is a total of 260 data instances/records. Each instance, comprised of a data label (i.e. whether man/woman or expert/novice. Note: only one of them at a time) and 7 x 4 attributes representing the seven features captured for each of the four DAOIs. Due to the large number of attributes to be used an initial attribute selection procedure was carried out using one of WEKA's implementations of data selection algorithms.

Two different experiments were conducted as follows:

### **1. Classification of the Expert/Novice**

### **2. Classification of Men/Women**

WEKA provides a large number of classification algorithms. In the experiments conducted the widely used Neural Network approach (named by WEKA as Multi Layer Perceptron, MLP), popular tree based approaches RepTree (a single classifier) and Random Forest(RF) and the popular Ensemble Algorithm, i.e. Bagging were used.

### 6.3. Experimental Results

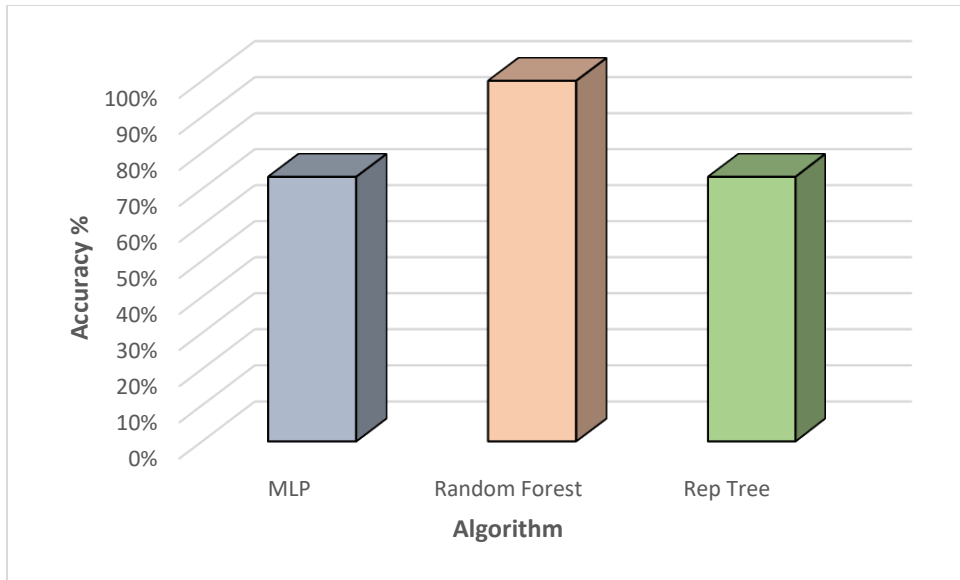
Following are the results obtained from WEKA. All details are provided so that the readers are able to make a clear judgment of the results.

#### 6.3.1 Results for the classification of experts from novice participants:

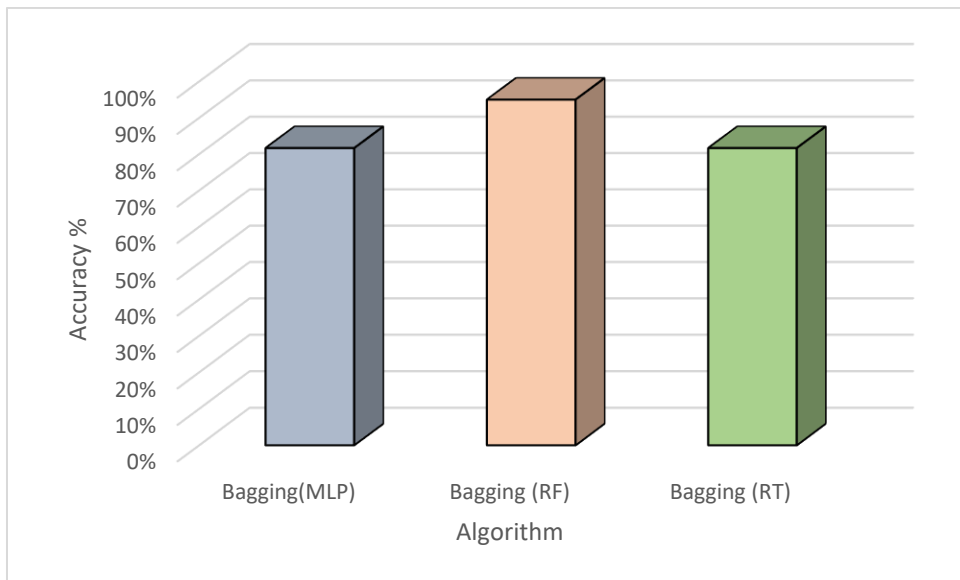
Table 6.2 summarises the results obtained by the various machine learning algorithms tested. Detailed results are presented in Appendix-2

	Accuracy	Precision	Recall	Area under ROC
<b>MLP</b>	73.33%	0.801	0.733	0.85
<b>Bagging(MLP)</b>	81.67%	0.820	0.817	0.911
<b>Random Forest</b>	100%	1.0	1.0	1.0
<b>Bagging (RF)</b>	95%	0.95	0.951	0.991
<b>Rep Tree</b>	73.33%	0.739	0.733	0.759
<b>Bagging (RT)</b>	81.67%	0.82	0.817	0.915

**Table 6.2: Results of classification of Experts vs Novice participants**



(a)



(b)

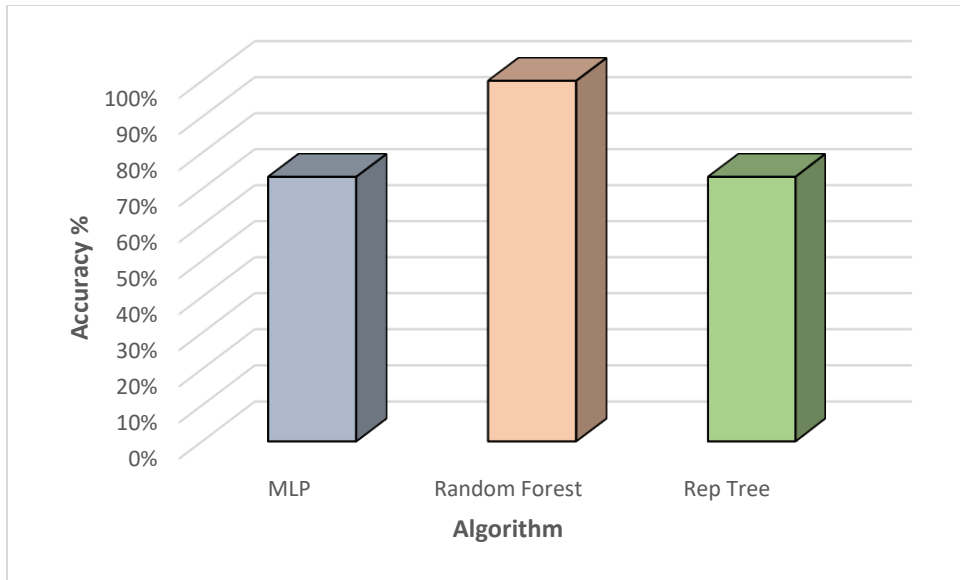
**Figure 6.1: Results of classification of Experts vs Novice participants  
(a) Without bagging (b) with bagging**

### 6.3.2 Results for the classification of Female from Male participants:

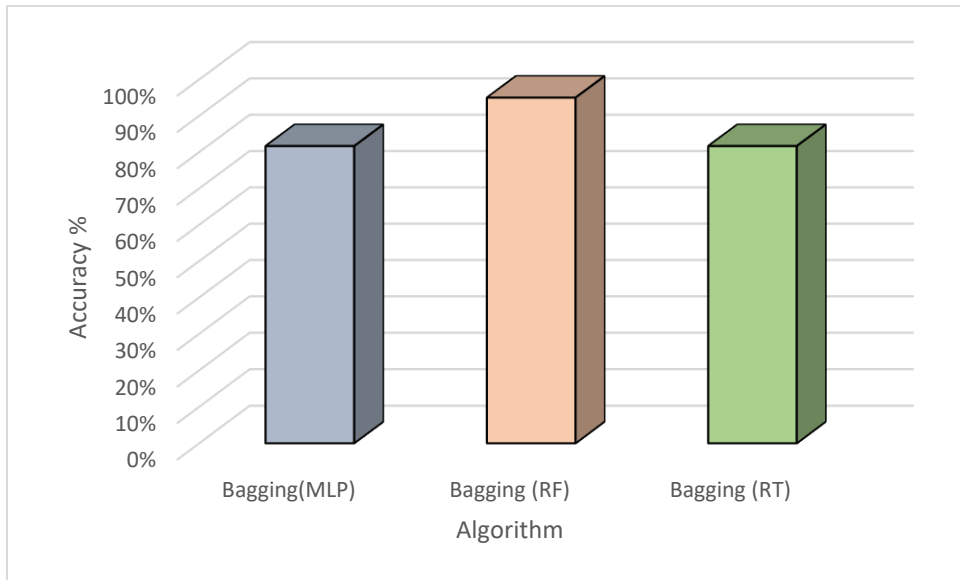
Table 6.3 and figure 6.1 summarise the results obtained by the various machine learning algorithms tested. Detailed results are presented in Appendix-3

	<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>Area under ROC</b>
<b>MLP</b>	50%	0.5	0.5	0.367
<b>Bagging(MLP)</b>	85%	0.854	0.85	0.93
<b>Random Forest</b>	100%	1.0	1.0	1.0
<b>Bagging (RF)</b>	100%	1.0	1.0	1.0
<b>Rep Tree</b>	50%	0.25	0.5	0.5
<b>Bagging (RT)</b>	60%	0.6	0.6	0.7

**Table 6.3: Results of classification of Female vs Male participants**



(a)



(b)

**Figure 6.2: Results of classification of Female vs Male participants  
(a) Without bagging (b) with bagging**

## 6.4. Analysis of Results

As mentioned in **section 6.2**, we examined or evaluate the Experimental data by using three different classifiers MLP, RF and Bagging

The experimental results tabulated in Table 6.2 shows that some classification algorithms are able to very accurately classify the two classes expert vs novice participants. This is particularly true with Random Forest which demonstrates a 100% accuracy. By nature, Random Forest is an Ensemble Learning algorithm where more than one decision tree is combined to make the overall classification and judgement.

Many studies such as [98] concluded that Random Forest is considered as the best performing ensemble classifier. Similarly, our Experimental results also indicate that the performance of Random Forest algorithm is significantly better than the rest.

In addition, the results as illustrated in Table 6.2, show that MLP and Rep Tree have significantly less accuracy result (7.3.3%) when compared with RF. However, using of Bagging with MLP and Rep Tree as base classifiers have increased the accuracy to 8.1.67% and 73.33% respectively, beyond what is capable by the base classifiers when they are used as a single classifier.

It is however noted that when Random Forest is used as the base classifier of Bagging the results become less accurate (81.67%) as compared to using the Random Forest as a single classifier as shown in table. This is due to the fact that Random Forest itself is an Ensemble Learning algorithm.

The Random Forest(RF) algorithm is powerful method performs significantly better than Bagging due to the extra randomness present in the process of building the model [74] [99]. In other words, the fundamental difference between RF and Bagging, that RF splits the node of a tree. The algorithm rather than looking for the best point to split the node among the entire set of variables, it randomly selects sub-features to search for.



Similar observations can be made for the accuracy of classification male vs female participants, with once again Random Forest performing best.

The results as illustrated in Table 6.3, again show that MLP and Rep Tree have significantly less accuracy result (50%) when compared with RF. However, using of Bagging with MLP and Rep Tree as base classifiers have increased the accuracy to 85% and 60% respectively, beyond what is capable by the base classifiers when they are used as a single classifier. It is however noted that when Random Forest is used as the base classifier of Bagging the results remained the same as when used as a single classifier.

It is noted that in both of the above classification tasks that the feature reduction process led to four of the seven attributes to be having a significant impact on the decision made. They are the FC, VD, VC and TFFF. The machine learning process adopted does not highlight which of these variables had the most discriminative power in the two classification tasks. However, combined, these four parameters have resulted in the ability to very accurately classify the two cases using the Random Forest classifier. The results indicate that the experts vs novice participant's and female vs male participant's observation behaviour, when defined by these four parameters together, show discriminative capabilities.

## **6.5. Summary & Conclusions**

This chapter has proposed the use of machine learning to classify the participants into respective groups of experts vs. novice and female vs male. It has been shown that in particular the classifier, Random Forest shows a 100% accuracy level. This concludes that there are inherent behavioural / eye gaze pattern differences between these groups. It has been possible to make use of attribute selection algorithms to determine which of the attributes results in this discriminative behaviour.

It is noted that in previous research, no attempt has been made to use machine learning in conducting the above investigations. Machine learning has not only allowed very accurate analysis in a simple manner, but has also made it possible to carry out the investigations with data gathered from a relatively small number of participants as compared to previous research.

# Chapter 7.

## Conclusion and future work

With the recent advances in GPU based computer processing capabilities and machine learning technology, in particular deep learning, the future of computer based CCTV analytic and forensic applications are now ready for a paradigm shift in technology and performance. With this in mind it is essential to understand better how human observers would carry out a given CCTV forensic task as this may provide some vital clues behind how and why humans are at present better equipped to carry out these tasks than computers. If algorithms that mimic human behaviour and systems that perform like humans can be created, the recent technological revolution may be even capable of outperforming human performance.

Traditionally eye tracking equipment has been used in analysing in detail human observer behaviour via capturing and measuring their eye gaze and movement patterns. Such research has been conducted mostly in the area of medical image analysis and inspection. In **Chapter-2** it was concluded that such research conducted within the area of video analytics is limited to few projects where the entire human object being observed and tracked has been considered as one object. It was argued that this approach is not suitable for the detailed analysis of observer behaviour as mostly within video forensic tasks that observers are given, clear instructions of the appearance of the object to be located and tracked is given and such information may refer to various body parts in different ways and with different levels of significance. The review of literature conducted concluded that all previous attempts in human observer visual attention analysis both within and outside CCTV video analytics and forensics used either statistical or conceptual approaches. It was shown that such approaches are limited in their analysis, in particular their ability to detect hidden patterns and make predictions. Thus, the detailed review of literature concluded that a novel way of breaking down the details of an object being tracked is required and novel approaches to data analysis are also critically needed.

In **Chapter-5**, a novel approach to articulating a human object to be identified and tracked in a visual surveillance task was designed. The novel design included the division of the Region-of-Interest into four parts that refers to the head, shirt and trouser/leg areas of the human and also to a fourth region that refers to any bag/luggage they carry/pull. This breakdown on the area of interest enabled the capture of more detailed eye tracking data that enabled the use of standard statistical analysis methods to perform a more detailed conceptual investigation into human observer behaviour.

Unfortunately, the use of statistical approaches was limited by the fact that more detailed features and patterns were difficult to identify and any predictions potentially possible were impossible to carry out. This led to the conclusion that more advanced data analysis approaches are required for detailed human observer analysis. Furthermore, the use of attribute selection algorithms and linear regression models were proposed in this chapter to replace statistical data analysis techniques which has been used to analysis gathered data. The attribute selection algorithms were used to remove redundant data collected during experiments, prior to creating models for human observer behaviour analysis using machine learning algorithms. Linear Regression was shown to be capable of accurately modelling human observer visual attention.

The work led to a number of significant findings about human observer behaviour analysis that was not possible via the use of traditional approaches, i.e. statistical and conceptual approaches. Detailed analysis of the results obtained indicated the potential behavioural differences between different participant groups, e.g. expert/novices, female/male, when carrying out specific tasks. This observation motivated further research into using more advanced machine learning algorithms for participant classification via the observed/measured visual attention features/attributes.

In **Chapter-6**, the use of advanced machine learning algorithms such as tree based and ensemble algorithms were proposed for participant classification. The research conducted concluded that clear differences exist between the behavioural patterns and such differences can be used to discriminate between the participant groups if appropriate machine learning algorithms are to be used for this purpose. In particular, the superior capability of the Random Forest learning algorithm to classify between expert/novice and female/male participant groups were clearly revealed.

## 7.1. Future Work

The outcomes of the research conducted in this thesis can be used to re-design existing state-of-the-art computer vision algorithms that contributes to the areas of video analytics and forensics. The research conducted in this thesis concluded that the visual attention given to different parts of a human body is different and in particular is very much dependent of the instructions given to the observer about the appearance of the individual to be identified and tracked, i.e. in terms of terminology used within computer vision research, person re-identification. The research outcomes of this thesis propose that a human object's area-of-interest usually consists of four parts, i.e. head, torso, legs and any object being carried/pulled. The research revealed that the 'head' area captures little or no attention in particular if the object is further away from the camera and the face is not clear or when the object's sex is not defined. The detailed analysis both in Chapters 5 and 6 revealed that the most attention is captured by the trouser (leg) area and then followed by the middle area (shirt). However, the presence of high contrast and bright colours could give more prominence to one area as compared to others and sometimes this information can over-ride the given instructions. Given these outcomes in mind in the research conducted by Muna Al-Rahbi [100] a novel approach to the design and implementation of a person re-identification algorithms was proposed. In this approach, a detected human object was divided into two regions, after ignoring the head region. The two regions referred to the middle area (shirt) and bottom area (trouser). Two 'agents' were designed to analyse the features of these areas using computer vision algorithms. Subsequently a multi-agent based system that was based on assigning different probabilities of significance to the upper and lower body areas in the final matched decision making, was proposed. The detailed experiments and evaluations carried out showed that this novel design not only allows one to scale this operation to multi-camera systems but also the system results in improved performance in terms to matching accuracy vital within the forensic application.

The experiments conducted within the research context of this thesis were limited to 20 participants. This could be considered as the minimal number sufficient to justify the results obtained. Given that with such a number of participants, results of significance

accuracy and confidence were obtained, indicates that this number is yet sufficient. However larger number of participants would have resulted in more accurate results from the algorithms adopted.

The experiments conducted in Chapter-6 revealed that it is possible to use machine learning algorithms to discriminate between groups of users, especially expert vs novice. However, in the experiments conducted the so-called experts are those who have had past experience in conducting image and video analysis, looking for details, discrepancies etc. The real experts in CCTV surveillance tasks are trained CCTV operators. Due to the security situation that prevailed during the time period this research was conducted it was not possible to obtain the cooperation from expert CCTV operators. In future, more useful results could be obtained by including a sufficient number of CCTV operators, as experts in the experiments conducted. Despite this fact, the fact that it has been possible to discriminate between the experts and novices users as defined in the experiments conducted within this thesis reveals that if and when real experts are considered as their behaviour is expected to be more discriminative, the classification will be more accurate.

In the recent past, traditional machine learning approaches have been challenged by the advances made in the area of Convolutional Neural Networks (CNN) and deep learning, supported by the recent advances of GPU based computing. In the future, the replacement of machine learning approaches that have been used in this thesis by CNN is proposed.

The research conducted in this thesis has been limited to investigating 13 different videos with 13 different instructions given to identify and track different individuals. The research conducted and presented in this thesis should be considered as a proof of the concept that machine learning can be used if more detailed analysis of human observer behaviour is required. This research can be extended to cover any complicated search on any complicated video, with due changes to the analysis process adopted in this thesis.

## References

- [1] S. F. Al Raisi and E. Edirisinghe, “A Machine Learning Based Approach to Human Observer Behaviour Analysis in CCTV Video Analytics & Forensics,” Proceedings of the 1st International Conference on Internet of Things and Machine Learning. ACM, 2017.
- [2] M. Huang, “Eye-Tracking Technology in Human-Computer Interaction,” 2012.
- [3] M. Obrist, R. Bernhaupt, E. Beck, and M. Tscheligi, “Focusing on elderly: an iTV usability evaluation study with eye-tracking,” *Interact. TV a Shar. Exp.*, vol. 4471, pp. 66–75, 2007.
- [4] S. Evans, J. Techdis, S. Minocha, and T. Open, “Accessibility , usability and safety of online environments : the implications for designing e-learning for older people,” pp. 1–3, 2013.
- [5] M. Horsley, M. Eliot, B. A. Knight, and R. Reilly, Current trends in eye tracking research. 2014.
- [6] A. Niedźwiecka, S. Ramotowska, and P. Tomalski, “Mutual Gaze During Early Mother-Infant Interactions Promotes Attention Control Development,” *Child Dev.*, vol. 00, no. 0, pp. 1–15, 2017.
- [7] O. Toubia, M. G. de Jong, D. Stieger, and J. Fuller, “Measuring Consumer Preferences Using Conjoint Poker,” *Mark. Sci.*, vol. 31, no. 1, pp. 138–156, 2012.
- [8] J. Zhang, M. Wedel, and R. Pieters, “Sales effects of attention to feature advertisements: a Bayesian mediation analysis,” *J. Mark. Res.*, vol. 46, no. 5, pp. 669–681, 2009.
- [9] P. Huddleston, B. K. Behe, S. Minahan, and R. T. Fernandez, “Seeking attention: an eye tracking study of in-store merchandise displays,” *Int. J. Retail Distrib. Manag.*, vol. 43, no. 6, pp. 561–574, 2015.
- [10] L. Roy, I. Vivier, and J., CuSToMER CaSE TV sports. 2008.



- [11] G. Anders, "Pilot's Attention Allocation during Approach and Landing -- Eye- and Head-Tracking Research in an A330 Full Flight Simulator," 11th Int. Symp. Aviat. Psychol., 2001.
- [12] B. Was, Christopher, Sansosti, Frank, Morris, *Eye-Tracking Technology Applications in Educational Research*. IGI Global, 2016.
- [13] A. Shareghi Najar, A. Mitrovic and N. Kouros, "Eye Tracking and Studying Examples: How Novices and Advanced Learners Study Sql Examples," *J. Comput. Inf. Technol. - CIT*, vol. 23, pp. 113–128, 2015.
- [14] T. Yamamoto and K. Imai-Matsumura, "Teachers' gaze and awareness of students' behavior: Using an eye tracker," *Innov. Teach.*, vol. 2, no. 6, pp. 1–7, 2013.
- [15] F. Y. Yang, C. Y. Chang, W. R. Chien, Y. T. Chien, and Y. H. Tseng, "Tracking learners' visual attention during a multimedia presentation in a real classroom," *Comput. Educ.*, vol. 62, pp. 208–220, 2013.
- [16] I. El Hadduoi and Mohamed.K, "Learner Behavior Analysis on an Online Learning Platform.," *Int. J. Emerg. Technol. Learn.*, vol. 7, no. 2, pp. 22–25, 2012.
- [17] S. Al-showarah, N. Al-jawad, and H. Sellahewa, "Effects of User Age on Smartphone and Tablet Use , Measured with an Eye-tracker via Fixation Duration , Scan-Path Duration , and Saccades Proportion."
- [18] K. Harezlak and P. Kasproski, "Application of eye tracking in medicine: A survey, research issues and challenges," *Comput. Med. Imaging Graph.*, 2017.
- [19] L. Rello and M. Ballesteros, "Detecting readers with dyslexia using machine learning with eye tracking measures," in *Proceedings of the 12th Web for All Conference on - Web for all* ,ACM, pp. 1–8,2015.
- [20] L. Dong, Y. Chen, A. G. Gale, B. Rees, and C. Maxwell-Armstrong, "Visual search behaviour during laparoscopic cadaveric procedures," *Prog. Biomed. Opt. Imaging - Proc. SPIE*, vol. 9037, p. 903719, 2014.
- [21] S. Voisin, F. Pinto, G. Morin-Ducote, K. B. Hudson, and G. D. Tourassi, "Predicting diagnostic error in radiology via eye-tracking and image analytics: preliminary investigation in mammography.," *Med. Phys.*, vol. 40, no. 10, p. 101906, 2013.

- [22] C. S. Yu, E. M. Y. Wang, W. C. Li, and G. Braithwaite, “Pilots’ visual scan patterns and situation awareness in flight operations,” *Aviat. Sp. Environ. Med.*, vol. 85, no. 7, pp. 708–714, 2014.
- [23] S. Almeida, A. Veloso, L. Roque, and Ó. Mealha, “The Eyes and Games : A Survey of Visual Attention and Eye Tracking Input in Video Games,” *Proc. SBGames*, pp. 1–10, 2011.
- [24] G. Roffo, M. Cristani, F. Pollick, C. Segalin, and V. Murino, “Statistical analysis of visual attentional patterns for video surveillance,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8259 LNCS, no. PART 2, pp. 520–527, 2013.
- [25] C. J. Howard, T. Troscianko, I. D. Gilchrist, A. Behera, and D. C. Hogg, “Suspiciousness perception in dynamic scenes: a comparison of CCTV operators and novices,” *Front. Hum. Neurosci.*, vol. 7, no. JUL, p. 441, 2013.
- [26] N. Robertson, I. Reid, and M. Brady, “Automatic Human Behaviour Recognition and Explanation for CCTV Video Surveillance,” *Secur. J.*, vol. 21, no. 3, pp. 173–188, 2008.
- [27] G. Graham, “The ( Change ) Blindingly Obvious : Investigating Fixation Behaviour and Memory Recall during CCTV Observation,” University of Portsmouth, PhD Thesis, 2016.
- [28] I. H. Witten, E. Frank, and M. a Hall, *Data Mining: Practical Machine Learning Tools and Techniques (Google eBook)*, 3rd ed. 2011.
- [29] S. Amershi and C. Conati, “Unsupervised and supervised machine learning in user modeling for intelligent learning environments,” *Int. Conference. Intell. User Interfaces*, pp. 72–81, 2007.
- [30] A. Lavecchia, “Machine-learning approaches in drug discovery: Methods and applications,” *Drug Discov. Today*, vol. 20, no. 3, pp. 318–331, 2015.
- [31] A. Jain, “Machine Learning Techniques for Medical Diagnosis : a Review,” 2nd Int. Conference on Science Technology and Management, pp. 2449–2459, 2015.
- [32] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, “Machine learning applications in cancer prognosis and prediction,” *Comput. Struct. Biotechnol. J.*, vol. 13, pp. 8–17, 2015.

- [33] D. Bone, M. S. Goodwin, M. P. Black, C. Lee, S. Narayanan, L. Angeles, I. Science, and K. Fu, "Pitfalls and promises," *J. Autism Dev. Disord.*, vol. 45, no. 101, pp. 1121–1136, 2016.
- [34] C. Perlich, B. Dalessandro, T. Raeder, O. Stitelman, and F. Provost, "Machine learning for targeted display advertising: Transfer learning in action," *Mach. Learn.*, vol. 95, no. 1, pp. 103–127, 2014.
- [35] J. Osaku, A. Asada, F. Maeda, Y. Yamagata, and T. Kanamaru, "Implementation of machine learning algorithm to autonomous surface vehicle for tracking and navigating AUV," 2013 IEEE Int. Underw. Technol. Symp. UT 2013, pp. 0–3, 2013.
- [36] A. Bouzalmat and J. Kharroubi, "Face Recognition Using Neural Network Based Fourier Gabor Filters & Random Projection," no. 5, pp. 376–386, 2014.
- [37] M. Grega, A. Matiolaski, P. Guzik, and M. Leszczuk, "Automated detection of firearms and knives in a CCTV image," *Sensors (Switzerland)*, vol. 16, no. 1, 2016.
- [38] S. Sathyadevan, A. K. Balakrishnan, S. Arya, and S. Athira Raghunath, "Identifying moving bodies from CCTV videos using machine learning techniques," in 1st International Conference on Networks and Soft Computing, ICNSC 2014 - Proceedings, 2014.
- [39] U. Vural and Y. S. Akgul, "A Machine Learning System For Human-in-the-loop Video Surveillance," *Int. Conf. Pattern Recognit.*, no. Icpr, pp. 1092–1095, 2012.
- [40] C. Rudin, "Predictive policing: Using machine learning to detect patterns of crime," *Wired*, pp. 1–5, 2013.
- [41] R. Blake and R. Sekuler, *Perception*, 5th ed. New York: McGraw-Hill, 2005.
- [42] D. Hubel, "The eye," *Eye Brain Vis.*, pp. 1–22, 1995.
- [43] Owlcation, "Perception Psychology - How We Understand Our World," *Psychology*. [Online]. Available: <https://owlcation.com/social-sciences/Perception-in-Psychology>. [Accessed: 29-May-2016].

- [44] D. T. Willingham, "Cognition, The Thinking Animal," [www.MySearchLab.com](http://www.MySearchLab.com) [online], vol. 2, p. 613, 2003.
- [45] E. B. Goldstein, *Sensation and Perception*. 2009. [online] <http://zhenilo.narod.ru/main/students/Goldstein.pdf>
- [46] I. L. Nilsson and W. V. Lindberg, *Visual Perception: New Research*, 1 edition. Inc, Nova Science Pub, 2008.
- [47] R. Szeliski, "Computer Vision : Algorithms and Applications," *Computer (Long. Beach. Calif)*., vol. 5, p. 832, 2010.
- [48] W. J. Scheirer, S. E. Anthony, K. Nakayama, and D. D. Cox, "Perceptual annotation: Measuring human vision to improve computer vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1679–1686, 2014.
- [49] R. J. K. Jacob, "The use of eye movements in human-computer interaction techniques: what you look at is what you get," *ACM Trans. Inf. Syst.*, 1991.
- [50] J. E. Hoffman, "Visual attention and eye movements," *Attention*. pp. 119–153, 1998.
- [51] K. Rayner, "Eye movements in reading and information processing: 20 years of research," *Psychol. Bull.*, vol. 124, no. 3, pp. 372–422, 1998.
- [52] D. Maltoni and A. K. Jain, *Biometric Authentication*. 2004.
- [53] H. Collewijn, *Vision research: A practical Guide to Laboratory Methods*. Oxford University Press, 1999.
- [54] Tobii Technology, "Tobii Eye Tracker." [Online]. Available: <https://www.tobiipro.com/product-listing/tobii-t60-and-t120/>. [Accessed: 13-Feb-2015].
- [55] Tobii Dynavox, "How eye tracking works." [Online]. Available: <https://www.tobiidynavox.com/about/about-us/how-eye-tracking-works/>. [Accessed: 01-Sep-2016].
- [56] Tobii Studio, "comprehensive Eye Tracking analysis & visualization software." [Online]. <https://www.tobiipro.com/product-listing/tobii-pro-studio/>

- [57] D. Noton and L. Stark, "Scanpaths in saccadic eye movements while viewing and recognizing patterns," *Vision Res.*, vol. 11, no. 9, pp. 929–942, 1971.
- [58] C. Holland and O. V. Komogortsev, "Biometric identification via eye movement scanpaths in reading," 2011 Int. Jt. Conf. Biometrics, IJCB 2011, pp. 1–8, 2011.
- [59] O. V. Komogortsev, A. Karpov, C. D. Holland, and H. P. Proenca, "Multimodal ocular biometrics approach: A feasibility study," 2012 IEEE 5th Int. Conf. Biometrics Theory, Appl. Syst. BTAS 2012, no. Btas, pp. 209–216, 2012.
- [60] P. Kasprowski and J. Ober, "Eye Movements in Biometrics," *Biometric Authentication*, no. August, pp. 248–258, 2004.
- [61] O. V. Komogortsev, U. K. S. Jayarathna, C. R. Aragon, and M. Mechehoul, "Biometric Identification via an Oculomotor Plant Mathematical Model," *Etra*, pp. 57–60, 2010.
- [62] T. Kinnunen, F. Sedlak, and R. Bednarik, "Towards task-independent person authentication using eye movement signals," *Proc. 2010 Symp. Eye-Tracking Res. Appl. - ETRA '10*, vol. 1, no. 212, p. 187, 2010.
- [63] R. S. Hessels, C. Kemner, C. van den Boomen, and I. T. C. Hooge, "The area-of-interest problem in eyetracking research: A noise-robust solution for face and sparse stimuli.," *Behav. Res. Methods*, pp. 1–19, 2015.
- [64] IBM, "IBM SPSS - IBM Analytics," IBM SPSS Software, 2016. [Online]. Available: <http://www.ibm.com/analytics/us/en/technology/spss/>. [Accessed: 27-Jan-2015].
- [65] "Statistics Help For Students." [Online]. Available: [http://statistics-help-for-students.com/How\\_do\\_I\\_report\\_a\\_1\\_way\\_between\\_subjects\\_ANOVA\\_in\\_APA\\_style.htm#.WYxWG1WGNFF](http://statistics-help-for-students.com/How_do_I_report_a_1_way_between_subjects_ANOVA_in_APA_style.htm#.WYxWG1WGNFF). [Accessed: 10-Aug-2017].
- [66] G. Band, "Herzberg Two Factor Theory among the Management Faculty in Nagpur City," pp. 13–21, 2016.
- [67] K. P. Murphy, "Machine Learning: A Probabilistic Perspective," MIT Press, p. 25, 2012.

- [68] A. L. Samuel, "Some Studies in Machine Learning Using the Game of Checkers," *IBM J. Res. Dev.*, vol. 44, no. 1.2, pp. 206–226, 1959.
- [69] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*, vol. 49. 2001.
- [70] E. Frank, M. A. Hall, and I. H. Witten, *Data Mining Practical Machine Learning Tools and Techniques*. 2016.
- [71] "MathsIsFun," *Correlation*, 2016. [Online]. Available: <http://www.mathsisfun.com/data/correlation.html>. [Accessed: 22-Jun-2017].
- [72] D. Borman, "Introduction to Statistic Web Book," *Pearson's r Correlation And Regression*. [Online]. Available: [http://www.derekborman.com/230\\_web\\_book/module4/correlation/index.html](http://www.derekborman.com/230_web_book/module4/correlation/index.html). [Accessed: 11-May-2017].
- [73] Wikipedia, "Statistical classification," *machine learning classification*. [Online]. Available: [https://en.wikipedia.org/wiki/Statistical\\_classification](https://en.wikipedia.org/wiki/Statistical_classification). [Accessed: 20-Apr-2016].
- [74] A. Liaw and M. Wiener, "Classification and Regression by randomForest," *R news*, vol. 2, no. December, pp. 18–22, 2002.
- [75] S.K. Jayanthi and S. Sasikala, "Reptree Classifier for Identifying Link Spam in Web Search Engines," *ICTACT J. Soft Comput.*, vol. 03, no. 02, pp. 498–505, 2013.
- [76] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996.
- [77] FastML, "Machine learning made easy: Intro to random forests." [Online]. Available: <http://fastml.com/intro-to-random-forests/>. [Accessed: 18-Aug-2017].
- [78] "WEKA; the University of Waikato," *Weka 3 - Data Mining with Open Source Machine Learning Software in Java*. [Online]. Available: <http://www.cs.waikato.ac.nz/ml/weka/index.html>. [Accessed: 27-Feb-2016].

- [79] B. Ratner, *Statistical and Machine-Learning Data Mining: Techniques for Better Predictive Modeling and Analysis of Big Data*, Third Edit. CRC Press, 2017.
- [80] J. G. Cromley and T. W. Wills, “Flexible strategy use by students who learn much versus little from text: Transitions within think-aloud protocols,” *J. Res. Read.*, vol. 39, no. 1, pp. 50–71, 2016.
- [81] A. Whitehead, “The use of Think Aloud Protocol to Investigate Golfers Decision Making Processes by,” no. June, 2015.
- [82] N. Ram and P. McCullagh, “Self-Modeling: Influence on Psychological Responses and Physical Performance,” *J. Appl. Sport Psychol.*, vol. 17, no. 2, pp. 220–241, 2011.
- [83] F. M. Weaver and John Carroll, “Crime Perceptions in a Natural Setting by Expert and Novice Shoplifters,” vol. *Social Psy*, pp. 48, 349–359., 1985.
- [84] Tobii Technology, “Retrospective Think Aloud and Eye Tracking Comparing the value of different cues when using the retrospective think aloud method in web usability test ing,” 2009.
- [85] A. Hyrskykari, S. Ovaska, P. Majaranta, K.-J. Rähkä, and M. Lehtinen, “Gaze Path Stimulation in Retrospective Think-Aloud,” *J. Eye Mov. Res.*, vol. 2, no. 4, pp. 1–18, 2008.
- [86] H. B. Zaman, M. H. M. Saad, M. A. Saghafi, and A. Hussain, “Review of person re-identification techniques,” *IET Computer Vision*, vol. 8, no. 6, pp. 455–474, 2014.
- [87] M. E. Irhebhude, “Object detection , recognition and re-identification in video footage,” PhD Thesis, Department of Computer Science, Loughborough University, UK, 2015.
- [88] C. J. Howard, I. D. Gilchrist, T. Troscianko, A. Behera, and D. C. Hogg, “Task relevance predicts gaze in videos of real moving scenes,” *Exp. Brain Res.*, vol. 214, no. 1, pp. 131–137, 2011.
- [89] M. S. Castelhana, M. L. Mack, and J. M. Henderson, “Viewing task influences eye movement control during active scene perception,” *J. Vis.*, vol. 9, no. 6, pp. 1–15, 2009.

- [90] Tobii Technology, “Tobii Studio Software.” [Online]. Available: <https://www.tobii.com/learn-and-support/learn/steps-in-an-eye-tracking-study/setup/installing-tobii-studio/>. [Accessed: 02-Jan-2016].
- [91] D. W. Hansen and Q. Ji, “In the eye of the beholder: a survey of models for eyes and gaze,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 478–500, 2010.
- [92] Tobii Technology, “Tobii Studio 3.2 User Manual,” p. 154, 2012.
- [93] H. Kenneth, N. Marcus, A. Richard, D. Richard, J. Halszka, and W. J. van De, *Eye Tracking: A Comprehensive Guide To Methods And Measures*. OUP Oxford, 2011, 2011.
- [94] A. Poole and L. J. Ball, “Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects,” *Encycl. Human-Computer Interact.*, pp. 211–219, 2005.
- [95] D. Cyr and M. Head, “The impact of task framing and viewing timing on user website perceptions and viewing behavior,” *Int. J. Hum. Comput. Stud.*, 2013.
- [96] D. Cyr, M. Head, H. Larios, and B. Pan, “Exploring Human Images in Website Design :A Multi-Method Approach ,” vol. 33, no. 3, pp. 1–32, 2009.
- [97] R. Nakashima, K. Kobayashi, E. Maeda, T. Yoshikawa, and K. Yokosawa, “Visual search of experts in medical image reading: The effect of training, target prevalence, and expert knowledge,” *Front. Psychol.*, vol. 4, no. APR, pp. 1–8, 2013.
- [98] P. Panov and S. Džeroski, “Combining Bagging and Random Subspaces to Create Better Ensembles,” in *International Symposium on Intelligent Data Analysis*, 2007, pp. 118–129.
- [99] A. Liaw and M. Wiener, “Classification and Regression by randomForest,” *R news*, vol. 2, no. 3, pp. 18–22, 2002.
- [99] L. Breiman, “Random Forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001.



- [100] M.S. Al-Rahbi, “Agent-based framework for person re-identification” PhD Thesis, Department of Computer Science, Loughborough University, UK, 2017.

## **Appendix 1**

**Results obtained by the Statistical Analysis of Visual Attentional Patterns for CCTV Footage**

## Video 2

**Description: Find a man wearing white shirt carrying black laptop bag.**

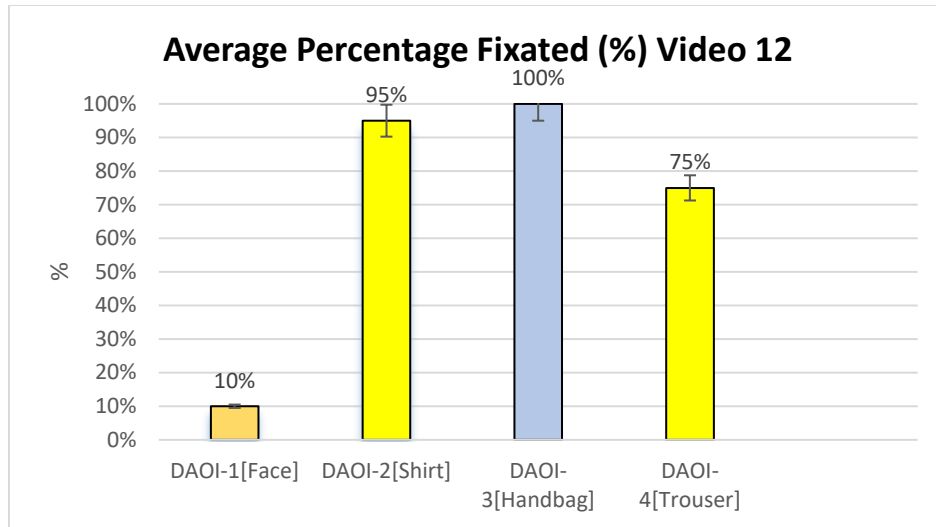
For this video footage, we identified or designed four areas of interest upper part-head (DAOI 1), middle part-shirt (DAOI 2), laptop bag (DAOI 3) and lower part-pants (DAOI 4), as shown in the in the following figure below.



**Figure 1: Four DAOIs**

### **Percentage Fixated**

On average, the DAOI 3 (laptop bag) had the highest percentage fixated (100%) then DAOI 2 (shirt) 95%. While the upper part (head) had the lowest percentage fixated (10%). Which means the most appealing part were DAOI 3 and less appealing part was DAOI1.

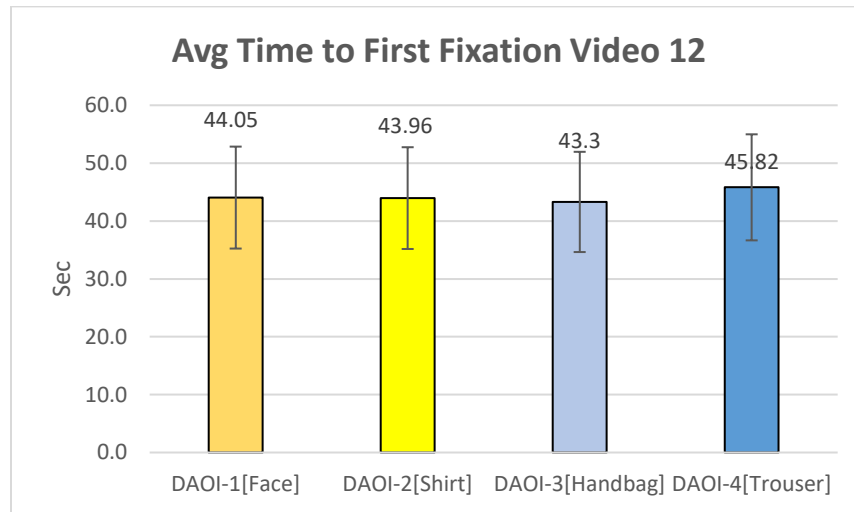


**Figure 2: Bar Graph for Percentage Fixated (%)**

According to the result most people have a tendency of focusing on the middle section (DAOI 2) as compared to other parts. There was a statistically significant difference on the percentage fixated between groups at  $p\text{-value} < 0.05$  as determined by one-way ANOVA ( $F(3,76) = 40.14, p = 0.000$ ). There is sufficient empirical evidence that most people fixated on the middle part while the lowest proportion of people are fixed on the DAOI 1 (face). The Tukey HSD test indicated that mean score for the DAOI 1 ( $M=1.00, SD=.31$ ) was significantly different than the AOI 2 ( $M=.95, SD=.22$ ), AOI 3 ( $M=1.00, SD=0.00$ ) and AOI 4 ( $M=.75, SD=.44$ ). Also, the DAOI 4 was significantly differ from AOI3.

### **Time to First Fixation**

On average, the study participants took the longest time for the first fixation on the lower part (DAOI 4) and the short time for the first fixation on the lower part (DAOI 3). Figure 3 reflects the same findings as summarized by the statistics.

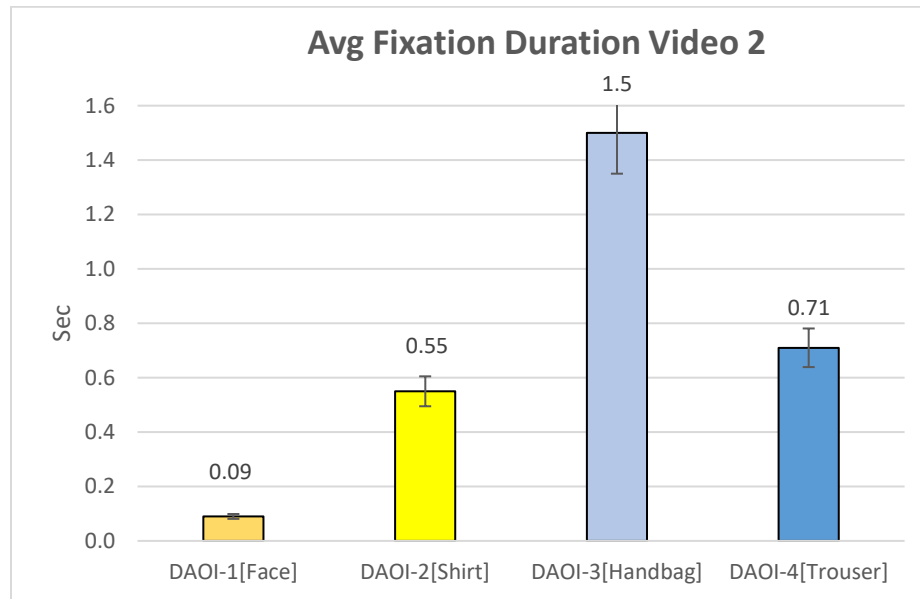


**Figure 3: Bar Graph for Time to First Fixation**

There was a statistically significant difference in the Time to first Fixation of the four areas at p-value < 0.05 as determined by one-way ANOVA [ $F(3, 52) = 14.7, p = 0.00$ ]. There is sufficient empirical evidence that the DAOI 3 has the lowest time to first fixation while the lower part (DAOI 4) has the shortest time to first fixation. A Post hoc comparison using the Tukey HSD test indicated that mean score for the AOI4 (M=45.82, SD=1.18) was significantly different than the AOI 2 (M=43.96, SD=1.44) and DAOI 3 (M=43.3, SD=0.71). However, DAOI 1 (M=44.1, SD=0.61) did not significantly differ from DAOI 2 (M=43.96, SD=1.44), DAOI 3 (M=43.62, SD=0.71) and DAOI 4 (M=45.82, SD=1.18). In addition to that, DAOI 3 did not significantly differ from the DAOI 2.

## Fixation duration

On average, the study participants observed the laptop bag (DAOI 3) for the longest duration while the upper part-head (DAOI 1) was observed for the shortest duration.

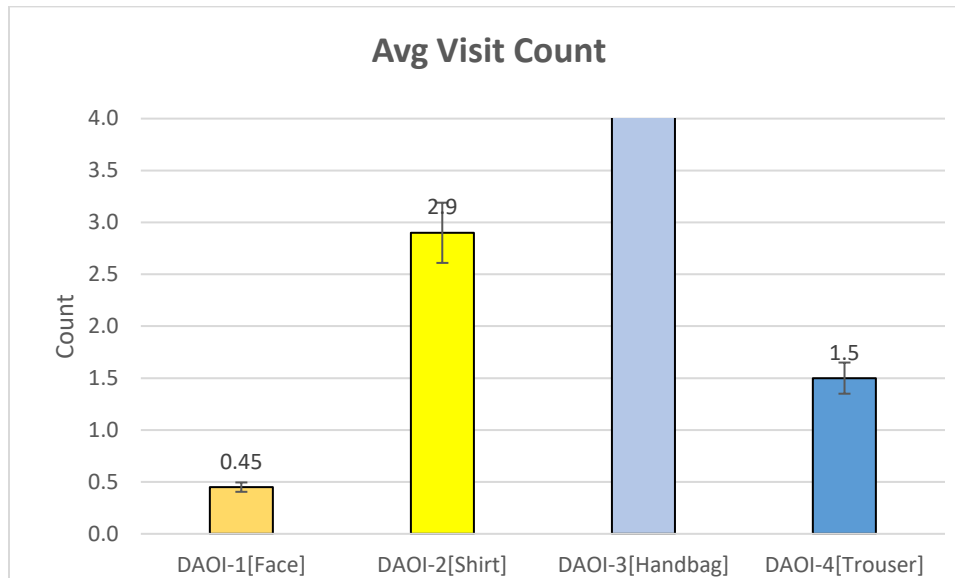


**Figure 4: Bar Graph for Duration of Fixation**

According to the result people spend the highest amount of time observing the DAOI 3. There was a statistically significant difference on total fixation duration for the four areas at  $p\text{-value} < 0.05$  as determined by one-way ANOVA [ $F(3, 76) = 11.37, p = 0.00$ ]. There is sufficient empirical evidence that the DAOI 3 has the long fixation duration while the upper part has the shortest observation or fixation length. The Tukey HSD test indicated that mean score for the DAOI4 ( $M=0.71, SD=0.78$ ) was significantly different than the DAOI 2 ( $M=0.55, SD=0.52$ ) and DAOI3 ( $M=1.5, SD=1.16$ ). However, the DAOI1 ( $M=0.0880, SD=0.36$ ) did not significantly differ from the DAOI 2, DAOI 3 and DAOI 4.

## **Total Visit Counts**

On average, the study participants visited the DAOI 3 (laptop bag) for the highest number of times while the upper part-head was visited for the least number of times. Figure 5 reflects the same findings as summarized by the statistics.



**Figure 5: Bar Graph for Visit Counts**

There was a statistically significant difference in the visit counts for the three areas at  $p$ -value  $< 0.05$  as determined by one-way ANOVA [ $F(3, 76) = 38.5, p = 0.00$ ]. There is sufficient empirical evidence that the DAOI 3 has the higher visit counts while the upper part has the least visit counts. The Tukey HSD test indicated that mean score for the DAOI 3 ( $M=10.10, SD=5.6$ ) was significantly different than the DAOI 1 ( $M=.45, SD=1.61$ ), DAOI2 ( $M=2.90, SD=1.94$ ) and DAOI4 ( $M=10.10, SD=5.6$ ). However, the DAOI1 did not significantly differ from the DAOI 2 and DAOI 4. Also, DAOI 4 did not significantly differ from the DAOI 2.

### Time to 1st fixation

For this we calculate the percentage of people that have seen the DAOI in a certain order as shown the following table. As shown in the figure 3 the majority of participant (75%) had seen bag (DAOI 3) first as shown in the table below.

Video 2								
Sample	DAOI Time to First Fixation				DAOI view order			
	DAOI 1	DAOI 2	DAOI 3	DAOI 4	1	2	3	4
G1-F01		43.25	42.95			2	1	
G1-F02		42.8	43			1	2	
G1-F03		43.91	42.79	45.67		2	1	3
G1-F04	44.4	43.24	42.9	45.15	3	2	1	4
G1-F05			43.2	48.41			1	2
G1-M01			42.73	44.41			1	2
G1-M02		42.97	42.96			2	1	
G1-M03		43.37	42.8	45.5		2	1	3
G1-M04			44.4				1	
G1-M05		43.48	42.78	44.96		2	1	3
G2-F01	47.16	43.33	42.73	47.91	3	2	1	4
G2-F02	43.6	42.73	42.86		3	1	2	
G2-F03		42.82	43.19	46.65		1	2	3
G2-F04			42.81	45.96			1	2
G2-F05		43.08	42.93	44.96		2	1	3
G2-M01		45.23	42.8	46.59		2	1	3
G2-M02			42.92	44.16			1	2
G2-M03		42.94	42.71	44.74		2	1	3
G2-M04		43.2	43.73	45.3		1	2	3
G2-M05		43.08	43.64			1	2	

**Table 1: Analysis of Order of Fixation for video 2**



*Video 3*

**Description: Find man wearing red half sleeve t-shirt and pants**

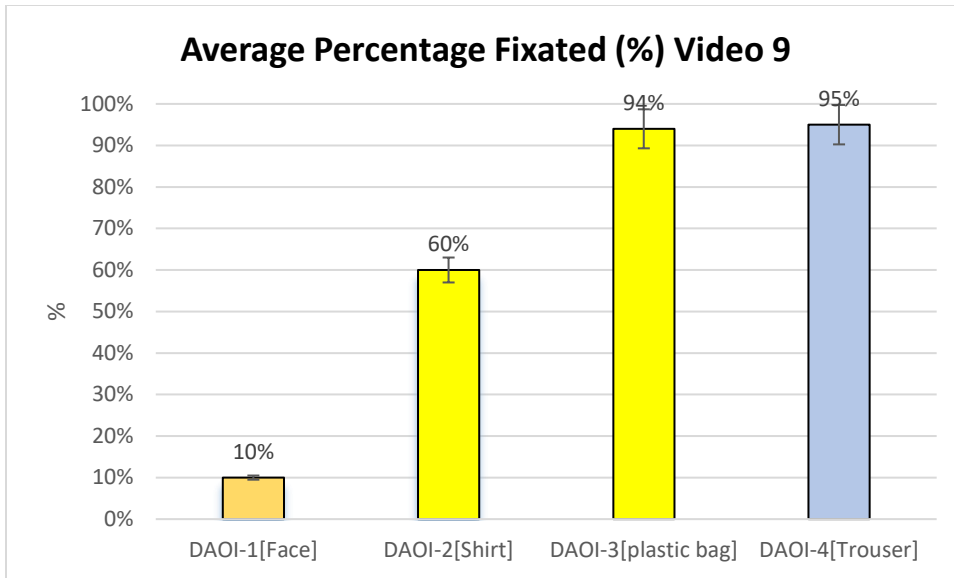
For this video footage (video 3), we identified four areas of interest upper part-Face (DAOI 1), middle part-shirt (DAOI 2), Plastic bag (DAOI 3) and lower part-pants (DAOI 4), as shown in the in the following figure below.



**Figure 6: Four DAOIs**

**Percentage Fixated**

On average, the DAOI 2 (target shirt) had the highest percentage fixated (100%). While target face (DAOI 1) had, the lowest percentage fixated just (35%). Figure 7 reflects the same findings as summarized by the statistics.



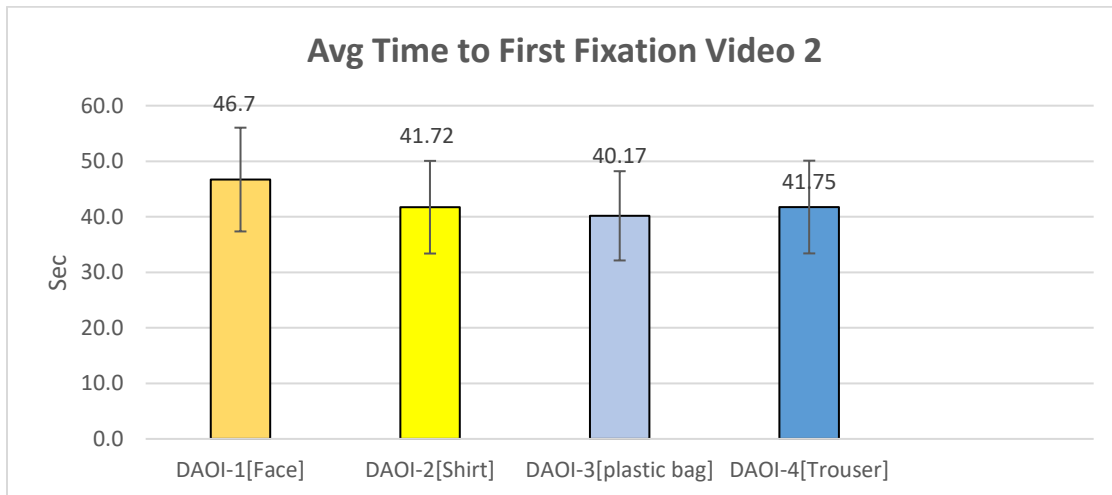
**Figure 7: Bar Graph for Percentage Fixated (%)**

According to the result most people have a tendency of focusing on the DAOI 2. There was a statistically significant difference on the percentage fixated between groups at  $p$ -value  $< 0.05$  as determined by one-way ANOVA [ $F(4,95) = 14.014, P=0$ ]. There is sufficient empirical evidence that most people fixated on the DAOI 1 while the lowest proportion of people are fixed on the DAOI 1

. The Tukey HSD test indicated that mean score for the DAOI 1 ( $M=0.35, SD=0.48936$ ) was significantly different than the DAOI2 ( $M=1, SD=0$ ), DAOI3 and ( $M=0.85, SD=0.36635$ ). Also, DAOI 4 ( $M=0.5, SD=0.51299$ ) was significantly different than the DAOI 2, DAOI 3 and DA. However, the DAOI 2 and DAOI 4 did not significantly differ from the DAOI 1 and DAOI 3 respectively.

## **Time to First Fixation**

On average, the study participants took the longest time for the first fixation on the upper part -face (DAOI 1). Figure 8 reflects the same findings as summarized by the statistics.



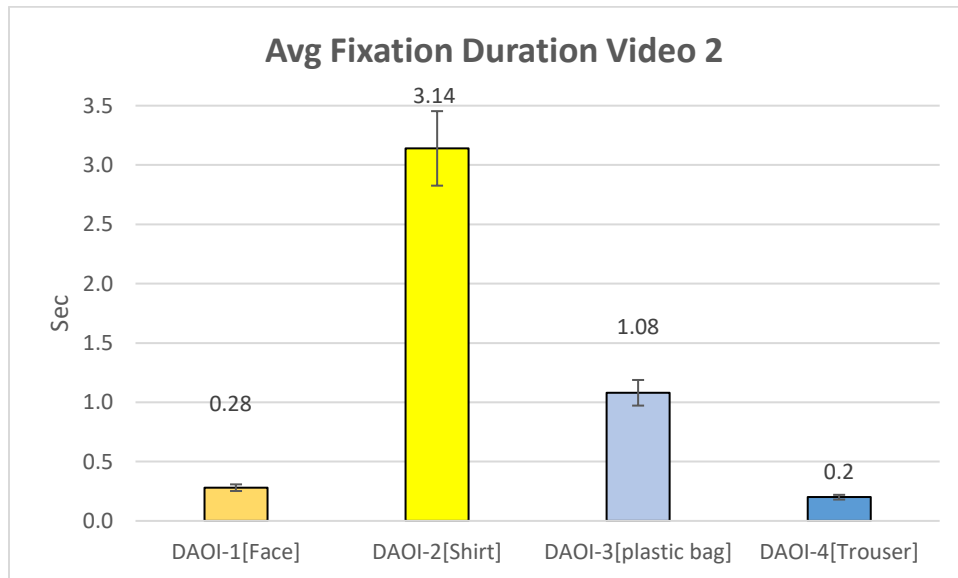
**Figure 8: Bar Graph for Time to First Fixation**

There was a statistically significant difference in the Time to first Fixation of the five areas at  $p\text{-value} < 0.05$  as determined by one-way ANOVA [ $F(4,69) = 3.607, P=0$ ]. There is sufficient empirical evidence that the DAOI 1 has the highest time to first fixation.

A Post hoc comparison using the Tukey HSD test indicated that mean score for the DAOI1 ( $M=46.6629, SD=4.73896$ ) was significantly different than the DAOI 2( $M=41.724, SD=3.93124$ ) and DAOI 3( $M=40.1659, SD=3.16318$ ). However, there was no statistically significant difference between the DAOI 4 ( $M=41.752, SD=4.48848$ ) and DAOI 1, DAOI 2 and DAOI 3. Also, DAOI 3 did not significantly differ from the DAOI 2.

## **Fixation duration**

On average, the study participants observed the DAOI 2 for the longest duration while the lower part (DAOI 4) was observed for the shortest duration.

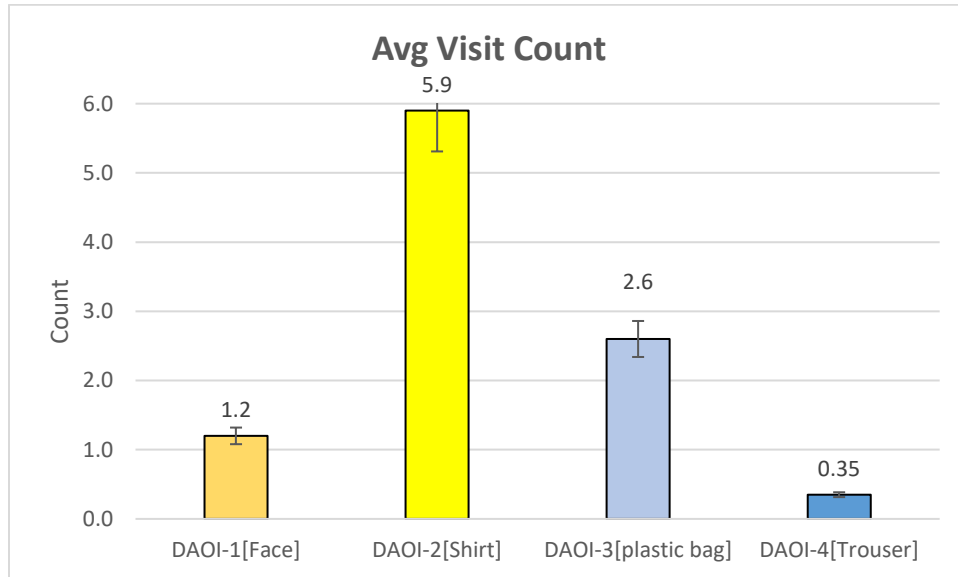


**Figure 9: Bar Graph for Duration of Fixation**

According to the result, people spend the highest amount of time observing the DAOI 2. There was a statistically significant difference in total fixation duration in the five areas of  $p\text{-value} < 0.05$  as determined by one-way ANOVA [ $F(4, 95) = 29.2, p = 0.00$ ]. Therefore, there are statistically significant differences across the five areas of interests. There is sufficient empirical evidence that the DAOI 2 part has a longer fixation duration while the DAOI 4 has the shortest observation or fixation length. The Tukey HSD test indicated that the mean score for the DAOI 1 ( $M=0.63, SD=0.19$ ) was significantly different than the DAOI 2 ( $M=0.63, SD=0.19$ ) and DAOI5 ( $M=0.63, SD=0.19$ ). However, the DAOI 1 did not significantly differ from the DAOI 3 ( $M=0.63, SD=0.19$ ) and DAOI 4 ( $M=0.63, SD=0.19$ ).

## Total Visit Counts

On average, the study participants visited the DAOI 2 for the highest number of times while lower part was visited for the least number of times. Figure 10 reflects the same



**Figure 10: Bar Graph for Visit Counts**

There was a statistically significant difference in the visit counts for the five areas at  $p$ -value  $< 0.05$  as determined by one-way ANOVA [ $F(4, 95) = 3.4, p = 0.00$ ]. There is sufficient empirical evidence that the DAOI 2 has the higher visit counts while the DAOI 4 has the least visit counts. The Tukey HSD test indicated that the mean score for the DAOI 4 ( $M=0.63, SD=0.19$ ) was significantly different than the DAOI 2 ( $M=5.9, SD=0.19$ ). However, the DAOI 4 did not significantly differ from the DAOI 1 ( $M=1.2, SD=0.19$ ) and DAOI 3 ( $M=2.6, SD=0.19$ ).

The table shows the percentage of people that have seen the DAOI in a certain order.

Video 3						
Sample	DAOI Time to First Fixation			DAOI view order		
	DAOI 1	DAOI 2	DAOI 3	1	2	3
G1-F01			17.25			1
G1-F02		16.75			1	
G1-F03			16.92			1
G1-F04		18.22	19.56		1	2
G1-F05			19.56			1
G1-M01		20.82	17.05		2	1
G1-M02		18.8	19.56		1	2
G1-M03			20.33			1
G1-M04		22.01	16.9		2	1
G1-M05		21.93	20.32		2	1
G2-F01	18.91	16.96	20.33	2	1	3
G2-F02		16.75	20.52		1	2
G2-F03			20.73			1
G2-F04		17.6	16.83			1
G2-F05			20.33			1
G2-M01			20.36			1
G2-M02		20.56	19.7		2	1
G2-M03		20.49	20.14		2	1
G2-M04			19.93			1
G2-M05		16.75	17.13		1	2

Avg Time To First Fixation		
DAOI 1	DAOI 2	DAOI 3
18.91	18.97	19.13

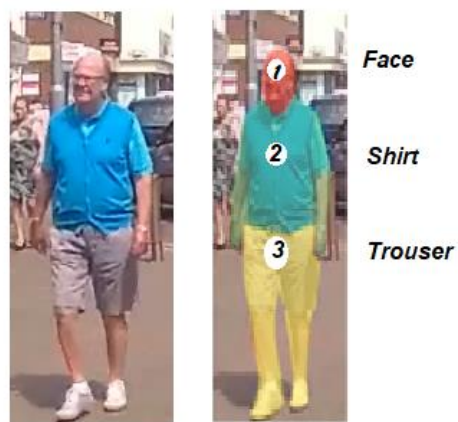
Order No	DAOI 1	DAOI 2	DAOI 3
1st	0%	30%	70%
2nd	5%	25%	20%
3rd	0%	0%	5%
Not seen	95%	45%	5%

**Table 2: Analysis of Order of Fixation for video 3**

*Video 4*

**Description: Find man wearing red half sleeve t-shirt and pants**

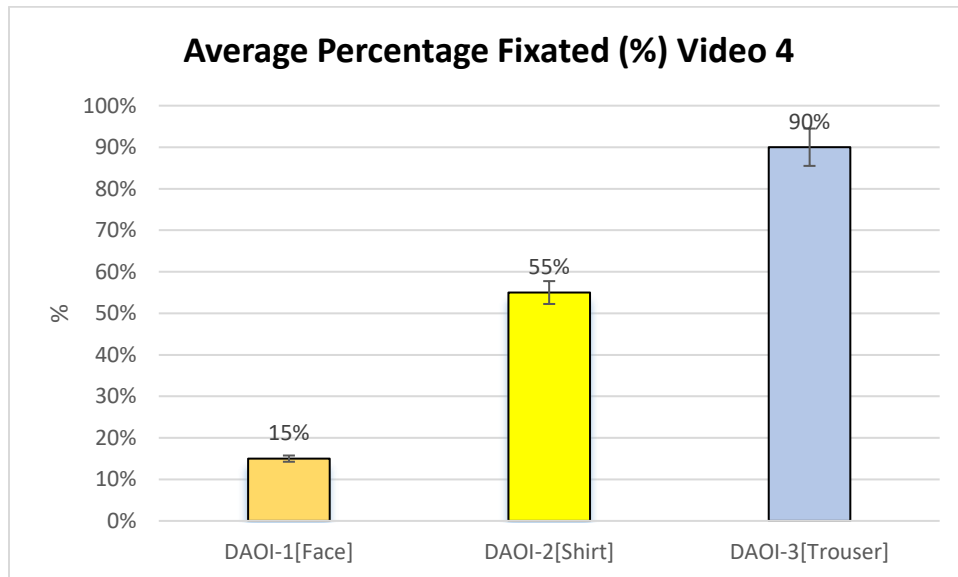
For this video footage, we identified or designed three areas of interest Face (DAOI 1), Shirt (DAOI 2) and Trouser (DAOI 3), as shown in the in the following figure.



**Figure 10: Four DAOIs**

## Percentage Fixated

On average, the DAOI 3 (trouser) had the highest percentage fixated (90%), while DAOI 2 (shirt) had 55%. The target face (DAOI 1) had the lowest percentage fixated (15%).



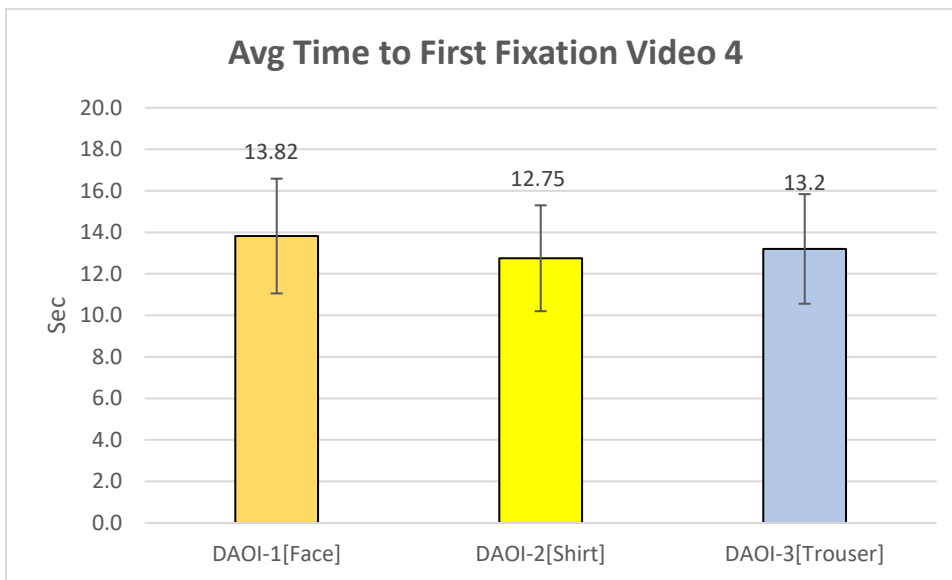
**Figure 11: Bar Graph for Percentage Fixated (%)**

According to the result most people have a tendency of focusing on the middle section (DAO I 2) as compared to other parts. There was a statistically significant difference on the percentage fixated between groups at  $p\text{-value} < 0.05$  as determined by one-way ANOVA ( $F(3,40) = 24.556, p = 0.00$ ). There is sufficient empirical evidence that most people fixated on the DAOI 3 (trouser), while the lowest proportion of people are fixed on the DAOI 1. The Tukey HSD test indicated that mean score for the DAOI 1 ( $M=0.1, SD=0.29$ ) was significantly different than the DAOI 2 ( $M=0.6, SD=0.43$ ) and DAOI 3 ( $M=0.95, SD=0.1.7$ ). The DAOI 3 was significantly different than the DAOI 2 and DAOI 1.



## Time to First Fixation

On average, the study participants took the longest time for the first fixation on the upper part (DAOI 1) and the short time for the first fixation on the lower part (DAOI 2). Figure 12 reflects the findings as summarized by the statistics.



**Figure 12: Bar Graph for Time to First Fixation**

There was no a statistically significant difference in the Time to first Fixation of the three areas at p-value < 0.05 as determined by one-way ANOVA [ $F(2, 30) = 0.34, p = 0.72$ ].

There is no statistically significant difference between your three DAOIs.

The below table shows the percentage of people that have seen the DAOI in a certain order. As shown in the figure the majority of participant (95%) had not seen target face (DAOI 1) and 70% had seen first DAOI 3 (trouser) as shown in the figure.

Video 4						
Sample	DAOI Time to First Fixation			DAOI view order		
	DAOI 1	DAOI 2	DAOI 3	1	2	3
G1-F01		12.7			1	
G1-F02		10.64			1	
G1-F03		10.34	11.16		2	1
G1-F04	10.2	12.03		2	1	
G1-F05		13.48	13.5		2	1
G1-M01		10.2	12.74		2	1
G1-M02		13	13.34		2	1
G1-M03		10.11			1	
G1-M04		13.21			1	
G1-M05		13.02	13.2		1	2
G2-F01						
G2-F02		12.07			1	
G2-F03		16.31	10.52		1	2
G2-F04	17.1	5.75	12.67	1	3	2
G2-F05		17.09	16.22		2	1
G2-M01		16.12	15.93		2	1
G2-M02		17.76	10.22		2	1
G2-M03	14.1	14.71	15.23	3	2	1
G2-M04			14			1
G2-M05		10.39			1	

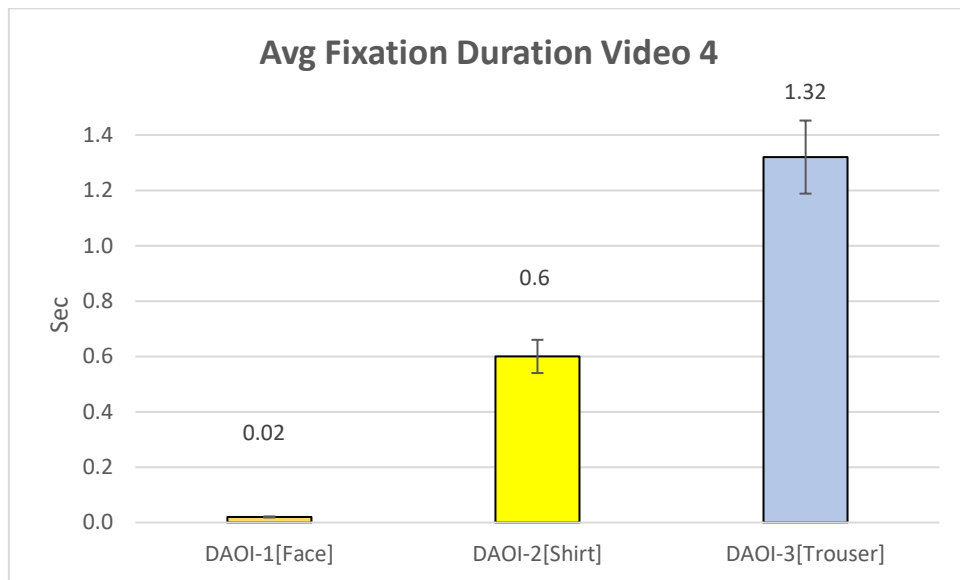
Avg Time To First Fixation		
DAOI 1	DAOI 2	DAOI 3
13.8	12.71833	13.2275

Order No	DAOI 1	DAOI 2	DAOI 3
1st	10%	45%	45%
2nd	5%	40%	15%
3rd	10%	10%	0%
Not seen	75%	5%	40%

**Table 3: Analysis of Order of Fixation for video 4**

## **Fixation duration**

On average, the study participants observed the trouser (DAOI 3) for the longest duration while the upper part-head (DAOI 1) was observed for the shortest duration. Figure 13 reflects the same findings as



**Figure 13: Bar Graph for Duration of Fixation**

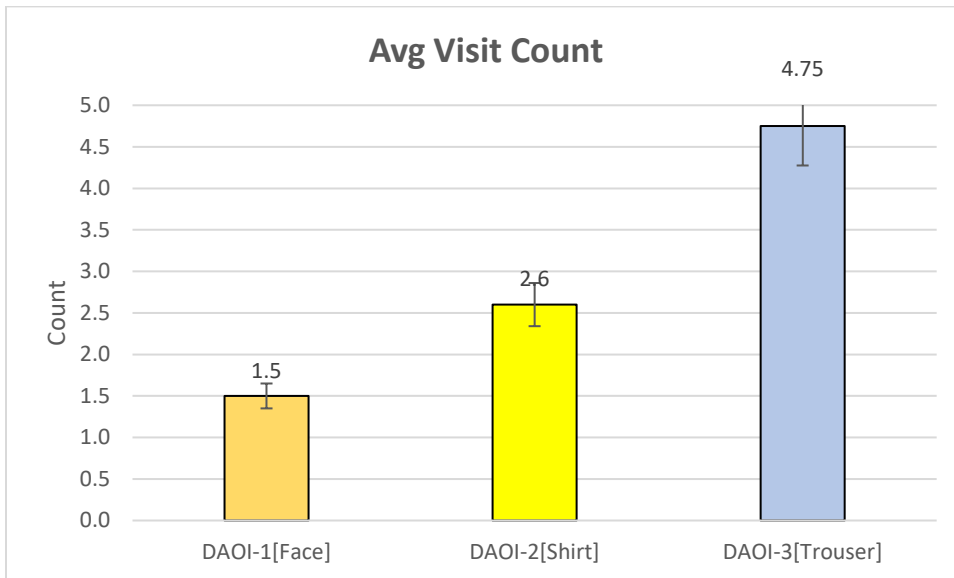
According to the result people spend the highest amount of time observing the lower part. There was a statistically significant difference on total fixation duration for the three areas at  $p$ -value  $< 0.05$  as determined by one-way ANOVA ( $F(2, 57) = 7.17, p = 0.0$ ).

Therefore, there are statistically significant differences across the three groups. There is sufficient empirical evidence that the lower part has the long fixation duration while the upper part has the shortest observation or fixation length.

The Tukey HSD test indicated that mean score for the DAOI 3 ( $M=1.1605, SD=0.72$ ) was significantly different than the DAOI 1 ( $M=0.014, SD=.0060$ ) and DAOI 2 ( $M=0.29, SD=0.6$ ). However, the DAOI 1 did not significantly differ from the DAOI 2.

## **Total Visit Counts**

On average, the study participants visited the DAOI 3 (trouser) for the highest number of times while the upper part-face was visited for the least number of times. Figure 14 reflects the same findings as summarized by



**Figure 14: Bar Graph for Visit Counts**

There was a statistically significant difference in the visit counts for the three areas at p-value  $< 0.05$  as determined by one-way ANOVA ( $F(2, 57) = 54.7, p = 0.0$ ). There is sufficient empirical evidence that the lower part (DAOI 3) has the higher visit counts while the upper part (AO11) has the least visit counts.

The Tukey HSD test indicated that mean score for the DAOI 1 ( $M=0.1, SD=0.308$ ) was significantly different than the DAOI 2 ( $M=1.2, SD=1.436$ ) and DAOI 3 ( $M=3.35, SD=1.927$ ). While the mean score for the DAOI 2 was significantly different than the DAOI 3.

### *Video 5*

**Description: Find a person wearing a pink short and white sleeveless top.**

For this video footage, we identified or designed four areas of interest around the target and one area of interest in person look just similar to the target.

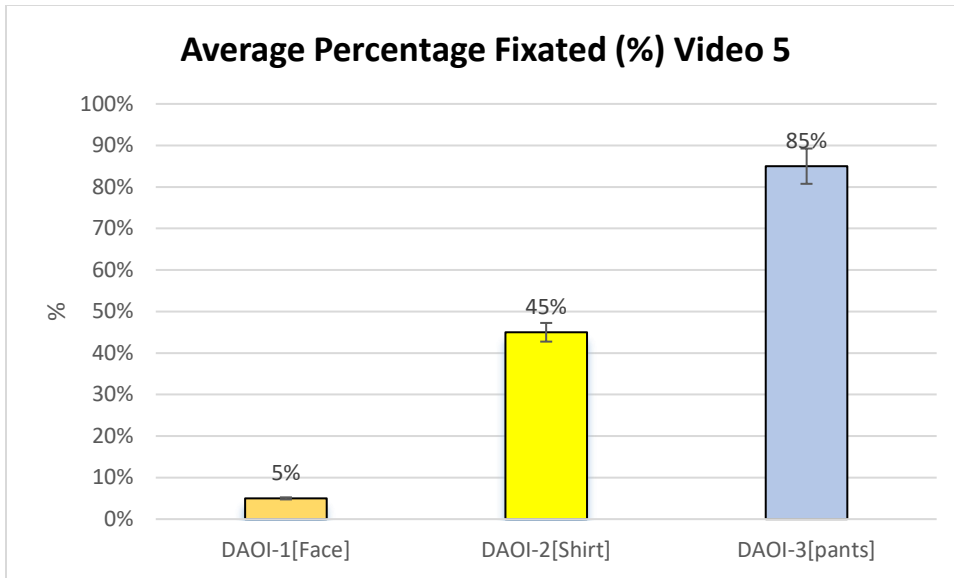


**Figure 15: Four DAOIs**

### **Percentage Fixated**

On average, the DAOI 3 (pants) had the highest percentage fixated (85%). While target face (DAOI 1) had the lowest percentage fixated just (5%). The DAOI 3 was the most appealing parts all the participants (85%) fixated on them and less appealing part was DAOI 1 only 5% fixated on it.

Figure 16 reflects the same findings as summarized by the statistics.

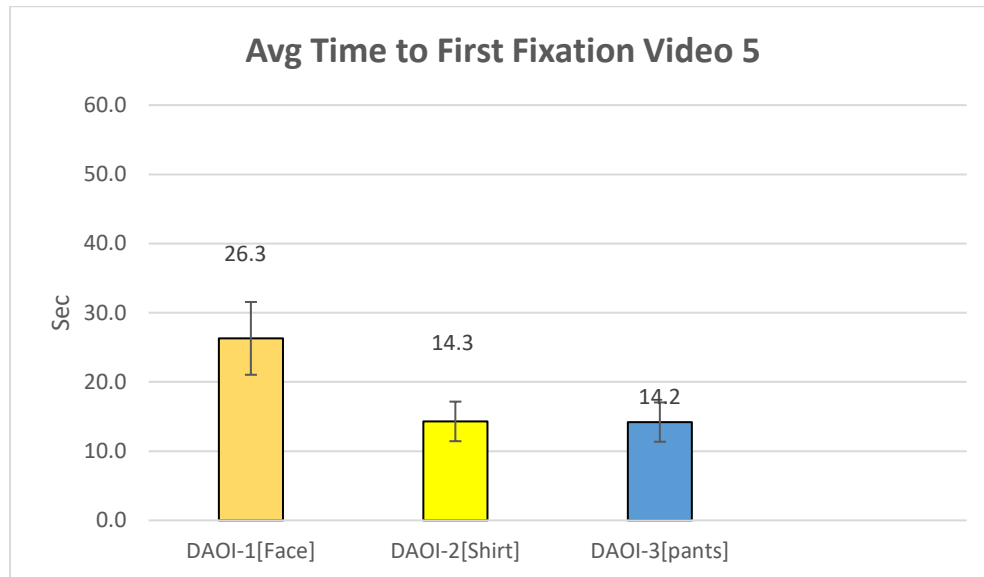


**Figure 16: Bar Graph for Percentage Fixated (%)**

According to the result most people have a tendency of focusing on the DAOI 3. There was a statistically significant difference on the percentage fixated between groups at  $p$ -value  $< 0.05$  as determined by one-way ANOVA [ $F(3,78) = 13, P=0$ ]. There is sufficient empirical evidence that most people fixated on the DAOI 3 while the lowest proportion of people are fixed on the DAOI 1. The Tukey HSD test indicated that mean score for the DAOI 1 ( $M=0.39, SD=0.43689$ ) was significantly different than the DAOI 2 ( $M=1, SD=0$ ), DAOI 3 and ( $M=0.5, SD=0.51299$ ). Also, DAOI 4 ( $M=0.5, SD=0.51299$ ) was significantly different than the DAOI 2 and DAOI 3.

## Time to First Fixation

On average, the study participants took the longest time for the first fixation on the upper part -face(DAOI1). Figure 3 reflects the same findings as summarized by the statistics.



**Figure 17: Bar Graph for Time to First Fixation**

There was a statistically significant difference in the Time to first Fixation of the five areas at  $p\text{-value} < 0.05$  as determined by one-way ANOVA [ $F(4,59) = 3.806, P=0$ ]. There is sufficient empirical evidence that the DAOI 1 has the highest time to first fixation. A Post hoc comparison using the Tukey HSD test indicated that mean score for the DAOI 1 ( $M=46.6629, SD=4.73896$ ) was significantly different than the DAOI 2 ( $M=41.724, SD=3.93124$ ), and DAOI 3 ( $M=40.8505, SD=4.05013$ ). However, there was no statistically significant difference between the DAOI 2 ( $M=41.752, SD=4.48848$ ) and DAOI 3.

The below table illustrate and calculate the percentage of people that have seen the DAOI in a certain order as shown in the following table. As shown in the table the majority of participant (90%) had not seen target face (DAOI 1) but in another hand 70% of people had seen first DAOI 4. The most appealing parts were DAOI 4 than DAOI 2, the participants focus at least once on those areas.

Video 5						
sample	DAOI Time to First Fixation			DAOI view order		
	DAOI 1	DAOI 2	DAOI 3	1	2	3
G1-F01			3.3			1
G1-F02		0.5			1	
G1-F03			9.92			1
G1-F04		13.56	19.56		1	2
G1-F05			19.56			1
G1-M01		19.82	12.05		2	1
G1-M02		18.8	19.56		1	2
G1-M03			10.33	1		
G1-M04		17.55	16.9		2	1
G1-M05		21.51	20.52		2	1
G2-F01		10.11	16.83		2	1
G2-F02		10	20.5		1	2
G2-F03			21.73			1
G2-F04	26.3	9.5	10.3	3	2	1
G2-F05			20.33			1
G2-M01			6.16			1
G2-M02		21.56	13.7		2	1
G2-M03		20	2.611		2	1
G2-M04			13.99			1
G2-M05		9.3	12.13		1	2

Avg Time To First Fixation		
DAOI 1	DAOI 2	DAOI 3
26.3	14.35083	14.20953

Order No	DAOI 1	DAOI 2	DAOI 3
1st	5%	25%	70%
2nd	0%	35%	20%
3rd	5%	0%	0%
Not seen	90%	60%	10%

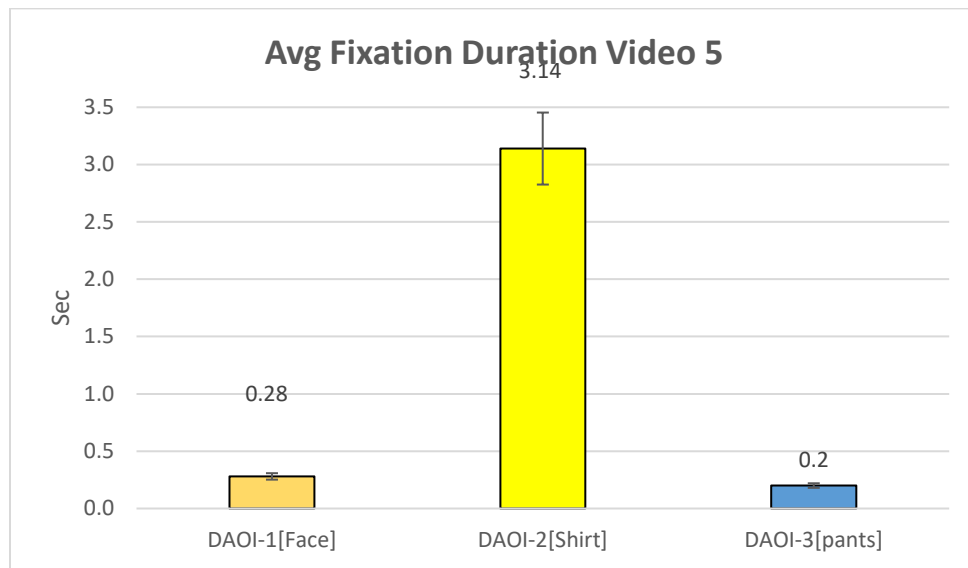
**Table 4: Analysis of Order of Fixation for video 5**



## Fixation duration

On average, the study participants observed the DAOI 2 for the longest duration while the upper part-head (DAOI 1) and DAOI 3 were observed for the shortest duration.

Figure 18 reflects the same

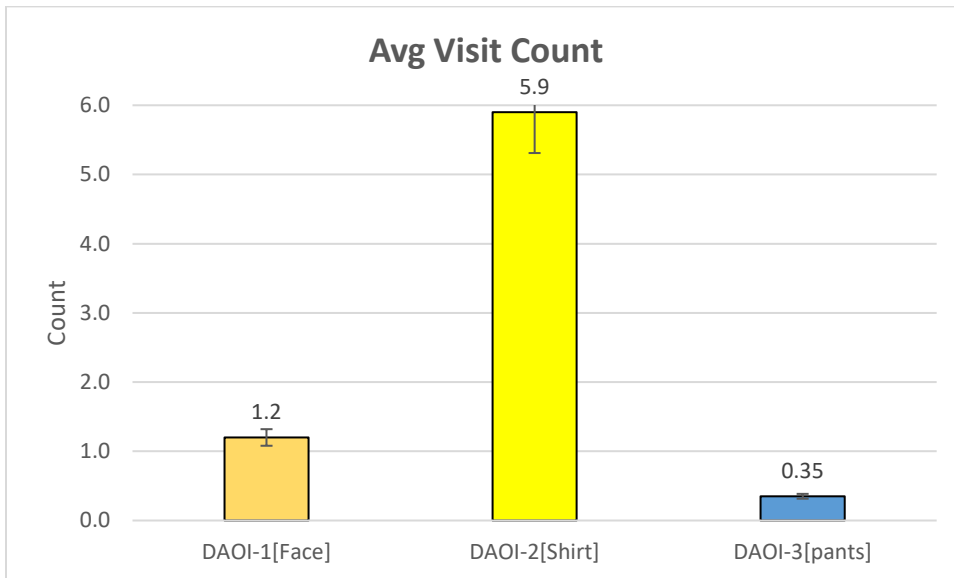


**Figure 18: Bar Graph for Duration of Fixation**

According to the result, people spend the highest amount of time observing the DAOI 2. There was a statistically significant difference in total fixation duration in the 3 areas of  $p\text{-value} < 0.05$  as determined by one-way ANOVA [ $F(4, 95) = 29.2, p = 0.00$ ]. Therefore, there are statistically significant differences across the five areas of interests. There is sufficient empirical evidence that the DAOI 2 part has a longer fixation duration while the DAOI 1 has the shortest observation or fixation length. The Tukey HSD test indicated that the mean score for the DAOI1 ( $M=0.63, SD=0.19$ ) was significantly different than the DAOI2 ( $M=0.63, SD=0.19$ ) and DAOI 3 ( $M=0.63, SD=0.19$ ). However, the DAOI1 did not significantly differ from the DAOI3 ( $M=0.63, SD=0.19$ ).

## Total Visit Counts

On average, the study participants visited the DAOI 2 for the highest number of times while the lower part-head DAOI 3 was visited for the least number of times. Figure 19 reflects the same findings as summarized by the statistics.



**Figure 19: Bar Graph for Visit Counts**

There was a statistically significant difference in the visit counts for the 3 areas at p-value  $< 0.05$  as determined by one-way ANOVA [ $F(4, 95) = 3.4, p = 0.00$ ]. There is sufficient empirical evidence that the DAOI 2 has the higher visit counts while the DAOI 3 has the least visit counts.

### Video 8

**Description: Find a man wearing dark blue sport trousers**

For this video footage, we identified or designed three areas of interest Face (DAOI 1), Shirt (DAOI 2) and Trouser (DAOI 3), as shown in the in the following figure.

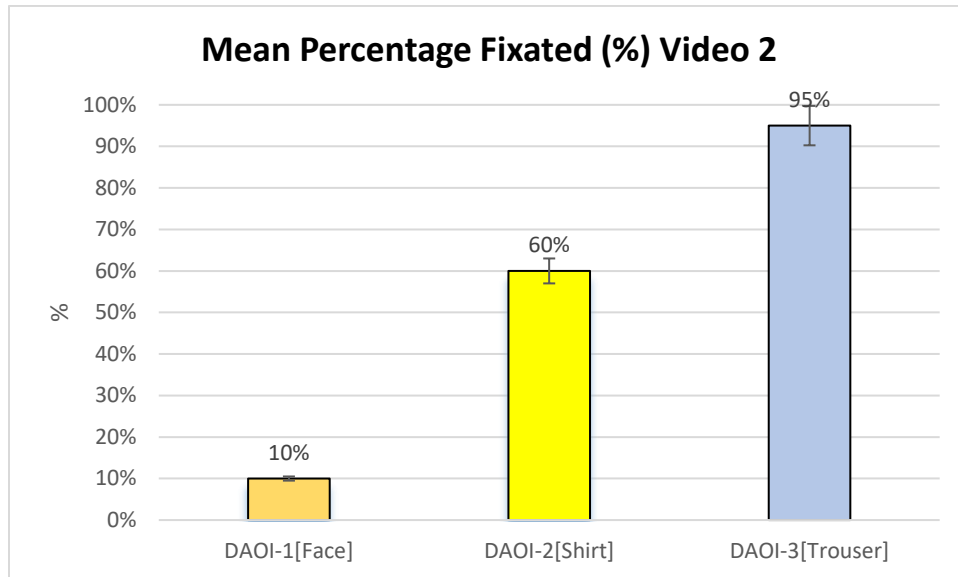


**Figure 20: DAOIs**

### **Percentage Fixated**

On average, the DAOI 3 (trouser) had the highest percentage fixated (95%), while DAOI 2 (shirt) had 60%. The target face (DAOI 1) had the lowest percentage fixated (10%).

Figure 21 reflects the same findings as summarized by the statistics.

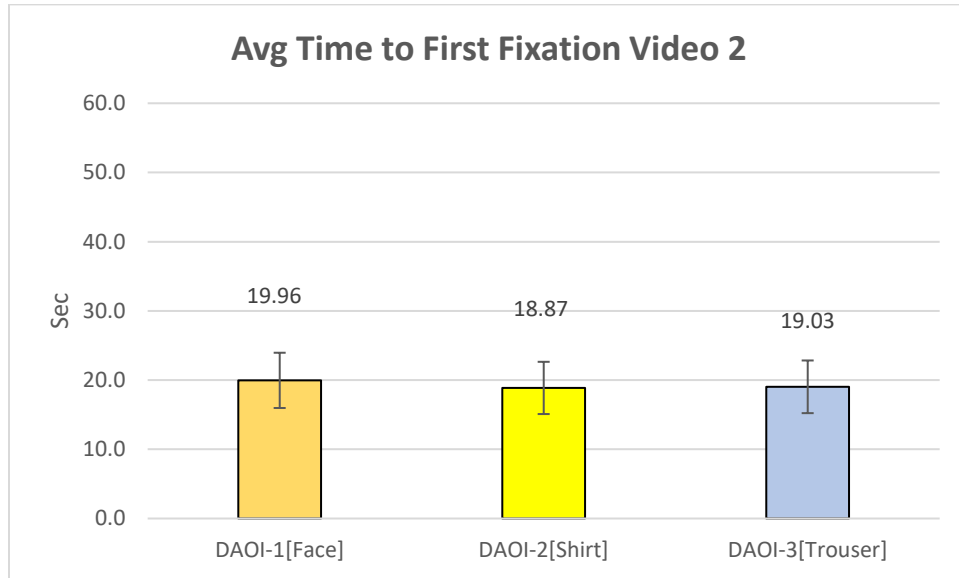


**Figure 21: Bar Graph for Percentage Fixated (%)**

According to the result most people have a tendency of focusing on the middle section(AO2) as compared to other parts. There was a statistically significant difference on the percentage fixated between groups at  $p\text{-value} < 0.05$  as determined by one-way ANOVA ( $F(3,57) = 27.556, p = 0.00$ ). There is sufficient empirical evidence that most people fixated on the DAOI 3 (trouser), while the lowest proportion of people are fixed on the DAOI 1. The Tukey HSD test indicated that mean score for the DAOI 1 ( $M=0.1, SD=0.31$ ) was significantly different than the DAOI2( $M=0.6, SD=0.51$ ) &DAOI 3( $M=0.95, SD=0.22$ ). The DAOI 3 was significantly different than the DAOI 2 & DAOI 1.

## **Time to First Fixation**

On average, the study participants took the longest time for the first fixation on the upper part (DAOI1) and the short time for the first fixation on the lower part (DAOI2). Figure 22 reflects the same findings as summarized by the statistics.



**Figure 22: Bar Graph for Time to First Fixation**

There was no a statistically significant difference in the Time to first Fixation of the three areas at p-value < 0.05 as determined by one-way ANOVA [ $F(2, 30) = 0.34, p = 0.72$ ].

There is no statistically significant difference between your three DAOIs.

The below table shows the percentage of people that have seen the DAOI in a certain order. As shown in the figure the majority of participant (95%) had not seen target face (DAOI 1) and 70% had seen first DAOI 3 (trouser) as shown in the figure.

Video 8						
Sample	DAOI Time to First Fixation			DAOI view order		
	DAOI 1	DAOI 2	DAOI 3	1	2	3
G1-F01			17.25			1
G1-F02		16.75			1	
G1-F03			16.92			1
G1-F04		18.22	19.56		1	2
G1-F05			19.56			1
G1-M01		20.82	17.05		2	1
G1-M02		18.8	19.56		1	2
G1-M03			20.33			1
G1-M04		22.01	16.9		2	1
G1-M05		21.93	20.32		2	1
G2-F01	18.91	16.96	20.33	2	1	3
G2-F02		16.75	20.52		1	2
G2-F03			20.73			1
G2-F04		17.6	16.83			1
G2-F05			20.33			1
G2-M01			20.36			1
G2-M02		20.56	19.7		2	1
G2-M03		20.49	20.14		2	1
G2-M04			19.93			1
G2-M05		16.75	17.13		1	2

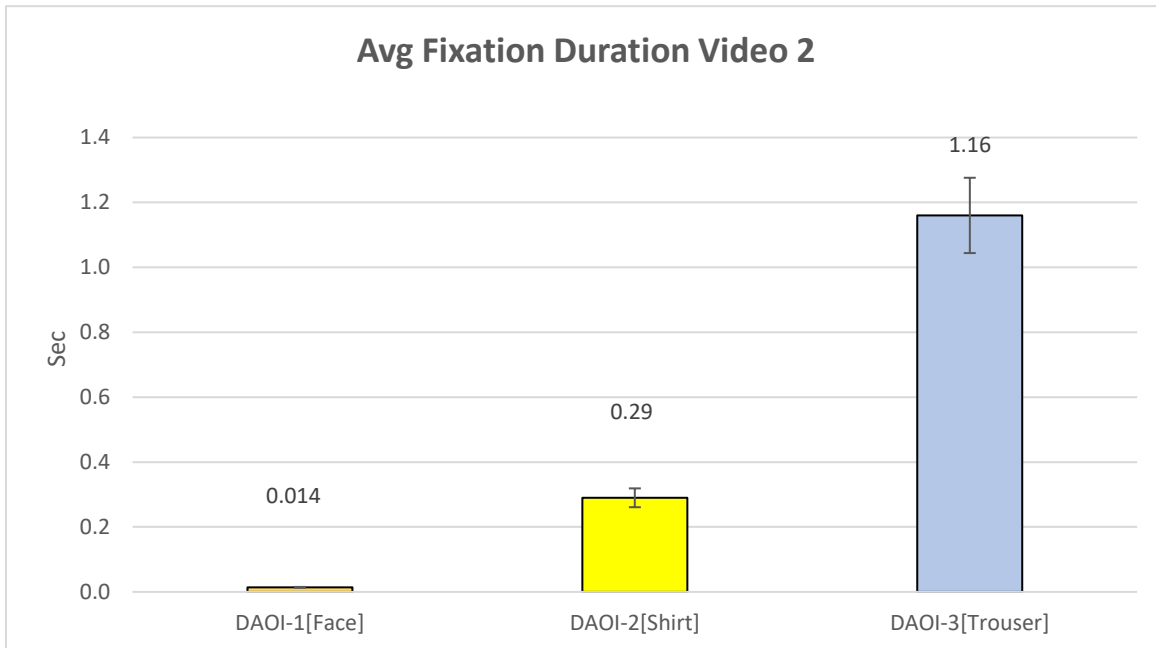
Avg Time to First Fixation		
DDAOI 1	DDAOI 2	DDAOI 3
18.91	18.97	19.13

Order No	DAOI 1	DAOI 2	DAOI 3
1st	0%	30%	70%
2nd	5%	25%	20%
3rd	0%	0%	5%
Not seen	95%	45%	5%

**Table5: Analysis of Order of Fixation for video 8**

## Fixation Duration

On average, the study participants observed the trouser (DAOI 3) for the longest duration while the upper part-head (DAOI 1) was observed for the shortest duration. Figure 23 reflects the same findings as summarized by the statistics.



**Figure 23: Bar Graph for Duration of Fixation**

According to the result people spend the highest amount of time observing the lower part. There was a statistically significant difference on total fixation duration for the three areas at  $p\text{-value} < 0.05$  as determined by one-way ANOVA ( $F(2, 57) = 7.17, p = 0.0$ ).

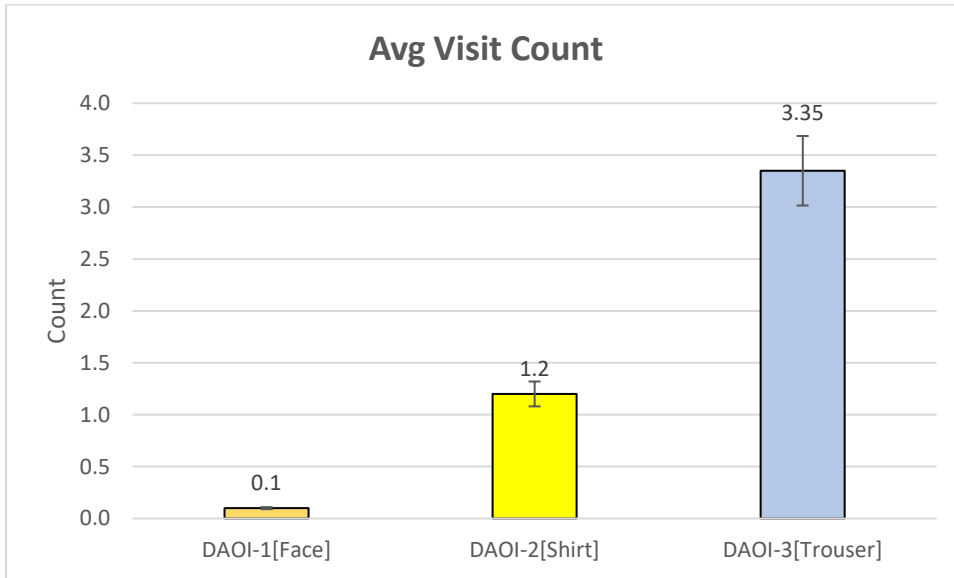
Therefore, there are statistically significant differences across the three groups. There is sufficient empirical evidence that the lower part has the long fixation duration while the

upper part has the shortest observation or fixation length. The Tukey HSD test indicated that mean score for the DAOI 3 ( $M=1.1605$ ,  $SD=0.72$ ) was significantly different than the DAOI 1 ( $M=0.014$ ,  $SD=.0.060$ ) and DAOI 2 ( $M=0.29$ ,  $SD=0.6$ ). However, the DAOI 1 did not significantly differ from the DAOI 2.



## **Total Visit Counts**

On average, the study participants visited the DAOI 3 (trouser) for the highest number of times while the upper part-face was visited for the least number of times. Figure 24 reflects the same findings as summarized by



**Figure 24: Bar Graph for Visit Counts**

There was a statistically significant difference in the visit counts for the three areas at p-value  $< 0.05$  as determined by one-way ANOVA ( $F(2, 57) = 54.7, p = 0.0$ ). There is sufficient empirical evidence that the lower part (DAOI 3) has the higher visit counts while the upper part (AO11) has the least visit counts.

The Tukey HSD test indicated that mean score for the DAOI 1 ( $M=0.1, SD=0.308$ ) was significantly different than the DAOI 2 ( $M=1.2, SD=1.436$ ) and DAOI 3 ( $M=3.35, SD=1.927$ ). While the mean score for the DAOI 2 was significantly different than the DAOI 3.

## **Appendix 2**

**Results obtained by the various machine learning algorithms tested for classification of expert verse novice participants**

## Multi-Layer Perceptron (MLP)

Weka Summary		
Correctly Classified Instances	44	73.3333 %
Incorrectly Classified Instances	16	26.6667 %
Kappa statistic	0.4743	
Mean absolute error	0.3563	
Root mean squared error	0.4223	
Relative absolute error	71.3418 %	
Root relative squared error	84.5008 %	
Total Number of Instances	60	

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.966	0.484	0.651	0.966	0.778	0.534	0.85	0.856	Novice
	0.516	0.034	0.941	0.516	0.667	0.534	0.85	0.87	Expert
<b>Weighted Avg.</b>	0.733	0.252	0.801	0.733	0.72	0.534	0.85	0.863	

### Confusion Matrix

<b>a</b>	<b>b</b>	<b>&lt;-- classified as</b>
28	1	a = Novice
15	16	b = Expert

## Multi-Layer Perceptron (MLP) with Bagging

Weka Summary		
Correctly Classified Instances	49	81.6667 %
Incorrectly Classified Instances	11	18.3333 %
Kappa statistic	0.6341	
Mean absolute error	0.3454	
Root mean squared error	0.3821	
Relative absolute error	69.1553 %	
Root relative squared error	76.457 %	
Total Number of Instances	60	

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.862	0.226	0.781	0.862	0.820	0.637	0.911	0.927	Novice
	0.774	0.138	0.857	0.774	0.814	0.637	0.911	0.908	Expert
Weighted Avg.	0.817	0.180	0.820	0.817	0.817	0.637	0.911	0.917	

### Confusion Matrix

<b>a</b>	<b>b</b>	<b>&lt;- classified as</b>
25	4	a = Novice
7	24	b = Exnert

## Random Forest

Weka Summary		
Correctly Classified Instances	60	100 %
Incorrectly Classified Instances	0	0 %
Kappa statistic	1	
Mean absolute error	0.1743	
Root mean squared error	0.1991	
Relative absolute error	34.9042 %	
Root relative squared error	39.835 %	
Total Number of Instances	60	

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	Novice
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	Expert
<b>Weighted Avg.</b>	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	

### Confusion Matrix

<b>a</b>	<b>b</b>	<b>&lt;-- classified as</b>
29	0	a = Novice
0	31	b = Expert

## Random Forest with Bagging

Weka Summary		
Correctly Classified Instances	57	95 %
Incorrectly Classified Instances	3	5 %
Kappa statistic	0.9	
Mean absolute error	0.2577	
Root mean squared error	0.2887	
Relative absolute error	51.6022 %	
Root relative squared error	57.7677 %	
Total Number of Instances	60	

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.966	0.065	0.933	0.966	0.949	0.901	0.991	0.992	Novice
	0.935	0.034	0.967	0.935	0.951	0.901	0.991	0.992	Expert
<b>Weighted Avg.</b>	0.950	0.049	0.951	0.950	0.950	0.901	0.991	0.992	

### Confusion Matrix

<b>a</b>	<b>b</b>	<b>&lt;- classified as</b>
28	1	a = Novice
2	29	b = Exnert

## Rep Tree

Weka Summary		
Correctly Classified Instances	44	73.3333 %
Incorrectly Classified Instances	16	26.6667 %
Kappa statistic	0.4684	
Mean absolute error	0.3805	
Root mean squared error	0.4362	
Relative absolute error	76.1771 %	
Root relative squared error	87.281 %	
Total Number of Instances	60	

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.793	0.323	0.697	0.793	0.742	0.473	0.759	0.676	Novice
	0.677	0.207	0.778	0.677	0.724	0.473	0.759	0.735	Expert
<b>Weighted Avg.</b>	0.733	0.263	0.739	0.733	0.733	0.473	0.759	0.707	

### Confusion Matrix

<b>a</b>	<b>b</b>	<b>&lt;- classified as</b>
23	6	a = Novice
10	21	b = Expert

## Rep Tree with Bagging

Weka Summary		
Correctly Classified Instances	49	81.6667 %
Incorrectly Classified Instances	11	18.3333 %
Kappa statistic	0.6341	
Mean absolute error	0.3407	
Root mean squared error	0.3762	
Relative absolute error	68.2103 %	
Root relative squared error	75.279 %	
Total Number of Instances	60	

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.862	0.226	0.781	0.862	0.820	0.637	0.910	0.911	Novice
	0.774	0.138	0.857	0.774	0.814	0.637	0.910	0.918	Expert
<b>Weighted Avg.</b>	0.817	0.180	0.820	0.817	0.817	0.637	0.910	0.915	

### Confusion Matrix

<b>a</b>	<b>b</b>	<b>&lt;- classified as</b>
25	4	a = Novice
7	24	b = Expert



## **Appendix 3**

**Results obtained by the various machine learning algorithms tested for classification of Female verse Male participants**

## MLP

Weka Summary		
Correctly Classified Instances	7	50 %
Incorrectly Classified Instances	7	50 %
Kappa statistic	0	
Mean absolute error	0.537	
Root mean squared error	0.6911	
Relative absolute error	107.4055 %	
Root relative squared error	138.2233 %	
Total Number of Instances	14	

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.429	0.429	0.500	0.429	0.462	0.000	0.367	0.472	F
	0.571	0.571	0.500	0.571	0.533	0.000	0.367	0.459	M
<b>Weighted Avg.</b>	0.500	0.500	0.500	0.500	0.497	0.000	0.367	0.466	

### Confusion Matrix

<b>a</b>	<b>b</b>	<b>&lt;-- classified as</b>
<b>3</b>	<b>4</b>	<b>a = F</b>
<b>3</b>	<b>4</b>	<b>b = M</b>

## MLP with Bagging

Weka Summary		
Correctly Classified Instances	17	85 %
Incorrectly Classified Instances	3	15 %
Kappa statistic	0.7	
Mean absolute error	0.2805	
Root mean squared error	0.337	
Relative absolute error	56.0933 %	
Root relative squared error	67.3945 %	
Total Number of Instances	20	

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.800	0.100	0.889	0.800	0.842	0.704	0.930	0.942	F
	0.900	0.200	0.818	0.900	0.857	0.704	0.930	0.935	M
<b>Weighted Avg.</b>	0.850	0.150	0.854	0.850	0.850	0.704	0.930	0.939	

### Confusion Matrix

a	b	<-- classified as
8	2	a = F
1	9	b = M

## Random Forest

Weka Summary		
Correctly Classified Instances	20	100%
Incorrectly Classified Instances	0	0%
Kappa statistic	1	
Mean absolute error	0.208	
Root mean squared error	0.2256	
Relative absolute error	41.6 %	
Root relative squared error	45.1132 %	
Total Number of Instances	20	

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	F
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	M
Weighted Avg.	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	

### Confusion Matrix

<b>a</b>	<b>b</b>	<b>&lt;-- classified as</b>
<b>10</b>	<b>0</b>	<b>a = F</b>
<b>0</b>	<b>10</b>	<b>b = M</b>

## Random Forest with Bagging

Weka Summary		
Correctly Classified Instances	20	100%
Incorrectly Classified Instances	0	0%
Kappa statistic	1	
Mean absolute error	0.3027	
Root mean squared error	0.3162	
Relative absolute error	60.54 %	
Root relative squared error	63.2479 %	
Total Number of Instances	20	

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	F
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	M
Weighted Avg.	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	

### Confusion Matrix

<b>a</b>	<b>b</b>	<b>&lt;-- classified as</b>
<b>10</b>	<b>0</b>	<b>a = F</b>
<b>0</b>	<b>10</b>	<b>b = M</b>

## REP Tree

Weka Summary		
Correctly Classified Instances	10	50%
Incorrectly Classified Instances	10	50%
Kappa statistic	0	
Mean absolute error	0.5	
Root mean squared error	0.5	
Relative absolute error	100%	
Root relative squared error	100%	
Total Number of Instances	20	

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	1.000	0.500	1.000	0.667	0.000	0.500	0.500	F
	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.500	M
Weighted Avg.	0.500	0.500	0.250	0.500	0.333	0.000	0.500	0.500	

### Confusion Matrix

a	b	<-- classified as
10	0	a = F
10	0	b = M

## REP Tree with Bagging

Weka Summary		
Correctly Classified Instances	13	60 %
Incorrectly Classified Instances	7	40 %
Kappa statistic	0.4	
Mean absolute error	0.448	
Root mean squared error	0.4547	
Relative absolute error	89.6018	
Root relative squared error	90.9403	
Total Number of Instances	20	

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.536	0.1	0.733	0.5	0.625	0.436	0.71	0.834	F
	0.73	0.5	0.543	0.7	0.75	0.436	0.74	0.792	M
Weighted Avg.	0.633	0.3	0.638	0.6	0.688	0.436	0.725	0.713	

### Confusion Matrix

<b>a</b>	<b>b</b>	<b>&lt;-- classified as</b>
<b>5 5</b>	<b> </b>	<b>a = F</b>
<b>1 9</b>	<b> </b>	<b>b = M</b>

## **Appendix 4**

### **Conference paper**

**A Machine Learning Based Approach to Human Observer Behaviour Analysis in  
CCTV Video Analytics & Forensics**



# A Machine Learning Based Approach to Human Observer Behaviour Analysis in CCTV Video Analytics & Forensics

Seema F. Al Raisi

Department of Computer Science,  
Loughborough University, UK  
S.Al-Raisi@lboro.ac.uk

Eran Edirisinghe

Department of Computer Science,  
Loughborough University, UK  
E.A.Edirisinghe@lboro.ac.uk

## ABSTRACT

Human observer behaviour analysis in image and video inspection in many areas of practical application is conducted based on using data captured by eye tracking devices. Such data is analysed using statistical approaches leading to the creation of useful information and the ability to make decisions about the content. CCTV observer behaviour analysis is one example of a most widely used application. Unfortunately, the information and knowledge that such statistical approaches to data analysis can create is rather limited, especially the trends and patterns of data cannot be easily analysed. Thus, important information and knowledge that the data can provide may not be identifiable. In this paper, we proposed a novel approach to human observer eye tracking data analysis based on machine learning algorithms. Further, in order to conduct a more detailed and practically useful data analysis, we specifically analyse the attention human observers given instructions to search for specified content. We provide experimental results to demonstrate the significance and novelty of the information and knowledge that this novel approach to data analysis can provide. To the authors' knowledge, there is no work in literature that has proposed the use of machine learning in eye tracking data analysis.

## KEYWORDS

Machine Learning, CCTV surveillance, eye-tracking system.

## 1 INTRODUCTION

Previous work on human observer behaviour analysis and classification based on data captured via eye tracking devices has been limited to direct use of statistical approaches and the drawing of graphs for the visualization of differences for the ease of human interpretation. These approaches are not only tedious, but will also not be able to identify the presence of fine detailed discriminative features between data captured from different groups nor identify complex patterns and trends that may be present in the data. To the authors' knowledge, all existing studies of CCTV observer performance analysis use statistical and/or conceptual data analysis approaches, using tools such as ANOVA that does not allow observer behaviour pattern analysis. It is noted that, machine learning has not been used in relation to the human observer performance analysis of CCTV video footage. Such analysis could lead to gathering information that was previously considered to be impossible to be gathered via statistical approaches. This is the key contribution of the research conducted in this paper.

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org). IML '17, October 17–18, 2017, Liverpool, United Kingdom © 2017 Association for Computing Machinery. ACM ISBN 978-1-4503-5243-7/17/10...\$15.00 <https://doi.org/10.1145/3109761.3158376>

Further, such studies conducted using statistical and/or conceptual data analysis approaches are limited in their analysis as the investigations are carried out with regards to observing eye tracking data captured in visually identifying particular human objects based on their entire body areas that significantly constrains the potential to carry out detailed behaviour analysis. The purpose, of using machine learning techniques for CCTV observer performance analysis is to better analyse, identify, learn, understand, make predictions and uncover hidden patterns in the eye tracking data.

Given the above reasons the research proposed and presented in this paper could immensely benefit the video analytic community in the future by providing the means to use the automatically generated human performance patterns in making modern video analytic and forensic algorithms, smarter and more time-efficient, especially by making them mimic the behaviour of a human.

For clarity of presentation, this paper is divided into several sections. A part from this section that provides an introduction to the research problem being addressed section-2 provides a brief review of literature on eye-tracking based human observer analysis of data captured by video and CCTV capture devices. Section-3 presents the proposed methodology for experimental set-up for data capture, preparation and representation for subsequent analysis. Section-4 provides the machine learning based data analysis results and a detailed analysis leading to novel results. Finally, section-5 concludes with a view to further work.

## 2 LITERATURE REVIEW

In 2013, Roffo et al. showed that understanding the way people visually analyse video sequences, beyond the content they observe in a video, is vital for the understanding and the prediction of people's activities. The analysis presented in this study was based on eye tracking data on CCTV video sequences. Because of the extensively investigated higher capabilities of expert operators in predicting violence in surveillance footage, the main goal of the research proposed was to understand how expert CCTV operators analyse such videos, and if there is a difference between expert operators and novice participants. It is noted that all the analysis was carried out using statistical approaches [1].

In 2013 [2] Iain Gilchrist et al. proposed that CCTV operators sometimes may be able to predict trouble, and trouble hotspots, very rapidly and they can take less than two-second delay before making any decision. The study suggests that this waiting strategy permits them to take an action and identify supplementary visual information, and

may be a thoughtful approach or method to help the operators make a correct decision and reduce the number of false alarms. In this study the CCTV operators were requested to view 80 clips of recorded footage (such as a night-time view of a car park, a shopping street underpass, a nightclub entrance and a cash point) for one minute each. They were instructed to observe the clips for behaviour believed to be suspicious enough in order to alert the relevant authorities and used a joystick device to specify the perceived level of suspicious behaviour. The eye tracking sensor data were used to calculate the relationship between these ratings and the CCTV operators' patterns of gaze. The research results were compared against a group of untrained observers (novices). A statistical data analysis tool was used to analyse the operator's performance. The study established that trained operators spend more time than untrained observers when determining whether a scene is suspicious. The collected data also suggested that trained operators moved their eyes to the significant part of the scene earlier, and they followed very similar viewing patterns. They would normally look at specific areas of focus rather than moving quickly between random locations as was the case with untrained operators and were also much more consistent in spotting suspicious events in ambiguous footage. The data analysis of this research was carried out using statistical approaches.

Robertson's et al. research presented [3] found that explanation of global scene activity, mostly where interesting events had happened, can aid human observer analysis of high volumes of captured data. This was mainly achieved by using an extensible, rule-based method that was generated based on past studies of observer behaviour analysis using statistical methods. This research paper was concerned with creating high-level text reports and explanations of people's activity in video from single, static cameras, with the motivation being to allow surveillance analysts to provide situational awareness regardless of the presence of huge data. The paper focused on urban surveillance where the pictured person was shown in low/medium resolution. The final output required was text descriptions that explained the interactions that took place and described what was happening (observed human activity). The research also states that the whole system denotes a general technique for video understanding, which involves a guided training phase via an experienced analyst. It is however noted that all the data analysis conducted within the scope of the research presented in this paper was carried out using statistical approaches.

Dr. Gemma Graham in 2016 [4] has integrated eye tracking technology into many research projects. In one of her studies she investigated how people observe the CCTV footage[4], mostly when they are given instructions to focus on given features in the video and seeing different severity of crimes. In this research four experiments were carried out in order to investigate the following research questions: “(a) does instruction and/or event type impact where observers view during CCTV footage observation?; (b) how do task instructions and central and marginal information influence fixation behaviour during CCTV observation?; (c) what is the effect of change detection on memory recall during CCTV observation?; and (d) do verbalization, attentional set and/or repeated viewing improve change detection rates and memory recall for CCTV footage?” [4]. The study was conducted with a different number of participants for each experiment. During the experiments, the participants' eye movements were recorded by a professional eye tracking system. The observer performance was analysed by using a statistical data analysis tool called ANOVA. This research found that the instructions did not significantly affected gaze behaviour for dynamic scenes and any changes in detectors evoked more accurate detail from the CCTV footage as related to the non-detectors, but only in the case where the seriousness of the crime had increased. However, when the observer was repeatedly shown the CCTV footage the rates of change detection, improved greatly, but verbalisation made no difference in terms of change detection and the accuracy of memory recall. The last finding obtained from this research can provide help to inform training courses aimed at instructing users on how to optimally attend to surveillance video.

The above review of literature clearly demonstrates that all eye tracking related research has focused on the use of fundamental and applied statistics in the analysis of the captured data.

### 3 THE STUDY METHODOLOGY AND PROCEDURE

The following sections provide the methodologies adopted for the capture, pre-processing and representation of eye-tracking data so that subsequent analysis can be conducted based on machine learning approaches.

#### 3.1 Study Sampling / Participants

In this research, the experiments were conducted with 24 participants. It is noted here that our original intention was to use a substantially larger group of participants. However

having completed the experiments with 24 participants we found out that the machine learning algorithms are effectively capable of recognizing patterns in the data and discriminate between participants even with 24 participants. Therefore it was decided that the having only 24 participants is justifiable. The 24 participants were equally divided gender-wise into 12 female and 12 male participants. The 12 female and 12 male participants were again divided into two further groups, namely; a Novice group (6 M + 6 F) that included participants who have never had any specific experience in the visual image analysis of any kind and an Expert group (6 M + 6 F) that included participants having substantial exposure to visual image/video analysis/analytics. All members of the groups viewed the same pre-recorded videos and were given the same descriptions, per video.

#### 3.2 Chosen Stimuli / video

A total of 13 different CCTV video was used for this study. The length of time each video was presented to each of the 24 participants was one minute. Further, before each video clip is presented to a participant, written descriptions regarding target appearance to look for, were shown on the screen. The objects that were requested to be identified in the chosen videos used in the study are outlined in Table. 1

**Table 1: Objects that are requested to be identified and tracked by the participants**

Video 1	Video 2	Video 3	Video 4	Video 5
				
				
				

#### 3.3 Eye tracking sensor/technology, hardware & software environments

The required eye tracking of participants was carried out using a Tobii T60Eye Tracker[5]. The presentation of the stimuli was designed, presented and controlled by using the associated Tobii Studio Software[6]. The Tobii studio

project consisted of 13 written descriptions and pre-recorded videos. These descriptions and videos were displayed one after another on a screen.

3.3.1 Tobii T60 Eye Tracker



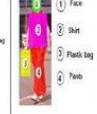


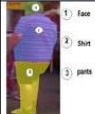

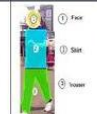
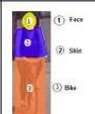




The required eye tracking of participants was carried out using a Tobii T60Eye Tracker[5]. The presentation of the stimuli was designed, presented and controlled by using the associated Tobii Studio Software[6]. The Tobii studio project consisted of 13 written descriptions and pre-recorded videos. These descriptions and videos were displayed one after another on a screen.

3.3.2 Dynamic Areas of Interest (DAOI) design

In order to investigate the human observer’s attention given to different parts of the human body, namely the head area, Shirt area, trouser/legs and any object being carried/pushed/pulled etc., (i.e. an area connected to the human body area that contains an object external to the human body), the DAOI tool provided by the Tobii Studio software was used for the creation of dynamic areas of interest within and around a target body.

Due to the human objects to be detected and monitored being of a dynamic/moving nature and very different to each other between the videos the design of the DAIOs in the thirteen different videos were different and hence the definition of the DAIOs was a tedious and time consuming task. Table 2 illustrates the definitions of the DAIOs for the objects intended to be detected and tracked in each of the videos to be monitored.

Table 2: DAOI for each video

Video 1	Video 2	Video 3	Video 4	Video 5
				
Video 6	Video 7	Video 8	Video 9	Video 10
				
Video 11	Video 12	Video 13		
				

3.4 Written descriptions of objects to be identified and tracked

There has been much research concluding that task instructions influence where an observer's eye is fixated while viewing static scenes [7] or a dynamic scene [8]. So in our experiment all the participants are provided with identical written instructions per human object to be identified and tracked in each video, displayed on the screen before the video is displayed. It is noted that in each instruction there are some important keywords participants need to follow in order to find the correct target. For example in video 1 the following instruction was given to the participants Find a **woman** wearing **white trousers** and carrying a **pink shoulder handbag**".

In this instruction, there are three important keywords "**woman**", "**white trousers**" and "**pink shoulder handbag**". The key-information that the participants will extract from the given descriptions that will be used in the search for the object. Since none of the instructions given it was not mentioned whether the person is static (not moving or stationary) or moving (walking), it was likely that all participants would attempt to analyze and look for both types of objects/people.

3.5 Study Method

The total duration of each participant’s subjective experiment session associated with the study carried out was approximately 30 minutes. The participant was requested to read 13 short written descriptions of a person and view 13 different videos (individually on the screen), while the eye tracker is recording the participant’s visual attention and making the Tobii studio software generate the related data. The experiment starts with a fixation cross sign displayed at the centre of the screen. This sign appears at the beginning of the experiment and after each piece of video footage. When the written description to find a target (person, man or woman) appears on the screen the participant is required to read it carefully and then press the space bar on the keyboard to allow a video to start. When the video starts, the participant given 1 minute to look at the video and find a person who matches the description. As soon as the participant finds the person, the participant requested to use the mouse to point at the target and click. This cycle repeated 13 times for the different videos; during this time, the sensor records eye tracking data with respect to the participant’s natural reactions to the content of the video given the specific instructions.

From the beginning of the experiment until the end, participants requested to verbalise, i.e. think-aloud, their thoughts during the process of carrying out the tasks. On the other hand, the researcher was observing and recording participant’s reactions and comments from the starting of the

A Machine Learning Based Approach to Human Observer Behaviour Analysis in CCTV Video Analytics & Forensics experiment until the end in order to support potential subsequent detailed analysis of eye tracking data. The figure 1 illustrates an overview of the experimental procedure adopted in collecting eye tracking data as described above.

### 3.6 Eye-movement Metrics

We used seven eye-tracking metrics in order to analyse our research objectives and questions.

**Times To First Fixation (Sec):** The time from the start of the stimulus display until the test participant fixates on the DAOI.

**Fixation Duration(sec):** Is the total amount of the time that gaze was fixated within the target area.

**Visit Duration(sec):**Duration of each individual fixation within an AOI or a DAOI.

**Visit Count:**Number of visits within an AOI or a DAOI.

**Percentage Fixated :**Percentage of participants that fixated at least once within an AOI or a DAOI.

**Fixation Count:**The number of fixations in particular AOI or DAOI.

**Time To First Mouse Click:** The time taken before the test participant clicks on an AOI or a DAOI for the first time.

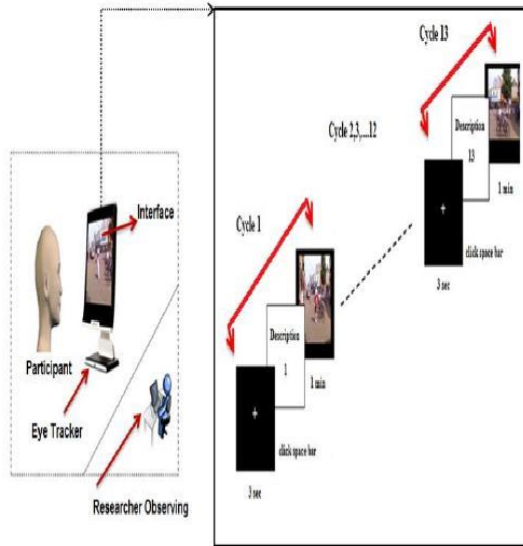


Figure 1: Overview of study setup with regards to data collection & information display

IML '17, October 17-18, 2017, Liverpool, UK

### 3.7 Data capture

During each subjective experiment, i.e. for the experiments of each of the thirteen videos be viewed by each of the 20 participants, seven attributes related to each of the four DAOIs are recorded. Table 3 lists the seven attributed indicating the attribute type as 'numeric'. The attribute 'type' play a key role in the modelling and machine learning experiments to be conducted in the following sections. All attribute values are automatically captured and provided by the Tobii Studio software.

Table 3. Attributes and attribute 'type' for each DAOI

Variables	Parameters	Variable Type
VD	Visit Duration	Numeric
VC	Visit Count	Numeric
TFFF	First Fixation Duration	Numeric
	Total Fixation Duration	Numeric
TFD	Fixation Count	Numeric
FC	Percentage Fixated	Numeric
PF	First Fixation Duration	Numeric

### 3.8 Data pre-processing or preparation

The data captured during the experiment is not fit for the purpose of modelling or machine learning. A close investigation of the captured data revealed that some data values are missing. The missing data is due to issues with regards to some human eye, registration issues with some subjects. When the conceptual and statistical analysis was done in the case of two users whose eyes did not get registered appropriately the entire records of the data were removed and thus excluded from consideration. While this approach is the best when the majority of the data captured for one specific user during a specific experiment is not recorded or lost, if only few values are missing, this approach could lead to substantial data loss. WEKA provides different algorithms to deal with missing data and in the work presented in this chapter, the possibility of using such algorithms were considered.

In addition to the above, not all attributed captured are required for the purposes of modelling and machine learning. Some attributes may not contribute to the models or to the machine learning to be carried out as they are redundant with respect to the creation of any additional knowledge/information. Having such attributes considered in the building of the models can have an adverse impact in terms of reducing the accuracy of the models created. Therefore, in this research the possibility of attribute selection was considered. WEKA has a number of attribute

selection algorithms implemented within itself, when we've tried and tested within the context of the experiments conducted.

**3.8.1 Removal of Missing Values**

In the experiments conducted the filter named RemoveWithValues listed under the preprocessing algorithms in WEKA (*Filters* → *Unsupervised* → *Instances* → *RemoveWithValues*) was used for the removal of missing values. Where data is values are missing in the records this filter removed the whole record, rather than attempting to fill the missing values, which often lead to inaccuracies in the substitute values generated for the missing values.

**3.8.2 Attribute or Feature Selection**

There are many features selection options/ filters implemented within WEKA. The most popularly used approaches are listed in Table 4. All of these approaches were tested on the data that had been captured for this research.

**Table 4. WEKA Feature Selection Methods**

	Method evaluator	Search Method
1	cfsSubsetEval Evaluator	Best First and Greedy Stepwise Search methods
2	ReliefFAttributeEval	Attribute ranking
3	Principal Components	Attribute ranking
4	WrapperSubsetEval	Best First and Greedy Stepwise Search methods

It was found that "cfsSubsetEval" with search method greedy stepwise filters improved the accuracy of data modelling therefore positively impacting on the effective use of data. Unfortunately the rest of the filters did not improve the accuracy of data modelling. It is noted that in the attribute reduction experiments conducted here data captured from all the videos for all DAOIs and all participants were considered. When considering data related to specific videos only it was observed that the features that remained were a subset of the features presented here. The following attributes Time to First Fixation (Sec), Fixation Count, Visit Duration (Sec) and Visit Count were selected as a result of the above selection procedure.

However, insignificant attributes were excluded as expected from use in modelling, thus impacting positively on the accuracy of modelling. One of the key attributes removed was the 'Percentage Fixated %' indicating that the percentage of participants fixating their view on a specific DAOI has no impact on the duration spent of inspecting the said DAOI. By the subject for any given video.

**3.8.3 Dataset Representation**

In this paper, due to limitation of space and purpose of clarity only two representative videos (video 1 and video 10) are used for the detailed presentation of results an analysis, even though the remaining videos were also tested in the research conducted and confirmed to behave in a similar manner.

Samples of data instances for all the DAOIs, i.e. DAOI 1, DAOI 2, DAOI 3 and DAOI 4, for the video 10 and Video 1 are presented in tables 5 and table 6 respectively. The data is arranged in .CSV file format and fed into WEKA for the purpose of modelling. It is noted that our intention is to model Fixation Duration (i.e. the dependent variable) based on the remaining four parameters, namely, Fixation Count, Visit Duration, Visit Count and Time to First Fixation (i.e. the independent variables) for each DAOI.

**Table 5: Sample data selection for video-10**

Time to First Fixation	Fixation Count	Visit Duration	Visit Count	Fixation Duration
1.93	1	0.93	1	0.93
1.94	1	0.93	1	0.93
1.98	1	0.94	1	0.94
1.92	1	0.93	1	0.93
2.02	1	0.94	1	0.94
1.91	2	0.91	2	0.91
1.96	3	0.91	3	0.91
1.88	1	0.91	1	0.91
1.99	1	0.92	1	0.92
1.89	2	0.92	2	0.92
1.98	1	0.93	1	0.93
1.86	1	0.91	1	0.91
1.87	1	0.74	1	0.74
1.88	1	0.79	1	0.79
1.89	4	0.93	3	0.93
2.02	1	0.92	1	0.92
1.92	3	0.93	3	0.93
1.88	1	0.71	1	0.71
1.94	3	0.92	3	0.92

Time to First Fixation	Fixation Count	Visit Duration	Visit Count	Fixation Duration
2.05	2	0.94	2	0.94
2.25	1	0.95	1	0.95
2.10	2	0.93	2	0.93
1.94	1	0.93	1	0.93
2.08	1	0.94	1	0.94
2.38	1	0.94	1	0.94
2.42	1	0.92	1	0.92
1.86	1	0.93	1	0.93
2.98	1	0.91	1	0.91
0.97	2	0.94	2	0.94
2.08	2	0.93	2	0.93
1.4	1	0.95	1	0.95
1.94	2	0.93	2	0.93

Time to First Fixation	Fixation Count	Visit Duration	Visit Count	Fixation Duration
1.9	1	0.9	1	0.9
2.04	1	0.92	1	0.92
1.93	1	0.89	1	0.89
1.92	1	0.98	1	0.98

Time to First Fixation	Fixation Count	Visit Duration	Visit Count	Fixation Duration
2.02	1	0.98	1	0.98

**Table 6: Sample data selection for video-1**

Time to First Fixation	Fixation Count	Visit Duration	Visit Count	Fixation Duration
42.84	2	0.97	2	0.97
42.36	1	0.88	1	0.88
47.77	1	0.93	1	0.93
48.54	5	0.94	4	0.94
49.99	4	0.9	4	0.9
37.74	6	0.43	5	0.43
48.25	2	0.88	1	0.88
39.69	2	0.5	2	0.5
49.23	4	0.43	4	0.43
39.93	2	0.89	1	0.89
48.68	4	0.94	2	0.94
49.54	4	0.85	2	0.85

Time to First Fixation	Fixation Count	Visit Duration	Visit Count	Fixation Duration
39.6	4	0.92	4	0.92
32.89	2	0.91	2	0.91
38.8	2	0.7	2	0.7
47.91	1	0.91	1	0.91
41.82	2	0.92	2	0.92
42.9	2	0.94	1	0.94
42.09	1	0.81	1	0.81
41.82	1	0.93	1	0.93
41.91	2	0.9	2	0.9
49.88	1	0.95	1	0.95
44.89	1	0.91	1	0.91
47.91	2	0.43	2	0.43
47.91	10	0.44	4	0.44
45.94	2	0.95	2	0.95

Time to First Fixation	Fixation Count	Visit Duration	Visit Count	Fixation Duration
48.01	3	0.91	3	0.91
28.13	17	0.98	9	0.98
27.27	22	0.97	9	0.98
44.04	5	1.02	2	0.4
41.95	4	0.93	1	0.95
49.9	4	0.93	3	0.97
47.93	7	0.98	4	0.92
29.98	10	0.95	5	0.94
31.79	8	0.64	4	0.27
39.54	54	0.61	7	0.98
39.9	5	0.91	2	0.98
27.99	14	1.02	8	0.93
40.02	8	0.82	5	0.95
44.9	3	0.42	2	0.27
47.94	7	0.48	1	0.97
46.94	3	0.66	1	0.26
27.07	11	0.99	6	0.2
41.93	4	0.42	3	0.94
32.99	9	0.46	5	0.25
34.97	10	0.42	6	0.35

Time to First Fixation	Fixation Count	Visit Duration	Visit Count	Fixation Duration
32.91	3	0.95	3	0.95
41.93	1	0.95	1	0.95
34.24	12	0.45	6	0.29
38.98	2	0.2	2	0.2
37.92	3	0.93	3	0.93
37.99	4	0.2	4	0.2
31.99	8	0.91	7	0.91
32.34	10	0.89	6	0.9
34.38	11	0.42	10	0.38
37.97	4	0.2	4	0.2
39.24	13	0.98	9	0.97
37.97	4	1.06	3	0.99
37.03	9	0.92	6	0.98
27.98	7	0.42	5	0.39

#### 4 EXPERIMENTAL RESULTS AND ANALYSIS

##### 4.1 Modelling Fixation Duration

Fixation Duration on a DAOI indicates a measure of visual importance of the DAOI in terms of the search task assigned to the subjects, given the videos and the descriptions given [9][10]. The purpose of this modelling exercise is to model the Fixation Duration, hence the visual significance given the task at hand and video of each DAOI based on the parameters that determine the visual attention pattern, such as the Fixation Count, Visit Count, Visit Duration and Time to First Fixation. Given that each video is different with the object of attention appearing in the scene at different times as compared to the viewing start time and the descriptions given are different, the modelling has to be carried out per video, per description but for all subjects.

video-1 and video-10 are presented and analysed. It is noted that similar modelling was carried out for all remaining eleven test video producing similar levels of accuracy of modelling.

Linear Regression is used to model the visual attention on different parts (DAOI 1, DAOI 2, DAOI 3 and DAOI 4) of the object being identified and tracked on both video 1 and video 10 and to thus identify Important features of the eye-gaze patterns and behaviour of participants when presented with the visual surveillance task. Basically, the data provided in table 7 and 8 for video-10 and video-1 respectively, for each DAOIs are the inputs to the linear regression modelling process. The data modelling was carried out using linear regression with Fixation Duration as the dependent variable and the remaining four attributes as the independent variables. Ten-fold cross validation was used to optimize prediction accuracy.

The tables 7 and 8 respectively lists the obtained models for each DAOI of video 1 and 10 and their correlation coefficient as generated by WEKA linear regression for video-1 and video-10 respectively. Note that FC, VD, VC and TTF are denoted as Fixation Count, Visit Duration, Visit Count and Time to First Fixation respectively.

**Table 7: The regression models for video-1**

	DAOI			
	DAOI 1	DAOI 2	DAOI 3	DAOI 4
Linear Regression Model	Fixation Duration= -0.1091 * FC+ 0.7547 * VD+ 0.1213 * VC + -0.0058 * TTF+ 0.3021	Fixation Duration= -0.0315 * FC+ 0.4714 * VD+ 0.065 * VC + -0.0309	Fixation Duration= -0.0541 * FC+ 0.7686 * VD+ 0.0721 * VC + -0.0017	Fixation Duration= -0.08 * FC+ 0.08 * VD+ 0.0966 * VC + -0.0391
Corr. Coeff.	0.9804	0.9619	0.9079	0.9786

**Table 8: The regression models for video-10**

	DAOI			
	DAOI 1	DAOI 2	DAOI 3	DAOI 4
Linear Regression Model	Only one person seen target face	Fixation Duration= -0.13 * FC + 1 * VD + 0.13 * VC + 0	Fixation Duration = 1 * VD + 0	Fixation Duration = 1 * VD + 0
Corr. Coeff.		1	1	1

##### 4.2 Interpretation and Analysis of Regression Models

In this section, a detailed interpretation and an analysis of the regression models are provided in an attempt to identify the observer behaviour via eye gaze attention patterns.

The resulting linear regression functions for DAOI 1, DAOI 2, DAOI 3 and DAOI 4 are the final modelling outcomes produced by WEKA. Following are the models obtained for video 1, with  $f(1)$  representing the FD of DAOI 1,  $f(2)$  representing FD of DAOI2,  $f(3)$  representing the FD of DAOI 3 and  $f(4)$  representing the FD of DAOI 4. Note also that FC, VD, VC and TTF have been replaced by the variables  $x(1)$ ,  $x(2)$ ,  $x(3)$  and  $x(4)$  respectively.

$$f(1)_{DAOI 1} = -0.1091 * x(1) + 0.7547 * x(2) + 0.1213 * x(3) - 0.0058 * x(4) + 0.3021$$

$$f(2)_{DAOI 2} = -0.0315 * x(1) + 0.4714 * x(2) + 0.065 * x(3) - 0.0309$$

$$f(3)_{DAOI 3} = -0.0541 * x(1) + 0.7686 * x(2) + 0.0721 * x(3) - 0.0017$$

$$f(4)_{DAOI 4} = -0.08 * x(1) + 0.08 * x(2) + 0.0966 * x(3) - 0.0391$$

(1)

Following are the models obtained for video-10 using the same notations for FD, FC, VD, VC and TTF (if relevant).

$$f(2)_{DAOI 2} = -0.13 * x(1) + 1 * x(2) + 0.13 * x(3) + 0$$

$$f(3)_{DAOI 3} = 1 * x(2) + 0$$

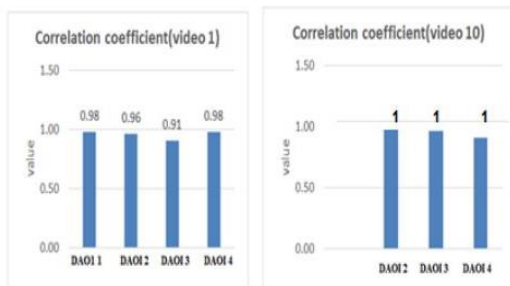
$$f(4)_{DAOI 4} = 1 * x(2) + 0$$

(2)

Given the correlation coefficients listed tables 9 and 10, obtained from the WEKA based modelling process for each DAOI of each of the two videos, it is seen that the Fixation Duration predictions obtained for all DAOIs in both videos

were very accurate, in particular for DAOI 1, DAOI 2 and DAOI 4 of video-1 (all having correlation coefficients  $\geq 0.96$ ) and DAOI 2, DAOI 3 and DAOI 4 of video-10 (all having correlation coefficients of precisely 1). The correlation coefficients for the two models are plotted in Figure 2 for the purpose of ease of comparison.

Overall the very high correlation coefficients obtained for all models justify the proposed use of a Linear Regression model for the prediction of the visual attention to the various parts of the human body.



**Figure 2: Correlation coefficient obtained for each DAOI of video-1 and video-10**

A detailed analysis of the models obtained for each of the DAOIs of video1 can be presented as follows, leading to the identification of a number of interesting behavioral patterns of observers when inspecting the video under the given instructions.

It is noted that when inspecting the video-1, the instruction given to all participants is, *“Find a woman wearing white trouser and carrying a pink shoulder handbag”*. In this specific video the object concerned, first appears in the scene in a far-away location and then comes closer to the camera and disappears at the front of the scene, close to the camera. For ease of reference for detailed analysis, Table 9 tabulates the coefficients obtained for each DAOI, for each independent variable, for video-1.

**Table 9: Models for video-1**

	$x(1)$ FC	$x(2)$ VD	$x(3)$ VC	$x(4)$ TFFF	C.C
DAOI-1 (Head area)	- 0.1091	0.7547	0.1213	-0.0058	<b>0.9804</b>
DAOI-2 (Upper body)	-0.0315	0.4714	0.0065	0	<b>0.9619</b>
DAOI-3 (Carried bag)	-0.0541	0.7686	0.0721	0	<b>0.9079</b>
DAOI-4 (Lower body)	-0.0800	0.0800	0.0966	0	<b>0.9786</b>

Given the instructions to the participants, it is noted that there are three keywords or important features to look for. First to identify that it’s a woman, and then also to identify the white trouser and pink shoulder handbag. Nothing has been mentioned about what the woman is wearing on the upper body.

The values tabulated in Table 9 show that the coefficient for the Fixation Count for all DAOIs is negative. This indicates that higher the Fixation Count, lower will be the Fixation Duration (due to negative correlation). Higher fixation count can indicate an area that is catching the observer’s attention, but intermittently. If the area was actually keeping the observer’s attention in-tact for a reason, then the FC should ideally be low as attention will result in focus for a longer period of time. Therefore, the negative correlation indicated in the models between FD and FC is accurate and justified.

It is also observed that the region of which the attention is mostly impacted by the fixation count is the ‘head area’. This observation is justified as the given description requests to look for a woman, and the actual person is a woman, but with short hair and facial features of a less feminine nature. The participants may have attempted to check for few times that it is indeed a woman as it is not necessarily clear by looking at the face that it is actually a woman. It is also noted that the least impact of FC on the FD is shown in the upper body area as the coefficient is the least. What this means is that there is less correlation between the FC and FD in the upper body, indicating that some people will visit this part more often than others despite that this changing nature of behaviour has less impact on the FD. At this moment as all groups of people, i.e. expert/novice, men/women, have all being considered together it is not possible to say whether this different in behaviour viewing the top body part, depends on the type of person observing.

It is seen that for all the DAOIs, the VD is the parameter that mostly impacts the Fixation Duration. This is indicated by the fact that the largest coefficient is for the VD parameter



A Machine Learning Based Approach to Human Observer Behaviour Analysis in CCTV Video Analytics & Forensics in all equations. The correlation is positive. The Visit Count (VC) is also positively correlated with the Fixation Duration.

When observing the parameters obtained for Time To First Fixation, it is seen that apart from the head area, TTFF do not impact the Fixation Duration of any of the other three regions. Even for the head area, the coefficient obtained is significantly small and can be considered as thus impacting insignificantly. However, what this shows is that the observers may have first tried to identify that it is a woman before they tried to look for other features. The Fixation Duration on the head area, thus depends slightly on the TTFF. Perhaps the individuals who showed a keener interest on the face, looked at it at an early stage and overall spent more time looking at this area, while identifying and tracking the individual. An interesting observation with regards to the ‘carried handbag’ area is that it is the region with the least correlation coefficient, or the least accurate model obtained. This is justified as the handbag is attractive and might receive different amounts of interest from the different observers, making the creation of a simple linear model difficult and hence ultimately less accurate. At this moment, the experiments conducted considers all participants together.

Further experiments are conducted to determine whether there is a different in the way different groups etc., men/women, experienced/novice observe the videos. For the remaining three areas the correlation coefficient of the models obtained are very accurate. This indicates that all observers behaved in a very similar way, when inspecting these regions.

It is seen that even though no description was given to the middle part of the human body being looked for participant attention is being received by this part. However, in this area the magnitude of the coefficients are the smallest, indicating that the Fixation Duration on this area is least impacted by, especially by FC, VD and VC. What this means is that this part of the body, even without being described with a feature is receiving attention from the participants, in a manner not similar to the other three regions. The fact that the lack of clarity whether it is a woman or a man by looking at the face may ensure added attention to this area is a way to justify this observation.

Given the above explanations, it is seen that the Linear Regression models obtained do not only illustrate the possibility to model observer behaviour accurately, but also allows a detailed behavioural analysis to be conducted. A detailed analysis of the models obtained for each of the DAOIs of video-10 can be presented as follows, leading to the identification of a number of interesting behavioral

IML '17, October 17-18, 2017, Liverpool, UK

patterns of observers when inspecting the video under the given instructions

It is noted that when inspecting the video-10, the instruction given to all participants is, “*Find a person wearing a blue top and white pants*”. In this specific video the object concerned, first appears very closely to the camera and disappears from the scene very quickly moving from the right side to the left of the scene.

For ease of reference for detailed analysis, Table 10 tabulates the coefficients obtained for each DAOI, for each independent variable, for video-10.

**Table 10: Models for video-10**

	$x(1)$ FC	$x(2)$ VD	$x(3)$ VC	$x(4)$ TTFF	<i>C.C</i>
DAOI-1 (Head area)	N/A	N/A	N/A	N/A	<b>N/A</b>
DAOI-2 (Upper body)	- 0.13	1.0	0.13	0	<b>1.0</b>
DAOI-3 (Carried bag)	0	1.0	0	0	<b>1.0</b>
DAOI-4 (Lower body)	0	1.0	0	0	<b>1.0</b>

The given description for video-10 requires the observers to only look to recognize the ‘person’ being searched for is wearing a ‘blue’ top and a ‘white trouser’. Although the said individual is wearing a handbag in her shoulder, it is not being requested to inspect this area at all. In the video the said individual only appears for a short duration, entering the scene from the right and leaving it on the left. The depth of view remains also constant and short distance from the camera. Therefore, whatever the observers’ do, an attempt will be made to very quickly identify the individual.

A closer look at the coefficients for each regression equation obtained for the different DAOIs it is observed that it has not been possible to model for the DAOI-1, the ‘head’ region. A closer investigation revealed that only one of the 20 participants ever decided to look at the face and hence the reason that WEKA has declined to create a model. The object concerned is moving fast and is close to the camera that most participants will be very much focused in looking to see whether the person is wearing a blue top and white trouser. It is not required to look at the head area that much as the search does not specify person being looked for is a man/woman.

It is noted that for video-10 the DAOI-2, DAOI-3 and DAOI-4 obtained correlation coefficients are 1. This demonstrates is that all participants behave very similarly when observing each of these areas. The most complicated model (yet with a 1.0 correlation coefficient) obtained is for the upper body area where the model indicated that FC, VD and VC all play a role in FD. The models obtained for DAOI-

3 and DAOI-4 are very simple with FD being directly defined by VD. Therefore, the above analysis shows that the use of Linear Regression to model human observer's attention has been very successful, but in addition also provides one the opportunity to study observer behaviour in much detail.

#### 4 CONCLUSIONS

The experiments conducted in this chapter proved that a simple machine learning model such as Linear Regression can be used to accurately model and analyse observer's attention while conducting a search for identifying and tracking a human with a given description on a given video. The model parameters allow one to analyse observer behaviour in detail, accurately. The additional behavioural analysis that the conceptual and statistical methods did not enable has been made possible by the use of machine learning. It is noted that more complicated machine learning models such as tree based and ensemble learning algorithms may have been able model human behaviour more accurately. However, as Linear Regression has been capable of doing the intended behavioural analysis to be conducted in this paper to a high level of accuracy, no attempt was made to consider such approaches .

In future it is possible to extend the use of more advanced machine learning algorithms to analyse eye tracking data. Such algorithms will help identify more complex and details patterns that may be practically very useful. Further the finding of this research could feed into computer vision research where such algorithms can be modified to mimic human behaviour thus potentially improving performance.

#### ACKNOWLEDGMENTS

The authors would like to acknowledge the support of the Ministry of Manpower of the Sultanate of Oman for providing the funding for this research study and Loughborough University UK for providing research support.

#### REFERENCES

- [1] G. Roffo, M. Cristani, F. Pollick, C. Segalin, and V. Murino, "Statistical analysis of visual attentional patterns for video surveillance," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8259 LNCS, no. PART 2, pp. 520–527, 2013.
- [2] C. J. Howard, T. Troscianko, I. D. Gilchrist, A. Behera, and D. C. Hogg, "Suspiciousness perception in dynamic scenes: a comparison of CCTV operators and novices.," *Front. Hum. Neurosci.*, vol. 7, no. JUL, p. 441, 2013.
- [3] N. Robertson, I. Reid, and M. Brady, "Automatic Human Behaviour Recognition and Explanation for CCTV Video Surveillance," *Secur. J.*, vol. 21, no. 3, pp. 173–188, 2008.
- [4] G. Graham, "The ( Change ) Blindingly Obvious : Investigating Fixation Behaviour and Memory Recall during CCTV Observation," 2016.
- [5] Tobii Technology, "Tobii Eye Tracker." [Online]. Available: <https://www.tobii.com/product-listing/tobii-t60-and-t120/>. [Accessed: 13-Feb-2015].
- [6] Tobii Technology, "Tobii Studio Software." [Online]. Available: <https://www.tobii.com/learn-and-support/learn/steps-in-an-eye-tracking-study/setup/installing-tobii-studio/>. [Accessed: 02-Jan-2016].
- [7] C. J. Howard, I. D. Gilchrist, T. Troscianko, A. Behera, and D. C. Hogg, "Task relevance predicts gaze in videos of real moving scenes," *Exp. Brain Res.*, vol. 214, no. 1, pp. 131–137, 2011.
- [8] M. S. Castelhana, M. L. Mack, and J. M. Henderson, "Viewing task in fluences eye movement control during active scene perception," *J. Vis.*, vol. 9, no. 6, pp. 1–15, 2009.
- [9] a Liaw and M. Wiener, "Classification and Regression by randomForest," *R news*, vol. 2, no. December, pp. 18–22, 2002.
- [10] J. S.K. and S. S., "Reptree Classifier for Identifying Link Spam in Web Search Engines," *ICTACT J. Soft Comput.*, vol. 3, no. 2, pp. 498–505, 2013.
- [11] E. Frank, M. A. Hall, and I. H. Witten, *Data Mining Practical Machine Learning Tools and Techniques*. 2016.
- [12] "MathsIsFun," *Correlation*, 2016. [Online]. Available: <http://www.mathsisfun.com/data/correlation.html>. [Accessed: 22-Jun-2017].