



# Smart Assistive Technology for People with Visual Field Loss

Thesis submitted in accordance with the requirements of the  
University of Liverpool for the degree of Doctor in Philosophy by

**Ola A. Younis**

December 2019





## **Dedication**

*For the wonderful gifts I had from GOD, I dedicate this work for my son Omar and my daughter Aya. I also dedicate this work to my jewel of husband, Mohammad Alomari, who helped me a lot in my PhD journey and made my life full of love and joyfulness, I am truly blessed.*



# Abstract

Visual field loss results in the lack of ability to clearly see objects in the surrounding environment, which affects the ability to determine potential hazards. In visual field loss, parts of the visual field are impaired to varying degrees, while other parts may remain healthy. This defect can be debilitating, making daily life activities very stressful. Unlike blind people, people with visual field loss retain some functional vision. It would be beneficial to intelligently augment this vision by adding computer-generated information to increase the users' awareness of possible hazards by providing early notifications.

This thesis introduces a smart hazard attention system to help visual field impaired people with their navigation using smart glasses and a real-time hazard classification system. This takes the form of a novel, customised, machine learning-based hazard classification system that can be integrated in a wearable assistive technology such as smart glasses. The proposed solution provides early notifications based on (1) the visual status of the user and (2) the motion status of the detected object. The presented technology can detect multiple objects at the same time and classify them into different hazard types.

The system design in this work consists of four modules: (1) a deep learning-based object detector to recognise static and moving objects in real-time, (2) a Kalman Filter-based multi-object tracker to track the detected objects over time to determine their motion model, (3) a Neural Network-based classifier to determine the level of danger for each hazard using its motion features extracted while the object is in the user's field of vision, and (4) a feedback generation module to translate the hazard level into a smart notification to increase user's cognitive perception using the healthy vision within the visual field.

For qualitative system testing, normal and personalised defected vision models were implemented. The personalised defected vision model was created to synthesise the visual function for the people with visual field defects. Actual central and full field test results were used to create a personalised model that is used in the feedback generation stage of this system, where the visual notifications are displayed in the user's healthy visual area.

The proposed solution will enhance the quality of life for people suffering from visual field loss conditions. This non-intrusive, wearable hazard detection technology can provide obstacle avoidance solution, and prevent falls and collisions early with minimal information.

# Acknowledgements

First of all, I would like to record my gratitude to my supervisor Dr Waleed Al-Nuaimy for his supervision, advice, and guidance from the very early stage of this research. Above all, he provided me with unflinching encouragement and support in various ways. I appreciate his time and efforts in reviewing my thesis and research papers.

I would also thank my co-supervisor Professor Fiona Rowe, for her ongoing encouragement, support and advice. I am indebted to her more than she knows. She was always beside me in my tough moments, supporting and encouraging me to finish this journey with the maximum level of professionalism and joyfulness at the same time. I am also grateful for her detailed and precisely pinpointed comments and suggestions while reviewing my thesis and the published work.

I would like to warmly thank my parents who have been a constant source of inspiration to me. It is their love and encouragement that has made this long journey full of joy and has encouraged me to be patient during the PhD period. My special gratitude is due to my brothers and sisters for their loving support. Finally, I would like to thank all the friends and colleagues at the University of Liverpool who constantly provided emotional support and took care of me in many aspects during the four years of studies.

## List of Publications

- Ola Younis, Waleed Al-Nuaimy, Fiona Rowe and Mohammad H. Alomari, “A Smart Context-Aware Hazard Attention System to Help People with Peripheral Vision Loss”, *Sensors*, April, **2019**, Vol. 19, issue 7, pp. 1630
- Ola Younis, Waleed Al-Nuaimy, Mohammad H. Alomari and Fiona Rowe, “A Hazard Detection and Tracking System for People with Peripheral Vision Loss using Smart Glasses and Augmented Reality”, *International Journal of Advanced Computer Science and Applications (IJACSA)*, Feb, **2019**, Vol. 10, issue 2, pp 1-9
- Ola Younis, Waleed Al-Nuaimy, Fiona Rowe and Mohammad Al-Omari, “Real-time Detection of Wearable Camera Motion Using Optical Flow”, in the *IEEE Congress on Evolutionary Computation (CEC)*. **2018**, pp 1 - 6
- Ola Younis, Waleed Al-Nuaimy, Majid Al-Tae and Ali Al-Ataby, “Augmented and Virtual Reality Approaches to Help with Peripheral Vision Loss”, in *IEEE 14<sup>th</sup> International Multi-Conference on Systems, Signals and Devices* **2017**, pp 303 - 307
- Ali Al-Ataby, Ola Younis, Waleed Al-Nuaimy, Majid Al-Tae and Baidaa Al-Bandar, “Visual Augmentation Glasses for People with Impaired Vision”, in *IEEE 9<sup>th</sup> International Conference on the Developments on eSystems Engineering* **2016**, pp 24 - 28

# Contents

<b>Abstract</b>	<b>i</b>
<b>Acknowledgements</b>	<b>ii</b>
<b>Publications</b>	<b>iii</b>
<b>Contents</b>	<b>iv</b>
<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>x</b>
<b>List of Abbreviations</b>	<b>xi</b>
<b>Glossary</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Peripheral Vision Loss . . . . .	4
1.2 Research Motivation . . . . .	9
1.3 Research Aims and Objectives . . . . .	10
1.4 Original Contributions . . . . .	11
1.5 Outline of the Thesis . . . . .	12
<b>2 Literature review</b>	<b>14</b>
2.1 Introduction . . . . .	14
2.2 Substitutive Intervention . . . . .	16
2.2.1 Assistive Technologies using Visual Feedback . . . . .	17
2.2.2 Assistive Technologies using Auditory Cues . . . . .	19
2.2.3 Assistive Technologies using Haptic Cues . . . . .	24
2.3 Compensation Interventions (Adaptive Behavioural Training) . . . . .	28
2.4 Restitutive Interventions . . . . .	29
2.5 Summary and Conclusions . . . . .	30
<b>3 User Requirements and System Design</b>	<b>32</b>
3.1 Introduction . . . . .	32
3.2 User Requirements Elicitation Study . . . . .	33

3.2.1	Questionnaire 1 . . . . .	34
3.2.2	Questionnaire 2 . . . . .	38
3.3	Augmented Reality and Smart Glasses . . . . .	42
3.4	System Design . . . . .	46
3.5	Discussion and Conclusion . . . . .	51
<b>4</b>	<b>Multiple-Object Detection and Tracking</b>	<b>53</b>
4.1	Introduction . . . . .	53
4.2	Datasets . . . . .	56
4.3	Object Detection Using Frame Differencing and Motion Compensation	57
4.3.1	Camera Motion Detection using Optical Flow . . . . .	58
4.3.2	Motion Detection using Stationary Camera . . . . .	63
4.3.3	Motion Compensation for Moving Camera . . . . .	64
4.3.4	Motion Compensation and Object Detection Evaluation . . .	67
4.4	Deep Learning Object Recognition . . . . .	68
4.5	Kalman Filter Object Tracking . . . . .	73
4.6	Assignment Algorithms (The Hungarian) . . . . .	77
4.7	Discussion and Conclusion . . . . .	80
<b>5</b>	<b>Dataset Creation and Hazard Classification</b>	<b>82</b>
5.1	Introduction . . . . .	82
5.2	Hazard Labelling . . . . .	84
5.3	Machine Learning-based Hazard Classification . . . . .	89
5.3.1	Performance Indicators . . . . .	90
5.3.2	Training and Testing Experiments . . . . .	92
5.4	Discussion and Conclusion . . . . .	97
<b>6</b>	<b>Vision Modelling</b>	<b>101</b>
6.1	Introduction . . . . .	101
6.2	Healthy Vision Model . . . . .	102
6.3	Personalised Vision Model . . . . .	104
6.3.1	Image Pre-processing . . . . .	104
6.3.2	Visual Feedback Design to Enhance Visual Perception . . . .	109
6.3.3	Visual Notification Examples . . . . .	112
6.4	Discussion and Conclusion . . . . .	116
<b>7</b>	<b>Conclusions and Future Work</b>	<b>120</b>
7.1	Overall Conclusion . . . . .	120
7.2	Detailed Conclusions . . . . .	121
7.3	Limitations and Strengths of the Proposed Technology . . . .	123
7.4	Suggestions for the Future Work . . . . .	125
	<b>References</b>	<b>129</b>
	<b>Appendices</b>	<b>155</b>

<b>A Ethical Approval Form</b>	<b>156</b>
<b>B Participant Information Sheet</b>	<b>158</b>
<b>C Participant Consent Form</b>	<b>161</b>
<b>D Questionnaire 1</b>	<b>163</b>
<b>E Questionnaire 2</b>	<b>168</b>



# List of Figures

1.1	The consequences of visual impairment on the quality of life. . . . .	3
1.2	The healthy visual field extent for both eyes. . . . .	4
1.3	Retinal anatomy. . . . .	5
1.4	The binocular and monocular visual field of both eyes . . . . .	6
1.5	Simple schematic for causes and results of visual field loss. . . . .	7
1.6	Research work organisation. . . . .	10
2.1	Literature review structure for low vision rehabilitation solutions . . .	16
2.2	Gottlieb and Eli Peli prisms glasses . . . . .	18
3.1	Object types preferences. . . . .	36
3.2	Notification timing preferences. . . . .	37
3.3	Notification format preferences. . . . .	37
3.4	System's output preferences. . . . .	40
3.5	Orcam MyEye 2.0 wearable camera. . . . .	43
3.6	Daqri smart glasses. . . . .	43
3.7	EyeTrek Insight EI-10 glasses attachment. . . . .	44
3.8	The OXSIGHT smart glasses. . . . .	45
3.9	The Epson Moverio BT-200 smart glasses product features. . . . .	46
3.10	The conceptual representation of our system. . . . .	47
3.11	Visual field test result examples. . . . .	48
3.12	Humphrey test printout. . . . .	49
3.13	System overview including the project's main components. . . . .	50
4.1	Moving object detection approaches . . . . .	54
4.2	Moving objects detection and tracking procedure. . . . .	58
4.3	The six-degrees of freedom for a wearable camera. . . . .	59
4.4	Optical flow techniques . . . . .	60
4.5	Frame sub-regions for camera motion classification. . . . .	62
4.6	Foreground detection using Mixture of Gaussian Segmentation. . . . .	64
4.7	Motion compensation and detection pipeline. . . . .	65
4.8	Moving object detection example after motion compensation. . . . .	66
4.9	Single Shot Detector (SSD) architecture. . . . .	69
4.10	A standard convolutional layer (left) and MobileNets depthwise separable convolution (right). . . . .	70

4.11	Accuracy vs time, with shapes represent object detector and colours represent feature extractor. . . . .	71
4.12	Memory usage in megabyte for each architecture/feature extractor configuration. . . . .	72
4.13	Memory usage in megabyte vs. time. . . . .	72
4.14	Detection performance for SSD+MobileNet on datasets used in this work. . . . .	74
4.15	A bipartite graph (a) with its weight matrix (b) and cost matrix (c). . . . .	78
5.1	Examples from the CamVid dataset . . . . .	83
5.2	Epson Moverio BT-200 smart glasses. . . . .	84
5.3	Examples captured by the Moverio BT-200 smart glasses with detection results. (a) indoor video frames; (b) outdoor frames. . . . .	85
5.4	Motion projection of a moving point example. . . . .	86
5.5	An example of the extracted features from one of our testing videos. . . . .	87
5.6	Motion features example extracted from the detection and tracking phases. . . . .	88
5.7	Hazard labelling according to region of interest. . . . .	89
5.8	A visual example of hazard classes. . . . .	90
5.9	The confusion matrix for classification results. . . . .	91
5.10	Neural network structure (Exp1). . . . .	93
5.11	ROC spaces for the NN optimisation (Exp1). . . . .	94
5.12	Correlation coefficients calculations results (Exp1). . . . .	95
5.13	Experiment 2 conceptual diagram. . . . .	96
5.14	Data representation example for Exp2. . . . .	97
5.15	Neural network structure (Exp2). . . . .	97
5.16	ROC spaces for the NN optimisation (Exp2). . . . .	98
5.17	Neural network training performance (Exp2). . . . .	98
5.18	Correlation coefficients results (Exp2). . . . .	99
6.1	Human's visual field extension with its corresponding visual acuity (left eye). . . . .	102
6.2	Resolution maps of the human visual field and the result of the transformation method on an example image. . . . .	103
6.3	The personalised vision model inputs. . . . .	105
6.4	The personalised vision model image processing. . . . .	107
6.5	Tunnel vision model example. . . . .	108
6.6	Left hemianopia vision model example. . . . .	109
6.7	Left central scotoma (AMD) vision model example. . . . .	110
6.8	Visualisation scheme examples. . . . .	111
6.9	Visual feedback schemes proposed by Zhao et al. . . . .	112
6.10	Hazard levels representation. . . . .	113
6.11	A system output example using tunnel vision model (1). . . . .	115
6.12	A system output example using tunnel vision model (2). . . . .	116
6.13	A system output example using left hemianopia model (1). . . . .	117

6.14	A system output example using left hemianopia model (2). . . . .	118
6.15	A system output example using AMD model (1). . . . .	119
6.16	A system output example using AMD model (2). . . . .	119
7.1	Future work integrating the user interface and smart glasses. . . . .	127
7.2	Time frame of future work. . . . .	128

# List of Tables

1.1	Examples of visual field loss . . . . .	8
3.1	Participants' ratings of the level of danger for different hazard scenarios	39
4.1	Performance of the camera motion classification . . . . .	63
4.2	Detection performance of SSD on different in-house datasets used in this work . . . . .	73
5.1	Hazard classes specifications . . . . .	88
5.2	Experiment 1 classification results . . . . .	100
5.3	Experiment 2 classification results . . . . .	100
6.1	Visual notification format . . . . .	114

# List of Abbreviations

**2D:** Two dimensions

**3D:** Three dimensions

**ACC:** Accuracy

**AMD:** Age-related macular degeneration

**AR:** Augmented reality

**AT:** Assistive technology

**CamVid:** Cambridge-driving Labelled Video Database

**FNR:** False negative rate

**FoV:** Field of vision

**FPR:** False positive rate

**FPS:** Frame per second

**HVFA:** Humphrey visual field analyser

**IMU:** Inertial measurement unit

**IoU:** Intersection over union

**KF:** Kalman filter

**MSE:** Mean square error

**MVC:** Motor vehicle collision

**NN:** Neural network

**OT:** Occupational therapist

**PV:** Peripheral vision

**PVL:** Peripheral vision loss

**QoL:** Quality of life

**RFID:** Radio frequency identification

**ROC:** Receiver operating characteristics

**RoI:** Region of interest

**SPC:** Specificity

**SSD:** Single shot detector

**TNR:** True negative rate

**TPR:** True positive rate

**UCD:** User-centred design

**VR:** Virtual reality

**VRQoL:** Vision related quality of life

**VRT:** Vision restoration therapy

**WHO:** World Health Organisation

**YOLO:** You only look once

# Glossary

**Assistive technology:** A particular type of technologies developed to aid people.

**Augmented reality:** A technology that places computer-generated contents on the top of one's normal vision creating a mixed (virtual and real) view.

**Binocular vision:** Vision wherein both eyes work together to deliver a 3D view of a scene.

**Bitemporal hemianopia:** A case of hemianopia where vision is impaired in the outer half of both the right and left visual field.

**Central vision:** One type of human's vision that extends less than 30 degrees around a fixation point.

**Context-aware systems:** A particular type of systems that can interact with the surrounding and change its outputs accordingly.

**Depth camera:** A special type of cameras that can capture depth information.

**Ego-motion:** The movement of a system (camera) in the three dimensions.

**Euclidean distance:** The distance of a standard straight line between two points in Euclidean space.

**Foveal vision:** Part of human's vision that extends less than 5 degrees around a fixation point.

**Frame rate (fps):** The number of frames a device can process in one second.

**Hemianopia:** A case of vision impairment where vision is lost in half of the visual field.

**Homography matrix:** In computer vision, it is a matrix used to relate two images in the same planner.

**Homonymous hemianopia:** A case of hemianopia where the vision is impaired in the same half of both the right and left visual field.

**Macular vision:** Part of human's vision that extends less than 18 degrees around a fixation point.

**Monocular vision:** Vision wherein each eye works separately to deliver a 2D view of a scene.

**Occupational therapy:** A healthcare profession that helps people who suffer from health problems perform daily activities.

**OpenCV:** A library of programming methods used mainly to perform computer vision algorithms.

**Parafoveal vision:** Part of human's vision that extends less than 8 degrees around a fixation point.

**Peripheral vision:** Part of human's vision that extends more than 18 degrees around a fixation point.

**Real-time systems:** A software system that operates and produce output in actual time.

**Regression analysis:** A statistical method used to study the relationship between two or more variables.

**RGB,RGB-D sensors:** A camera that captures coloured images (red, green and blue), or coloured -depth images.

**Scotoma:** A blind spot inside a healthy vision area.

**Smart glasses:** Computer-enabled eyeglasses equipped with a camera and display units.

**Tango device:** An augmented reality-based device developed by Google.

**User requirements elicitation study:** A study conducted before the development of a system aims to gain an initial understanding of users' needs.

**VGG network:** A deep-learning-based convolutional network trained by Oxford's visual geometry group for the object recognition task.

**Virtual reality:** The technology of creating virtual worlds that the user can interact with.

**Vision eccentricity:** A visual phenomenon represents the reduction of visual acuity concerning visual field extension.

**Visual acuity:** Vision resolution.

**Visual field:** Vision extension.



# Chapter 1

## Introduction

According to the World Health Organisation (WHO), 1.3 billion people have impaired vision worldwide [1]. In the UK, almost two million people are living with different degrees of sight loss and more than 360,000 are registered as blind or partially sighted [2]. Regrettably, these numbers are expected to increase.

Vision impairment consequences have a direct and negative impact on the individuals' quality of life (QoL) [3–5]. It is hard to have a universal definition for QoL that can be applied in different contexts. However, the WHO [6] defines QoL as: “An individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns. It is a broad-ranging concept affected in a complex way by the person's physical health, psychological state, personal beliefs, social relationships and their relationship to salient features of their environment.”

From this definition, researchers in the field of optometry and ophthalmology studied and evaluated vision-related quality of life (VRQoL) as a set of patients' concerns related to their QoL with vision impairment [7–9]. Previous research studies have shown that vision impairment is associated with a decrease in the employability rate, driving and self-navigation ability and scene perception. Furthermore, vision

impairment has been connected with an increase of Motor vehicle collision (MVC), depression rate, and falling while walking.

It has been reported that only 25% of registered blind and partially sighted people in the working-age are employed with a notable decrease in this proportion from 2015 to 2017 [10]. Driving a car and self-navigation difficulties are common in people with vision impairment. To legally be able to drive a car, a person should meet the minimum eyesight standards by having a good visual acuity and a sufficient field of vision (FoV) (in most countries the legal horizontal visual field required for driving is 120°). People with impaired vision could have lower vision acuity or narrower FoV, thus making it harder for them to be able to use their car [11]. Another increment factor in the QoL for people with visual impairment is their ability for scene perception and object recognition. This condition is associated more with visual field defects conditions [12, 13].

Studies show that people with impaired vision (especially visual field defects as a result of glaucoma) are more likely to be involved in MVC. In their retrospective cohort study [14], the authors mentioned that it is twice as likely for drivers with severe visual field defects to have at-fault MVC compared to those with less visual field impairment.

The prevalence of depression and anxiety disorders is significantly higher in people with vision impairment (especially in elderly) compared to healthy vision people [15–17]. Also, falling while walking occurs more in older people with low vision [18]. In their study about the possible causes of falls [19], the authors mentioned that two of the five factors identified to have a direct association with falls are the impact of sight impairment. As a result, the mentioned consequences of visual impairments may cause a further decline in the QoL and could lead to more mental and physical problems. Figure 1.1 shows the most critical consequences of visual impairment on the quality of life.

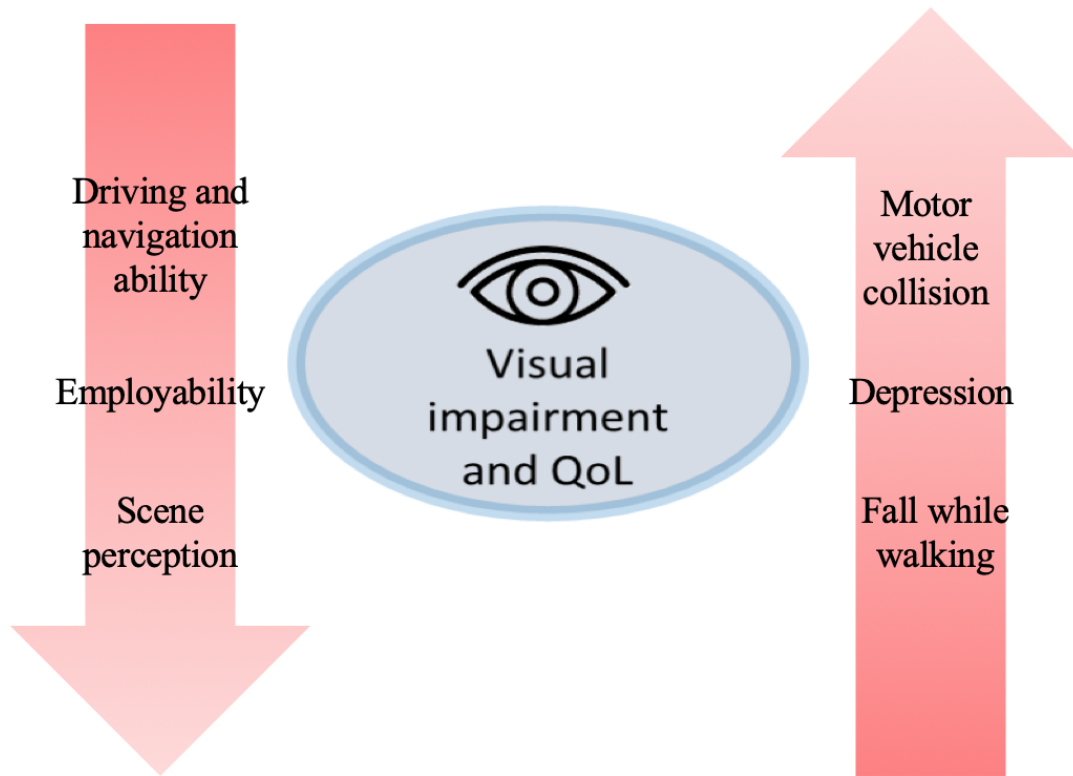


Figure 1.1: The consequences of visual impairment on the quality of life.

Globally, the main reasons of vision impairment are uncorrected refractive errors, cataract, age-related macular degeneration (AMD), glaucoma, diabetic retinopathy, corneal opacity and trachoma [1].

Visual impairment types could be classified -according to the kind of damage it causes- into two groups: visual acuity defects and visual field defects [20]. Reflective errors are the leading cause of low visual acuity, which can lead to several defects, such as nearsightedness (myopia), farsightedness (hypermetropia), and astigmatism. On the other hand, visual field loss could be classified according to the location of the vision loss, and into central and peripheral vision loss.

While most visual acuity problems are correctable using different techniques and traditional solutions such as eyeglasses, visual field defects are not easily rehabilitated. This is because most of these defects happen after brain injuries or eye

conditions where parts of the visual system become permanently diseased [21].

## 1.1 Peripheral Vision Loss

Two types of vision areas define a human's visual field as shown in Figure 1.2: central and peripheral. These areas are used to see and recognise different levels of details and information.

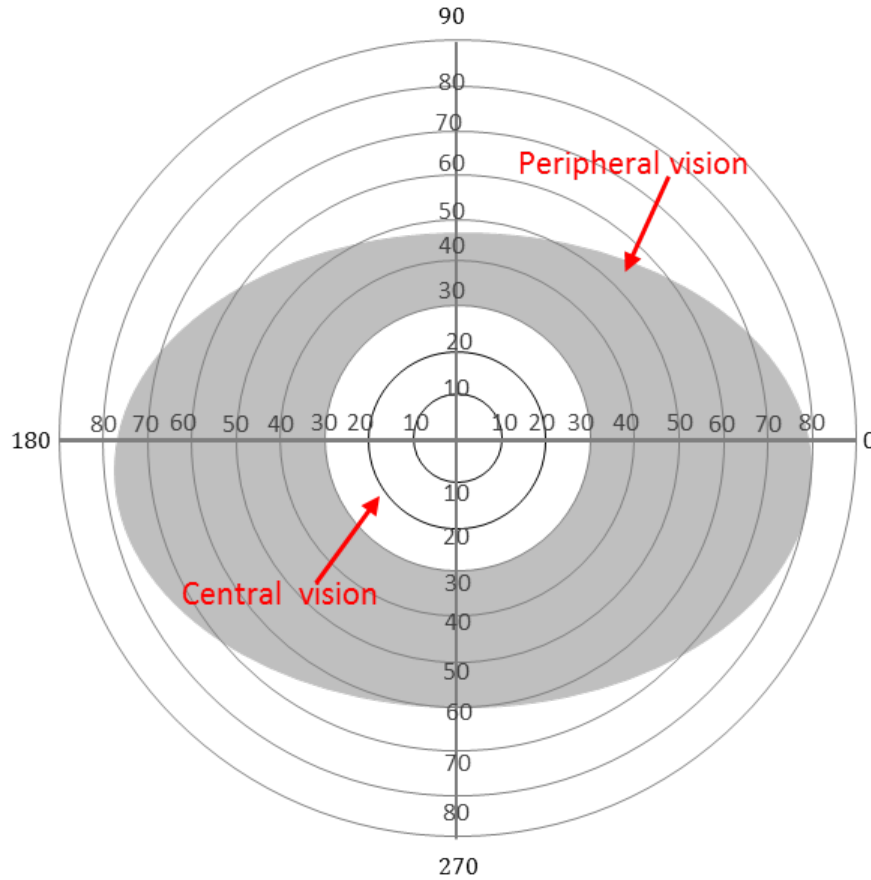


Figure 1.2: The healthy visual field extent for both eyes, in accordance with [22].

As shown in Figure 1.3, the retina is the light-sensitive tissue layer in the back of the eye that has two areas: the macular area which is responsible for the central vision and the peripheral area that is responsible for the peripheral vision. The figure also shows the two types of receptors included the retina, rods and cones.

The central vision contains the highest density of cones [24]; thus, it is capable of colour vision and is responsible for high spatial acuity (fine details). This proportion decreases when we move from the centre of the vision towards the peripheral vision, where we have a high concentration of rods and the lowest spatial resolution [25].

Our brain uses the most central visual field ( $5^\circ$ ) mostly for reading, focusing, drawing, crossing the road, and many other daily activities that require a deep understanding of specific details. On the other hand, the peripheral vision (PV) is used to detect larger contrasts, colours and motion and extends up to  $160^\circ$  horizontally and  $145^\circ$  vertically for each eye [26]. While the PV is inferior to the central vision in terms of detailed view, it is particularly useful to attract the brain's attention to the surrounding environment.

One of the critical roles that a human's PV provides is the ability to detect and avoid potential hazards in the surroundings. To explore the fine details about a specific object, humans use head movements to gather more information and increase their cognitive understanding.

The central vision is shown as a white circle in the middle covering  $30^\circ$  around the fixation point (assumed to be the centre of the figure), while the grey area represents the peripheral visual field. Both visions are used with different functionalities for

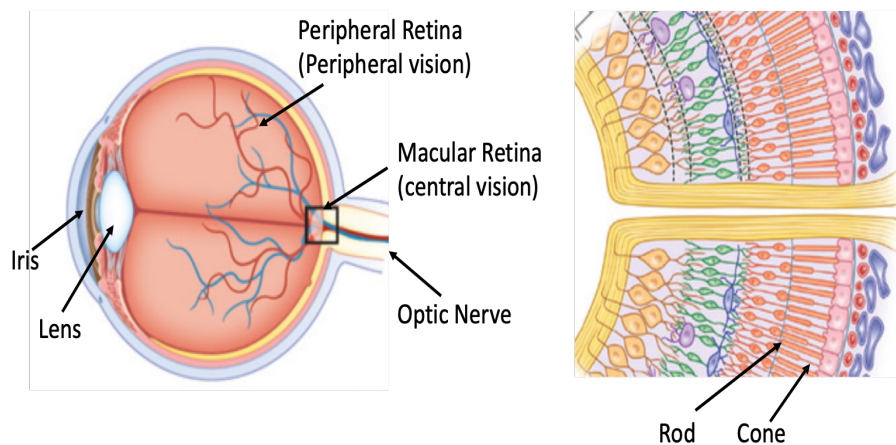


Figure 1.3: Retinal anatomy [23].

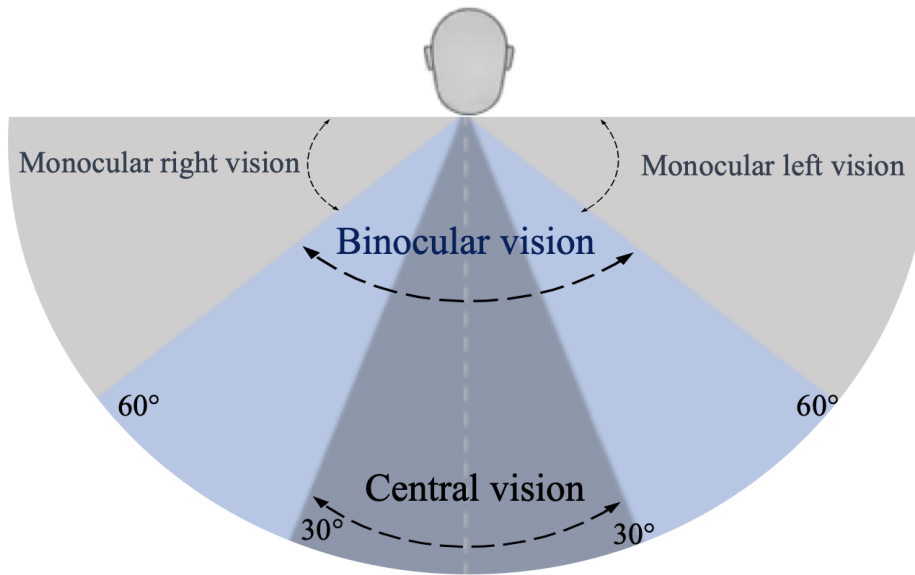


Figure 1.4: The binocular and monocular visual field of both eyes, in accordance with [27].

each of them to explore the real-world [28]. Due to retinal eccentricity [29], different degrees of resolution occur in different parts of the visual field areas. The more central the area, the more resolution for vision [30].

Figure 1.4 depicts the binocular (3D) and monocular (2D) visual field extension of both eyes. As shown in the figure, the visual fields of both eyes widely overlap in the central vision (binocular visual field). This overlap allows us to see fine details and recognise the 3D shape of objects. Binocular vision extends for both central and peripheral visual fields.

The leading causes for visual field loss are ocular problems (e.g. glaucoma, AMD, retinopathy, papilloedema, optic neuropathy) and brain injuries (e.g. stroke, tumour, trauma, inflammation). These conditions result in several types of defected vision such as constricted circular (tunnel) vision, scotoma, arcuate vision, altitudinal vision, hemianopia (bitemporal /homonymous) vision or quadrantanopia vision [31].

Figure 1.5 depicts a simple schematic of the leading causes and types of visual field loss.

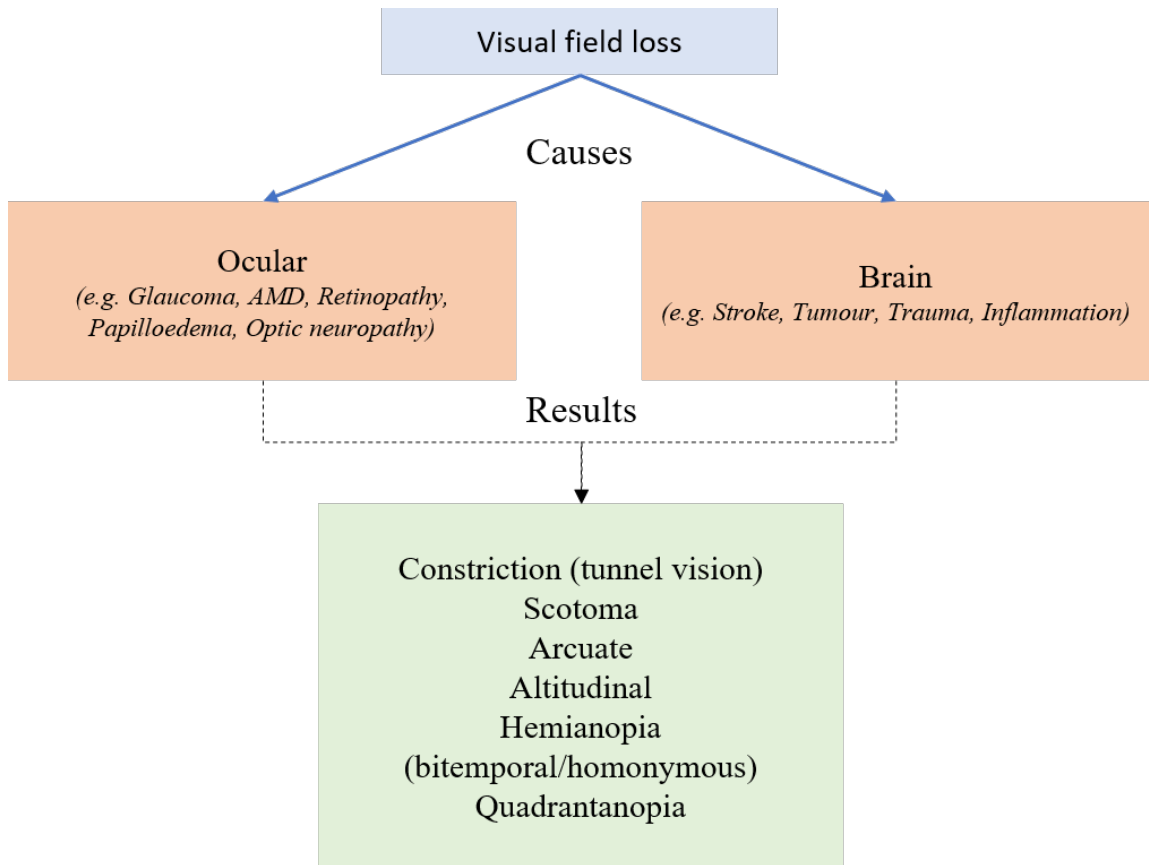








Figure 1.5: Simple schematic for causes and results of visual field loss.

In the case of peripheral vision loss (PVL), the outer visual field areas are impaired to varying degrees, while central vision may remain healthy. This defect can be debilitating and makes a person's daily life very stressful. People with PVL find it hard (or illegal in some cases) to drive and navigate [32–35], recognise their surrounding scene [28, 36] and avoid possible hazards. Understanding these challenges can help in rehabilitation and making the best use of the remaining vision. In this case, it is essential for the visually impaired people to continuously shift their focus around to have a full understanding of the surroundings and possible threats [37, 38]. Table 1.1 shows five types of visual field defects with a simulated view of each type.

Table 1.1: Examples of visual field loss

Condition	Description	Illustration example (left eye, right eye)
Healthy vision	An image captured by a wide-angle camera illustrating a healthy vision.	
Tunnel vision	This is the extreme case of peripheral vision loss. People with tunnel vision can see through a tiny circular area in their central vision ( $\approx 10^\circ$ ).	
Central Scotoma	Blind area in the middle of the visual field. The central vision (e.g. within $30^\circ$ ) is damaged due to AMD or papilloedema.	
Left Hemianopia	A homonymous loss to the same side (left or right) in both halves of the visual field.	
Altitudinal vision	Either the upper (superior) or lower (inferior) half of the visual field is affected.	
Quadrantanopia vision	Quadrant of the visual field is affected. Could be homonymous or bitemporal quadrantanopia.	



## 1.2 Research Motivation

It was shown in the previous section how visual field loss can affect the patient's daily tasks and decrease the overall quality of life. Understanding the surrounding environment and avoiding possible hazards in the correct time are crucial tasks in a person's life. People with visual field loss have good vision in some parts of their visual field, so it would be beneficial to use this vision smartly by adding computer-generated information to increase their awareness by providing early notifications about possible hazards.

This can be achieved by designing a system that implements computer vision algorithms in real-time, to provide useful information about any possible threats existing in the user's blind area. This will enhance functional vision by giving cues about the affected field without the need for the visually impaired people to shift their fixation point all the time. The additional signals must provide fast and trustful notifications that reflects the hazard type, danger degree, and most importantly, the location of that hazard.

Intelligent assistive technologies and mobile healthcare systems are developing rapidly. With the massive growth in the hardware and software sectors, wearable smart devices have become widely affordable. Vision assistance devices have been developed to be worn on several body parts such as the head, chest, fingers, feet, and ears.

In this research, the focus is on the design, development, validation, and evaluation of a smart hazard attention to help people with visual field loss. The approach in this research work is on designing a system that recognises objects in the user's visual field and classifies them to determine the possible danger level.

The system design process followed in this research can be depicted as shown in Figure 1.6. Starting from user requirements, we aim this process to make efficient use of the available smart glasses, computer vision algorithms, machine learning and

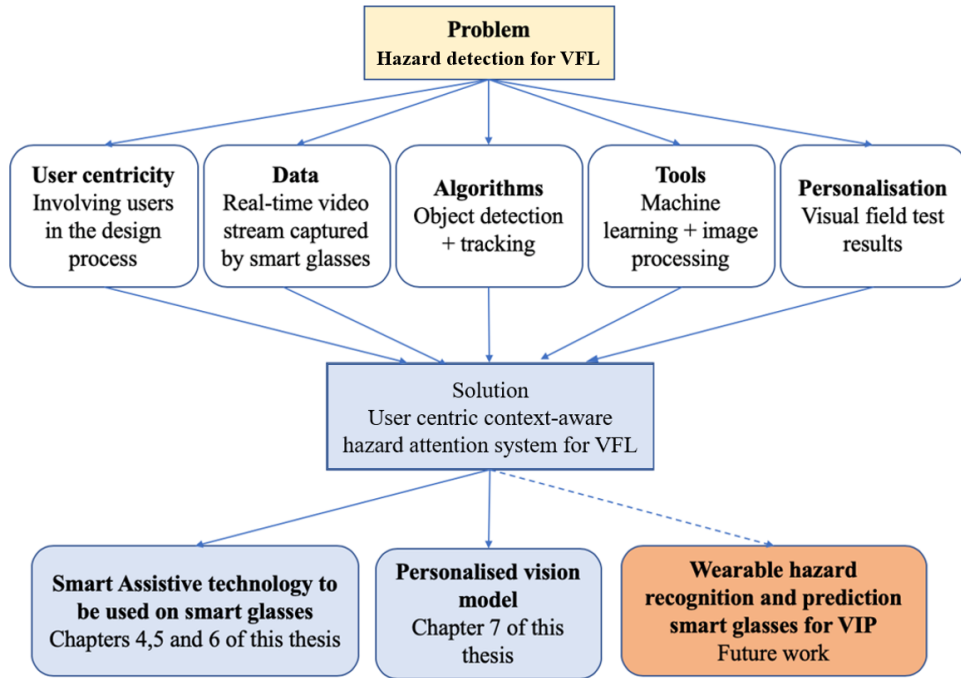


Figure 1.6: Research work organisation.

image processing tools and visual field test results to develop a user-centric, smart hazard attention system for people with visual field loss. This thesis presents the assistive technology and the personalised vision model that will be used in the future for designing wearable hazard recognition and prediction smart glasses for visually impaired people.

### 1.3 Research Aims and Objectives

The main aim of this research is to develop a user centric, context-aware hazard attention system for people with visual field loss. The outcome of this research will be mainly the development of smart assistive technology to detect, recognise, track and classify hazards in the user's peripheral area and produce a smart and meaningful notification to alert the user in the right location at the right time.

Initially, the target is to use computer-enabled smart glasses equipped with a

wide-angle camera. Our proposed system augments users' existing healthy vision with proper, meaningful and intelligent notifications to attract the user's attention to possible obstructions or hazards in their peripheral field of view.

As a summary, this research attempts to fulfil the following objectives:

1. To develop a smart assistive technology with the potential to aid people with vision loss that provides real-time alerts of possible obstacles or hazards.
2. To provide a wearable, affordable and unobtrusive solution that could be used on smart glasses for daily activities.
3. To use Artificial Intelligence techniques for detecting and classifying possible hazards around the visually impaired people and assign different danger level for each one.
4. To provide personalised alerts based on the user's particular visual field loss and preferences.
5. To assist vision specialists and ophthalmologists in their work by developing a vision system that displays what patients are seeing based on their visual field test results and share this with others to understand their defects and help them in their rehabilitation.

## 1.4 Original Contributions

The original contributions presented in this thesis can be summarised as follows:

1. Different from available obstacle detection solutions; a smart technology is developed to increase cognitive awareness for people who have vision impairment using computer vision and machine learning algorithms.

2. A vital user engagement with visually impaired users; which directed us into the daily challenges they face. Also, their invaluable suggestions and preferences for the feedback style and timing helped us to design a wearable, smart assistive technology for people with different types of visual field loss. We believe that such information would greatly benefit other researchers developing similar assistive applications for visually impaired people.
3. A unique egocentric indoor and outdoor hazard recognition dataset is created using a wearable camera. Detected objects in this data are classified using a deep learning object detector and tracked using Kalman Filter.
4. A motion model that describes the hazard type in the user's environment based on motion features extracted from the detection and tracking phases is presented. This model is used in the classification stage later.
5. Machine learning-based hazard classification system, using motion features for multiple hazards is proposed to provide a smart and early warning system to help people with peripheral vision loss.
6. A personalised vision model is implemented using the visual field test results to visualise the visual case for people with impaired vision.

## 1.5 Outline of the Thesis

The rest of this thesis is organised as follows:

- Chapter 2 provides an extensive literature review for the recent research on assistive technologies for visually impaired people.
- Chapter 3 explores the user's requirements and the fundamental needs, challenges and preferences gathered from a patient group.

- Chapter 4 discusses methods used for object detection and tracking as the first and second modules of the proposed system. In this work, two different approaches were used to detect the objects before tracking it. The pros and cons of each method are discussed and evaluated. The implementation and evaluation of the object tracking stage in addition to the assignment algorithm used to handle the multi-object tracking problem are presented.
- Chapter 5 demonstrates the datasets (private and public) used in this work. It discusses the reasons for creating our private dataset and the way we captured the photos and the human perception of hazards. It presents the motion features that were extracted to be used in the machine-learning phase to determine the level of danger for each object. Additionally, this chapter demonstrates the implementation and evaluation of the proposed hazard attention systems using machine learning algorithms. It compares the performance of the proposed method using two different approaches to predict the level of danger for each tracked object.
- Chapter 6 discusses the vision model (normal vision and defected vision). This part is used to visualise the different visual field loss types based on actual visual field tests. The feedback generation stage is presented in this chapter with three different visual field loss examples.
- Chapter 7 discusses the overall and detailed conclusions, in addition to the main limitations, the originality of the proposed work and suggestions for future work.

# Chapter 2

## Literature review

### 2.1 Introduction

People who have peripheral vision loss constitute about 25% of the total number of patients with low vision, while 75% have central vision loss [39]. As a result, the majority of rehabilitation methods and assistive technologies are directed to the people with central vision loss [40].

When reviewing the literature research work, we found that low vision rehabilitation solutions could be broadly classified into two main categories: (1) visual field loss aids, and (2) vision impairment (partial/total blindness) aids.

For visual field loss problems, scientists and clinicians followed three different approaches: (1) using prisms established in eyeglasses (2) training patients how to efficiently use the residual vision to compensate for vision loss, and (3) using assistive technologies that provide additional information to the users using their healthy vision. All these solutions aim to compensate the visual field and increase the individual's surrounding awareness.

On the other hand, assistive technologies (AT) that help with vision impairment are widely used and their advantages are well documented [37, 41, 42]. Assistive

technology is a term that could be used with any equipment or technology that enhances an individual's quality of life and increases his/her engagement and inclusion in society [43].

According to this definition, AT can be classified into two main groups: (1) traditional AT (e.g. prisms, occupational therapy, white canes, walkers, eyeglasses), and (2) mobile IT-based AT (e.g. navigation devices, screen readers, object and face recognition devices). Low vision and total blindness are used interchangeably in the field of AT to refer to any visual condition that impairs an individual's ability to perform daily tasks.

Since visually impaired people have difficulties using visually demanding devices, scientists started investigating other options for AT development. Non-visual sensory modalities such as speech recognition [44], text-to-speech [45], haptic feedback [46], multimodal input [47, 48] and gesture recognition [49] are used to make mobile devices and AT more accessible and suitable for visually impaired people.

While reviewing the literature, it was hard to find a comprehensive classification method to cover all previous work in the field of vision impairment rehabilitation solutions. Some reviews categorise the vision impairment aids according to the primary function they perform into: (1) navigation and wayfinding, (2) obstacle detection and (3) scene perception methods.

Other researchers divided these aids based on the capturing device into (1) sensor-based and (2) camera-based solutions. Feedback style could be used to classify visual assistance technologies into audio or haptic solutions. Further classification could be applied to these technologies and methods according to their weight, cost and coverage area.

In this chapter, we present and discuss the previous research work and the state-of-the-art methods to help with visual field loss in specific, and vision impairment in general. Figure 2.1 depicts the structure followed in this review.

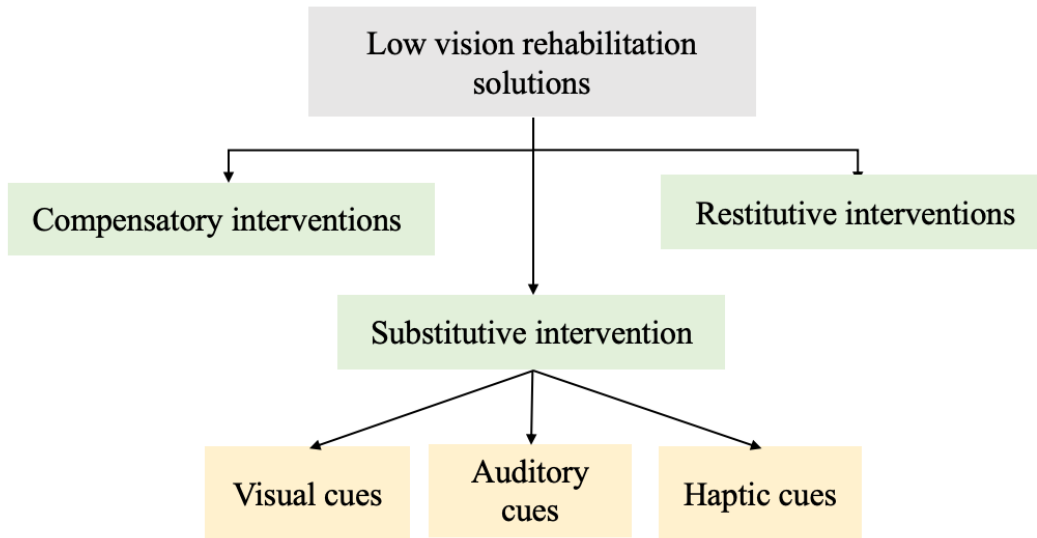


Figure 2.1: Literature review structure for low vision rehabilitation solutions

Firstly, substitutive interventions will be discussed covering aids that replace the lost vision with new information generated by an external source. Aids in this category are classified according to cues style they provide into: (1) visual-based, (2) auditory-based and (3) haptic-based solutions.

In auditory and haptic cues subsections, aids will be presented according to the function they perform as mentioned early in this section. In the second and third sections, compensation and restitution treatments are discussed, respectively.

## 2.2 Substitutive Intervention

As mentioned in the first chapter of this thesis, several ocular and brain causes could lead to visual field loss. Unfortunately, there are no guaranteed solutions, such as traditional eyeglasses or surgeries, to help with this condition. Consequently, ophthalmologists and scientists looked for other solutions to enhance visual perception by providing additional information about the surrounding environment.



Substitution interventions include solutions that help in vision loss adaptation using electronic aids, mechanical devices or any alteration to the visually impaired environment [50]. This section will study the reported research work that provided substitutive solutions for visual field loss people. Furthermore, solutions that proposed smart assistive technologies to help this group of low vision people will be presented.

### 2.2.1 Assistive Technologies using Visual Feedback

Early attempts for extending individuals FoV have been developed using eyeglasses [51]. The goal is expanding an individual's FoV to shift the peripheral field of view inward. This shifting with the users scanning would enhance the overall functional field [39].

#### Prisms

In 1979, Mehr & Quillman [52] presented the idea of using reversed telescopes to increase individuals visual field area in patients with retinitis pigmentosa (RP). Later, in 1998, Zlyk et al. [53] used amorphous lenses to compress images with a wide field of vision in the central vision to present more peripherally located information. However, both techniques were reported to be inadequately used by some peripheral vision loss users [54].

Another approach for expanding the FoV using prisms was presented by researchers several decades ago. Jose et al. [51] introduced this idea in their research work in 1976. The development prismatic systems applications to expand visual field has challenged clinicians and researchers for many decades.

The visual field awareness system (Gottlieb lens) [55] is one of the well-known methods for using prisms with visual field loss people. The commercially available product that was proposed by Gottlieb et al. in 1992 abandons the traditional equal prism in both lenses and use a small round one on the side of the visual field. Users will become aware of their surroundings as displace objects in the blind field over

the seeing field.

The same idea was adopted by Peli's research group at Harvard medical school. In their work [56, 57], Peli et al. developed glasses with high-power prism segments that are located above and below the user's fixation point giving him/her a glance about missing information in the periphery area. By moving gaze between the overlaid prism shift and healthy vision, users would be able to increase their surrounding awareness, as shown in Figure 2.2 shows the two prism glasses.

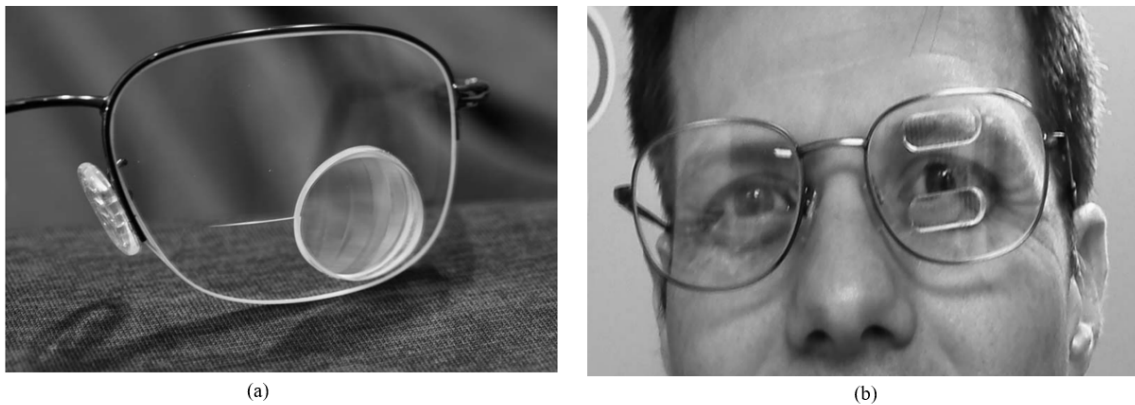


Figure 2.2: Gottlieb prism [55] (a) and Eli Peli prisms [56, 57] (b)glasses

Although the use of prism glasses helped increase visual field expansion to include  $\approx 20^\circ$ , the use of these glasses was poorly adopted by the patients [39, 54].

### Smart glasses

In 2001, a group at Harvard developed a device that produced augmented reality (AR) vision for people with severe peripheral vision loss (tunnel vision) [58]. The scheme comprises a wide-angle camera and one display unit that projects a processed image (cartoon style) from the camera on the healthy vision area. The device was tested on healthy and vision impaired people and results showed improvements of self-navigation and object finding.

Elango & Murugesan [59] proposed their work of using AR to extend visually affected patients' knowledge using the cellular neural network. The study presented

a model consisting of a monochrome camera, display unit and a portable processor to perform image processing.

There are two main weaknesses of these two solutions: (1) The displayed images simply replicate everything the camera can see into the user's central vision, consequently, producing unnecessary information. (2) The solutions created a double vision that could cause distraction and reduce the efficiency of the user's healthy vision.

More recently, Pundlik et al. [60] proposed their collision warning device for peripheral vision loss people. The proposed prototype consists of a portable, battery-powered video camera attached to a processing unit. The device detects obstacles in the user's environment and computes the time to collision (TTC) [61] for each object. Simple audio warning messages are generated only in the event of a possible collision. The proposed work was evaluated with 25 participants (12 Hemianopia and 13 tunnel vision). All participants walked through a pre-designed obstacle course under two conditions: with and without the proposed device. Results show a significant reduction in collisions by approximately 37% with the device.

This device is considered to be a pioneer solution that applies computer vision algorithms on wearable devices for peripheral vision loss people. However, its main weakness is that it only detects stationary obstacles, where the real world scenarios are more complex. Obstacles may move around the visually impaired, changing their level of danger. Moreover, this device does not distinguish between healthy and blind areas assuming that the user has total blindness.

### 2.2.2 Assistive Technologies using Auditory Cues

#### Navigation and wayfinding

Many research studies were reported in the literature discussing navigation and mapping problems for visually impaired people using audio-based feedback sys-

tems [62–64].

In 2007, the stereo vision-based electronic travel aid (SVETA) system was proposed to help blind people in their navigation [65]. The authors used a stereo camera to capture the surrounding environment and determine the location and distance of the obstacles in the user’s navigation path. The output of the proposed system is delivered through stereo earphones. The input sensor and the output unite are all connected to a compact computing device that is placed in a specially designed pouch. While the system’s tests proved its applicability in helping blind users, outdoor environment challenges and the slow system’s performance could be considered as critical issues for the presented method.

In 2014, Fiannaca et al. presented Headlock [66]; a wearable device to assist blind people in traversing open spaces. The system used Google glasses and OpenCV blob detection algorithm to detect doors and guide the blind person towards them with minimum veering and the shortest path. Although the presented work provided quantitative and qualitative results after testing the system’s usability with blind subjects, limiting the object detection and navigation to doors only make it inefficient for hazard avoidance or general blind navigation systems.

Positioning and mapping techniques are used to enhance visually impaired mobility. An indoor navigation system for visually impaired and elderly people based on radio frequency identification (RFID) was proposed by Tsirmpas et al. [67]. The authors used passive RFID tags by installing them in different locations in the user’s environment. The proposed model unites a reader and ultra-sonic finder to detect both tags and obstacles using a wearable prototype.

In their study, the authors suggest cells of  $40 \times 40$  cm to install the RFID tags, which is considered to be a short range and consequently require adding more tags in large environments. This weakness of RFID-based systems makes it costly to be considered as a navigation system for visually impaired people.

Indoor navigation system based on visual simultaneous localisation and mapping (SLAM) algorithm was developed by Jinqiang et al. [68] in 2018. The system addresses three main problems: (1) user localisation, (2) goal recognition, and (3) navigation and obstacle avoidance.

The proposed device consists of different sensors such as depth and fisheye cameras, an ultrasonic rangefinder and AR glasses with an embedded processing unit. The proposed navigation system requires a sighted person to wear it before the visually impaired person can use it. This step is necessary to build the "virtual-blind-road". The needed information for goal navigation and obstacle avoidance is delivered using both images on the AR glasses and audio feedback through the earphone.

### **Obstacle detection and avoidance**

An obstacle stereo feedback (OSF) system [69] was implemented using a depth sensor and computer vision algorithms to guide blind or impaired vision users in indoor navigation. The hand-free system uses depth information for obstacles in front of the user and produces acoustic notification when necessary. The developers also provided Head-Related Transfer Functions (HRTF) to their system to create a more realistic and 3D stereo sound environment that represents the detected obstacles. Moving objects were not tracked in this system, limiting the detection stage to stationary obstacles only. Also, the user's motion was not considered in this work. This could affect the of the hazard detection in general.

Kang et al. [70–72] proposed a new method for detecting obstacles called "deformable grid (DG)". The authors used this method with their wearable prototype, which consists of a monocular camera, WIFI module, battery and a Bluetooth earphone — all attached to standard eyeglasses. The camera sends the captured video to a laptop to perform the obstacle detection and avoidance method.

Auditory feedback regarding the estimated collision risk was implemented and

tested on blindfolded participants. Experimental results show that the proposed technique outperforms other conventional methods for obstacle avoidance. However, with the presented prototype design, it is not suitable to be used by visually impaired people.

A group of researchers from Munich developed a lightweight device to help visually impaired people during their everyday activities [73]. This wearable device uses two depth cameras for data collection and a real-time depth processing algorithm extracts information from the video stream to produce acoustic outputs. The use of this low power, low latency sensor is useful to develop a user-friendly device that performs real-time processing. However, the clinical tests for this system revealed that real-life scenarios are far more complicated and need more sophisticated systems and algorithms to deal with dynamic motion and multiple object detection.

A novel navigation assistant system for blind people was implemented in work proposed by Tapu et al. [74]. The proposed method (denoted DEEP-SEE) detects both moving and stationary objects using the you only look once (YOLO) object recognition method [75]. Based on two convolutional networks, their system tracks the detected objects in real time and solves the occlusion problem. The system then classifies the object based on its location, type and distance and notifies the user using acoustic warning message prioritisation based on the object semantic interpretation.

Detecting traversable area and avoiding obstacles for visually impaired people was proposed by Yang et al. [76]. The authors presented a sensor combination, multi-thread assistance framework integrating wearable smart glasses, inertial measurement unit (IMU) sensor, and the Intel RealSense RS410 depth camera. Although the proposed work enhanced the pathfinding task for blind and visually impaired people, the system did not provide any information about the type of the detected objects or the motion model of the dynamic objects in the user's environment.

Many other research studies were reported in the literature discussing obstacle

detection and avoidance methods for visually impaired people using audio-based feedback systems [77–80].

### **Scene perception**

Yang et al. [81] developed a novel model for navigation-related scene perception system to help people with low vision. The proposed work unifies terrain awareness for best traversable areas, obstacles, sidewalks, stairs, water hazards, pedestrians and vehicles in one method that operates in a real-time navigational assistance framework.

The proposed wearable device composed of a pair of smart glasses and a portable processor. The smart glasses consist of an RGB-D sensor, RealSense R200 camera, and a set of bone-conducting earphones. A deep-learning based network was developed to provide an efficient semantic understanding regarding the surrounding environment. The detection results are then transferred to the visually impaired using the bone conduction headphones for both terrain awareness and collision avoidance.

Mekhalfi et al. [82] developed a prototype for blind people that combines object recognition, obstacle detection/avoidance and navigation modules. Their prototype integrates lightweight components including camera, IMU, and laser sensors. The proposed system performs different navigation and recognition process, such as object recognition, path planning, ego-motion computation, and object detection. Distance information is provided by the laser unit. Users can communicate with the device through speech recognition unite. Audio feedback is generated and transmitted back to the user for navigation instructions and obstacle’s warning messages.

A machine learning based ”coarse recognition” method was implemented to detect static and dynamic objects in an open environment rather than focusing on specific classes of objects. This recognition phase was evaluated on an indoor and outdoor datasets and results showed good results in recognition rate.

We believe this system is beneficial for blind people in their navigation and scene

perception problems. However, the incorporation of different sensory modules raises several issues regarding its wearability criterion. Furthermore, the overwhelming feedback about different objects in the surrounding environment would disturb the user with unnecessary information.

Recently, a context-aware indoor mapping system was developed by a group of researchers [83] based on Tango device, semantic maps editors, and obstacle detection algorithms. This system (ISANA; intelligent situation awareness and navigation aid) uses information from the mentioned modules to compute a safe navigation path for blind people. Other audio-based scene perception solutions were reported in the literature such as Hands On [84] and the work proposed by Chae et al. [72] and Yang et al. [81].

### 2.2.3 Assistive Technologies using Haptic Cues

Visually impaired people depend on their hearing sense to extract information regarding their surroundings. Therefore, scientists tried to use other sensory cues to deliver information without disturbing the user's hearing. Haptic feedback is the use of touch to interact with the user and it has two types: kinesthetic (force) and tactile (touch) feedback.

#### Navigation and wayfinding

Many research studies were reported in the literature to address the navigation and wayfinding problem for visually impaired people using haptic/tactile feedback [85–88].

Electro-neural vision system (ENVS) is a prototype developed by Meers & Ward [89] in 2005 for blind people. The system uses depth information extracted by a stereo camera to provide 3D perception and GPS navigation using the electro-tactile interface. The head-mounted camera feeds the main algorithm with real-time images, which then used to a disparity depth map. Transcutaneous electro-neural stimu-



lation (TENS) unit converses this information into electrical pulses that stimulate nerves in the hand skin.

Outdoor experiments were performed by a blindfolded user, and results showed that the proposed system could provide sufficient information for obstacles avoidance and navigation by interpreting sensory data via the electro-tactile data gloves.

Amemiya & Sugiyama [90, 91] proposed a hand-held force feedback device for helping visually impaired pedestrians based on the "pseudo-attraction force technique". The goal of their device was to guide individuals in their navigation and helping them avoid dangerous obstacles.

The proposed device used asymmetric acceleration to guide the user to the right navigation direction. By accelerating more in the correct direction, the device allows the user to experience the kinesthetic illusion of being pushed or pulled towards the right path and avoiding the collision [92]. Experimental testing was applied to the proposed work for evaluating navigation direction. The device was tested by 23 visually impaired participants who confirmed the usability of the proposed solution [92].

Sharma et al. [93] proposed a sensors based, low-cost smart stick to help visually impaired people in their navigation. The stick can detect static and dynamic obstacles and estimate their distance to provide both audio and vibration feedback. The developed system consists of a microcontroller, an ultrasonic sensor, and master-slave Bluetooth modules. The system can only detect obstacles in front of the user, which is considered to be a significant shortage in its implementation.

### **Obstacle detection and avoidance**

Cardin et al. [94] developed a sonar-based scanning system that senses the surrounding environment looking for obstacles and sending position information of the closest object to the user using vibrotactile feedback.

The prototype consists of four sonar sensors, a microcontroller, vibrators and

a personal digital assistant (PDA) attached to a jacket to be worn by the user. The sensors are located at the shoulders height and placed to cover a large area ( $90^\circ$ ) in front of the user. The microcontroller calculates the approximate distance of the nearest obstacle and then converts the distance to a pulse width modulation (PWM) signal. This signal is then sent to the vibrators to generate suitable vibration feedback (different vibration speeds representing the detected distance).

The prototype was tested on five participants in a controlled indoor environment. The authors reported promising results as users were able to walk through the testing area, localise themselves and to distinguish obstacles from the left and the right. However, this solution needs a pre-training session and could be used as a complement to the traditional white cane [94]. The authors reported other issues in the system's wearability such as the interference of hands and their wrong detection as obstacles.

A mobile kinect sensor was used to develop an obstacle detection and warning system for visually impaired people [95]. The device is composed of two modules: obstacle detection and obstacle warning. A portable laptop uses RGB and depth images captured by the Kinect, in addition to accelerometer data to detect obstacles in front of the user. A tactile—visual substitution system was used to deliver obstacle warning messages to the user. Although experimental results showed that participants were able to (1) correctly react to navigational cues and, (2) warning messages provided by the system, these modules are expensive and require considerable processing power.

Other previous research work addressing obstacle detection using haptic/tactile feedback has been reported with similar methods [89, 96–98].

### **Scene perception**

Using haptic feedback to help people with vision impairment in their navigation, as well as dealing with a dynamic environment was reported by several research studies in the literature [99, 100].

ARIANNA [101] is a path recognition system used to help in indoor navigation for people with impaired vision. It is a flexible application that could be used on smartphones and portable devices with AR capabilities. This system uses computer vision algorithms to detect tapes deployed in the users' pathway to easily navigate them through the use of vibration signals as feedback. The authors also presented a second version for their system by enhancing the tracking performance [102].

A novel and smart indoor mobile assistive technology for blind people using Tango devices was introduced by Li et al. [103]. They delivered a full system from obstacle detection and tracking phases to the final notification. The system uses the indoor map editor to extract semantic features from the geometric map for global path mapping. This step is continuously updated and enhanced in real time with the obstacle detection and avoidance algorithm they created to correct the path of the projected obstacle if the user faces any possible threat. Finally, a smart cane prototype was designed and implemented for human-machine interface and communication. Limiting the navigation process to indoor environment with only pre-planned routes restricts the number of users who could benefit from this system.

In 1999, Abowd et al. [104] introduced the concept of context-awareness in computing. They described it as the ability of computer systems to simulate real human communication options by gathering data about its surrounding environment at any given time. Also, these systems should adjust their interactions based on the collected data accordingly. Context-aware approaches use software and hardware for data collection. They perform real-time data analyses and processing smartly.

From the previously mentioned studies, we can see the shortage of outdoor context-aware hazard detection systems for impaired vision people. Current systems do not distinguish between the detected objects in their level of danger, providing the same feedback style for any object detected. Moreover, most of these proposed systems assume total blindness of the user's vision. People with peripheral vision

loss retain healthy vision with good central acuity. This highlights the need for smart outdoor hazard detection and classification systems that work in real time and deliver smart notifications that could alert the user as early as possible about developing hazards.

## 2.3 Compensation Interventions (Adaptive Behavioural Training)

Low vision rehabilitation is a complicated process that requires a multidisciplinary effort including various professions like ophthalmologists, orthoptists, occupational therapists, orientation and mobility instructors, social workers, teachers and others [105].

Compensatory interventions are a set of treatments that help visually impaired people to compensate or adapt for their visual impairments, which will eventually help them to perform everyday tasks more easily [50]. Ong et al. [106, 107] reported significant improvements in eye-search and reading-writing activities using free, online treatments such as Eye-Search [108] and Read-Right [109]. Audio-visual stimulation of the visual field has been used in the field of compensatory intervention [110, 111]. The research in the literature reported a potential for further development in this area [50].

The majority of research studies in this field are concerned with two main goals: (1) increasing saccadic movements [112, 113] and (2) improving eye movements and scanning into the defected field [114–117]. On the other hand, some research studies focus on the specific help that occupational therapists (OT) provide for visual field loss people in their daily life.

The OT role is multifaceted and includes different tasks such as environmental assessments in the patient’s environment and providing a personalised training plan

for enhancing residual vision-related skills [118, 119]. For people with peripheral vision loss, OT's assist them to compensate for the loss in their side vision by training them to move their head and eyes into the blind areas in their visual field [105]. Additionally, other interventions administered by OT's may include strategies to address independent mobility and instrumental activities of daily living (IADL) training [105, 120].

## 2.4 Restitutive Interventions

As reported in the first chapter, visual field loss has been considered to be non-restorable. For several decades, researchers used to think that restitution interventions have limited effect in visual rehabilitation [121, 122]. More recently, research work suggested that with the correct use of specific interventions, it is possible to expand the visual field after brain or optic nerve injuries [123, 124].

Restitution treatments are a set of interventions where the defected visual field is trained or stimulated repeatedly [125]. One of the most reported restitution treatments in the literature is vision restoration therapy (VRT). The goal of VRT is to expand the normal visual field by stimulating the boundaries between the healthy and damaged areas [50]. It is a non-intrusive (does not require surgery or medication) and personalised therapy for the individual vision case. NovaVision<sup>®</sup> [126] is a commercially available treatment that uses visual stimuli targeting the defected areas, trying to support the brain to strengthen the visual information processing of residual vision.

A number of studies reported visual field expansion after VRT treatment (Schmielau and Wong [127], Marshall et al. [128] and Gall and Sabel [129]), which consequently, increased the quality of life measurements.

On the other hand, Reinhard et al. [130], Roth et al. [114] and Pollock et al. [125]

conclude that VRT is an ineffective procedure when compared to other rehabilitation procedures such as placebo, control or no treatment taking into account the visual field outcomes [50].

## 2.5 Summary and Conclusions

From previous research, it can be shown that assistive technologies can help visually impaired people in their navigation, obstacle detection and scene perception tasks. Visual field loss solutions were developed to extend the visual field by adding extra information in the healthy vision areas. However, these solutions are still primitive and do not incorporate state-of-the-art techniques in computer vision and machine intelligence fields. Most of the available assistive technologies to help people with visual field loss use equipment such as prisms to enhance the healthy vision. Although this solution can help with the visual field extension, literature studies had reported serious problems that affect the healthy vision for users using such devices.

On the other hand, massive growth has been reported in the field of assistive technologies for visually impaired people using different types of sensors, computer vision and machine learning algorithms. Most of the systems mentioned above focused on technical aspects and assumed the users to be blind. As people with visual field loss have good vision in some parts of their visual field, it would be beneficial to use this vision smartly by adding computer-generated information to increase the patient's awareness by providing early notification about possible hazards.

The need for a wearable assistive technology that is unobtrusive with physical convenience and utilises the healthy vision for the peripheral vision loss is highly needed. It is also essential to involve the potential users in the design and implementation phases of the developed solution.

Overall, there have been limited solutions that use computer vision algorithms

to help people with visual field loss. It is intended that the research work presented in the chapters to follow in this thesis will tackle these issues by advancing the state of the art technologies that can be used to provide smart assistive technology that:

1. Differentiates between healthy and defected vision areas;
2. Uses computer vision to understand the surrounding dynamic environment;
3. Considers the human perception of hazards;
4. Supplements the user's knowledge with the necessary information to avoid possible risks taking into account the personal visual impairment case;
5. Propose a wearable, friendly, affordable and smart solution that could be used by visually impaired people in their daily tasks.

The following chapter will discuss the method used to collect users' requirements and needs. In specific, patients' questionnaires and discussion outcomes will be presented and analysed. In addition, the main components and system design are discussed.

# Chapter 3

## User Requirements and System Design

### 3.1 Introduction

While working on the design of the assistive technology proposed in this thesis, a user-centred design (UCD) approach was adopted. Preece et al. [131] defined UCD as an iterative methodology where the designers focus on the user's needs and preferences in every phase of the project. Understanding the end-user requirements before and during designing an assistive technology is very crucial for enhancing usability issues and ensuring that the proposed solution will meet their needs.

By starting with users, their problems, priorities, hopes, challenges, and needs, we tried to discover what is most desirable and essential for them and built the assumptions and solutions based on this. Interacting with the users is in accordance with the ethical approval from the Research Ethics Committee at the Faculty of Science and Engineering, University of Liverpool, UK (Reference: 1982) [Appendix A]. Participant information sheet [Appendix B] and a formal consent form [Appendix C] for participating in the study was obtained.



Several research papers explored the visually impaired people requirements for assistive technologies. In their research paper [132], Jafri and Khan presented their obstacle detection and avoidance application for visually impaired people based on the results they obtained from a semi-structured interview. While the human guide was superior to the white cane as a navigation aid, the participants mentioned that this method caused them problems as they depend entirely on the guide who may not provide accurate warnings about obstacles. In the same study, moving and minimal obstacles were the most difficult to detect and avoid during the indoor navigation.

The lack of information that describes the physical environment is one of the core challenges for visually impaired people navigation. This was mentioned by many users [133, 134] as they need a clear description for indoor and outdoor main landmarks that would help them building a mental map. In their comprehensive study about computer vision algorithms for assistive technology, Leo et al. [135] highlighted several open challenges for developing assistive technology for visually impaired people. Object detection and tracking problems are examples, especially for egocentric video streams.

Based on this, it was decided to explore the requirements and preferences for people with visual field defects using our surveys.

In the following sections, explanatory studies used in this project will be discussed to understand the potential users' needs. The system design and primary phases, which were built based on the reviews and feedback gathered from the questionnaires' participants are presented in the following section.

## 3.2 User Requirements Elicitation Study

Two questionnaires [Appendices D-E] were conducted with the potential users who were recruited via the Institute of Population Health Sciences at the University of

Liverpool. The VISION group is a representative patients group; where participants are coming from different places all over the UK representing other local groups. The participants had different types of visual field defects, including hemianopia, double vision and tunnel vision.

The questionnaires were designed to understand the daily challenges and needs for them and to gain some insight into their preferences for wearable assistive technology. Only eleven participants were able to complete these questionnaires which was one of the challenges in the recruitment process of participants. The questionnaires were prepared in accordance with the ethical approval and the project's design phases.

The project's idea and potential outcomes were presented to the participants before completing the questionnaires.

The first questionnaire was conducted during the second year of the project to gain a fundamental understanding of the participants' main challenges. Also, it helped to choose the right system design that best fit their needs. The developed solution was discussed with participants after the second year. Then the second questionnaire was conducted to select the best notification (or what is known as feedback) style and format to be used in the system. The project goals were presented; then, participants started answering the questionnaire points. In some cases, extra information was discussed with participants to demonstrate the idea of the problem or to distinguish between similar alternatives. The answers were examined manually, and additional comments or opinions were noted. The results of the two questionnaires were used to enhance and further improve the development of the project's phases.

### 3.2.1 Questionnaire 1

Five participants were able to answer the questionnaire, covering the following topics:

(1) which hazard types are the most essential to be notified about, (2) which object

types participants would like to have information about, (3) what is the best warning time for notifications, (4) what are the hazard notifications levels and (5) what are the preferred notification styles.

### **(1) Hazard types**

The goal of the first question was to determine the hazards types that they think are the most critical (dangerous). Two options were given to answer this question: stationary, or moving hazards. The purpose of this question was to select the best object detection technique to be used in this project. It was found that 80% of the participants prefer having notifications about moving rather than stationary hazards. This result guided us to choose the correct object detection technique for the first phase of the project. Since the majority of the participants preferred having notifications for moving objects, motion detection method for the object detection phase was first used.

### **(2) Object types**

The purpose of this question was to define the types of objects the participants think are essential to have notifications about them. When they were asked to specify the types of objects they are interested in, cars, people and bicycles were the most chosen options. Figure 3.1 shows participant's preferences according to the this question.

As shown in Figure 3.1, all the participants prefer to have a notification about cars. People and bicycles are the second preferred categories, as selected by 80% of participants. These answers guided us to the need of having an object recognition method in addition to the object detection one.

### **(3) Notifications timing**

The third question was about the warning timing. It was essential for us to determine how fast the participants preferred to see the system's output. They were asked, "how early do you prefer to get a warning notification (in seconds)"? Three

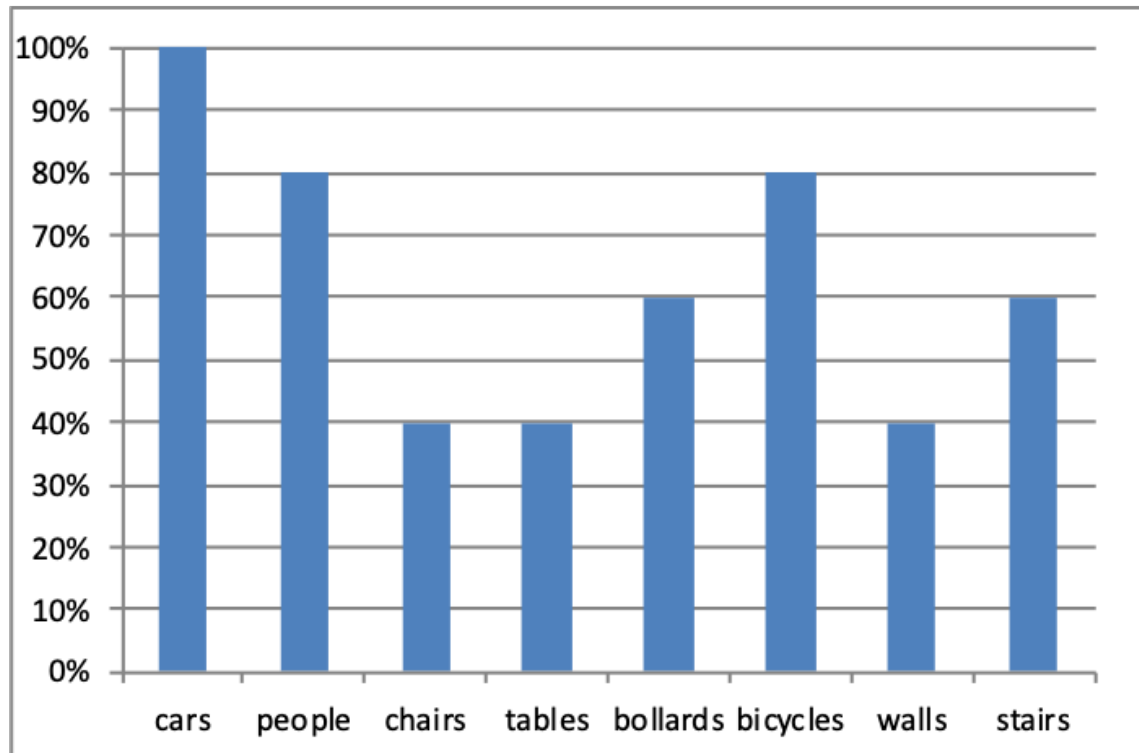


Figure 3.1: Object types preferences.

participants chose 2-3 seconds, while only one chose a longer (5-10 seconds) option. One of the participants preferred to have a notification as soon as possible, as shown in Figure 3.2.

#### (4) Notification levels

This question aimed to investigate levels of notifications the participants prefer to have. The majority (75%) of the participants opted for a multi-level option compared to a single option. This preference gave us a clue that the proposed system should include several types of notifications representing different levels of danger.

#### (5) Notification format

The last question in this questionnaire was about the notification format with three possible styles: acoustic, visual or hybrid notification. The purpose of this question was determining the most practical output design that would suit the visually impaired person. The answer to this question is critical in our project since it is

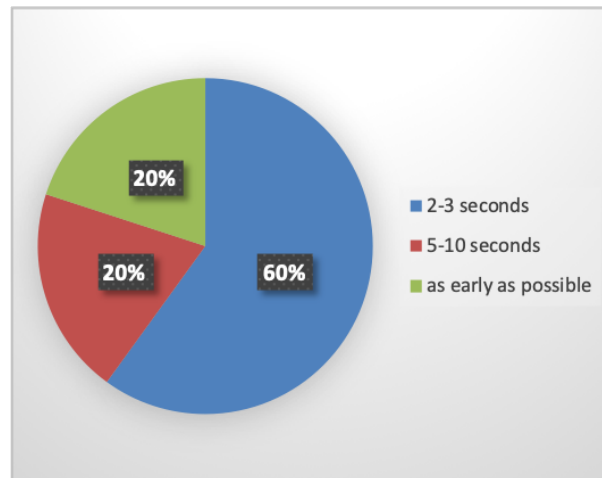


Figure 3.2: Notification timing preferences.

essential to deliver the warning notification in a meaningful way that would increase the participant's awareness about the surroundings.

Figure 3.3 depicts the participants' preferences for the notification's format. As shown in the figure, the majority of the participants prefer visual notifications over the acoustic warnings. This result guided us to the need for distinguishing blind and healthy vision areas for the people with visual field loss. None requested hybrid feedback.

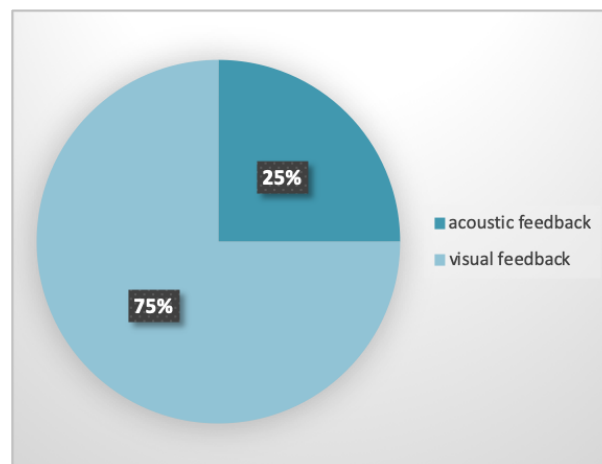


Figure 3.3: Notification format preferences.

The answers for this questionnaire were the initial starting points for this project. From this questionnaire, we knew that the project design should start with an object

detection and recognition phase. Besides, fast outputs (notifications) should appear visually on a display unit for users to be aware of any possible moving hazards around them.

In this context, smart glasses would be used for capturing real-time videos and processing them to detect possible hazards and display a proper visual notification on the glasses display unit. The processing unit integrated into the smart glasses' frame would process the video stream while the participant is moving and should update its output according to the dynamic changes in the participant's environment.

### 3.2.2 Questionnaire 2

Six participants were able to answer the second questionnaire, which was conducted after the second year of this project. The first goal of this questionnaire was to discuss the developed assistive technology, its design, main processes and the basic version of the visual notification output.

The second goal was to gather more information about the participant's preferences for the feedback generation phase, which was planned to be the last stage in this project. The questionnaire was structured around the following topics: (1) basic visual impairment history of the participants and the assistive technology devices they use, (2) hazards priorities, and notification format and (3) an open discussion regarding participants opinions, impression, ideas and feedback about the proposed technology.

#### (1) Visual impairment history and assistive technology use

Five of the participants suffer from hemianopia, which is one of the visual field loss types. One of the participants was a double vision patient. He did not have any problem with his visual field, but his visual acuity (central vision) was low. The purpose of this part was to investigate the participants' familiarity with assistive technologies in general and wearable vision assistance devices in particular.

While participants used portable electronic devices, only 16% of them used navigation aids in their daily life. This means that participants do not depend on their smartphones or other mobile devices to help them in their mobility. The majority of participants (83%) wore eyeglasses, and some of them were familiar with smart glasses and happy to use them.

## (2) Hazards prioritisation and the notification format

This section of the questionnaire had two parts: hazard prioritisation and notification formatting. The first part aimed to discover participants prioritisation about hazards danger levels. For this question, the participants were asked to rank five predefined hazard scenarios from low danger to high danger level hazards. Table 3.1 shows how the participants rated the defined hazards scenarios.

Table 3.1: Participants' ratings of the level of danger for different hazard scenarios

Hazard scenario	Level of danger				
	v.low	low	neutral	high	v.high
Static hazards outside your pathway	4	2	0	0	0
Static hazards in your pathway	0	2	0	1	3
Moving objects not in your way	0	2	4	0	0
A person moving towards you	0	0	1	3	2
Object moving towards you	0	0	0	2	4

Objects that move towards the participant or the participant's pathway were selected to have the highest danger level, while static hazards (obstacles) located outside the participant's pathway were given the lowest danger level. Other hazard scenarios vary between low, natural, and high danger levels. This rating helped us in the machine learning phase (Chapter 5 of this thesis) to train a Neural Network classifier to classify the detected objects into one of the mentioned five hazard types.

The second part of this section aimed to find the most useful notification style that would increase the participant's engagement with the proposed technology. Figure 3.4 shows participants' preferences for the system's output (in this question, the answers from both first and second questionnaires were combined). As seen in this

figure, the majority of the participants preferred visual format for the system's output compared to other types as audio or touch (vibration) format. Other participants chose a hybrid format as visual and touch or visual and sound styles. These preferences prove that people with visual field loss like to overcome their vision impairment and use their healthy vision in their daily tasks.

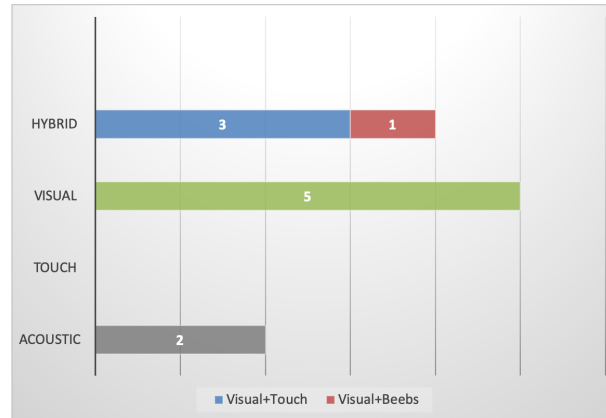


Figure 3.4: System's output preferences.

From these answers, it was concluded that using visual feedback (notification) is the most recommended choice. The majority of participants chose this type before developing the proposed system, and again after demonstrating the basic version of the assistive technology with smart glasses.

Most of the participants (83%) who chose visual notifications preferred to have one shape with different colours. Regarding the notification timing, 66% of the participants proffered to see one notification (the highest priority) at a time while the rest 33% accepted multiple notifications at the same time. To determine the best information that each notification will deliver for the participant, participants were asked to choose from three different types: (1) hazard direction only (2) hazard type and direction and (3) hazard direction and speed. Half of the participants opted for the second and third options, while no one chose the first one.

Appearance and disappearance styles of the notification were discussed with the participants to determine the best format. It was found that 83% of the participants



prefer the notification to appear in regular time intervals for as long as the hazard is present while only 16% prefer to see it only once. Regarding the automatic or manual disappearance of the notification, the answers were balanced with 50% for the first option and 50% for the second one.

### **(3) Participants open discussion**

This part of the questionnaire aimed to gather participants thoughts and opinions regarding the project in general, and the developed technology and smart glasses in specific. A basic system demonstration was conducted with the participants to explain the main functionalities and modules. Participants were able to try the smart glasses and see two different notifications that were displayed on the see-through display units of the smart glasses. They wore it and moved within the meeting room to have an idea about its weight and usability. The meeting room was  $4 \times 7$  meters space, with tables and chairs placed randomly.

Three participants commented on the user interface with the smart glasses. They preferred having different levels of commands between the participant and the smart glasses. Starting from simple options at the beginning to more complicated options, once the participant learnt how to adopt the system.

Regarding the project idea and design, all participants were delighted with the solution provided. They were very excited to try the final product and see the results in real-time while navigating indoor and outdoor. Some of the participants commented on the smart glasses by saying it is light-weight with an accepted design.

One participant stated “It is a fantastic idea and great use of technology. Could the system be simple at first then give more information once the user adapted.”

Another participant mentioned that “Would need more information and understanding about the device/idea. However, the concept is excellent and looks forward to seeing the final development of the system.”

According to Mr.Smith, “It is a great and very useful idea. The device is light-

weight and could be used with a personal iPhone.”

### 3.3 Augmented Reality and Smart Glasses

Augmented reality (AR) is relatively a new technology in the information visualisation field. The main idea of AR is superimposing computer-generated information, images or animations over real-world images or videos [136, 137]. Most AR implementations are used by mobile applications and smart devices such as Sony<sup>®</sup> smart eyeglass and the Microsoft<sup>®</sup> Hololens.

Although AR and virtual reality (VR) are similar in the general idea of having an alternative (computer-generated) vision, they follow entirely different approaches. AR systems use the individuals’ normal vision and add more helpful information to extend their knowledge, whereas VR creates an artificial environment where users have a synthesised vision [137].

Wearable computers and head-mounted display devices and technologies are steadily gaining publicity. Smart glasses in particular are more popular due to their entertainment functions and techniques [136]. Al-Ataby et al. and Younis et al. suggested the use of augmented reality concepts to help people with visual field defects using smart glasses [138, 139]. Starting with the announcement of Google<sup>®</sup> glasses in 2012, several smart glasses were developed to use both artificial and augmented reality concepts. The first scientific review for the clinical and surgical applications of smart glasses in healthcare systems was presented by Stefan et al. [140].

OrCam<sup>®</sup> MyEye 2.0 [141] is a smart technology that uses computer vision algorithms with the addition of wearable platforms to help people with vision problems. Their main goals are to improve individuals independence navigation and help visually impaired people to read by themselves.

The design is straightforward, lightweight, efficient and could be clipped onto a



Figure 3.5: Orcam MyEye 2.0 wearable camera [141].



Figure 3.6: Daqri smart glasses [143].

pair of glasses. Utilising any surface, the attached camera can read text instantly using the person's gesture and generates a loud voice using a small speaker for the user. The system also can recognise faces, products, and money notes in real time. Figure 3.5 shows the design for Orcam.

Much like Microsoft<sup>®</sup> HoloLens [142], Daqri<sup>®</sup> smart glasses [143] implement augmented reality concepts in manufacturing, medical remote experts, field services, maintenance, and repair sectors.

Their design is robust and consists of a wearable head bond that contains a video camera, a display unit and mini portable computer including Intel core m7 processor as shown in Figure 3.6.



Figure 3.7: EyeTrek Insight EI-10 glasses attachment [144].

The EyeTrek Insight EI-10 [144] is the latest generation of Olympus<sup>®</sup> optical solutions for smart glasses. Inspired by Google glasses [145] design, the small display unit superimposes the user's FoV by computer-generated information without blocking the normal vision.

The difference between this design and the Google glasses design is that EI-10 can easily be connected to any pair of regular eyeglasses. Its lightweight, powerful operating system, and efficient display unit make it an excellent choice for business applications as shown in Figure 3.7.

Hicks et al. [146] from Oxford University used AR and the Epson Moverio smart glasses to aid visually impaired people. OXSIGHT [147] glasses help people with different types of vision loss to regain control of their vision. The glasses provide two options to help with different vision loss types: OXSIGHT crystal and OXSIGHT prism. Both OXSIGHT crystal and OXSIGHT prism expand users' horizontal visual field up to 68°, resulting in a real difference in the daily life of peripheral vision loss people. OXSIGHT crystal comes with extra features for Unrestricted peripheral awareness. The glasses provide a more unobstructed view of the surrounding environment using different computer vision algorithms to make the edges of objects sharper and help in safe and secure navigation.

In this project, Epson<sup>®</sup> Moverio BT-200 [148] smart glasses were used to capture the video stream and display the final visual notifications. The Moverio has been

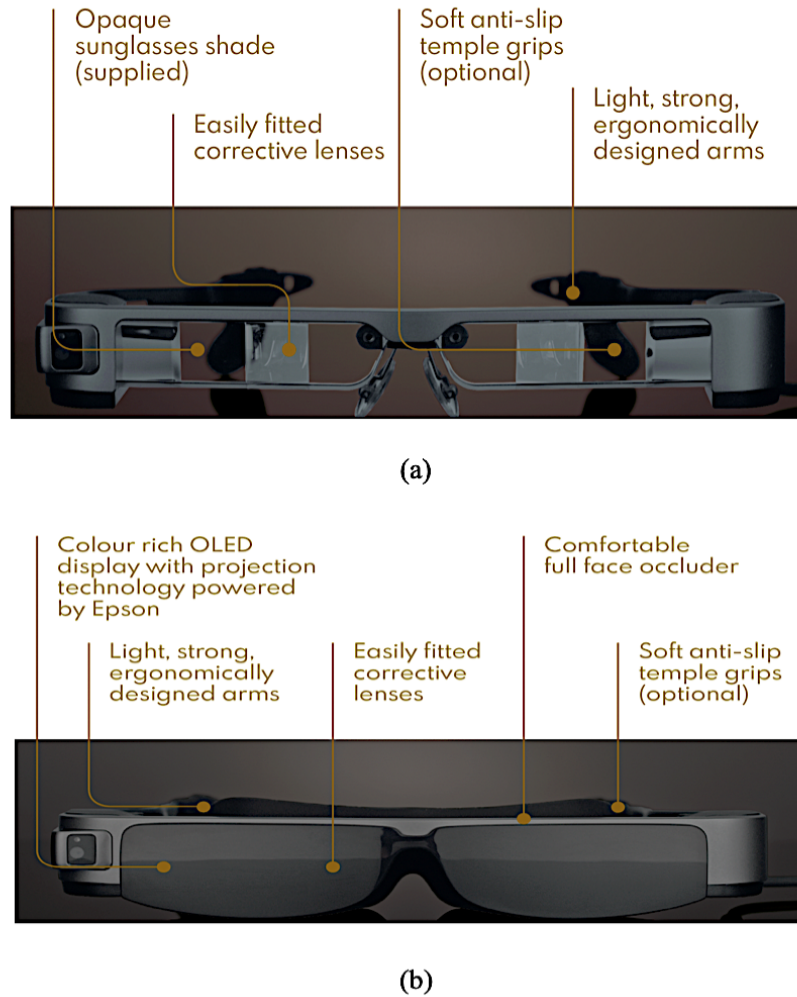


Figure 3.8: The OXSIGHT [147] smart glasses. (a) OXSIGHT crystal, (b) OXSIGHT prisms

chosen for many reasons. (1) It has AR products and supports new application development, (2) it is lightweight and stylish, (3) it has a reasonable price (around £600), and (4) it contains two see-through display units. These features and many others encouraged us to use this product in our project. Figure 3.9 shows the Moverio BT-200 product features and specifications. Although the Moverio has a narrow field of view ( $\approx 49^\circ$  horizontally), it is still an excellent option to be used for the data collection stage. In future work, this could be replaced by another wide-angle camera.

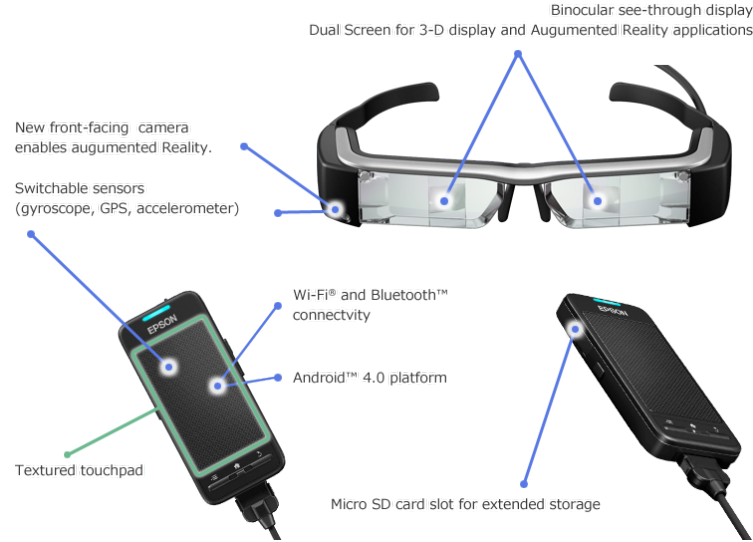


Figure 3.9: The Epson Moverio BT-200 smart glasses product features [148].

### 3.4 System Design

Based on the mentioned requirements, preferences and challenges, our design was developed in a smart way that would provide early, meaningful and straightforward notifications to extend the user's mental map<sup>1</sup>.

Possible hazards were defined as any moving or stationary object that visually impaired people are not able to see or recognise and could collide with while walking. The system starts scanning the real-time video stream to search for objects and then tracks their movement. Motion features such as object speed, direction and location in addition to the object type and other features were used to determine the level of danger for each detected object.

As shown in Figure 3.10, the goal of the system was to generate notifications for the user to become aware of potential hazards. In this case, the system prioritises the detected objects to produce useful notifications without overloading the user with too much information. Visual field test results were used to delineate both healthy and defected vision areas.

<sup>1</sup>A mental map is a personal representation of the surrounding world that helps in navigation and object localisation [149].

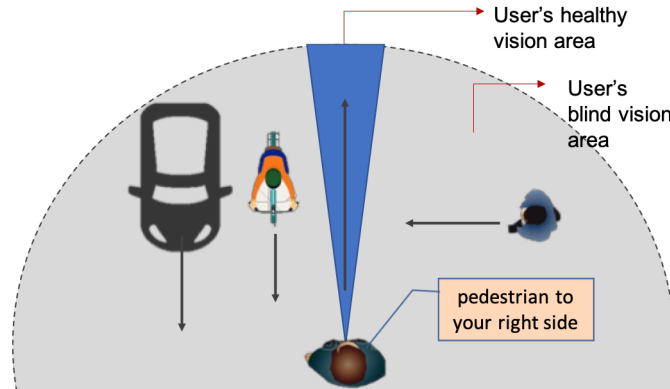


Figure 3.10: The conceptual representation of our system.

Figure 3.11 shows three examples of visual field test results for peripheral vision loss. The left column shows the central field test (left/right eye results), while the right column shows the full field test (both eyes together, covering about 160°).

The Humphrey visual field analyser (HVFA) is a measurement tool used by eye specialists to test the human central visual field. It delineates both healthy and defected vision areas. The grayscale map of HVFA test shown in Figure 3.12 presents the visual sensitivity across the user's central visual field. Dark regions reflect lower visual sensitivity and light regions indicate a higher visual sensitivity. The comparison plots and map keys can be used to interpret the result.

The first row shows severe glaucoma (tunnel vision), the second row shows left hemianopia (blind to the left side, healthy in the right side) and the third row is an example of central scotoma (blind area). More details about these tests will be discussed in Chapter 6 of this thesis.

The system uses the visual field test results to search for possible threats in the user's blind area and classifies these threats based on their danger level. The smart glasses is used to display essential notifications outputs in the user's healthy vision area. Figure 3.13 shows an overview of the proposed system and the main modules used in this project.

Object detection and recognition is the first stage, where objects are detected

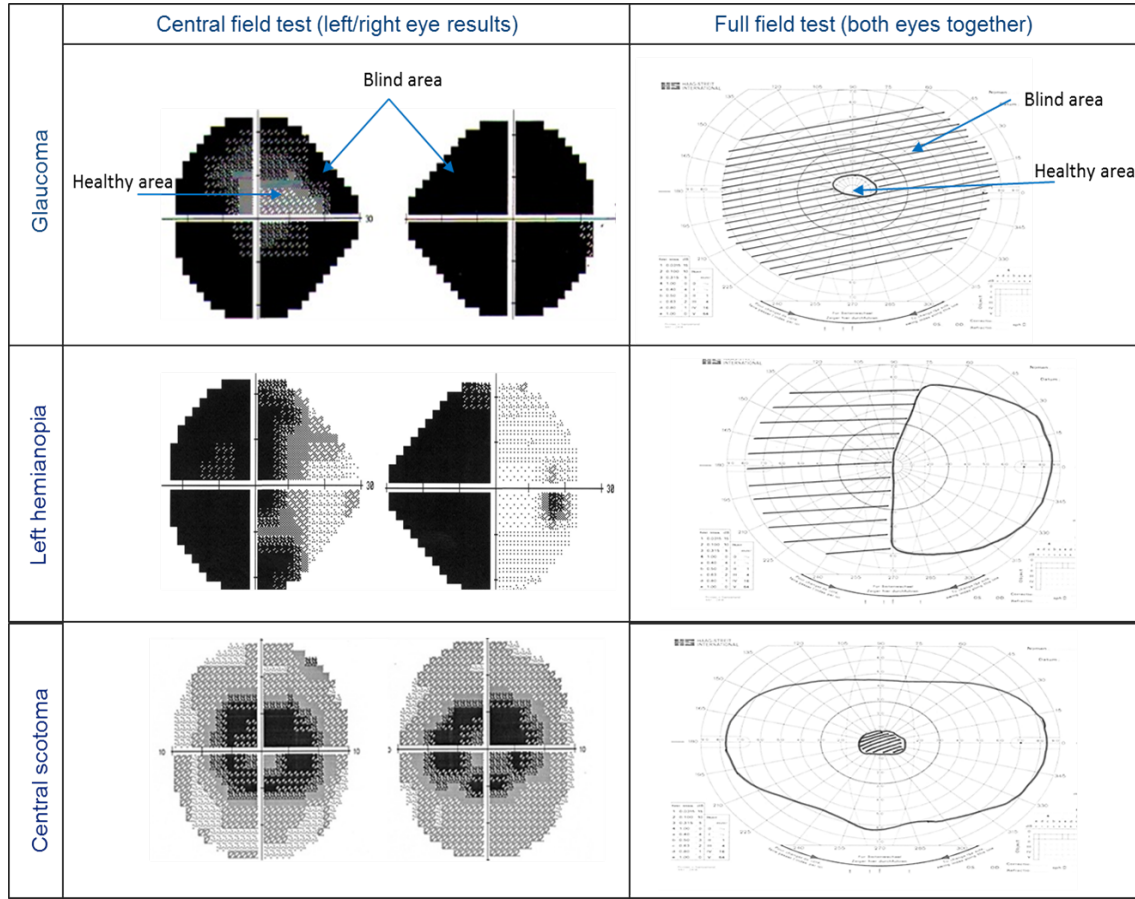


Figure 3.11: Visual field test result examples.

using a deep learning object classifier to determine the type and location. Motion features are extracted using the moving objects' tracking module to determine the age (the appearance duration over frames), speed and direction for each detected object. This information is processed and used to determine the level of danger for each identified object using a neural network classifier.

Objects in the peripheral vision of visually impaired people manifest themselves in different ways such as hazards, obstructions, surprises or immediate dangers. For example, someone walking in the street may not be aware of a cyclist/pedestrian walking on the other side of the road or of dangers such as a car crossing their walking route, street bollard, overhanging cables, trees or bushes to the side of the road.



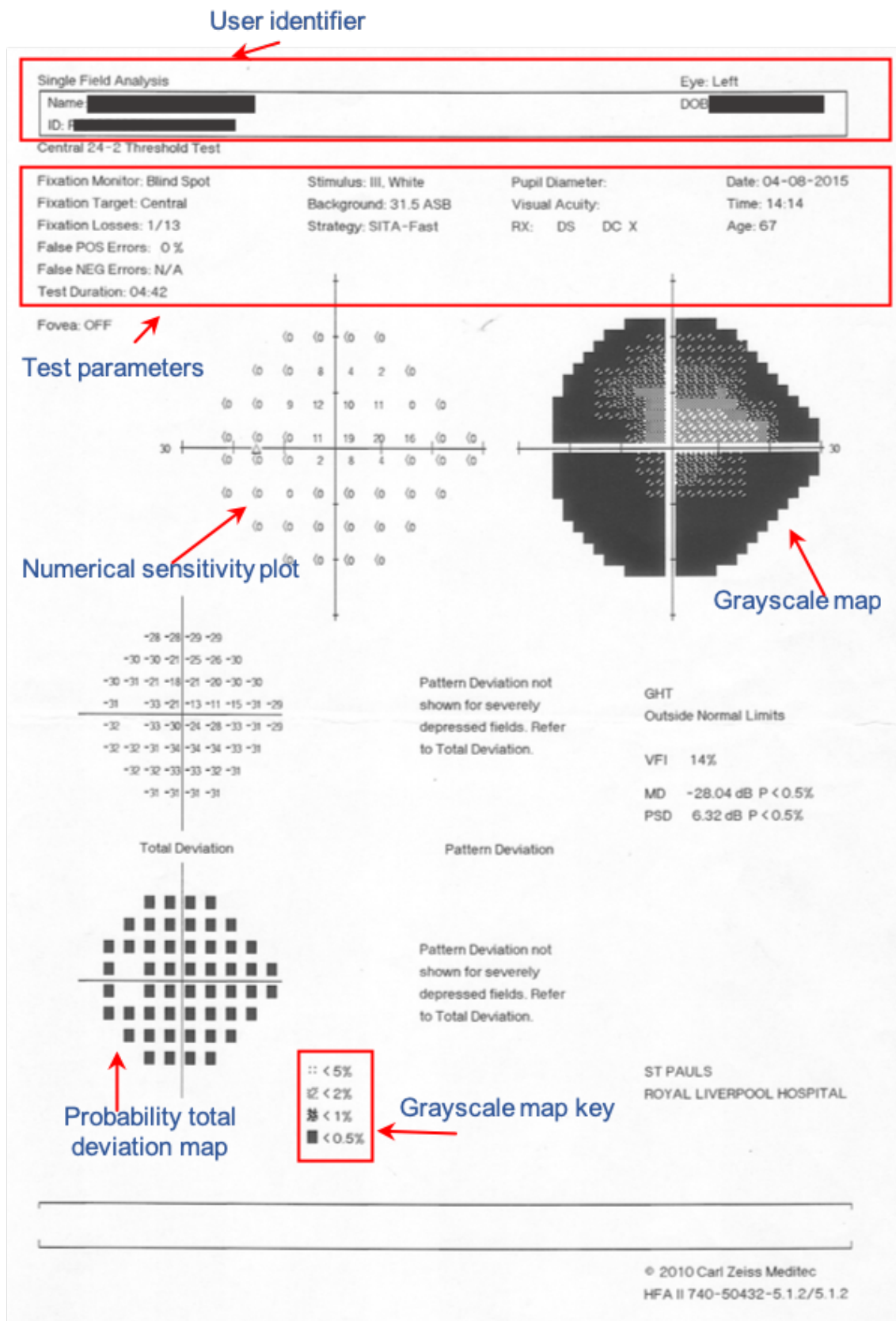


Figure 3.12: Humphrey test printout.

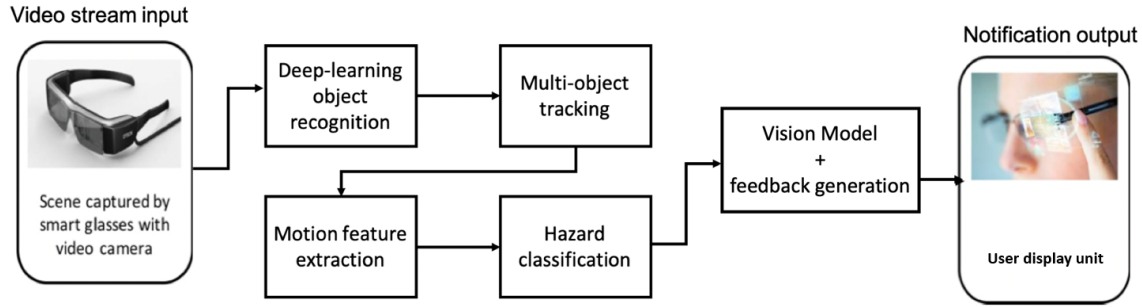


Figure 3.13: System overview including the project's main components.

Not all activities in the periphery are equally crucial to the visually impaired people. Therefore, a system that prioritises all these activities is needed to alert the users to turn their heads to the most immediate threat to see it through their healthy vision.

Based on the mentioned user's preferences and needs, five hazard classes were defined as follows (the class number represents the danger level, one is the lowest, five is the highest):

- Class 1: static object not in the user's pathway,
- Class 2: moving objects not related to the user (any type),
- Class 3: static object in the user's pathway,
- Class 4: person moving towards the user (or user's pathway),
- Class 5: object moving towards the user (or user's pathway).

The visual field has different levels of visual sensitivity depending on where the image lies relative to the fovea or fixation point [21]. This inspired us to define the user's navigation route as the depth extent of the central vision and a small part of the macular vision ( $\approx 10^\circ$ ) around the fixation point. While the fixation point will vary, images are treated as centred around the fixation point.

## 3.5 Discussion and Conclusion

The results of the users' questionnaires highlighted the need for using a wearable smart device, that is capable of detecting different types of hazards and produce a proper visual notification.

The results showed that the device needs to distinguish between detected hazards based on the level of danger for each one, and reflect this on the notification format.

From the first questionnaire, it was concluded that to be able to classify the hazards around the user, an object tracker should be implemented to track moving and stationary objects. Tracking should be done while the user is moving or in a motionless state. While the object or/and the user are moving, the level of danger for each hazard could change. Therefore, the system should keep checking the level of danger for each object during the time it is in the user's environment. To achieve this, objects need to be detected and recognised before being tracked.

Both questionnaires showed that participants prefer visual notifications over other formats for displaying the output. The reason behind this preference is that visual field loss people have both healthy and blind areas. Therefore, they tend to use their healthy area to see what they are missing in the blind area.

This choice guided us to the need for using smart glasses showing computer-generated information (notification) to alert the user's attention to possible threats, but without blocking the healthy vision field.

The second questionnaire focused on participants predilections regarding the feedback generation phase, where the final output of the proposed system is developed. In addition to these preferences, the participants provided us with valuable opinions and discussion about the technology, design, idea and device specifications that were used in this project.

Based on visual field loss studies, it was found that each type has different healthy/defected areas. Visual field tests can clearly distinguish between these parts.

Consequently, it was decided to use these tests to determine the best location for displaying the visual notification for the user to see and avoid possible hazards.

The results are compatible with the findings of the previous study conducted by Zhao et al. [150]. The proposed work studied low vision people's visual perception on augmented reality glasses. To determine the ability of visually impaired people to recognise and response for visual notifications on see-through glasses, the authors conducted a study with 20 low vision participants and 18 controls with healthy vision.

These results of the previous study proved that low vision participants were able to use commercial smart glasses (Moverio BT-200) to identify basic shapes (such as triangle, circle and square) while sitting and moving. The study also reported a similar negative effect on the participant's walking speed for both sighted and low vision participants.

All these findings yielded to three main guidelines for designing virtual notifications for low vision people: (1) Basic shapes are more natural to be identified and interpreted, (2) white and yellow colours are more readable compared to other colours, and (3) the size of the displayed elements should be large enough (they mentioned it should be larger than 100 pixels) to be recognised.

The following chapter will discuss the methods used in this work to detect objects in the user's environment. Different approaches were adopted to detect and track multiple objects at the same time. The pros and cons of each method used are presented and discussed the following chapter.

## Chapter 4

# Multiple-Object Detection and Tracking

### 4.1 Introduction

As mentioned in the previous chapter, object detection and tracking are the first two steps in the proposed technology. The output of these steps will be used to extract motion features that will be used in the hazard classification phase. Detecting and tracking a moving objects requires distinguishing them from the surrounding background and constantly tracking them in all video frames to determine their trajectories and other motion features.

All object tracking methods necessitate an object detection mechanism, which could be activated in each frame or in the first appearance of the target object. Although several methods have been used for moving object detection, six main approaches are widely used by most applications [151], as shown in Figure 4.1.

Generally, one can consider the first three methods (background subtraction, frame differencing and optical flow estimation) as traditional motion detection approaches, and the next three methods (soft computing based) as non-traditional

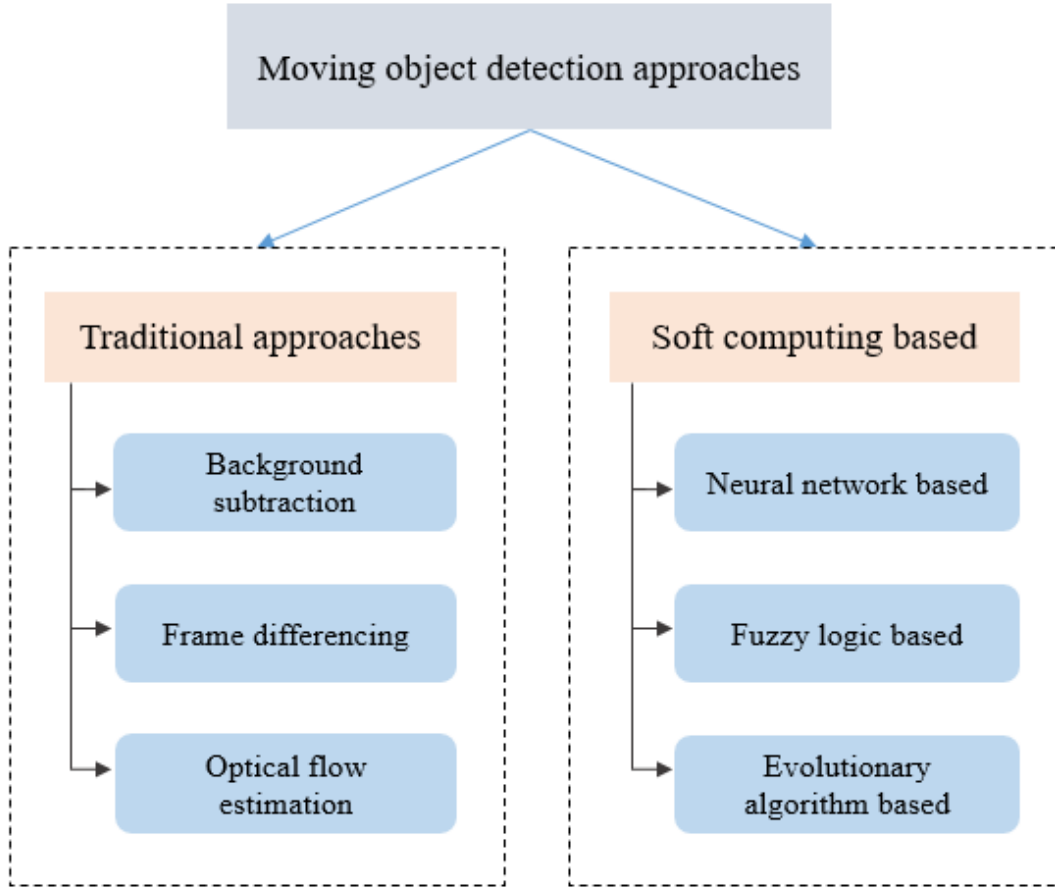


Figure 4.1: Moving object detection approaches

methods.

Background subtraction [152] is a widely used technique for generating a foreground mask using a static camera. Its procedure includes two main steps, background initialisation, and background update. Moving objects are segmented by subtracting each frame from the background model. Changes in lights, illumination, or background can affect the performance of this method and reduce its efficiency [151].

Frame differencing [153] is similar to the background subtraction method in terms of computing the frame difference, but it does not require building a background model. Simply, subtracting the current frame from the previous one, and the result

will be the moving objects. It is easy to implement and gives high accuracy with a static background. Pixel-wise subtraction version of this method is called temporal differencing [154].

Optical flow estimation [155] is a powerful technique that has been applied to numerous applications. Flow vectors are used to detect moving objects by segmenting regions with different optical flow estimation [156]. Different techniques (differential methods, region-based matching, energy-based and phase-based) were implemented using optical flow in motion detection applications. The main goal of the traditional motion detection methods is to estimate the 2D motion vectors from spatiotemporal patterns of the image [157]. Although this method is considered to gain complete knowledge about moving objects, its high computation cost and noise sensitivity make it inefficient for real-time and resource-constrained applications [151].

With the massive growth in hardware facilities and software capabilities, the number of object detection and tracking applications rapidly increased. In addition to this increment, new challenges arose such as moving camera, dynamic background, rapid light changes and shadow detection.

The complexity of these challenges can not be solved using traditional object detection and tracking techniques. Soft computing-based methods are used nowadays to handle these issues and produce more reliable results. Kaushala et al. proposed a comprehensive review [158] of techniques that use soft computing for detecting, recognising and tracking multiple objects in real-time applications.

Each approach has its pros and cons that make it suitable for a specific application than other methods. Despite the technical differences between these approaches, they all share the same general procedure for segmenting the object from its background.

In this project, different methods of object detection and recognition were used. This chapter will discuss these methods and explain the pros and cons of each one. In the first phase of this project, moving objects were detected using frame dif-

ferencing method. Motion compensation was integrated into the motion detection phase to recover image distortion caused by camera movement. This method was used in the early stage of the project development and was changed later for two reasons. First, after applying this method on video streams captured by a wearable camera, it was found that the false motion detection error rate was high and unacceptable, especially when the camera was moving. The second reason for changing the object detection method was the need for identifying the object type to be used in the hazard classification phase. This need arose after conducting our user explanatory study with visually impaired subjects. Therefore, a deep learning-based object recognition method was then applied to overcome these issues. The second section will describe the datasets used in this chapter. The third section will discuss the motion detection and compensation techniques used at the beginning of this project. Deep learning-based object recognition phase is presented in the fourth section of this chapter. Finally, the Kalman filter (KF) and the Hungarian algorithm for multi-object tracking are discussed in the fifth and sixth sections respectively.

## 4.2 Datasets

This section describes the details of the different datasets that have been used in this chapter for multi-object detection and tracking stages.

- **Camera motion detection:** An in-house dataset has been used to evaluate the camera motion classification method. 6300 images ( $640 \times 480$ ) captured by the Epson Moverio BT-200 smart glasses have been used for this experiment. The images contained six motion classes: forward, backward, left, right, up and down. A private expert<sup>1</sup> labelled the ground truth for the classification phase. Data were split into 70% training, 10% validation and 20% testing.

---

<sup>1</sup> A postdoc from the computer vision group in the computer science department, University of Liverpool



- **Motion compensation and object detection using frame differencing:**

A high-quality ( $900 \times 640$ ) publicly available video captured in street using a wearable camera was used in this evaluation with a total of 3650 frames (30 frames/second). Also, sequence *ContinuousPan* under PTZ category from ChangeDetection [159] dataset has been used for evaluating the object detection using frame differencing method. 1700 images ( $704 \times 480$ ) containing five classes (static, hard shadow, outside region of interest, unknown motion and motion) were used. The publicly available ground truth from the dataset website was used to evaluate the motion detection method.

- **Multi-object detection/tracking using SSD and Kalman filter:** Two datasets were used to evaluate the deep learning-based object detection and Kalman filter based object tracking methods. The first dataset was *Seq06R0* from CamVid [160, 161]. A total of 2311 images ( $960 \times 720$ ) were used to train a neural network classifier for hazard classification. The detector classifies the detected object into one of 21 classes. The second dataset used in this step was an in-house data captured by the Epson Moverio BT-200 smart glasses. A total of 6832 images ( $640 \times 480$ ) containing the same 21 classes. More details about these two datasets will be discussed in Chapter 5.

### 4.3 Object Detection Using Frame Differencing and Motion Compensation

According to the work proposed by Panchal et al. [162], frame differencing approach has a high accuracy rate and performs well when the background is static. It also requires low to moderate computational time and resources, making it suitable for augmented reality applications on smart glasses.

Since we have a wearable camera in this system, camera motion is often synony-

mous with head motion. This movement affects the whole processing phase directly from the object detection stage to the notification generation stage.

When the camera (head) moves, motion compensation should be applied before the frame differencing method to reduce false detections due to camera movement. To determine the camera status (stationary or moving), a motion detection technique was developed to define the head motion type (static, moving or rotating). If the head is static (stationary camera), a frame differencing method is applied directly. Otherwise, camera motion is compensated, and then the result is used in the frame differencing step as shown in Figure 4.2.

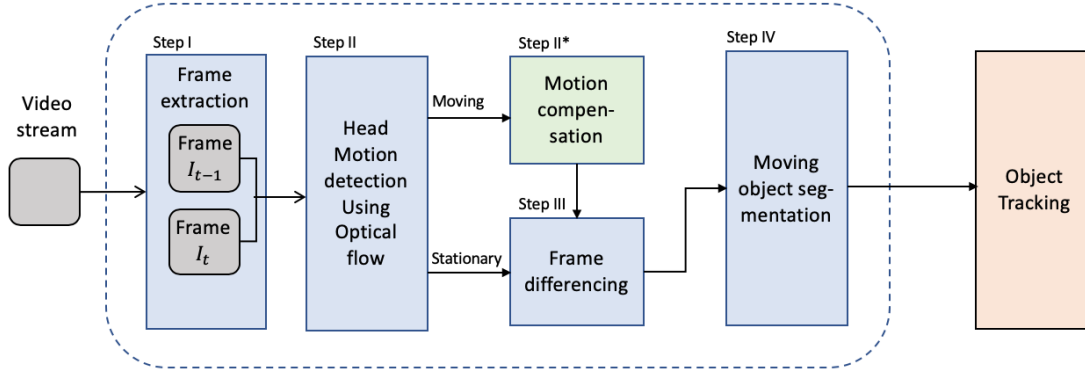


Figure 4.2: Moving objects detection and tracking procedure.

### 4.3.1 Camera Motion Detection using Optical Flow

The camera motion status can be used to determine if the person - wearing smart glasses - is moving or not, and to define the type of this motion. In the case of a wearable camera, six degrees of freedom are expected based on head movements as shown in Figure 4.3.

The head can move in a forward/backward, left/right and up/down translation. In terms of rotation, pitch motion represents the rotation around the x-axis, yaw rotation is a movement around the y-axis, and finally, a roll is a rotation around

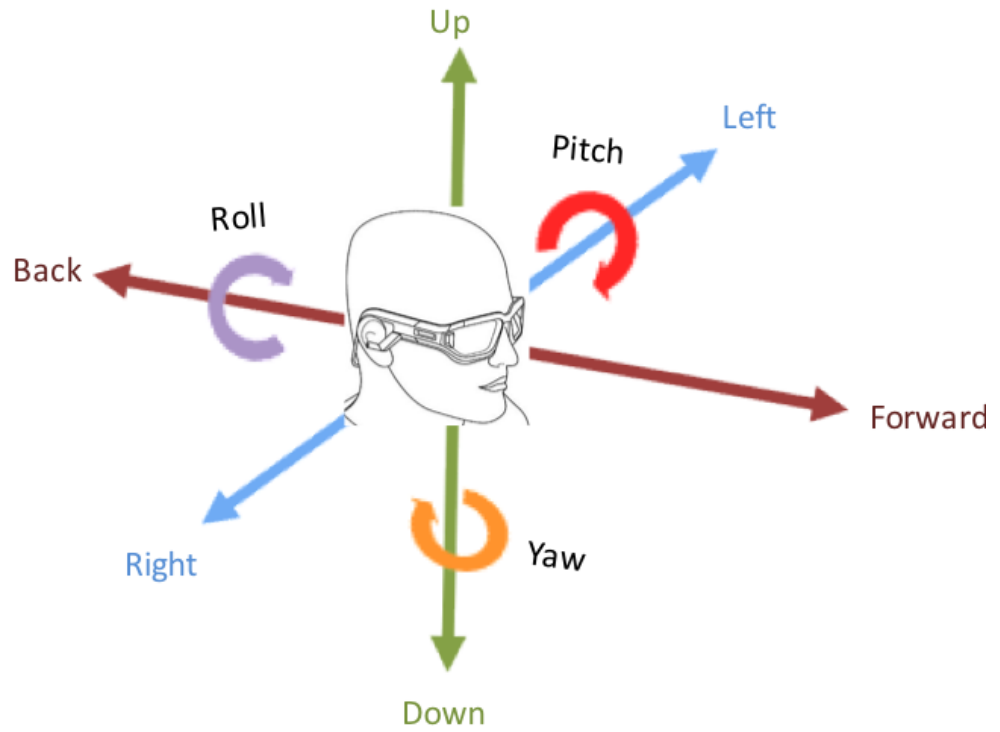


Figure 4.3: The six-degrees of freedom for a wearable camera.

the z-axis. In this work, all translation motion types (left/right, up/down and forward/backwards) were covered. The mentioned motion types can be summarised as follows:

1. Stationary camera (S): static background, moving objects.
2. Translation/rotation right (TRR), moving translation/rotation left (TRL): background change in horizontal direction.
3. Translation/rotation up (TRU), moving translation/rotation down (TRD): background change in vertical direction.
4. Moving forward (MF) or moving backward (MB): fast changes in the background and foreground.

Optical flow is defined as the apparent motion of image pixels, which is computed over two consecutive frames. It often serves as a reasonable estimation for determining objects or camera movement. By calculating pixel displacement, motion vectors (velocity and direction) can be computed.

Barron et al. [157] presented a comprehensive study discussing technical performance issues of optical flow estimation techniques. According to the proposed work and as shown in Figure 4.4, optical flow methods can be classified into four categories:

- Differential techniques [163–169],
- Region-based techniques [170–172],
- Energy-based techniques [173],
- Phase-based techniques [174, 175].

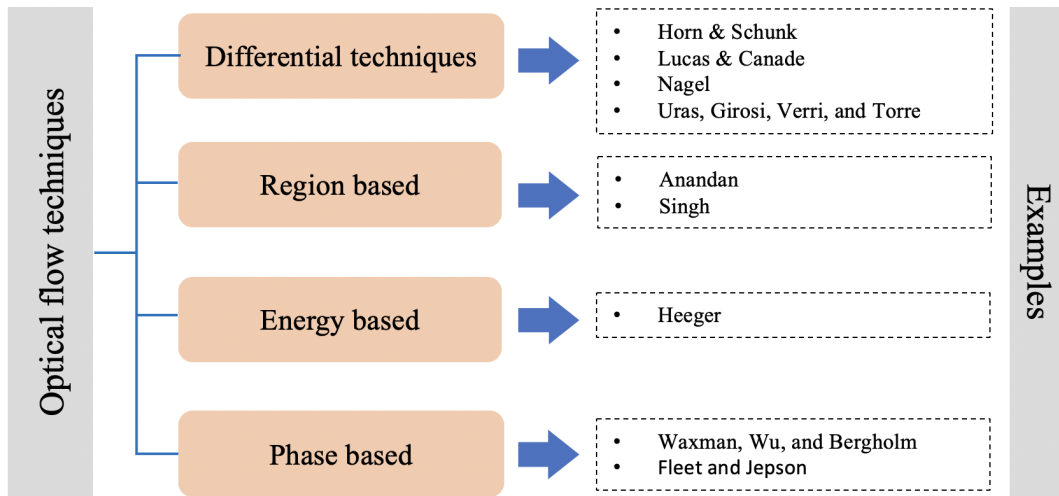


Figure 4.4: Optical flow techniques

Optical flow estimation works under two main assumptions: (1) the same object remains constant pixel intensities over consecutive frames, (2) motion is consistent

over neighbouring pixels [157]. In this work, Lucas - Kanade's [164] method was used to estimate the camera motion by calculating the optical flow for a set of predefined points. The first step is to define a set of key-points in each frame  $I_t$  and look for their corresponding location in the next frame  $I_{t+1}$ .

The Lucas-Kanade with pyramid method [164] was used to repeatedly calculate the optical flow for sparse feature points over a time window  $T$  where  $T = FR/2$  (half the video frame rate). The speed ( $|V_n|$ ) and direction ( $\theta_n$ ) for the key-points were calculated for each frame then averaged per frame to find ( $|V|$ ) and ( $\theta$ ). Then ( $V$ ) values were averaged over  $T$  to determine the current camera motion type.

Given:

$$|V| = \frac{\sum_{n=1}^N |V_n|}{N} \quad and \quad \theta = \frac{\sum_{n=1}^N |\theta_n|}{N} \quad (4.1)$$

If the velocity exceeds a pre-defined threshold  $V_{th}$ , the camera is considered to be moving, and a motion type is calculated. It was assumed that translation/rotation happens when the direction of motion  $\theta$  of the detected points falls along the horizontal axis between  $-45^\circ$  to  $45^\circ$  for TRL or between  $-135^\circ$  to  $135^\circ$  for TRR.

For TRU case, It was assumed that the direction of motion ( $\theta$ ) of the detected points falls along the vertical axis between  $-45^\circ$  to  $-135^\circ$ . Finally, it was assumed that TRD case happens when the direction of motion of the detected points falls along the vertical axis between  $45^\circ$  and  $135^\circ$ .

A neural network (NN) classifier was used for camera motion classification using the calculated average speed and direction. Each frame was divided into nine sub-regions as shown in Figure 4.5. The main goals of dividing frames into nine subregions are (1) to simplify the motion flow calculations, (2) to reduce the effect of moving objects, (3) and to provide a better representation for the camera motion using more key-points that are widely spanning over all sub-regions.

An in-street , 15 frame per second (FPS) video of about 7 minutes duration

was captured using the Epson Moverio BT-200 smart glasses. The NN model uses eighteen inputs (9-speed - direction pairs for the corresponding sub-regions) and six targets (static, left, right, up, down and forward). Several experiments were carried out to find the optimum NN configuration, and the camera motion cases were detected with 95% average accuracy.

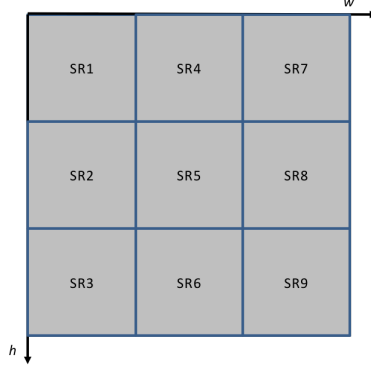


Figure 4.5: Frame sub-regions for camera motion classification.

In this work, classification result is considered to be true positive (TP) if the camera motion prediction matches with the ground truth value. The number of hidden neurons was varied from 1 to 20 to optimise the training process. Many training/testing experiments were carried out and the simulation outputs were in the range from 0 to 1 for each target. A decision threshold was used such that all outputs greater than the threshold are considered as 1 (positive) and all outputs smaller than the threshold are considered as 0 (negative). The threshold value was varied from 0.3 to 0.7 for each number of the hidden neuron. The used metrics are: true positive rate (TPR), specificity (SPC), positive predictive value (PPV), negative predictive value (NPV), false negative rate (FNR) and accuracy (ACC) as following:

$$TPR = \frac{TP}{TP + FN}, \quad (4.2)$$

$$SPC = 1 - FPR, \quad (4.3)$$

where,

$$FPR = \frac{FP}{FP + TN}, \quad (4.4)$$

$$PPV = \frac{TP}{TP + FP}, \quad (4.5)$$

$$NPV = \frac{TN}{TN + FN}, \quad (4.6)$$

$$FNR = \frac{FN}{FN + TN}, \quad (4.7)$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.8)$$

All NN configurations were compared via the ROC curve analysis and it was found the best performance (highest TPR and lowest FPR) was achieved with the 0.4 threshold value using a NN of 14 hidden neurons as shown in Table 4.1.

Table 4.1: Performance of the camera motion classification

Class	TPR	SPC	PPV	NPV	FPR	FNR	ACC
Forward	0.91	0.97	0.83	0.99	0.03	0.09	0.97
Static	0.94	0.94	0.88	0.97	0.06	0.06	0.94
Left	0.90	0.96	0.81	0.98	0.04	0.10	0.95
Right	0.96	0.99	0.98	0.98	0.01	0.04	0.98
Up	0.95	0.99	0.97	0.99	0.001	0.05	0.99
Down	0.93	0.99	0.93	0.99	0.002	0.07	0.99

### 4.3.2 Motion Detection using Stationary Camera

Object detection phase is where all critical objects are defined by their location to be tracked and classified later. This step needs the output from the previous stage (camera motion detection) to determine the best technique for moving object detection. Background subtraction method was used in the case of a stationary

camera to model the static background and segment the foreground.

The Gaussian mixture-based background/foreground segmentation algorithm [176] was used to model the background and detect moving objects. After applying the foreground mask on each input frame, moving objects were displayed as white blobs in the foreground image. Useful features (centre, size, location) were extracted after contouring the detected objects to be used in the tracking process. Figure 4.6 shows the mentioned steps.

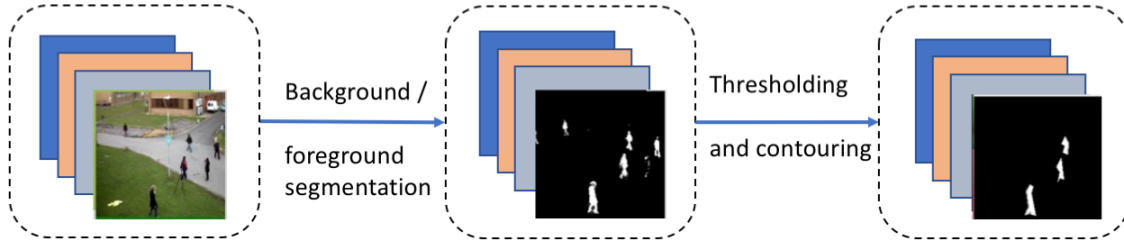


Figure 4.6: Foreground detection using Mixture of Gaussian Segmentation.

### 4.3.3 Motion Compensation for Moving Camera

In the moving camera rotation scenario, a motion compensation step was performed before detecting moving objects. The motion caused by the camera was compensated using a homography matrix ( $H$ ) that aligns the previous frame with the current one. Moving objects detection using motion compensation pipeline is shown in Figure 4.7.

The first step is to define key-points in the current frame ( $I_t$ ) to track their corresponding location in the previous frame ( $I_{t-1}$ ). Shi-Tomasi corner [177] detection algorithm was used to find the most prominent points in each frame. The corresponding location for the detected points was computed using Lucas-Kanade optical flow in pyramids method [164].

After defining the new location for each point in the frame ( $I_{t-1}$ ), a perspective transformation between the two frames was calculated to determine the homography matrix ( $H$ ). Random sample consensus (RANSAC) [178] was used to compute the



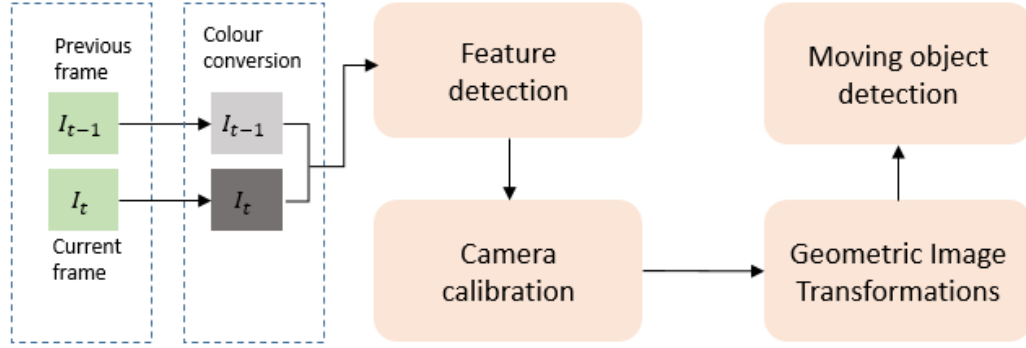


Figure 4.7: Motion compensation and detection pipeline.

homography matrix ( $H$ ). The planar homography relates the transformation between two planes (up to a scale factor) as follows:

$$s \begin{bmatrix} x'_i \\ y'_i \\ 1 \end{bmatrix} = H \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \quad (4.9)$$

with maximum allowed re-projection error to treat a point pair as an inlier as :

$$\| \text{dstPoints}_i - (H * \text{srcPoints}_i) \| > \text{RANSACReprojThreshold} \quad (4.10)$$

This matrix ( $H$ ) is used to compensate the camera motion by aligning the previous frame to the current frame using the flowing equation:

$$\hat{I}_{t-1} = H I_{t-1} \quad (4.11)$$

Figure 4.8 shows an example output of the motion compensation method. Image

(a) shows frame ( $I_{t-1}$ ), and image (b) shows frame ( $I_t$ ). Image (c) shows the warped frame using the homography matrix  $H$  calculated based on the optical flow from the two consecutive frames. Image (d) depicts the thresholding result for frame subtraction ( $c-b$ ), and image (e) shows the final output where a moving object with maximum area is detected. Red arrows show the optical flow results for the detected points. Black sides (right and top) represent the translation that occurred due to camera movement. The new images were almost identical, and the frame subtraction method detected the moving object clearly, as shown in Figure 4.8-(e).

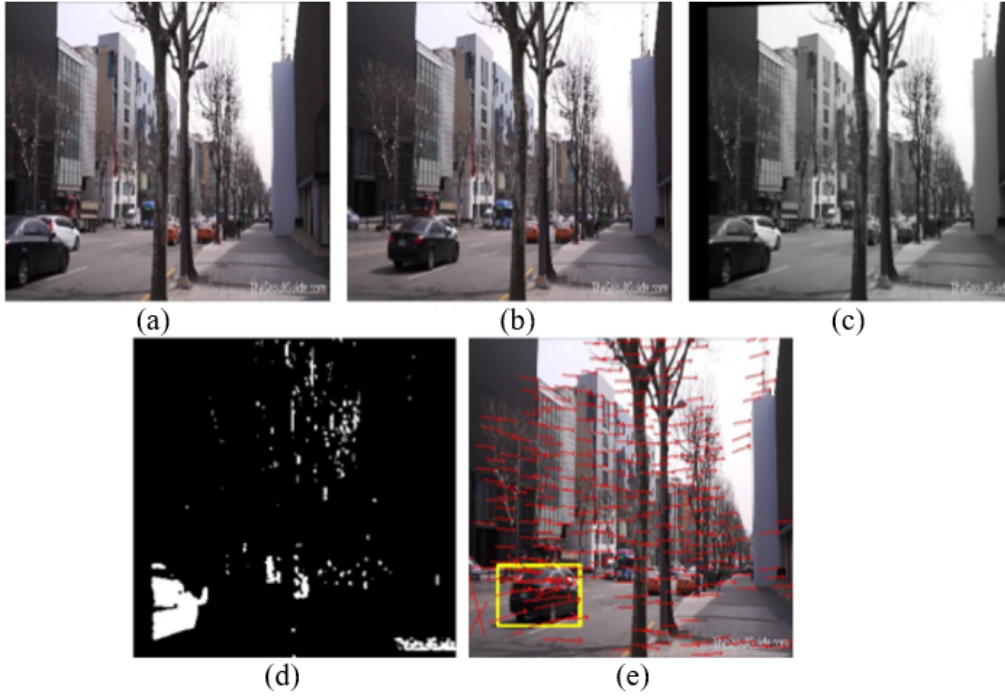


Figure 4.8: Moving object detection example after motion compensation.

It is worth mentioning that multiple noise results were expected because of the accuracy of the homography matrix used for translation. This accuracy has a strong correlation with the number of the key points used to compute the optical flow, which is a trade-off between accuracy and computation load. An additional threshold based

on blob's area was applied to extract the significant objects only.

#### 4.3.4 Motion Compensation and Object Detection Evaluation

Since the purpose of this system is to detect moving objects for people with vision impairment using smart glasses, the performance of the proposed work should be tested on moving camera videos.

To test the effectiveness of the motion compensation method, it was applied it on a video [179] containing scenes from a continually moving camera that rotates horizontally and vertically on the side of a street. Different types of moving objects appeared in this video such as cars, pedestrians, bikes and others. Detection after post-processing (performing some morphological transformations to filter out small noises) was considered to optimise the detection process. Moving object detection with a rotating camera using the motion compensation method provided good results. Around 48% of the detected objects have been filtered out without affecting the detection accuracy.

A publicly available dataset from Changedetection.net [159]- sequence (continuousPan) was used under the category PTZ was used to evaluate the object detection method after motion compensation. A detection rate of 93% was achieved on this sequence that contains 1700 frames ( $480 \times 704$ ).

This sequence was chosen because it contains scenes from a continuously moving camera. The camera pans horizontally at slow speed. Moving objects (such as cars and trucks) were seen moving fast.

Object detection methods (e.g. St-Charles et al. [180], Maddalena et al. [181], Allebosch et al. [182], Sajid et al. [183] and others) use intersection over union (IoU) metric to evaluate the precision and recall rates of their methods using the following

equation:

$$IoU = \frac{Areaofoverlap}{Areaofunion} \quad (4.12)$$

This value represents how much the predicted boundary box of the method overlaps with the ground truth boundary box. Most of the object detection methods predefine an IoU threshold=0.5 in classifying whether the prediction is a true positive or a false positive. In this work, no IoU threshold was defined and prediction was classified as true positive if the moving object was detected, regardless the bounding box. It is important to detect an approximate location which is as close as possible to the real moving object. This explains the high recall rate for this test comparing to other work. Our method achieved 93% recall, 98% specificity, 2% false positive, 7% false negative and 63% F score rates.

## 4.4 Deep Learning Object Recognition

As mentioned in the introduction of this chapter, traditional object detection approaches failed to handle moving camera challenges. The need for identifying the types of detected objects necessitated the use of object recognition method instead of the traditional moving object detection techniques.

The main goals of this stage were to obtain (1) the types of the detected objects and (2) current locations of these objects. In the related literature, researchers used you-only-look-once (YOLO) [75], faster- recurrent convolutional neural networks (R-CNNs) [184], and single-shot detectors (SSD) [185] for object detection using deep convolutional neural networks. In our work, it was found that YOLO method needs a powerful graphics processing unit to perform the classification process, which is not available in smart glasses. On the other hand, the Faster R-CNNs method is quite slow (on the order of seven frames per second), and this affects the whole process of hazard classification.

A research group originally developed SSD at Google<sup>®</sup>. The method can detect multiple objects at the same time in an image using a single deep neural network [185]. It outperforms other object recognition approaches such as YOLO and R-CNNs in terms of processing speed and accurate accuracy.

Using small convolutional filters, SSD predicts not only the class scores but also the offsets for a fixed set of small bounding boxes. Its unique architecture allows for high detection accuracy by producing predictions of different scales and separating them by aspect ratio [185]. Figure 4.9 depicts the original SSD network architecture showing its feature layers.

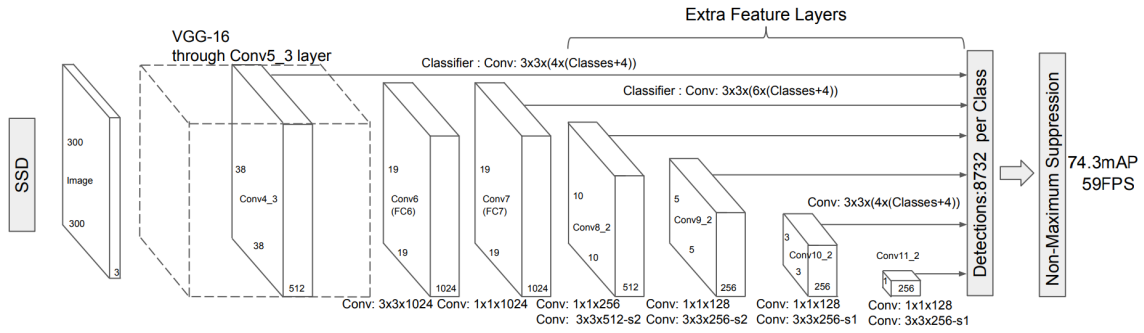


Figure 4.9: Single Shot Detector (SSD) architecture [185].

As seen in Figure 4.9, SSD method uses the VGG network [186] as a feature extractor. This architecture can be substantial in the order of 200-500MB, which makes it unsuitable for real-time applications. Therefore, a lightweight network architecture called MobileNets [187] was used because it is designed for resource-constrained devices. The main difference between this architecture and other traditional CNNs is the usage of depthwise separable convolution, as shown in Figure 4.10. The left side depicts a standard convolutional layer with batch normalisation (BN) and rectified linear unit (ReLU). The right side shows MobileNets architecture with depthwise and pointwise layers followed by batch normalisation and ReLU [187].

A combined version of SSD and MobileNets, which is called MobileNets SSD was

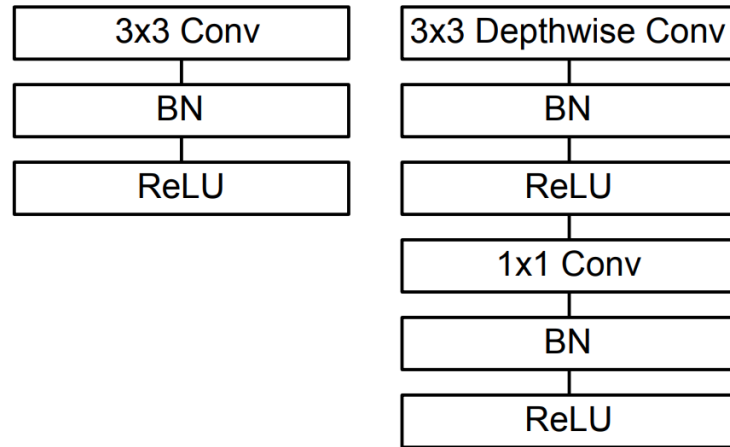


Figure 4.10: A standard convolutional layer (left) and MobileNets depthwise separable convolution (right) [187].

used. The module was trained on common objects in context (COCO) dataset [188] and then fine-tuned on a pascal visual object classes (VOC) [189] dataset to achieve better accuracy rates.

This framework was implemented using the OpenCV 3.3 deep neural network (DNN) module to create the real-time object detector that is capable of detecting 21 classes including airplanes, bicycles, birds, boats, bottles, buses, cars, cats, chairs, cows, dining tables, dogs, horses, motorbikes, people, potted plants, sheep, sofas, trains, and TV monitors. This framework was found to be the best choice to cover the object types mentioned in the user’s requirements chapter.

Huang et al. [190] compared modern object detectors (R-CNN, R-FCN [191] and SSD) using a unified implementation to draw an analogy between them in terms of speed/memory/accuracy balance for a given application and platform. According to Huang’s findings, SSD has lower sensitivity rate to the quality of feature extractor compared to other detectors. This finding makes SSD a competitive choice for resource-constrained mobile devices.

Figure 4.11 shows the trade-offs between detection accuracy (overall mAP) and GPU time. SSD models with MobileNet feature extractor achieved the highest ac-

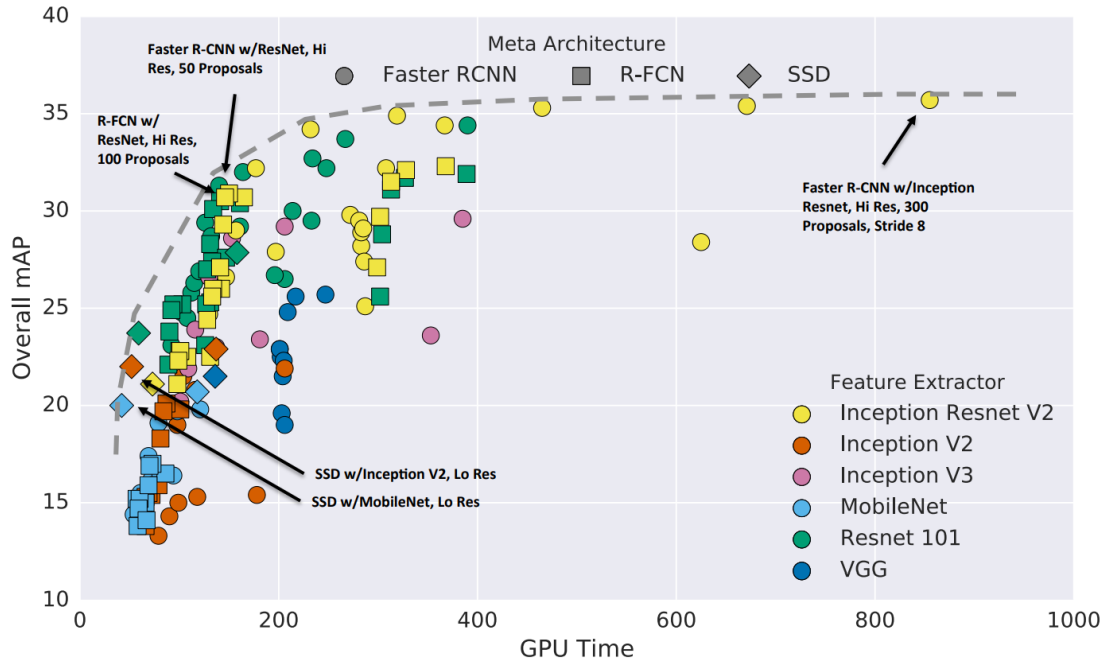


Figure 4.11: Accuracy vs time, with shapes represent object detector and colours represent feature extractor [190].

curacy among the fastest models.

SSD models are reported to have the lowest memory usage compared to other object detectors using all feature extractors, as seen in Figure 4.12. Furthermore, SSD+MobileNet configuration has the lowest computation time (GPU and CPU) compared to other configurations, as illustrated in Figure 4.13.

All these results made the usage of SSD and MobileNet the best choice for our application, which requires good object detection accuracy, fast computation speed and low memory usage.

The detection stage starts by processing each frame to extract objects' blobs. These blobs are then sent to the deep learning module to recognise the type of each detected blob. The final check is to filter out the objects with low confidence to reduce the number of false detections.

The detection performance on the datasets used in this work was lower than the

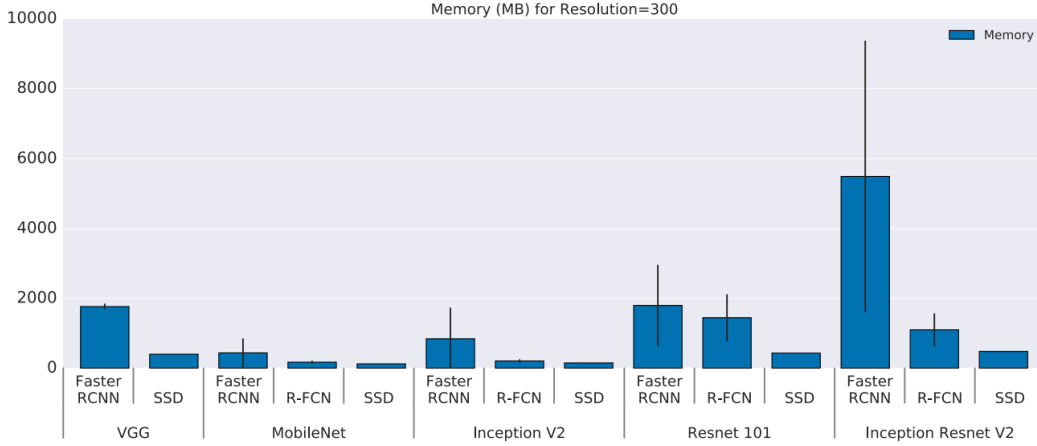


Figure 4.12: Memory usage in megabyte for each architecture/feature extractor configuration [190].

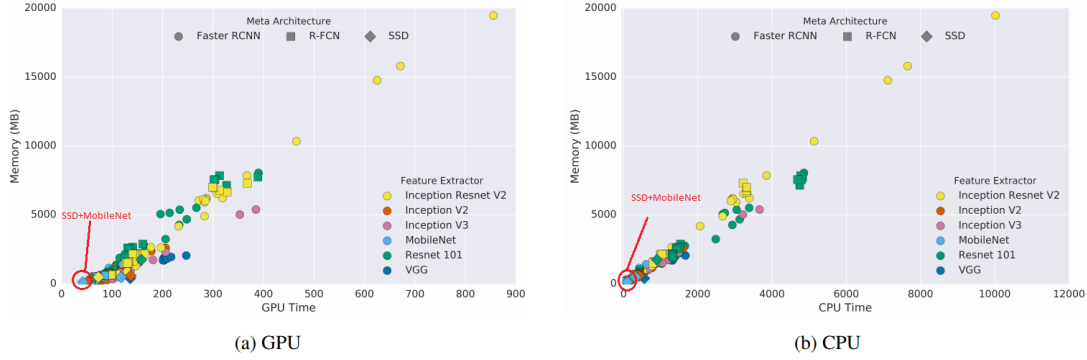


Figure 4.13: Memory usage in megabyte vs. time [190].

one reported by the developers of SSD [186]. This reduction is due to the use of lower resolution images captured by a wearable camera. Appeared objects in the data used in this work exposed to different types of deformation and light changes. Table 4.14 shows the detection performance details on the used data, where detection rate (DR) represents the ration between number of actual objects (Actual) and number of detected objects (Detected) as follows:

$$DR = \frac{Detected}{Actual} \quad (4.13)$$



As shown in the table, the lowest detection rate reported in dataset Eps9, which was captured indoor. Images in this dataset include scenes from a shopping mall where people and clothes are seen the most. The lightening was low in a crowded place, with fast camera and objects movement.

Table 4.2: Detection performance of SSD on different in-house datasets used in this work

Dataset	Resolution	Actual Objects	Detected Objects	DR	Description
CamVid	$960 \times 720$	14930	8955	60%	In-street images captured by a wearable camera.
Eps4	$640 \times 480$	1950	975	50%	Outdoor images captured by the Moverio smart glasses.
Eps9	$640 \times 480$	1668	556	30%	Indoor images captured by the Moverio smart glasses.
Eps15	$640 \times 480$	2106	1053	50%	Outdoor images captured by the Moverio smart glasses.

Figure 4.14 summarises the detection performance of SSD on the dataset used in this work.

## 4.5 Kalman Filter Object Tracking

Since the system had detected and recognised moving objects in the previous stage, the approximate location for each object was known. For each detected object, the location information was used to initialise a Kalman filter (KF) to predict its motion over time. KF is a recursive estimator [192] that predicts the state of the system  $X_t$  at time  $t$  based on information from the previous state  $X_{t-1}$  using the following

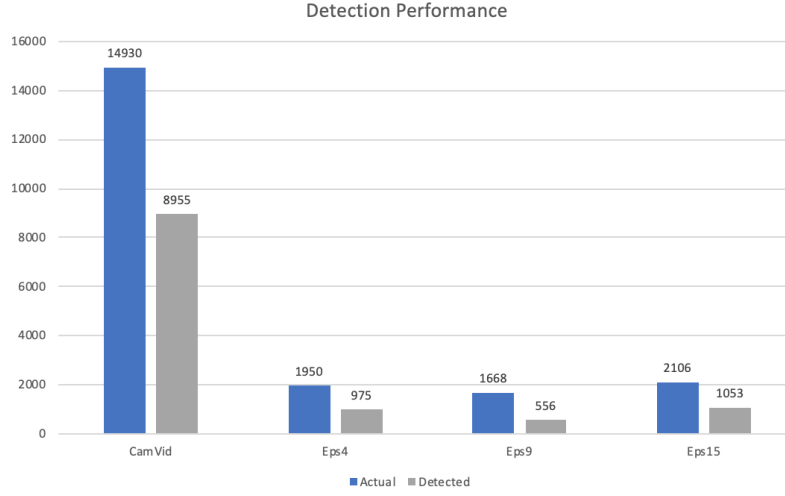


Figure 4.14: Detection performance for SSD+MobileNet on datasets used in this work.

equation:

$$X_t = F_t X_{t-1} + G w_{t-1}, \quad (4.14)$$

where  $F_t$  refers to the state transition model that describes the change which happens to the state between time  $t-1$  and  $t$ .  $w_t$  is the process noise for the process transition model which is assumed to be Gaussian white noise with covariance  $Q_t \delta(t-j) = E[w_t w_j^T]$ .

Then, the measurements' vector  $Z_t$  is computed using the following equation:

$$z_t = H_t X_t + V_t, \quad (4.15)$$

where  $H$  is the observation model and  $V_t$  is the observation noise at time  $t$  which is assumed to be Gaussian white noise with covariance  $R_t \delta(t-j) = E[v_t v_j^T]$ .

The KF estimation process has two phases: the prediction and the update. In the prediction phase, the filter uses the initial estimate state  $X_0$  and its associated variance of uncertainty (covariance) matrix  $Q_0$  to create an estimate of the current

state. For a better and more accurate estimation, the update phase computes the KF gain and uses the measurements vector from the current state to enhance the prediction result in the next state  $X_t$ .

The KF is used in this work to estimate the detected object's location and speed. Thus, the state of each object is represented as:

$$X = \begin{bmatrix} px \\ py \\ \dot{x} \\ \dot{y} \end{bmatrix}, \quad (4.16)$$

where  $px$ ,  $py$  are the centre of mass coordinates for each object and  $\dot{x}$ ,  $\dot{y}$  are the velocity components.

In the prediction phase, the system predicts both the state vector  $X_t$  and the covariance state  $P_t$  using the following equations:

$$\hat{X}_{t|t-1} = F_t \hat{X}_{t-1|t-1}, \quad (4.17)$$

$$\hat{P}_{t|t-1} = F_t \hat{P}_{t-1|t-1} F_t^T + G Q_t G^T. \quad (4.18)$$

where the transition model  $F$  is:

$$F = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4.19)$$

and

$$G = \begin{bmatrix} 1/2\Delta t^2 & 0 \\ 0 & 1/2\Delta t^2 \\ \Delta t & 0 \\ 0 & \Delta t \end{bmatrix} \quad (4.20)$$

follows logically from the Newtonian equations of motion.

To update the predicted state and covariance, three values should be computed:

(1) the error innovation (the difference between actual and predicted measurements) as:

$$\tilde{y}_t = z_t - H\hat{x}_t \quad (4.21)$$

(2) the measurement innovation covariance (the sum of predicted and measurement covariance) as:

$$S_k = H\hat{P}_{t|t-1}H^T + R \quad (4.22)$$

(3) Kalman gain (the ration between the predicted and actual covariance) as:

$$K_t = \hat{P}_{t|t-1}H^TS_t^{-1} \quad (4.23)$$

Finally, the system corrects the state vector  $X$  and the covariance matrix  $P$  using the following KF update equations:

$$\hat{X}_{t|t} = \hat{X}_{t|t-1} + K_t\tilde{y}_t, \quad (4.24)$$

$$\hat{P}_{t|t} = \chi_t\hat{P}_{t|t-1}\chi_t^T + K_tRK_t^T \quad (4.25)$$

where  $\chi_t = I - K_tH$ .

These two phases are applied for all detected objects over time to update the

motion model of each detected object.

## 4.6 Assignment Algorithms (The Hungarian)

One of the well-known challenges for tracking multiple objects at the same time is the assignment problem (deciding which detection refers to which tracked object). The presented system uses the Hungarian algorithm [193] for best assignments between detected and estimated measurements.

Figure 4.15 depicts the main idea of the Hungarian algorithm. As shown in (a), if we have a bipartite graph  $G = i, v, E$ , with  $n$  number of vertices in each partition of the graph, and  $k$  number of Edges ( $E$ ) where each edge has a non-negative weight as shown in (b), then the Hungarian algorithm will solve this problem with minimum cost  $c_{i,v}$  for all vertices. Missing edges will have zero weight and infinity cost as shown in (c) [194].

In object tracking problems, the goal of the Hungarian algorithm is to find the best assignment that has the lowest cost between detections and tracks. The cost, in this case, represents the Euclidean distance between these two sets of variables. Initially, the system defines a tracker instance for each detected object. The tracker object includes a KF and other motion features history for each object. Each time, the system detects new objects, and the multi-object tracking algorithm updates its state to include the new/old objects using Algorithm 1.

At this stage, all the detected objects have been tracked and our system can now determine the type, position and speed for each one of them. Objects with low type confidence were filtered out to reduce false alarms.

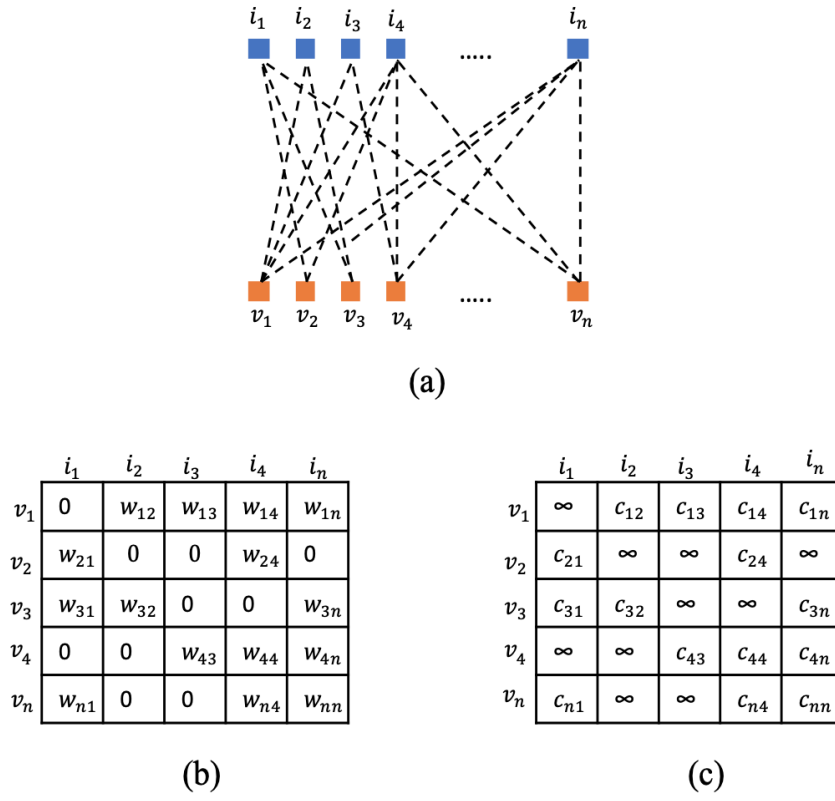


Figure 4.15: A bipartite graph (a) with its weight matrix (b) and cost matrix (c).

---

**Algorithm 1:** Multi-object tracking update procedure.

---

**Input:***D*: the detected objects' positions.*T*: the previous tracks.

```

for d in D do
  if T.size()=0 then
    for d in D do
      trackk ← newTrack(d)
      append trackk in T
    end
  else
    N ← len(T)
    M ← len(D)
    cost ← newcost[N, M]
    for t in T do
      for d in D do
        cost[d, t] ← distance(t, d)
      end
    end
    Assignments ← new Assignments[.,.]
    for t in T do
      for d in D do
        assign t to the correct d
        append correct assignment to Assignments
      end
    end
    for a in Assignments do
      identify num of assignment tracks
      if (cost[current assignment] is high)
        unsigned current assignment
      end
    end
    for t in T do
      if (t is not tracked for long time)
        delete (t)
      end
    end
    for d in D do
      if (t not in Assignments)
        trackk ← newTrack(d)
        append trackk in T
      end
    end
    for i in Assignments do
      Update i. KF
      Update i. Motion Features
      Update i. previous tracking
    end
  end
end

```

---

## 4.7 Discussion and Conclusion

This chapter discusses the methods used in this work for object detection and tracking steps. Several approaches are reported in the literature to detect moving objects in real-time applications and recognise their types.

In the early stages of this project, a motion detection using frame differencing method was used. A camera motion detection method was developed using optical flow and machine learning algorithms to detect if the camera is stationary or moving. In case of a moving camera, a motion compensation technique was applied before applying a frame differencing method to identify the independently moving objects. Despite that this method showed high motion detection accuracy, it was found that it is inefficient in the situations where the camera is moving due to the high error rates.

After discussing the project idea with participants who have different types of visual field defects, they highly recommended to include the object's type in the hazard classification phase. According to them, the object's type has a high correlation with the level of danger for each detected object. Based on this, a deep learning-based real-time object detection module was used and replaced the original motion detection methods. MobileNet SSD module that was pre-trained and tested to include 21 different classes that could exist in the user's environment was used. This method was chosen based on literature reviews, which recommended it for resource-constrained devices.

Moving objects can not be treated all in the same way. We can say that one of the apparent indicators of the extent of object hazard is its movement speed and direction. The object is tracked while it exists in the user's environment to extract this information.

Kalman Filter multi-object tracker was used to track each detected object. The purpose of this phase was to analyse the motion model for each object, which will be used later in the hazard classification phase. As Kalman filtering is all about matrices and vectors' operations, from the simple addition of two vectors to the



inversion of a matrix, it is believed that it would run in real-time applications. However, the performance of KF is highly correlated with the used processing unit. In the proposed work, the technology that would work on a wearable device to track the moving objects was presented. Since the object tracker is used to determine the motion model for each detected object, it is possible to skip some frames for detecting and tracking the object if the process would slow down the hazard detection phase.

In the following chapter, motion feature extraction and the hazard classification phases will be discussed. Also, the datasets used in this work and labelling mechanism will be explored and analysed.

# Chapter 5

## Dataset Creation and Hazard Classification

### 5.1 Introduction

In this work, it is assumed that the user will wear smart glasses while walking indoors and outdoors during the day. Therefore, it was important to use videos captured by a moving camera with variable weather and light conditions.

First, a publicly available dataset from the Cambridge-driving labelled video database (CamVid) [160, 161] was used. This dataset was chosen for two reasons: (1) it was captured using a fixed camera on a moving vehicle to show the drivers view and (2) it contains a set of different objects that our object recognition system is trained to detect. Unfortunately, the field of view information for the used camera in this video was not found. Thus, a standard dash camera field of view of  $120^\circ$  was assumed.

The proposed system was tested on the sequence Seq06R0 that contains high-quality 30 FPS footage captured in street view. It shows a video taken while the vehicle is moving amongst other vehicles (moving and stationary). Some of these vehicles were moving towards the camera, and others were moving away. Also, some pedestrians were seen crossing the road and walking aside the moving vehicle.

Figure 5.1 shows different examples from the used CamVid dataset. In this

example, several outdoor object types (moving and stationary) were seen, such as cars, pedestrians, bicycles and trees.



Figure 5.1: Examples from the CamVid dataset [160, 161].

For more realistic evaluation covering all possible hazard classes, a private dataset was created. The proposed system was trained and evaluated using the new dataset, which uses videos captured by a wearable camera with different conditions (light changes, object deformation, and object occlusion).

Moverio BT-200 smart glasses was used to capture in-street videos with the help of a sighted participant<sup>1</sup>. Figure 5.2 demonstrates the main specifications for the Epson Moverio BT-200 smart glasses. For this purpose, two indoor videos with a total of 23 seconds (644 frames) and seven outdoor videos with a total of 221 seconds (6188 frames) were captured. Figure 5.3 shows some examples of indoor and outdoor videos with the detection results (each detected object is labelled above the surrounding box). In these examples, the blurriness and the deformation of the detected objects can be seen. These issues are due to camera movement and shakiness while the user is moving. For each detected object, its hazard class was marked by an expert<sup>2</sup> to be used in the training and testing stages.

<sup>1</sup>Video capture is in accordance with the ethical approval from the research ethics committee at the Faculty of Science and Engineering, University of Liverpool, UK (Reference: 1982).

<sup>2</sup> A postdoc from the computer vision group in the computer science department, University of Liverpool

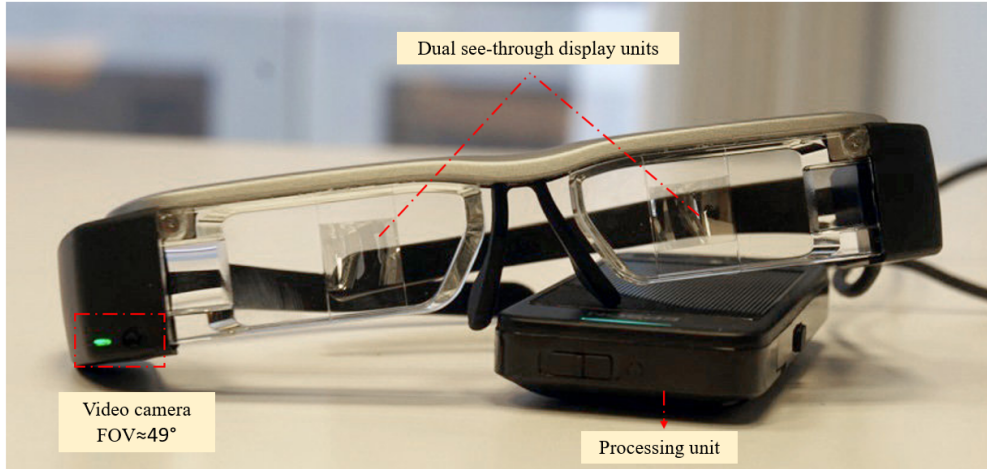


Figure 5.2: Epson Moverio BT-200 smart glasses.

## 5.2 Hazard Labelling

Unfortunately, no public or private datasets classifying the possible hazards in the visually impaired people's everyday life were found. Therefore, a new private dataset was created to be used in this work. The five hazard classes mentioned in chapter 3, which were ranked by the participants according to their level of danger, were used to label the detected objects into one of five categories.

Motion features extracted from both the detection and tracking phases were used to assign a class type for each object. The purpose of this step was to collect information about how each object is behaving while it is in the user's environment.

From the detection stage, the system recognises the type of the detected object and the confidence of that recognition and saves this information into a global feature array. The tracking method will access the same information to add the following object's features:

1. Age: a feature that represents the appearance duration (number of frames) for the tracked object;
2. Current and estimated next location. This information is important to distinguish between moving objects and static obstacles. It was extracted from the tracking record for each object;



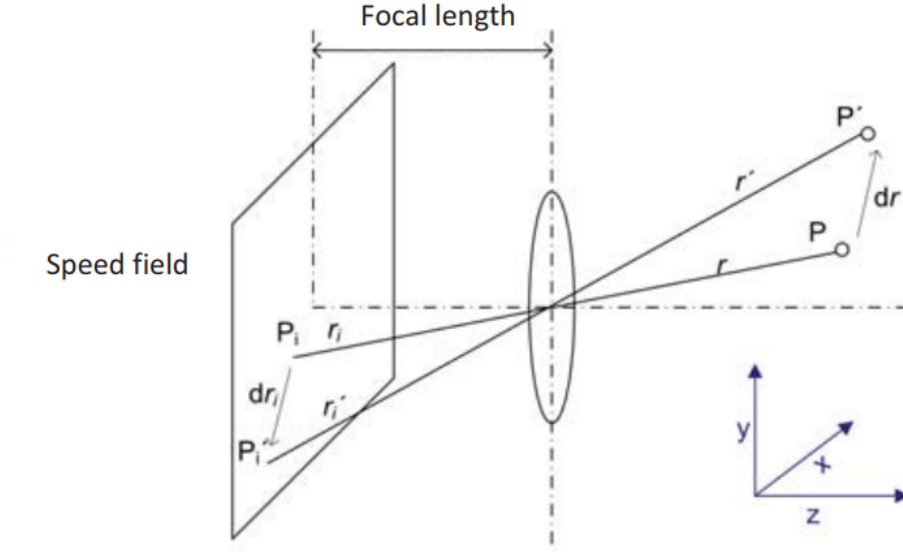


Figure 5.4: Motion projection of a moving point example. Image courtesy [195].

displacement  $dr$  was calculated as:

$$dr = \sqrt{(v_x)^2 + (v_y)^2} \quad (5.1)$$

where  $v_x$  and  $v_y$  are the displacement in both axes. Then, the speed equation was used as:

$$v = \frac{dr}{dt} \quad (5.2)$$

The frame rate ( $f_s$ ) was used to calculate the time information where the speed was measured as the object displacement over two consecutive frames. So, the speed becomes:

$$v = \frac{dr}{1/f_s} \quad (5.3)$$



The direction of movement ( $\theta$ ) of each detected object was computed as:

$$\theta = \text{atan2}(v_y, v_x) \quad (5.4)$$

Object's speed and direction information were extracted from the object detection phase. If the detector fails to locate the object, this information was estimated using the Kalman Filter prediction phase mentioned in chapter 4 (equations 4.17 and 4.18).

Figure 5.5 shows a sample motion feature for one of our testing videos, where these features were extracted over two consecutive frames. The moving object (type 15: person) was moving towards the camera. As seen, the detector recognised the object type as a person, and the tracker estimated the speed and direction for that person.

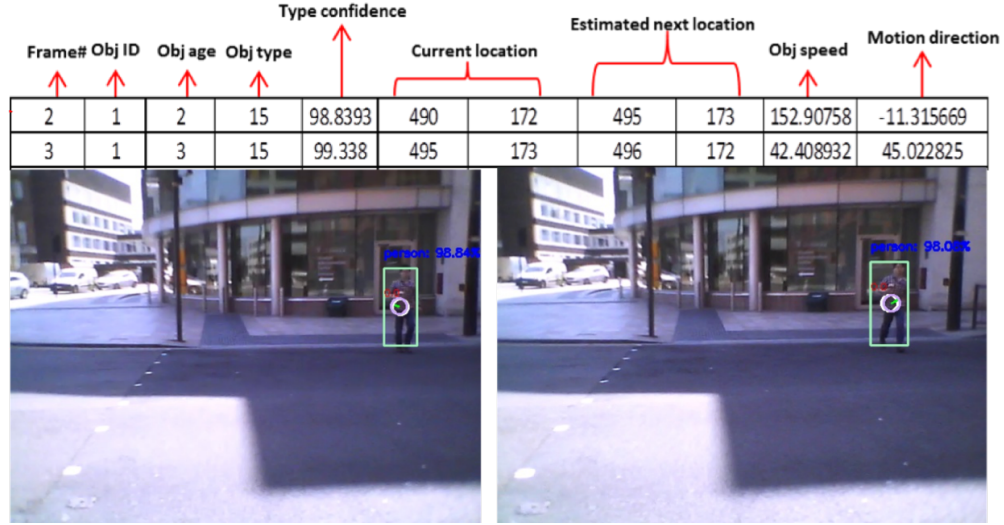


Figure 5.5: An example of the extracted features from one of our testing videos.

Figure 5.6 shows an example of the CamVid dataset. In this example, three cars were detected and tracked (red arrows), and their motion features were extracted. The fourth record in this frame refers to an object that was tracked by Kalman Filter but not detected due to the occlusion problem.

To label the data into the predefined hazard classes, a region of interest (RoI) area was defined as  $10^\circ$  around the camera fixation point (centre of the horizontal field of view, with accordance to the description in 6.1). If the object was located



Figure 5.6: Motion features example extracted from the detection and tracking phases.

inside this area, it was considered to be in the user's pathway. Figure 5.7 shows the procedure for defining RoI based on the camera's field of view.

After extracting the requisite features, data was labelled by an expert according to the hazard classes mentioned in section 3.4. Table 5.1 summarises the used specifications for the labelling process.

Table 5.1: Hazard classes specifications

Class	Number of samples	Kinematic state	Type	Location
c1	1022	static	any	outside RoI
c2	1103	moving	any	outside RoI
c3	258	static	any	inside RoI
c4	475	moving	person	inside RoI
c5	716	moving	not person	inside RoI



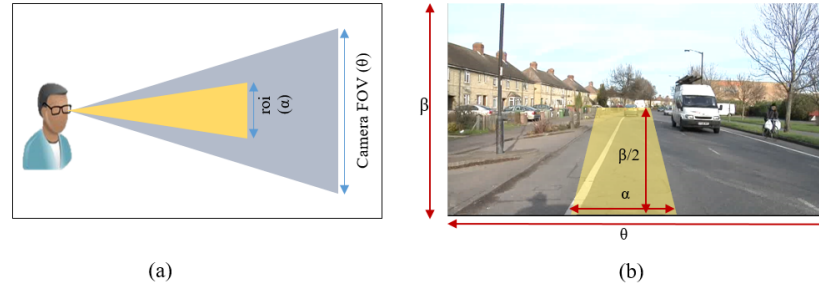


Figure 5.7: Hazard labelling according to region of interest.

### 5.3 Machine Learning-based Hazard Classification

The purpose of this stage was to classify the detected objects into one of the five hazard classes. As mentioned in the users' requirements section, visually impaired peoples' needs and challenges differ in terms of the object type, motion type and other physical features. Generating feedback for each detected object will annoy the users and disturb their healthy vision with unnecessary information. In contrast, displaying notifications with different priorities will enhance users hazard perception.

For these reasons, participants' choices were grouped into five hazard classes that were described in Section 3.4. Figure 5.8 shows a visual example of these classes. (a) The seat represents class 1; (b) The approaching bus and the pedestrian are examples of class 2; (c) The street polar is an example of class 3; (d) The crossing pedestrian is an example of class 4; (e) The approaching bus is an example of class 5.

In this work, the neural network (NN) algorithm was used to classify the detected object into one of the five hazards types mentioned before. The number of input features in the NN determines the number of input nodes, while the number of different output classes defines the number of output nodes.

The training dataset was used with the aid of the back-propagation learning algorithm to establish direct input-output connections. The network starts with no hidden nodes, then adds them gradually. Every new node was connected to every input node and to every pre-existing hidden node. Training was carried out using the training vectors, and after each pass, the weights of the new hidden nodes are

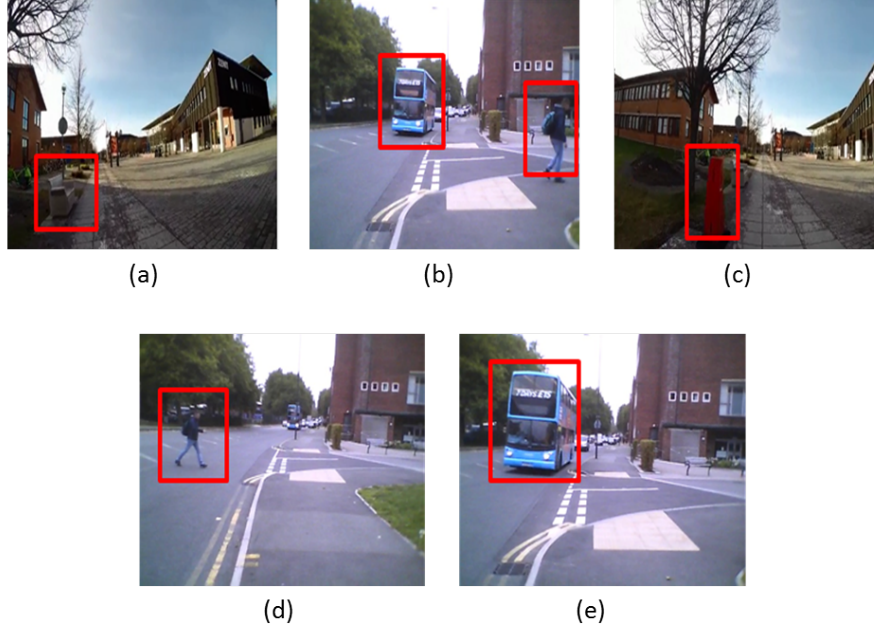


Figure 5.8: A visual example of hazard classes.

adjusted [196].

### 5.3.1 Performance Indicators

The most common way of measuring the performance of a classifier is the confusion matrix as depicted in Figure 5.9. This matrix helps the reader to understand indicators definitions while showing the classification results.

For this purpose, the following indicators were used to evaluate the classification performance: True Positive Rate (TPR), False Positive Rate (FPR), True Negative Rate (TNR), False Negative Rate (FNR), accuracy (ACC), specificity (SPC), sensitivity, and the Mean Square Error (MSE). Since the system design aims to determine if a given set of motion vector  $M_j(t)$  for object  $j$  at time  $t$  belongs to a particular hazard class  $HC_n$  (positive) or not (negative), for each hazard class, these indicators are defined as follows:

$$\mathbf{TPR}_n = \frac{TP_n}{TP_n + FN_n} \quad (5.5)$$

where  $TP_n$  (True Positives) is the total number of cases for which the system correctly classifies  $M_j(t)$  to  $HC_n$ ,  $FN_n$  (False Negatives) is the number of cases where the

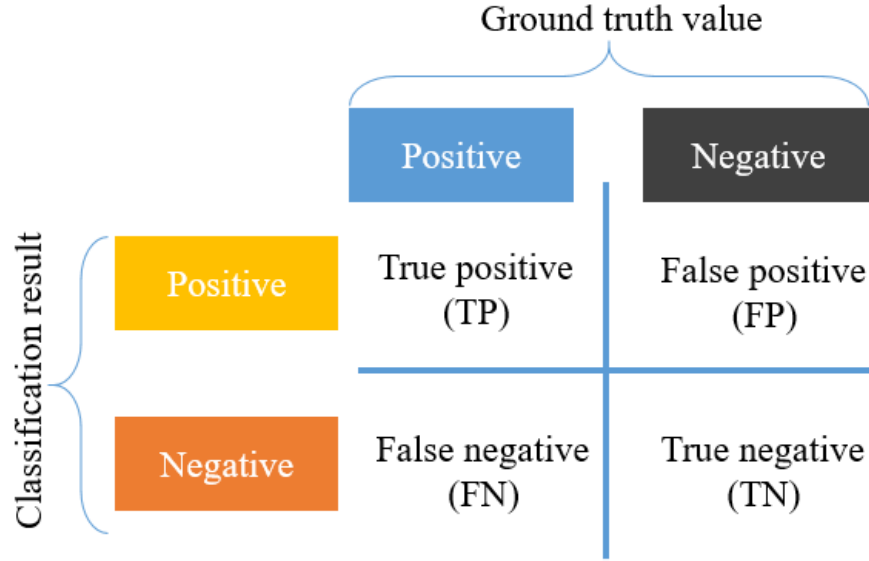


Figure 5.9: The confusion matrix for classification results.

system incorrectly classifies  $M_j(t)$  not to belong to  $HC_n$ , and  $n = 1, 2, \dots, 5$ . This ratio indicates the system's ability to correctly identify positives (**Sensitivity**).

$$\mathbf{FPR}_n = \frac{FP_n}{FP_n + TN_n} \quad (5.6)$$

where  $FP_n$  (False Positives) is the total number of cases for which the system incorrectly classifies  $M_j(t)$  to  $HC_n$ , and  $TN_n$  (True Negatives) is the total number of cases for which the system correctly classifies  $M_j(t)$  not to belong to  $HC_n$ .

$$\mathbf{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (5.7)$$

where the summation  $(TP + FP + TN + FN)$  represents the total number of samples.

**Specificity** is an indicator of the system's ability to correctly identify negatives and defined as  $1 - FPR = TNR$ .

False Negative Rate (FNR) (or miss rate) is an indicator that shows the number of cases for which the system incorrectly classifies  $M_j(t)$  not to belong to  $HC_n$ , compared to the total instances in  $M_j(t)$  is truly not belongs to  $HC_n$ .  $FNR$  can be calculated as:

$$\mathbf{FNR} = \frac{FN}{FN + TN} \quad (5.8)$$

The Mean Square Error (MSE) is an estimator that calculates the average squared difference between the estimated values (predicted by the system) and the observed values (stated in the ground truth).

MSE can be computed using the following equation:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n \left( \frac{p_i - o_i}{\sigma_i} \right)^2 \quad (5.9)$$

where:

- $p$  : is the predicted value;
- $o$  : is the observed value;
- $\sigma$  : is total number of values.

### 5.3.2 Training and Testing Experiments

The presented experiments (training and testing) in this work were performed on a MacBook<sup>®</sup> laptop (2.7 GHz Intel<sup>®</sup> Core i5 processor, 8 GB RAM) which was able to process the high resolution CamVid videos (30 FPS) with an average of 0.2160s per frame and an average of 0.1932s for videos captured by the Moverio BT-200 smart glasses.

#### Experiment 1: Hazard classification based on spatial motion features

A three-layers NN model was created with seven inputs to the input layer representing the detection and motion features: object type, detection confidence, object age, object location  $(p_x, p_y)$ , object speed, and motion direction). The output layer has five nodes representing the five hazard classes, as described in Figure 5.10.

For each detected object, the classifier classifies its hazard class based on its motion features. Some of these objects may change class over time, depending on the way it is moving around the user. For each object, its class was determined for every frame in which the system can detect it. It is important to mention here that these classes do not reflect the actual degree of danger but only determines

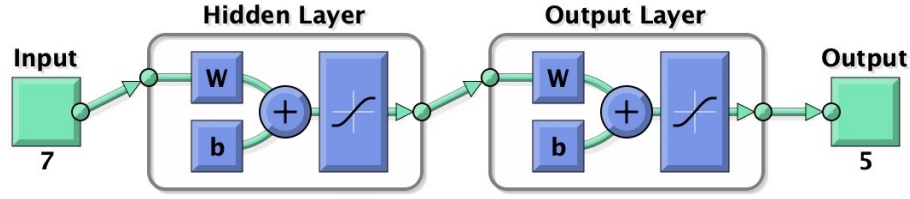


Figure 5.10: Neural network structure (Exp1).

the hazard type. The degree of danger will be determined later in the notification module of the system.

NN map non-linear inputs to the input layer through adjustable weights among the hidden layer into the desired targets at the output layer. Figure 5.10 shows an example of a three-layer NN model of seven inputs, ten hidden neurons, and five outputs.

All training and testing experiments were carried out using the MATLAB NN toolbox with the aid of the back-propagation learning algorithm [197]. To optimise the model performance, the number of hidden neurons was incremented from 1 to 20, and at each value of hidden neurons, ten experiments were carried out using a different set of randomly mixed samples consisting of 80% of the samples for training; 5% for validation, and 15% for testing. The average MSE for each of the ten experiments was calculated to evaluate the performance per specific number of hidden neurons.

The datasets described earlier in this chapter were used to evaluate the classification model. A total of 3536 samples were used, and the best NN configurations were found to provide the lowest FPR and the highest TPR for all the hazard classes using 19 hidden layers and a 0.3 decision threshold. The best NN configuration was found to provide the highest TPR of 90% with the lowest FPR of 7%. An average of 13% FNR was achieved. This value represents the average values for all classes together.

The average MSE value for the five classes was 8.7655%. Figure 5.11 shows the receiver operating characteristics (ROC) space for the configuration optimisation, where (a) shows results for class 1, (b) results for class 2, (c) results for class 3, (d)

results for class 4, and (e) results for class 5.

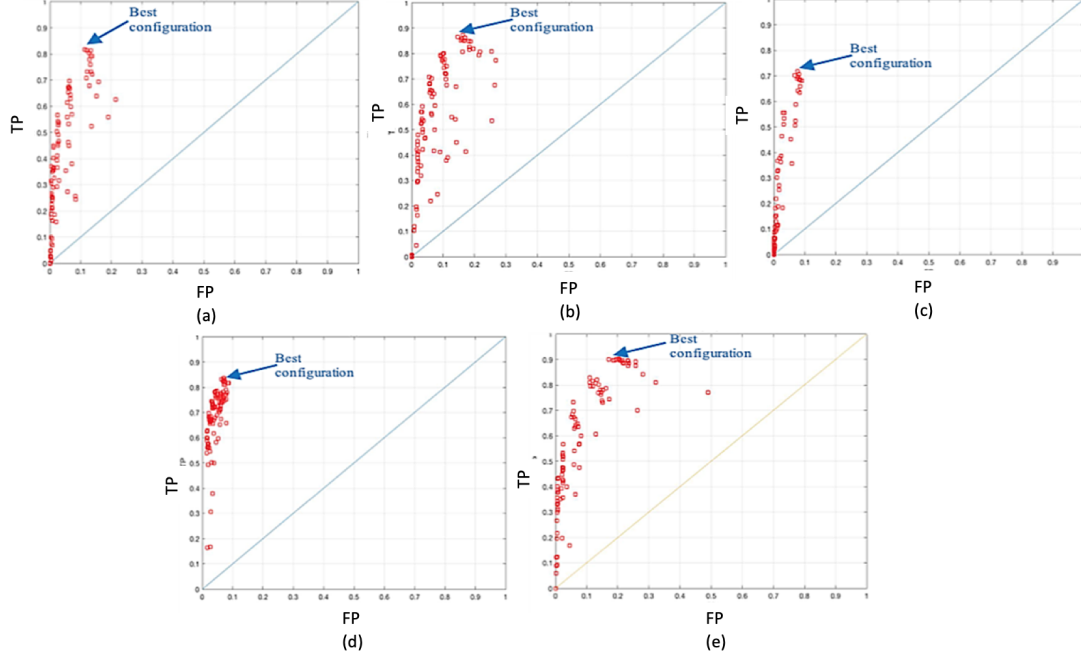


Figure 5.11: ROC spaces for the NN optimisation (Exp1).

A regression analysis was applied to the classification results to understand the relationship between the predicted hazard class (dependent variable) and the extracted features (independent variables).

In statistics, the coefficient of determination denoted  $R^2$ , is a statistical measure representing the relationship between the regression predictions and real data points [198]. Its value extends from 0 to 1, with  $R^2 = 1$  indicating that the regression prediction model fits the data correctly.

The best coefficient of determination for all phases (training, validation and testing) was  $R = 0.72$ , meaning that the regression model can reasonably predict the hazard class correctly. The regression performance for training, validation, and testing are presented in Figure 5.12.

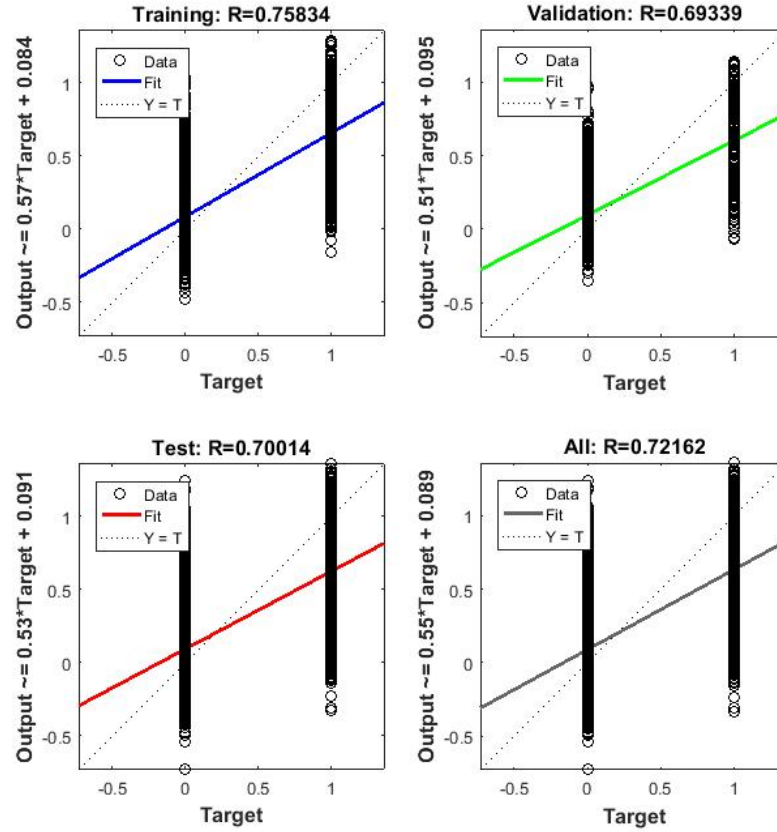


Figure 5.12: Correlation coefficients calculations results (Exp1).

### Experiment 2: Hazard classification based on temporal motion features

The second experiment for hazard classification used temporal motion features for each object to determine its class type. This method was inspired by the human's visual perception of moving objects in the peripheral environments.

Our brain detects objects in the peripheral field and evaluates if and how they are moving around us to determine any possible immediate or imminent threats. To mimic this process, temporal motion information for each detected object was tracked over a time frame ( $t$ ) for further hazard classification.

Figure 5.13 depicts the conceptual diagram of the second experiment of hazard classification. As shown in the figure, the same procedure of object detection and tracking as in the first experiment was applied. The difference is in the motion features extraction and classification phases, where these features are saved and

processed temporally (over a time period ( $t$ )) instead of a single case each time.

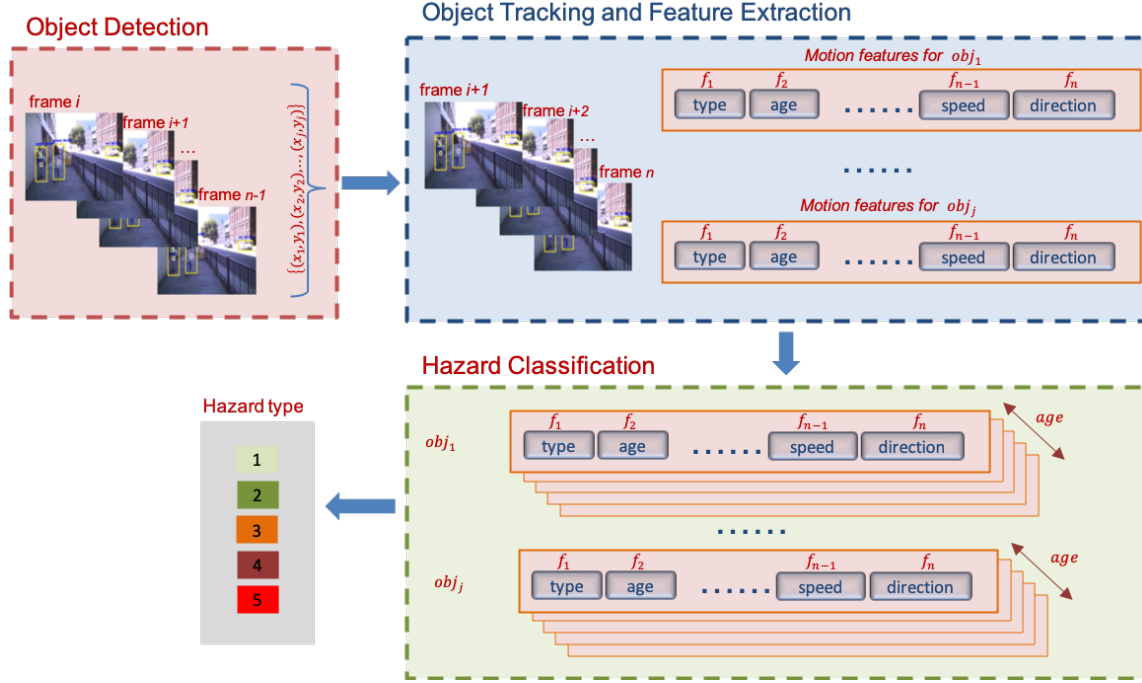


Figure 5.13: Experiment 2 conceptual diagram.

A history of five frames was used in the classification phase to create the motion vector  $M_j(t)$  for object  $j$  at frame  $t$  as:

$$M_j(t) = [T_j, Con_j, Curx_j, Cury_j, Ex_j, Ey_j, S_j, D_j] \quad (5.10)$$

where  $T_j$  is the object type,  $Con_j$  is the confidence,  $Curx_j$  and  $Cury_j$  are the  $x$  and  $y$  components of the current position,  $Ex_j$  and  $Ey_j$  are the  $x$  and  $y$  components of the estimated position,  $S_j$  is the speed and  $D_j$  is the motion direction with respect to the camera. The temporal motion vector  $TM_j(t)$  for object  $j$  at frame  $t$  has been formed using five frames as:

$$TM_j(t) = [M_j(t-1), M_j(t-2), M_j(t-3), M_j(t-4)] \quad (5.11)$$

Figure 5.14 shows an example of data representation for the second experiment. As seen in the figure, the temporal data for object two was processed every five



frames. Therefore, the input layer contains 36 variables instead of 7 as in the first experiment.

frame	ID	age	type	conf	prev x	prev y	cur x	cur y	speed	direction	Hazard level				
											1	2	3	4	5
3	2	2	7	99.1	297	374	304	377	228.24	-23.21	1	0	0	0	0
4	2	3	7	99.7	304	377	305	378	42.38	-45.02	1	0	0	0	0
5	2	4	7	99.6	305	378	303	379	67.01	-153.51	1	0	0	0	0
6	2	5	7	99.5	303	379	301	381	84.77	-135.07	1	0	0	0	0
7	2	6	7	99.3	301	381	299	382	67.01	-153.51	1	0	0	0	0

7	99.1	99.7	99.6	99.5	99.3	297	304	305	303	301	374	377	378	379	381	304	305	303	301	299	377	378	379	381	382	228	42	67	85	67	-23	-45	-154	-135	-154	1
---	------	------	------	------	------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----	----	----	----	-----	-----	------	------	------	---

Figure 5.14: Data representation example for Exp2.

A NN model was created to classify the temporal models with 36 input variables and a target function of five nodes (hazard classes) as seen in Figure 5.15.

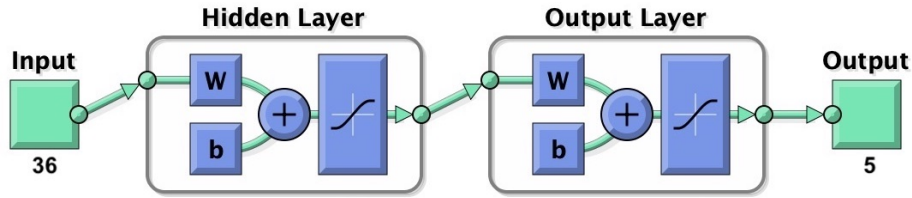


Figure 5.15: Neural network structure (Exp2).

The model optimisation for the five classes 1, 2, 3, 4 and 5 are shown in Figures 5.16(a)-(e), respectively. The best NN configuration was found to provide the highest TPR of 97.15% with the lowest FPR of 7.9% for the hazard outputs using 18 hidden neurons and a 0.3 threshold for each output class.

As shown in Figure 5.17, the best validation MSE was 5.68%, and the average testing MSE was 6.415%. Regression performance results for training, validation, and testing for the second experiment are presented in Figure 5.18.

## 5.4 Discussion and Conclusion

This work is part of a larger project for developing a user-centred, wearable assistive device for people with visual field defects. In this work, an assistive technology

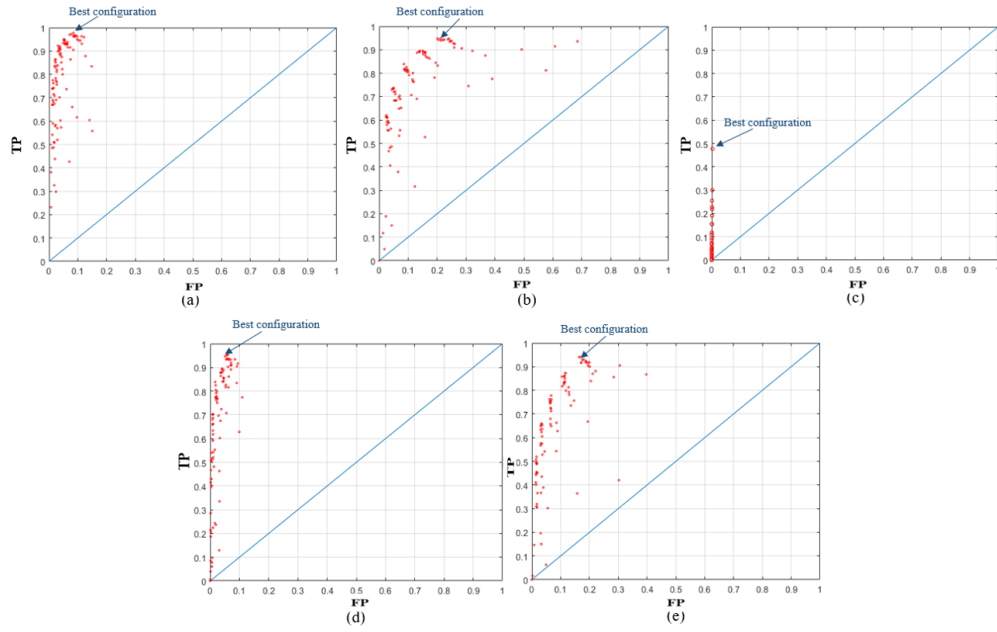


Figure 5.16: ROC spaces for the NN optimisation (Exp2).

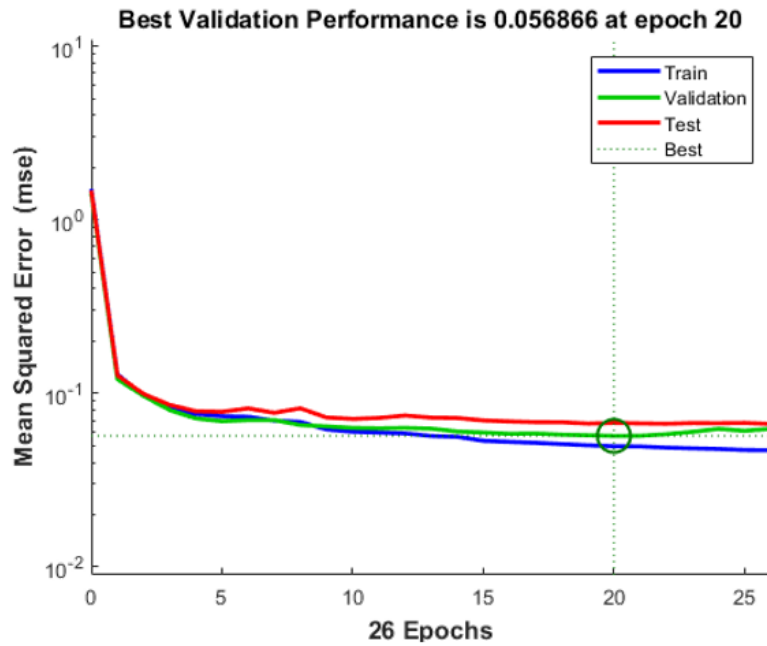


Figure 5.17: Neural network training performance (Exp2).

for people with peripheral vision loss was presented. Therefore, the performance of hazard detection and classification subsystems was analysed and evaluated based on users' recommendations.

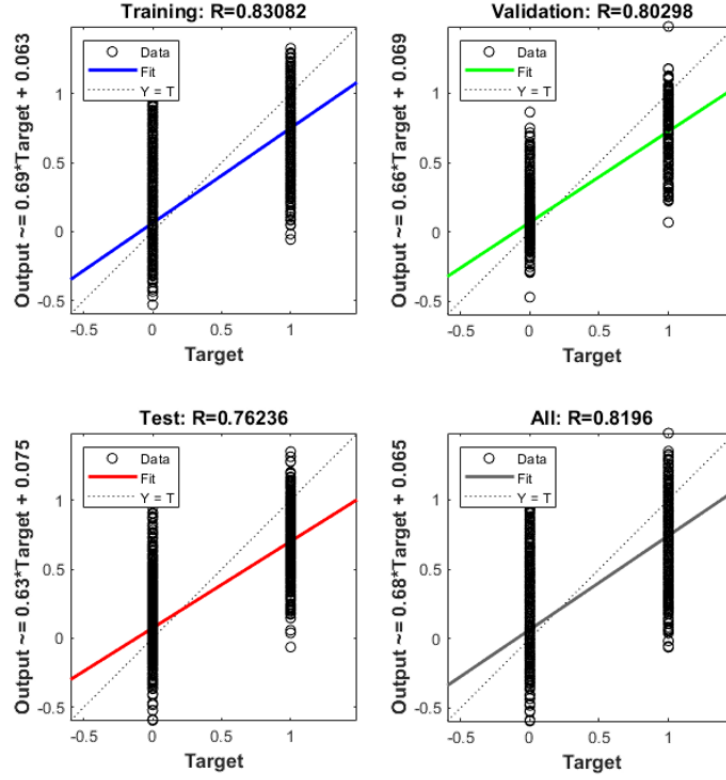


Figure 5.18: Correlation coefficients results (Exp2).

The proposed system was implemented on the Moverio BT-200 smart glasses, which captures videos at 15 FPS. The average processing time of a single frame at the glasses was 0.49 seconds. Based on this, the glasses can process at least two FPS, which means input frame rate could be reduced to guarantee real-time feedback generation. Thus, the glasses, process one frame per seven captured frames without affecting the overall detection accuracy and is considered sufficient.

Findings in this chapter show that the second experiment configuration outperforms the first one. The TPR for most of the hazard classes increased, and the FNR decreased. Tables 5.2 and 5.3 show the average TPR, SPC, FPR, FNR and ACC values for the best configuration for each experiment.

The overall system performance was improved using the temporal model. FPR (wrong alarms) and FNR (missed hazards) values decreased for most of the hazard classes. It was noted that class 3 has lower TPR compared to other hazard classes. This was because the number of samples for this class was the lowest compared to

Table 5.2: Experiment 1 classification results

Class	TPR	SPC	FPR	FNR	ACC
c1	0.92	0.88	0.12	0.08	0.89
c2	0.89	0.83	0.17	0.11	0.85
c3	0.64	0.97	0.03	0.36	0.94
c4	0.97	0.90	0.10	0.03	0.91
c5	0.92	0.85	0.15	0.08	0.86

Table 5.3: Experiment 2 classification results

Class	TPR	SPC	FPR	FNR	ACC
c1	0.97	0.92	0.08	0.03	0.93
c2	0.95	0.79	0.21	0.06	0.85
c3	0.68	0.99	0.01	0.32	0.99
c4	0.95	0.94	0.06	0.05	0.94
c5	0.93	0.82	0.18	0.07	0.85

other types (see Table 5.1).

The definition of this class is a static object in the user’s pathway. Since the data used in this project was collected using wearable smart glasses and the person usually avoids obstacles while walking, it was hard to include a large number of samples from this class. Although the FPR for this class was the highest between all other classes, we believe it is still acceptable since users will be able to see it easily by moving their gaze. This problem can be solved using obstacle detection and avoidance algorithms such as the solutions presented by Pundlik et al. [61] and Balakrishnan et al. [65].

As the idea of hazard classification using wearable cameras is considerably new, we can argue that this is a good result in this field. These results are promising and could be used to determine hazard classes in real-time applications to help people with impaired vision in their daily activities.

# Chapter 6

## Vision Modelling

### 6.1 Introduction

Foveated rendering is an image processing technique to create images with full resolution in the eye fixation point and progressively fewer details outside [199]. This technique has been used recently in the virtual reality products to lower computation cost and speed up the processing time. Researchers applied this technique to imitate a human's healthy vision, showing full resolution in the user's central vision and lower resolution in the peripheral vision.

The main aim of this chapter is to model healthy and defected vision using image processing techniques. These models will be used to understand visually impaired people's vision and to test the visual feedback phase proposed by our work.

As shown in Figure 6.1, the spatial accuracy of the normal vision decreases as a function of vision eccentricity. Researchers explain this degrade by reducing contrast sensitivity, which is described as a progressively less detailed, texture and contours [200]. The horizontal axis represents the visual field extent for the left eye and the vertical axis represents the visual acuity associated with the visual field area. As shown in the figure, the visual acuteness decreases from 100% in the foveal visual field (less than  $10^\circ$ ) to a deficient degree (less than 1%) in the peripheral visual field (more than  $50^\circ$ ).

In this work, the standard vision is synthesised using different Gaussian levels

to reflect the spatial accuracy of the normal vision, as shown in Figure 6.1. In Section 6.3, the personalised defected vision model is presented using three different examples: tunnel vision, left hemianopia and central scotoma (AMD). Videos captured by the Moverio BT-200 smart glasses were used in these examples to reflect real-world scenarios.

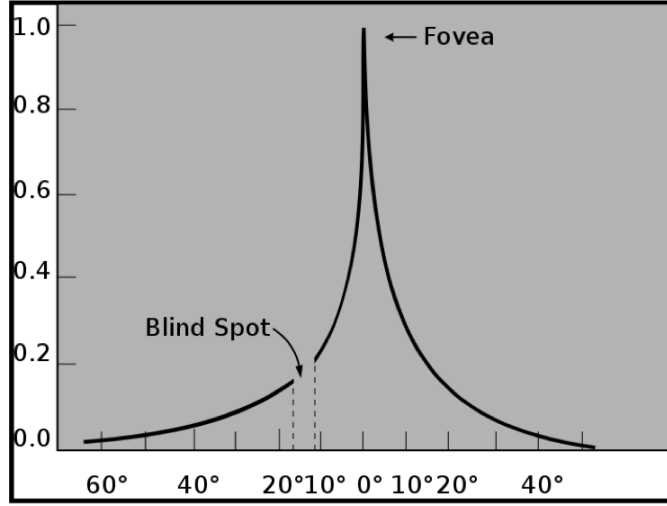


Figure 6.1: Human’s visual field extension with its corresponding visual acuity (left eye) [201].

## 6.2 Healthy Vision Model

To demonstrate the healthy vision model, a Gaussian blurring filter was used with the standard deviation corresponding to the distance away from the fixation point (the centre of the image).

Inspired by Figure 6.1, the visual field was divided into four areas, foveal ( $5^\circ$ ,  $\sigma = 0$ , full resolution), parafoveal ( $8^\circ$ ,  $\sigma = 2$ ), macular ( $18^\circ$ ,  $\sigma = 3$ ) and peripheral vision (more than  $18^\circ$ , lowest resolution,  $\sigma = 4$ ).

Figure 6.2 shows contour graphs of the resolution maps covering angles of  $49 \times 42$  and the transformed image on a  $640 \times 480$  image size captured using the Epson Moverio BT-200. Part a is the resolution map according to Figure 6.1, and part b is the original image. The four different resolution levels applied to the original image

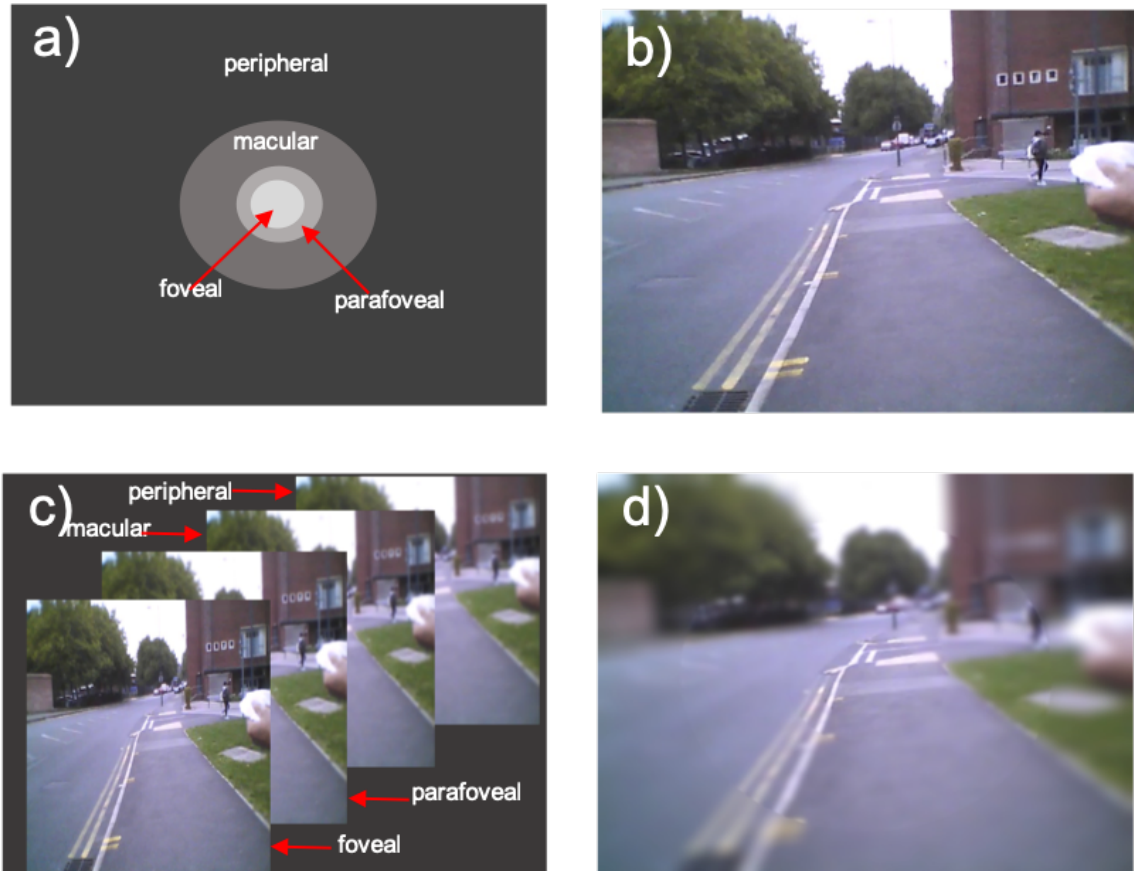


Figure 6.2: Resolution maps of the human visual field and the result of the transformation method on an example image.

are shown in part c. The rendered full image after applying the four resolution levels is displayed in part d showing how a normal (healthy) vision person sees the world.

## 6.3 Personalised Vision Model

To model the defected vision for visually impaired people, both central and full visual field test results were used. The grayscale map of HVFA test as (shown in Figure 3.12) was used to extract information for the central visual field, while Goldmann test was used to extract information for full visual field.

In contrast to HVFA, Goldmann visual field (GVF) perimetry is not popular and few skilled perimetrists can perform it. It requires manual mapping for the visual field without the help of a computer tool. An isopter is drawn around the points that have been tested to show the visual field extension. Different isopters could be seen representing the size of the stimulus and the attenuation of the light. scotomata (decreased sensitivity) areas are shown as a shaded isopter with a solid colour [202].

The personalised vision model uses the grayscale map to create a real-time synthesised view for users with defected vision. The purpose of this model is to (1) show the actual visual sensitivity for each user according to their real visual field test results, and (2) to be used by the hazard detection system to produce the early visual notification in the correct place.

### 6.3.1 Image Pre-processing

Figure 6.3 presents the three inputs used in this work to create a personalised vision model. The top image shows central and full visual field tests results. Both of these tests were collected from our patient group subjects and validated by Professor Fiona Rowe. The bottom image shows the input frame that will be used in all the following examples. The model was applied on an outdoor video captured by the Moverio BT-200 smart glasses.

For the central visual field, the test checks the visual sensitivity for each eye's visual field (30° around the fixation point) and displays different grey levels for



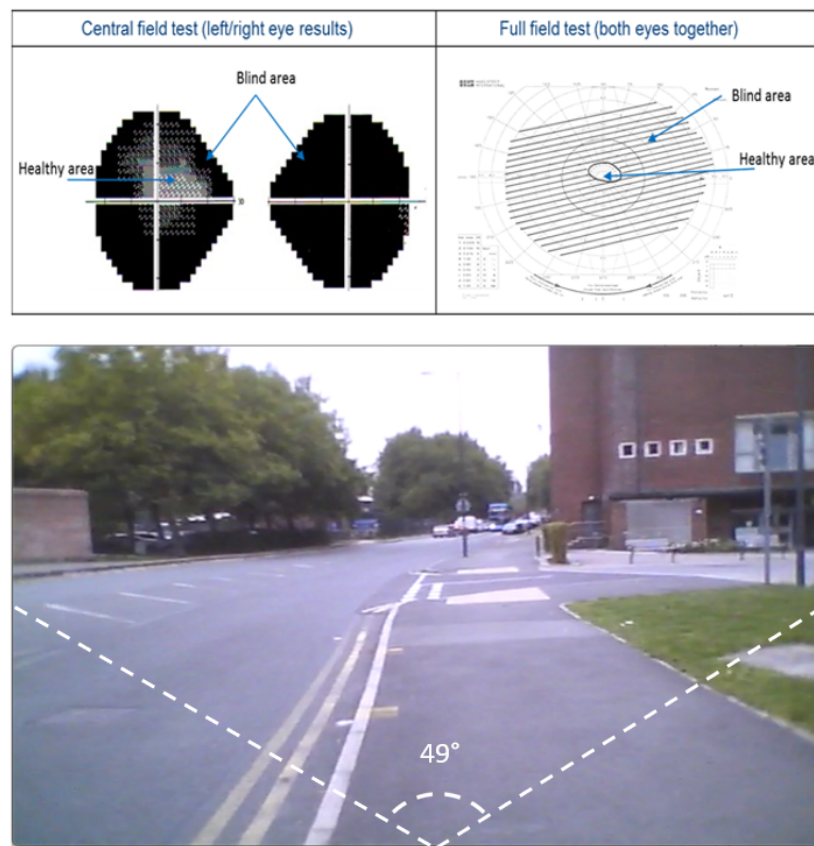


Figure 6.3: The personalised vision model inputs.

each location representing different visual sensitivity. A full field test result covers  $\approx 160^\circ$  and shows binocular vision test results.

Figure 6.4 presents the main steps to create the personalised vision model according to the user's visual field test results. The result of each step is shown in the figure. The first step is to load the visual field tests printout file with a predefined test mask according to the eye used in the test. Hence the HVFA tests  $30^\circ$  around a fixation point for each eye, a zero-padding around the test result is needed to fill the untested areas in the image in the case of peripheral vision loss.

To extract the grayscale map, the initial test is cropped before starting a series of general image transformations. Transformations include:

1. Removing the vertical and horizontal axes. This step is performed manually to produce a homogeneous shape for the grayscale plot.
2. Removing small noises by convolving the image with a normalised box filter. It takes the average of all the pixels under a kernel area and replaces all elements with the average value. A kernel size of  $3 \times 3$  was used in this step.
3. Smoothing the resulted image using a low-pass filter to remove large noises such as big edges. A  $5 \times 5$  averaging filter kernel was used in this step.
4. Combining the result in the zero-padding image.
5. Resizing the combined image to the same final image size and sampling the resulting image according to the given grayscale map key (see Figure 3.12). The pixel values were adjusted according to the blue chart in step 5 of Figure 6.4.
6. Applying the result of step 5 on the final image/video frame (after splitting it into R,G,B channels) using pixel multiplication method.
7. Combining the three channels (RGB) into one final image and display it.

### **Tunnel vision example**

Figure 6.5 depicts a tunnel vision model example. The left side part (A) shows both central and full-field tests results. The printouts show that the patient suffers

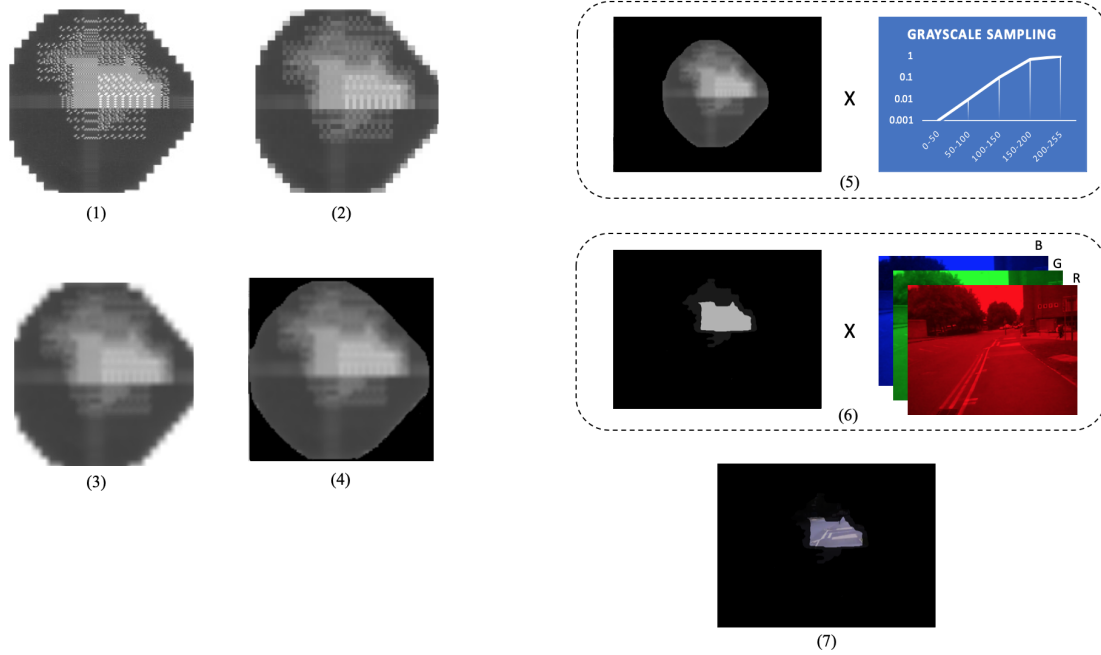


Figure 6.4: The personalised vision model image processing.

from severe glaucoma, with total blindness in the right eye. His left eye has a small healthy vision in the centre of his visual field. Image A-bottom shows the full field result.

The second step is to apply image processing methods described earlier in this section to extract the final visual field case from these tests results. Image B-1 shows the output after removing the vertical and horizontal axes. Images B-2 and B-3 show the output after applying average and smooth filters, respectively. Finally, image B-4 depicts the same image after sampling its values according to the given key.

The output of all these methods is applied to each video frame, as shown in Figure 6.5-C. The black areas in this image represent the total blindness, where the healthy vision area is shown in the centre of the visual field.

### Left hemianopia example

Figure 6.6 depicts a left hemianopia vision model example. The left side image (A) shows both central and full-field tests results. The printouts show that the participant has total blindness in the left side of both eyes. Half of the right side field of the left eye is defected too. His right eye has a healthy right side visual field,

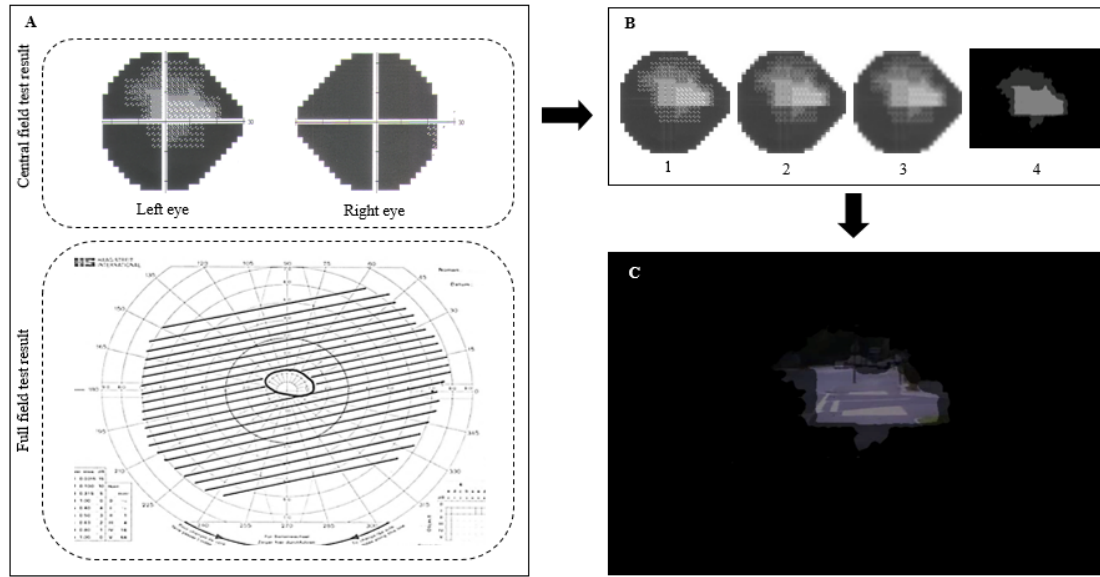


Figure 6.5: Tunnel vision model example.

leading to left hemianopia case, as shown in image A-bottom.

Image B-1 shows the output after removing the vertical and horizontal axes. Images B-2 and B-3 show the output after applying average and smooth filters, respectively. Finally, image B-4 depicts the same image after sampling its values according to the given key.

The output of all these methods is applied to each video frame, as shown in Figure 6.6-C. The black areas in this image represent the total blindness, where the healthy vision area is shown on the right side of the visual field.

### Central scotoma (AMD) example

Figure 6.7 depicts an AMD vision model example. Left side image (A) shows both central and full-field tests results. The printouts show that the participant suffers from a central scotoma, with total blindness in the central vision for both eyes. These images represent the results for  $10^\circ$  only. Image A-bottom shows the full field result.

Image B-1 shows the output after removing the vertical and horizontal axes. Images B-2 and B-3 show the output after applying average and smooth filters, respectively. Finally, image B-4 depicts the same image after sampling its values

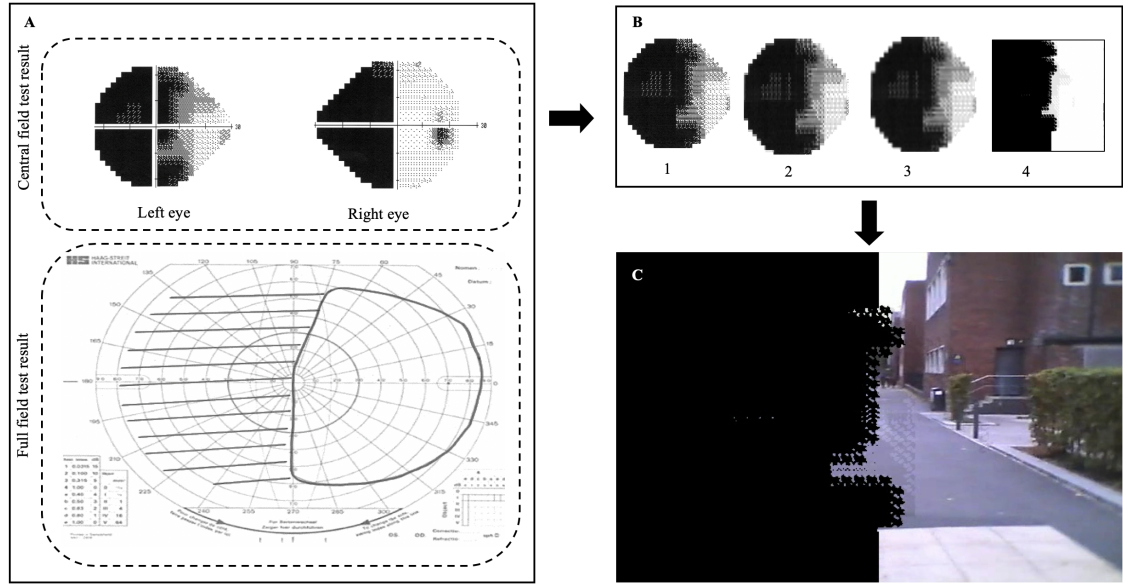


Figure 6.6: Left hemianopia vision model example.

according to the given key. The output of all these methods is applied to each video frame, as shown in Figure 6.7-C. The black areas in this image represent the total blindness, where the healthy vision area is shown in the peripheral visual field.

### 6.3.2 Visual Feedback Design to Enhance Visual Perception

Using visual feedback to enhance the user's visual perception is used in many research fields, such as 3D video games [203], endoscopic surgery [204], sports [205], care driving [206], and low vision rehabilitation [150].

With the recent growth in AR research and applications, researchers developed different visualisation schemes to help users to accomplish daily tasks with the help of computer-generated information. This includes but is not limited to:

- **Navigation and wayfinding**, such as the work proposed by Virtual Cable [207], Charissis and Naef [208] and Sato et al. [209].
- **Hazard perception**, such as visual longitudinal and lateral driving assistance system proposed by Tonnis et al. [210] and the work developed by Sauerbrey and Jens [211].

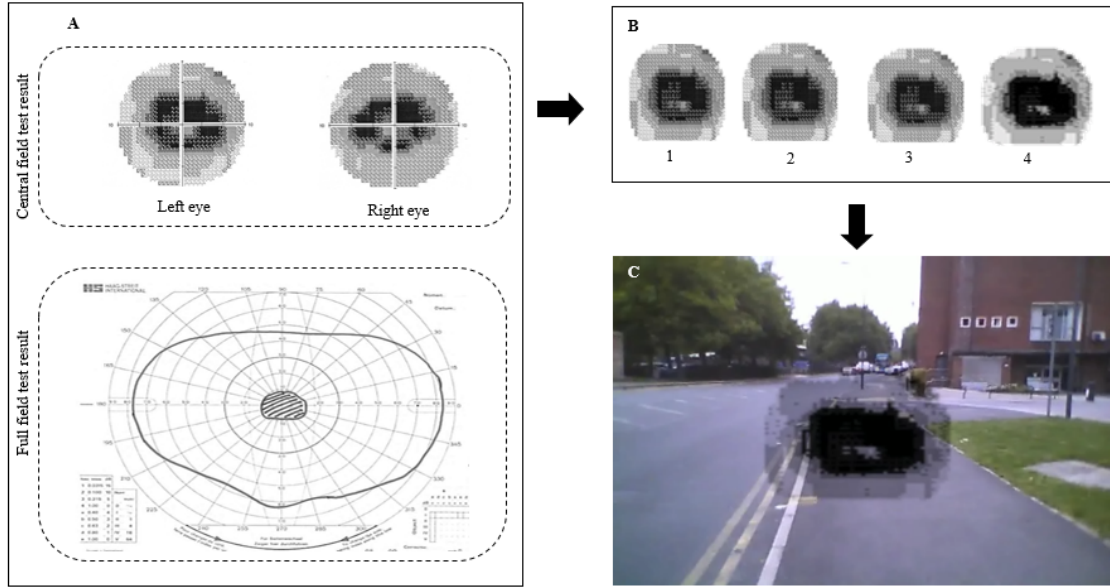


Figure 6.7: Left central scotoma (AMD) vision model example.

- **Task localisation**, such as the work proposed by Henderson et al. [212], Henderson and Feiner [213] and Lee and Akin [214].

Figure 6.8 shows examples from the mentioned studies. The first row depicts visual feedback schemes used for navigation and wayfinding. Different virtual cues to help to identify the navigation pathway in addition to other road objects are shown. The second row presents three examples of visualisation techniques used to alert/warn users about hazards. For this purpose, different styles of visual notifications are displayed on the top of the users' view to point their direction to possible threats. The last row shows examples of AR visualisation methods used in maintenance and repair tasks. Different types of visual schemes were used to add extra information or to highlight a specific area in the user's FoV.

Selecting the best style of this feedback is a challenging task. Users requirements vary depending on the goal of the feedback and according to their visual conditions. Across low vision people, one user's needs may differ to another user's needs. Moreover, it is pervasive for one person's needs to be different from day-to-day or even throughout one day!



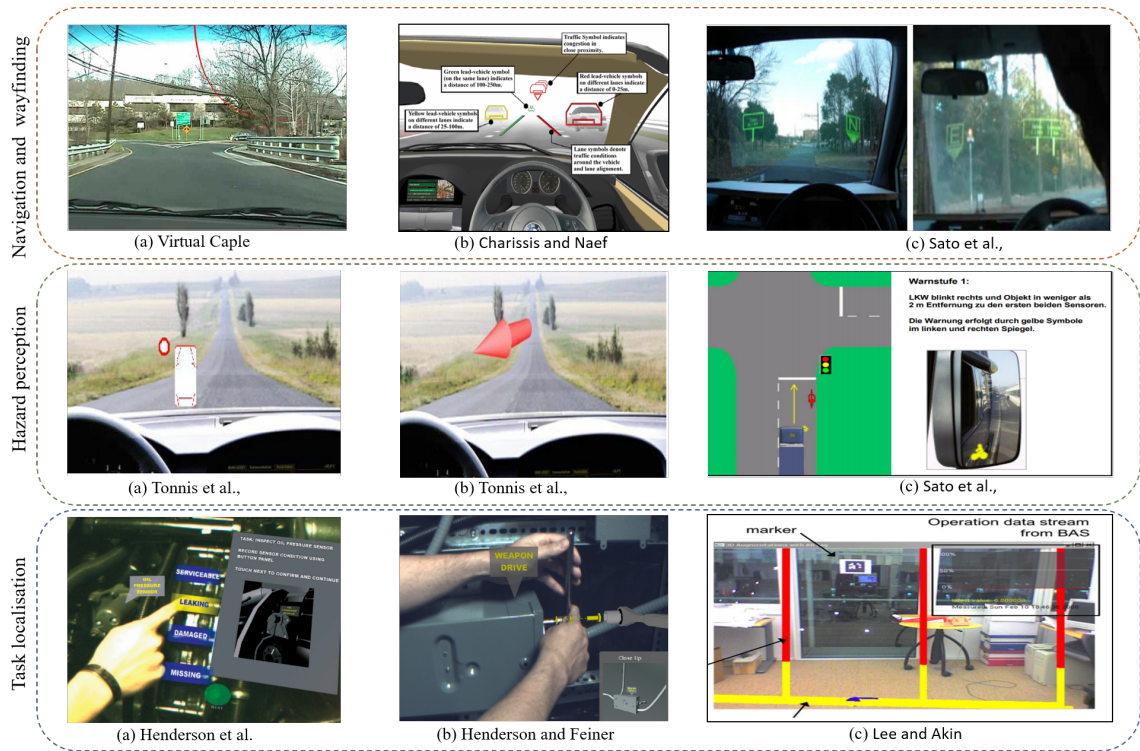


Figure 6.8: Visualisation scheme examples.

Using AR to develop visual feedback for low vision people is a very challenging task. Zhao et al. [150] used commercially available smart glasses to evaluate different visual notification styles. The goal of their work was to test what the low vision people could see and study their response.

The authors conducted an experimental study with 20 low vision participants and 18 sighted controls. They asked them to identify and respond to two different types of visual notifications; basic shapes (triangle, square, circle) and texts.

Figure 6.9 depicts these schemes. Image (a) shows an example of a triangle on black and white backgrounds with different thickness levels. Image (b) shows text examples with varying colours on black and white backgrounds.

According to their findings, basic shapes were more easily identified by low vision people [150]. Using bright colours such as white and yellow is recommended more than other colours such as red and blue. These findings agree with our questionnaire results mentioned in chapter 3 of this thesis.

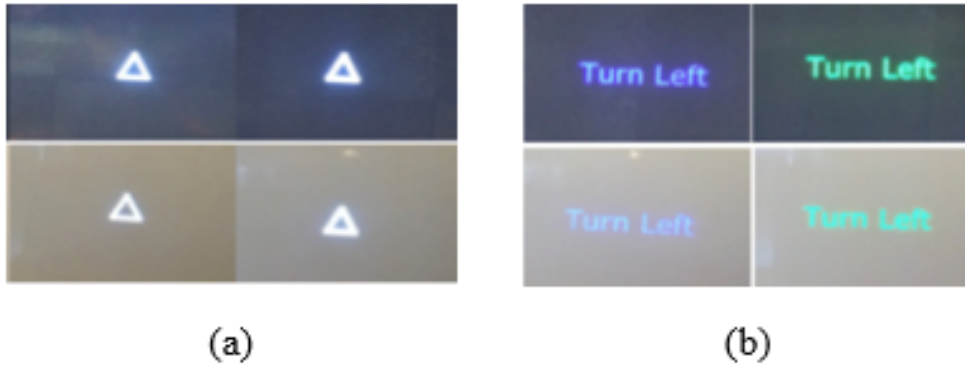


Figure 6.9: Zhao et al. [150] visual feedback schemes.

Consequently, the feedback generation phase was developed to reflect these recommendations and results using an arrow with three features:

1. Width: the wider the arrow, the higher the danger level it represents;
2. Speed: the faster the arrow moves, the higher the danger level it represents;
3. Colour: all notifications have yellow colour, except hazard level five, which has yellow and white colours.

Figure 6.10 shows the hierarchical representation of the different notification representations for each hazard class and Table 6.1 depicts a full description about the visual notifications used in this project.

### 6.3.3 Visual Notification Examples

In this section, three examples (tunnel vision, left hemianopia, and AMD) will be discussed to demonstrate the output of the proposed defected vision model.

For the examples used in this section, an illustration image will be shown at the beginning to describe the starting status. The original and defected images will follow, showing the captured frame and the modelled frame, respectively. Original frame images, including detection (bounding boxes) and tracking (coloured circles)



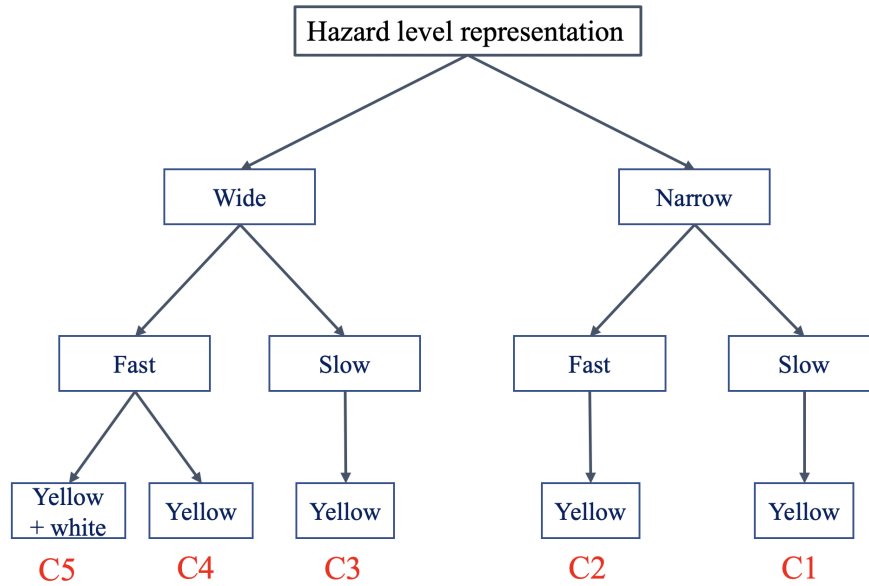







Figure 6.10: Hazard levels representation.

outputs, in addition to the region of interest (RoI) mentioned in Section 5.2, Figure 5.7 (illustrated as a yellow trapezium) are demonstrated.

Figure 6.11 shows an output example using a tunnel vision personalised model. The first image shows an illustration of the examples. In the first row (frame 152) example, the original image shows that the system detected a walking pedestrian at the left side, walking towards the user. According to the hazard representation mentioned in Table 6.1, this is class 4 hazard, and the output will be wide yellow arrows. The arrow's orientation points to the left side to nudge the user to move his/her head to that direction to see the hazard. After 129 frames, the position and type of the detected hazard were changed. The pedestrian is moving away from the user. This means it is now a class 2 hazard presented as narrow yellow arrows pointing to the right side.

Figure 6.12 shows the second example of a tunnel vision model. In this example, the system detected two objects in frame 302; a person on a bicycle in the user's pathway (class 3), and a stationary object (bus) outside the user's path (class 1).

Table 6.1: Visual notification format

Class	Representation	Description
Class 5		White and yellow, wide and fast arrows represent hazard class with the highest danger level (moving, inside/towards RoI, not a person)
Class 4		Yellow, wide and fast arrows represent hazard class with a high danger level (moving, inside/towards RoI, person)
Class 3		Wide slow arrows represent hazard class with a middle danger level (static, inside RoI)
Class 2		Narrow fast arrows represent hazard class with a low danger level (moving, outside RoI)
Class 1		Narrow slow arrows represent hazard class with the lowest danger level (static, outside RoI)

Although class 3 has a higher priority than class 1, the system neglected its output and presented only a narrow yellow arrows pointing to the right side. The reason for this is that this hazard (c3) exists in the seeing area, which means that the user can see the cyclist.

In frame 388, the system was unable to detect the bus to the right side. At the same time, the cyclist moved and his position became outside the user's seeing area. As the cyclist is partially recognised by the user and his location is still so close to the user's path, a class 3 notification was produced (narrow yellow arrows pointing to the left side).

Figure 6.13 shows an output example using a left hemianopia vision personalised model. As shown in illustration 1, the system detected a moving bus to the left side of the yellow trapezium. Using the hazard types representation in Figure 6.10, the system generated a narrow fast arrows pointing to the left side. As the object kept moving towards the user, its hazard type changed from level 2 to 5. Consequently, a wide yellow and white arrows are presented in frame 370, nudging the user to turn to the left side to recognise the detected hazard.

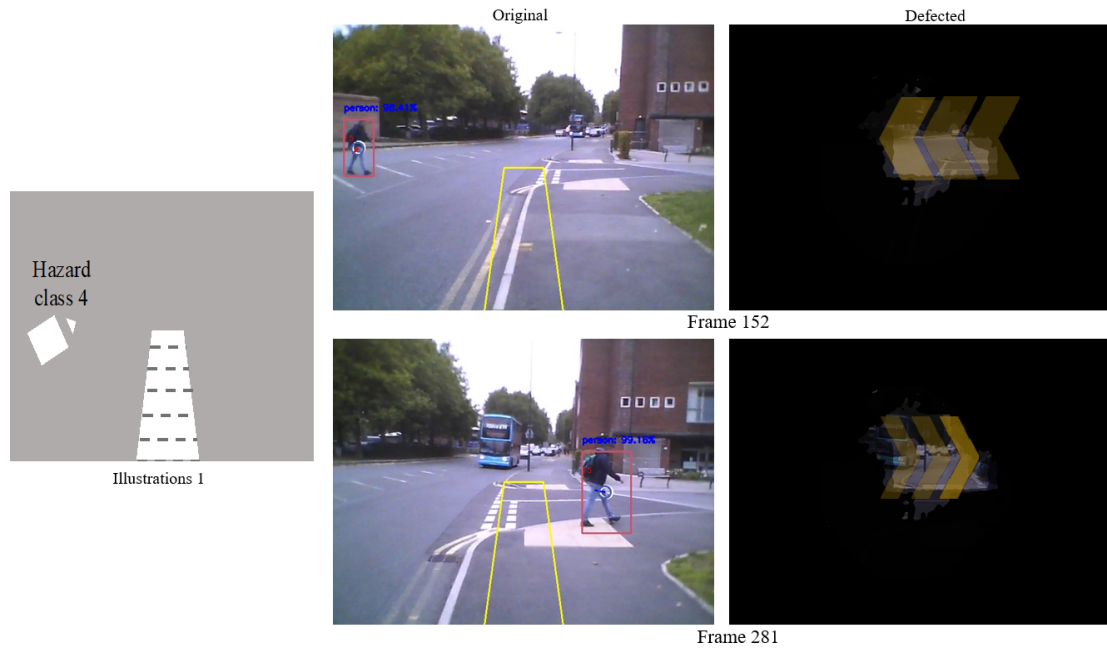


Figure 6.11: A system output example using tunnel vision model (1).

Figure 6.14 shows different example using the same vision model. As illustrated by illustration 2, the system detected two pedestrians to the left side and moving vehicles to the right side (frame 6). Since the right visual field is healthy, the user can see the buses but can not recognise the pedestrians. In this case, a class 4 hazard notification was generated as a wide moving arrows pointing to the left side.

After 40 frames, the pedestrians started moving away from the user. Therefore, the generated notification was downgraded from level 4 to level 2. Accordingly, the notification style changed to be narrow arrows pointing to the left side.

Figure 6.15 shows output example using AMD vision personalised model. This case is different compared to the two previous cases. The modelled vision is mostly healthy, with a central scotoma within the middle of the visual field. Therefore, the interpretation of the produced notification is changed.

In frame 244, the system detected a moving pedestrian in the centre of the user's path, as shown in illustration 1. The pedestrian at this point is not visible. A wide fast arrows were generated to present a level 4 hazard. After 27 frames, the pedestrian continued crossing the blind area, but he is moving away for the user's



Figure 6.12: A system output example using tunnel vision model (2).

path. Therefore, the hazard level changed from class 4 to class 2 (narrow fast arrows).

In Figure 6.16, the illustration shows that a pedestrian was detected moving towards the user (frame 740). The central scotoma prevents the user from seeing that person. Therefore, the system generated a wide fast arrows pointing to the right side of the central scotoma, telling the user to expect a person coming from that side. After a few frames, a black cover was placed in front of the camera, blocking most of the scene. The tracker was able to estimate the position of the detected object even when the detector failed to detect it. A notification showing level 4 hazard is presented in frame 745 at the right side of the central scotoma.

## 6.4 Discussion and Conclusion

The main aim of the work presented in this chapter was the development of a personalised vision model using visual field tests. Three different models were introduced, tunnel vision, left hemianopia and central scotoma (AMD) models.

In the second section, the standard vision model was proposed to synthesise a



Figure 6.13: A system output example using left hemianopia model (1).

human's healthy vision. The central vision was presented using full resolution, while the peripheral vision was rendered using different levels of Gaussian filters.

In the personalised vision model section, both central and full visual field test results were used to produce defected vision videos. Several image processing methods were used to extract the needed information from the test printouts and apply it on video frames in real-time. It is believe that by using this model, eye specialists can apply the visual field tests results on real-time videos to see what visually impaired people see. This will better help to understand visual problems and daily difficulties<sup>1</sup>.

In Section 6.3.2, visual feedback designs used to enhance visual perception in different research fields, especially in AR applications was discussed. This step is crucial in this work for selecting the most useful notification for visually impaired people. It was found by Zhao et al. [150] that basic shapes with bright colours are more appropriate to enhance visual perception for visually impaired people using AR applications. This supports the findings found in the second questionnaire, which

<sup>1</sup>More video examples could be found at my website: <https://pcwww.liv.ac.uk/~younis/>

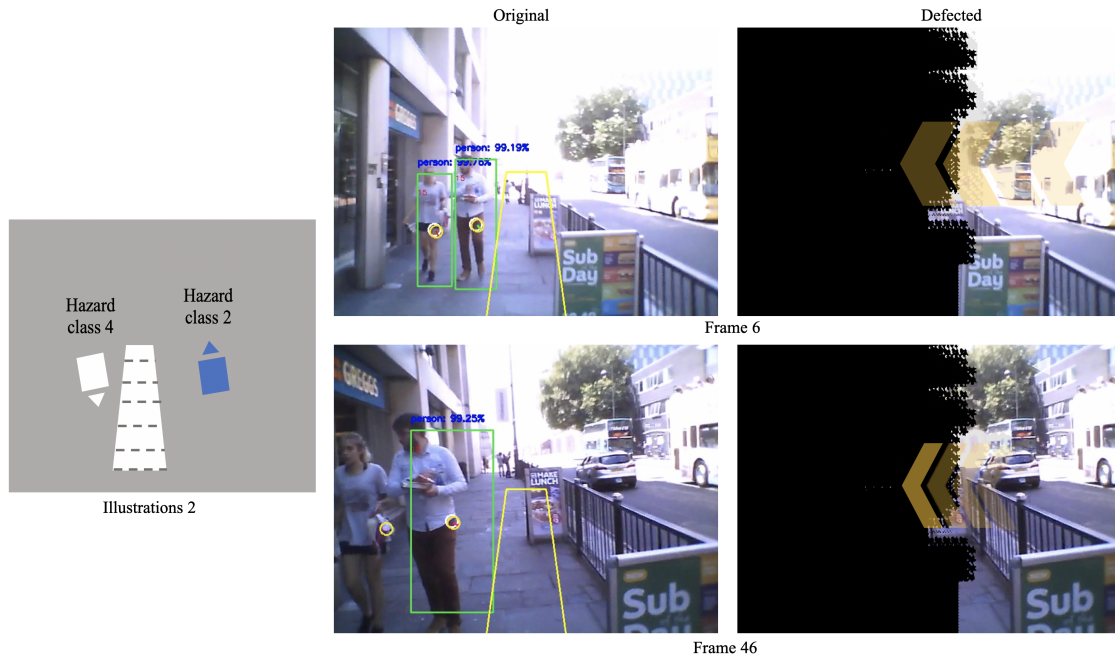


Figure 6.14: A system output example using left hemianopia model (2).

was conducted with our patients group.

According to this information, the visual feedback designs used in our system was presented. For the first time in visual rehabilitation solutions, a system that can detect and classify potential hazards around a visually impaired person, and provide visual cues personalised to the user's vision condition is presented.

Five different representations were used to reflect hazard levels, and an arrow shape was chosen to show the hazard direction. Various examples were given to demonstrate the system output using the three defected vision models.

These designs and colours could be customised according to the user's preferences. It is also possible for users to choose the type of information they want to be informed about. Currently, the system generates one notification at a time, showing a hazard with the highest danger level.



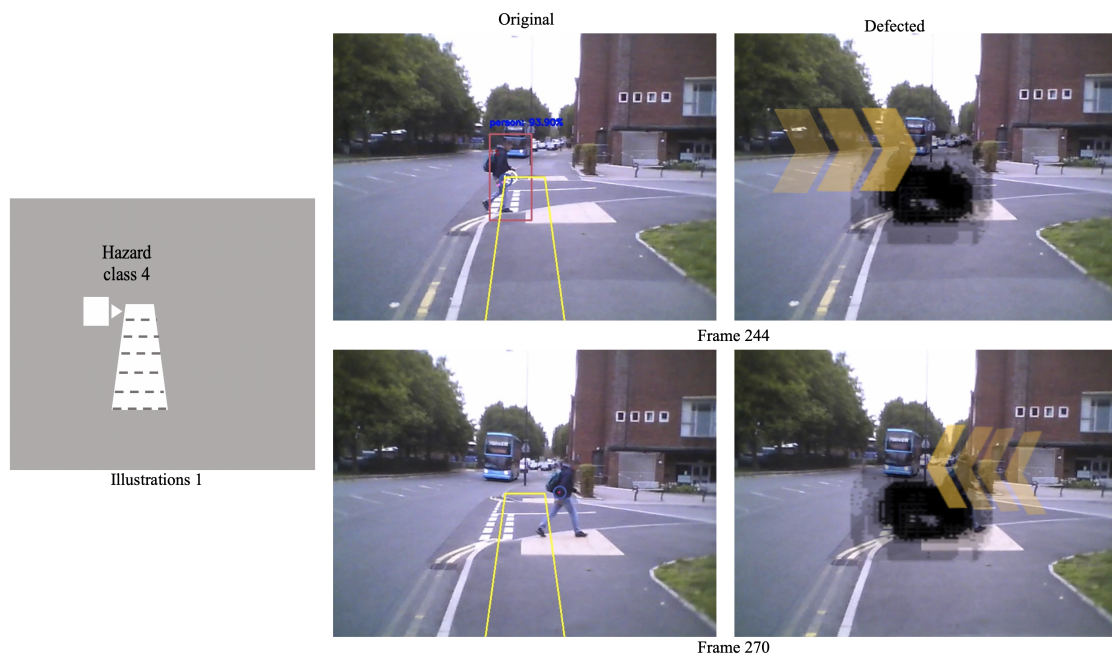


Figure 6.15: A system output example using AMD model (1).

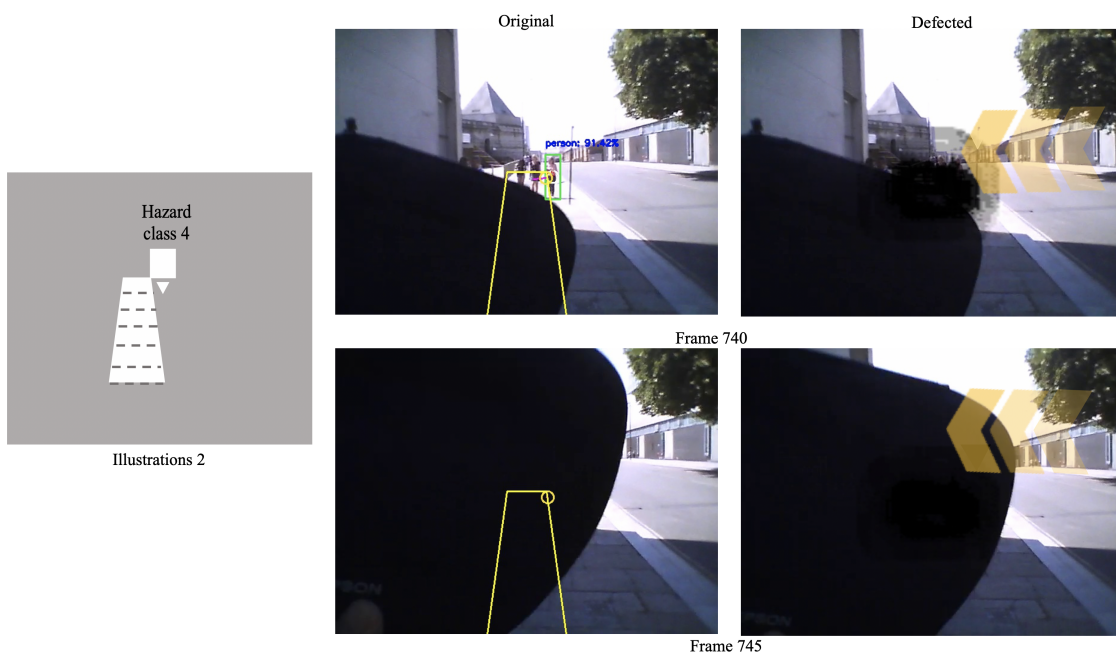


Figure 6.16: A system output example using AMD model (2).

# Chapter 7

## Conclusions and Future Work

### 7.1 Overall Conclusion

The main achievement of the research presented in this thesis can be described as the design and development of smart assistive technology to help visual field loss people, that can enhance hazard perception and provide early personalised visual feedback.

Visually impaired participants were involved from the early stages to help in the system design process, and to give their opinions and feedback throughout the development process. The presented work includes hazard detection and recognition, hazard tracking and hazards classification modules, which assigns a hazard level for each detected object according to its motion features. The same object could change its hazard level over time according to the way it is moving around the user.

The main goal of this system is to increase the user's awareness of the surrounding environment without interfering with the healthy vision. It is believed that this work is important because unlike other obstacle avoidance and navigation systems, the proposed technology is directed to the people who have partially healthy vision. The proposed work uses this vision and augments it with new, meaningful and smart notifications that appear only if necessary.

Also, by using visual field test results, it is believed that the proposed work can be personalised according to the visual case of each user.



This system has been tested on both publicly available and private datasets. The classification method showed promising results, proving that the system could genuinely classify any detected hazard into one of five predefined hazard classes. The proposed solution will enhance the quality of life for people suffering from visual field loss conditions. This non-intrusive, wearable hazard detection technology can provide obstacle avoidance solution, and prevent falls and collisions early with minimal information.

## 7.2 Detailed Conclusions

Concluding remarks on this research are listed as follows:

- People with visual field loss suffer from many daily challenges that decrease their quality of life. Because this group retains healthy vision in some parts of their visual field, it is possible to use this vision and superimpose it with computer-generated information to increase the overall visual perception. One of the main difficulties for visually impaired people is their low hazard perception, which increases their falls and collisions rates. Designing a smart assistive technology using computer vision algorithms in real-time to provide useful information about any possible threats existing in the user's blind area will enhance functional vision.
- Two questionnaires were conducted with visual field loss participants to study their challenges and preferences for smart assistive technology toward helping them to enhance their hazard perception. It was concluded that participants are highly interested in moving objects and consider them to have higher danger levels. It was also found that participants prefer having visual feedback integrated into their healthy vision to help them recognise possible hazards. Five different hazard classes were defined and ranked according to their danger level by the participants, who also helped in designing the best visual feedback format. The mentioned questionnaires are included in the appendices of this thesis.

- A deep learning-based object recognition module was used to detect the objects reported in the participants' questionnaire. Each detected object was tracked while it is in the camera's FoV to extract its motion features. These features are used in the hazard classification stage to assign the detected object into one of five pre-defined classes. The Hungarian algorithm was used to map the tracked objects with the new detections.
- Through our research collaboration with the Department of Health Services Research, we created our dataset for hazard detection and classification. Using Epson's Moverio BT-200 smart glasses, we captured indoor and outdoor videos. An expert labelled the data to one of five hazard classes. These classes were discussed with the visual field loss patient group.

A Neural Network-based classifier was trained and tested on the private and public datasets. It was concluded that it is possible to classify the detected object into one of the five hazards classes using its motion features such as type, current position, estimated position, speed, the direction of motion and age. It was also found that if we add the previous hazard classification results into the current features, it would noticeably enhance the classification performance.

For example, the average accuracy for the five hazard classes when we used temporal information (previous classification results) was 91% compared with 89% using conventional classification. The false Negative Rate average for standard classification technique was 0.132%. This value decreased when the temporal information was used at 0.1%.

- Chapter 6 discussed the model used for personalising defected vision using visual field test results. After developing three different vision models, it was concluded that the visual feedback style, location, colour, speed and interpretation should be customised based on the user's visual condition.

Selecting the best style for this feedback is a challenging task. Users requirements vary depending on the goal of the feedback and according to their visual condition. Across low vision people, one user's needs may differ to other users

needs. Moreover, it is pervasive for one person's needs to be different from day-to-day or even throughout one day! Therefore, it is crucial to give the end-user the ability to control all these settings and train her/him how to use the proposed technology.

- A participant study was conducted with a group of patients with different visual field defects to explore their preferences, suggestions and opinions about the notification style, frequency and other presentation features. After describing the project's idea and design, basic demography information and the visual impairment history of the participants were collected. The participants tried the Moverio BT-200 smart glasses with basic notifications in an indoor environment. Due to the ethical approval constraints, we were unable to perform outdoor experiments for the proposed system. Therefore, we presented the basic system concepts (a single notification for the highest hazard level) to the participants and collected their feedback through a questionnaire.

Regarding the project idea and design, all participants were delighted with the solution provided. They were very excited to try the final product and see the results in real-time while navigating indoor and outdoor. Some of the participants commented on the smart glasses by saying it was light-weight with an accepted design.

## 7.3 Limitations and Strengths of the Proposed Technology

There are limitations to the approaches adopted by our research work. We used a monocular camera to capture the videos used in all of our experiments, but this may falsely reflect the user's binocular vision. To have a realist view, either a binocular or a wide-angle camera should be used to cover the full FoV vision for a healthy person.

The presented evaluation in this work was performed on a MacBook laptop (2.7

GHz Intel Core i5 processor, 8 GB RAM) which was able to process high-resolution CamVid videos (30 FPS) with an average of 0.2160s per frame and an average of 0.1932s for videos captured by the Moverio BT-200 smart glasses.

To test the usability of the proposed work, it should be run on wearable smart glasses. Both sighted, and visually impaired subjects should evaluate a system that capture videos, perform the detection, tracking and classification process and provide visual notifications in real-time.

Although the presented evaluation proved that the proposed technology is capable of classifying hazards truly, an end-to-end system evaluation is needed to confirm the usefulness of the system.

The personalised defected vision model assumes the patient to have a similar visual condition in both eyes. However, each eye may have a different visual field case, resulting in two different visual models for each eye. With the current system, the result is an approximation for the full field test result using only one eye from the central visual field test.

There were also some strengths to our study design. Most of the systems mentioned in the literature focused on technical aspects to solve navigation and obstacle avoidance problems. The systems were designed without prior consultation with the potential users to understand their challenges and preferences. In our research, we adopted a user-centred design approach from the early product development stages to the end discussions after demonstrating a primary system output with visually impaired participants.

This involvement helped to define and rank five different hazard classes. Using machine learning algorithms, and for the first time, we were able to classify the detected objects according to their motion features. The classification helped in distinguishing between significant, important and good-to-know hazards, and presenting customised visual feedback.

One of the strengths of this work is the dataset captured by the smart glasses, which enabled us to train and test the hazard classification method. The dataset could be used by other research work to detect and recognise potential hazards for

video footage captured by a wearable camera.

Each visually impaired person has a different visual field case. By including his/her visual field test results in the proposed assistive technology, we are personalising our solution to each patient's condition.

Furthermore, the feedback style could be customised based on the user's preferences. For example, the user may prefer audio over visual feedback. Other options, such as feedback timing, input/output styles, and frequency could be tailored based on the user's needs. Since this work presents the hazard classification as assistive technology, we can say that it can be integrated with other wearable devices to help visually impaired people.

## 7.4 Suggestions for the Future Work

Some of the challenges that still need to be overcome, with suggested solutions and some ideas for further research are included in the following list:

- As concluded by chapter 4, it is essential to include an object recognition module to detect pre-defined objects. If we give the user the possibility to select the objects he/she would like to have notifications for, this will increase the system's usability and would reduce training time needed for new users.

In chapter 3, we mentioned the participants' discussion feedback. They prefer having different levels of commands between the user and the smart glasses. This started from simple at the beginning, to more complicated options once the user learnt how to adopt the system. These preferences could be implemented through a user-friendly interface as a mobile application that can be used on different portable platforms.

- Hazard classification results mentioned in chapter 5 reported reasonable classification rates for the five hazard classes. However, the features used in this phase depend only on videos captured by a wearable camera. Data fusion using

different input sensors such as gyroscope, accelerator and depth camera could be used in future to enhance feature extraction step.

Furthermore, Visual Odometry (VO) methods can be used to add the camera motion information to the extracted features, which will increase hazard classification performance.

- In this work, three vision models were used to visualise the defected model. In future work, this model will be generalised to include different visual field loss cases using central and peripheral vision test results. The work will be evaluated by patients and eye specialists to ensure its accuracy.
- Soon, it is planned to perform an experimental evaluation for the proposed system. For this purpose, navigation time and collision rates will be tested with and without the use of the presented smart technology. Different types and levels of visual hazard notifications will be used to determine the best format and location to be used with visually impaired people.

Hazard perception tests will be used to evaluate participants' results with the help of the proposed work compared to their normal perception (baseline). These tests can be performed by sighted people using different types of defected vision models, or by partially sighted people with different visual field defect cases. Multi-Luminance Mobility Test (MLMT) will be used for this purpose.

- As this work is a part of a larger project, the main target is to provide a wearable assistive technology to help partially sighted people in their navigation by increasing hazard perception. For this purpose, the proposed work will be implemented on smart glasses that include a powerful processing unit and a see-through display unit. The user would be able to wear the glasses and move freely while it is operating. Real-time notifications will be generated to nudge the users to turn their head to the correct direction in order to be able to see potential hazards.

Figure 7.1 summarises the points mentioned about the future work. The time

frame for the future work is shown in Figure 7.2.

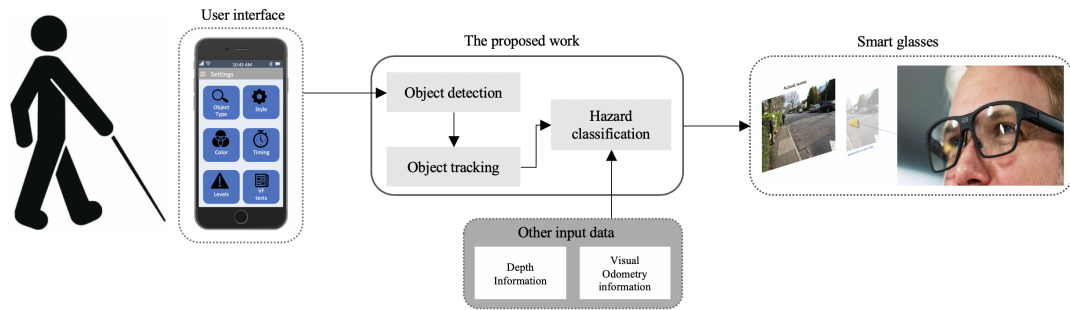


Figure 7.1: Future work integrating the user interface and smart glasses.

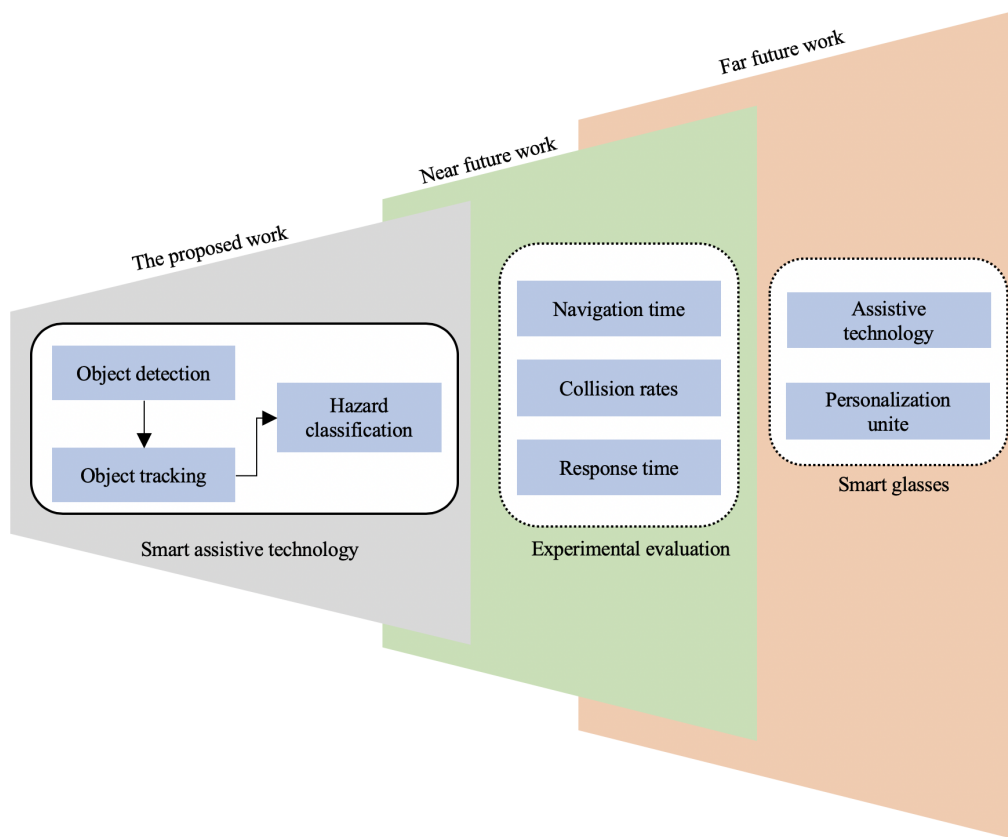


Figure 7.2: Time frame of future work.



# References

- [1] “Blindness and vision impairment,” 2018. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment>
- [2] “Blindness and vision loss,” 2018. [Online]. Available: <https://www.nhs.uk/conditions/vision-loss/>
- [3] R. L. Brown and A. E. Barrett, “Visual impairment and quality of life among older adults: an examination of explanations for the relationship,” *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, vol. 66, no. 3, pp. 364–373, 2011.
- [4] D. Peters, A. Heijl, L. Brenner, and B. Bengtsson, “Visual impairment and vision-related quality of life in the early manifest glaucoma trial after 20 years of follow-up,” *Acta Ophthalmologica*, vol. 93, no. 8, pp. 745–752, 2015.
- [5] Y.-C. Tseng, S. H.-Y. Liu, M.-F. Lou, and G.-S. Huang, “Quality of life in older adults with sensory impairments: a systematic review,” *Quality of Life Research*, vol. 27, no. 8, pp. 1957–1971, 2018.
- [6] “Health statistics and information systems,” 2015. [Online]. Available: <https://www.who.int/healthinfo/survey/whoqol-qualityoflife/en/>
- [7] K. Hirooka, S. Sato, E. Nitta, and A. Tsujikawa, “The relationship between vision-related quality of life and visual function in glaucoma patients,” *Journal of Glaucoma*, vol. 25, no. 6, p. 505, 2016.

- [8] M. Roh, A. Selivanova, H. J. Shin, J. W. Miller, and M. L. Jackson, “Visual acuity and contrast sensitivity are two important factors affecting vision-related quality of life in advanced age-related macular degeneration,” *Plos One*, vol. 13, no. 5, p. e0196481, 2018.
- [9] N. Jones, H. E. Bartlett, and R. Cooke, “An analysis of the impact of visual impairment on activities of daily living and vision-related quality of life in a visually impaired adult population,” *British Journal of Visual Impairment*, vol. 37, no. 1, pp. 50–63, 2019.
- [10] S. John, E. Emma, and W. Andy, “Employment status and sight loss,” 2017. [Online]. Available: <https://www.rnib.org.uk/sites/default/files/Employment%20status%20and%20sight%20loss%202017.pdf>
- [11] G. De Haan, B. Melis-Dankers, W. Brouwer, R. Bredewoud, O. Tucha, and J. Heutink, “Car driving performance in hemianopia: an on-road driving study,” *Investigative Ophthalmology and Visual Science*, vol. 55, no. 10, pp. 6482–6489, 2014.
- [12] M. Boucart, P. Despretz, K. Hladiuk, and T. Desmettre, “Does context or color improve object recognition in patients with low vision?” *Visual Neuroscience*, vol. 25, no. 5-6, pp. 685–691, 2008.
- [13] M. A. Cohen, G. A. Alvarez, and K. Nakayama, “Natural-scene perception requires attention,” *Psychological Science*, vol. 22, no. 9, pp. 1165–1172, 2011.
- [14] G. McGwin Jr, C. Huisinigh, S. G. Jain, C. A. Girkin, and C. Owsley, “Binocular visual field impairment in glaucoma and at-fault motor vehicle collisions,” *Journal of Glaucoma*, vol. 24, no. 2, p. 138, 2015.
- [15] G. I. Kempen, J. Ballemans, A. V. Ranchor, G. H. van Rens, and G. R. Zijlstra, “The impact of low vision on activities of daily living, symptoms of depression, feelings of anxiety and social support in community-living older adults seeking

- vision rehabilitation services,” *Quality of Life Research*, vol. 21, no. 8, pp. 1405–1411, 2012.
- [16] H. P. van der Aa, H. C. Comijs, B. W. Penninx, G. H. van Rens, and R. M. van Nispen, “Major depressive and anxiety disorders in visually impaired older adults,” *Investigative Ophthalmology and Visual Science*, vol. 56, no. 2, pp. 849–854, 2015.
- [17] C. Nollett, B. Ryan, N. Bray, C. Bunce, R. Casten, R. T. Edwards, D. Gillespie, D. J. Smith, M. Stanford, and T. H. Margrain, “Depressive symptoms in people with vision impairment: a cross-sectional study to identify who is most at risk,” *British Medical Journal*, vol. 9, no. 1, p. e026163, 2019.
- [18] A. Black and J. Wood, “Vision and falls,” *Clinical and Experimental Optometry*, vol. 88, no. 4, pp. 212–222, 2005.
- [19] C. Brundle, H. A. Waterman, C. Ballinger, N. Ollevent, D. A. Skelton, P. Stanford, and C. Todd, “The causes of falls: views of older people with visual impairment,” *Health Expectations*, vol. 18, no. 6, pp. 2021–2031, 2015.
- [20] R. Bourne, J. Jonas, S. Flaxman, J. Keeffe, J. Leasher, K. Naidoo, M. Parodi, K. Pesudovs, H. Price, R. White, T. Wong, S. Resnikoff, and H. Taylor, “Prevalence and causes of vision loss in high-income countries and in eastern and central europe: 1990–2010,” *British Journal of Ophthalmology*, vol. 98, no. 5, pp. 629–638, 2014.
- [21] H. Strasburger, I. Rentschler, and M. Jüttner, “Peripheral vision and pattern recognition: A review,” *Journal of Vision*, vol. 11, no. 5, pp. 1–82, 2011.
- [22] V. L. Clark and J. A. Kruse, “Clinical methods: the history, physical, and laboratory examinations,” *Journal of the American Medical Association*, vol. 264, no. 21, pp. 2808–2809, 1990.
- [23] “Normal retinal anatomy,” 2018. [Online]. Available: <http://www.retinareference.com/anatomy/>

- [24] C. A. Curcio, K. R. Sloan, R. E. Kalina, and A. E. Hendrickson, "Human photoreceptor topography," *Journal of Comparative Neurology*, vol. 292, no. 4, pp. 497–523, 1990.
- [25] P. M. van Diepen, M. Wampers, and G. d'Ydewalle, "Functional division of the visual field: Moving masks and moving windows," in *Eye guidance in reading and scene perception*. Elsevier, 1998, pp. 337–355.
- [26] D. Anderson and M. Patella, *Automated static perimetry*, 1st ed. Mosby Year Book, 1992.
- [27] I. Aprile, M. Ferrarin, L. Padua, E. Di Sipio, C. Simbolotti, S. Petroni, C. Tredici, and A. Dickmann, "Walking strategies in subjects with congenital or early onset strabismus," *Frontiers in Human Neuroscience*, vol. 8, p. 484, 2014.
- [28] P. Wang and G. W. Cottrell, "Central and peripheral vision for scene recognition: a neurocomputational modeling exploration," *Journal of Vision*, vol. 17, no. 4, pp. 9–9, 2017.
- [29] M. Carrasco, D. L. Evert, I. Chang, and S. M. Katz, "The eccentricity effect: Target eccentricity affects performance on conjunction searches," *Attention, Perception and Psychophysics*, vol. 57, no. 8, pp. 1241–1261, 1995.
- [30] R. Dougherty, V. Koch, A. Brewer, B. Fischer, J. Modersitzki, and B. Wandell, "Visual field representations and locations of visual areas v1/2/3 in human visual cortex," *Journal of Vision*, vol. 3, no. 10, pp. 586–598, 2003.
- [31] A. Denniston and P. Murray, *Oxford handbook of ophthalmology*, 1st ed. Oxford University Press, 2014.
- [32] A. R. Bowers, A. J. Mandel, R. B. Goldstein, and E. Peli, "Driving with hemianopia, i: Detection performance in a driving simulator," *Investigative Ophthalmology and Visual Science*, vol. 50, no. 11, pp. 5137–5147, 2009.

- [33] A. Bowers, A. Mandel, R. Goldstein, and E. Peli, “Driving with hemianopia, ii: lane position and steering in a driving simulator,” *Investigative Ophthalmology and Visual Science*, vol. 51, no. 12, pp. 6605–6613, 2010.
- [34] C. F. Alberti, E. Peli, and A. R. Bowers, “Driving with hemianopia: Iii. detection of stationary and approaching pedestrians in a simulator,” *Investigative ophthalmology and Visual Science*, vol. 55, no. 1, pp. 368–374, 2014.
- [35] A. R. Bowers, E. Ananyev, A. J. Mandel, R. B. Goldstein, and E. Peli, “Driving with hemianopia: Iv. head scanning and detection at intersections in a simulator,” *Investigative ophthalmology and Visual Science*, vol. 55, no. 3, pp. 1540–1548, 2014.
- [36] M. Thibaut, T. H. C. Tran, S. Szaffarczyk, and M. Boucart, “The contribution of central and peripheral vision in scene categorization: a study on people with central vision loss,” *Vision Research*, vol. 98, pp. 46–53, 2014.
- [37] M. Hersh and M. Johnson, Eds., *Assistive Technology for Visually Impaired and Blind People*, 1st ed. Springer-Verlag London, 2008.
- [38] M. Ervasti, M. Isomursu, and I. I. Leibar, “Touch-and audio-based medication management service concept for vision impaired older people,” in *proceedings of the IEEE international conference on RFID-technologies and applications (RFID-TA)*, 2011, pp. 244–251.
- [39] J. R. Ehrlich, L. V. Ojeda, D. Wicker, S. Day, A. Howson, V. Lakshminarayanan, and S. E. Moroi, “Head-mounted display technology for low-vision rehabilitation and vision enhancement,” *American Journal of Ophthalmology*, vol. 176, pp. 26–32, 2017.
- [40] C. Owsley, G. McGwin, P. P. Lee, N. Wasserman, and K. Searcey, “Characteristics of low-vision rehabilitation services in the united states,” *Archives of Ophthalmology*, vol. 127, no. 5, pp. 681–689, 2009.

- [41] K. L. Johnson, B. Dudgeon, and D. Amtmann, "Assistive technology in rehabilitation," *Physical Medicine and Rehabilitation Clinics*, vol. 8, no. 2, pp. 389–403, 1997.
- [42] M. J. Scherer and J. P. Lane, "Assessing consumer profiles of 'ideal' assistive technologies in ten categories: an integration of quantitative and qualitative methods," *Disability and Rehabilitation*, vol. 19, no. 12, pp. 528–535, 1997.
- [43] L. Hakobyan, J. Lumsden, D. O'Sullivan, and H. Bartlett, "Mobile assistive technologies for the visually impaired," *Survey of Ophthalmology*, vol. 58, no. 6, pp. 513–528, 2013.
- [44] T. Paek and D. M. Chickering, "Improving command and control speech recognition on mobile devices: using predictive user models for language modeling," *User Modeling and User-adapted Interaction*, vol. 17, no. 1-2, pp. 93–117, 2007.
- [45] E. Balajthy, "Text-to-speech software for helping struggling readers," *Reading Online*, vol. 8, no. 4, pp. 1–9, 2005.
- [46] S. Brewster, S. Brewster, F. Chohan, and L. Brown, "Tactile feedback for mobile interactions," in *proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2007, pp. 159–162.
- [47] S. Oviatt, "Multimodal system processing in mobile environments," in *proceedings of the 13th annual ACM symposium on user interface software and technology*. Citeseer, 2000, pp. 21–30.
- [48] A. A. Karpov and R. M. Yusupov, "Multimodal interfaces of human–computer interaction," *Herald of the Russian Academy of Sciences*, vol. 88, no. 1, pp. 67–74, Jan 2018.
- [49] L. Porzi, S. Messelodi, C. M. Modena, and E. Ricci, "A smart watch-based gesture recognition system for assisting people with visual impairments," in *proceedings of the 3rd ACM international workshop on Interactive multimedia on mobile & portable devices*, 2013, pp. 19–24.

- [50] C. Howard and F. J. Rowe, “Adaptation to poststroke visual field loss: A systematic review,” *Brain and Behaviour*, vol. 8, no. 8, p. e01041, 2018.
- [51] R. Jose and A. Smith, “Increasing peripheral field awareness with fresnel prisms,” *Optical Journal and Review of Optometry*, vol. 113, no. 12, pp. 33–37, 1976.
- [52] E. B. Mehr and R. D. Quillman, “Field expansion by use of binocular full-field reversed 1.3 x telescopic spectacles: a case report,” *American Journal of Optometry and Physiological Optics*, vol. 56, no. 7, pp. 446–450, 1979.
- [53] J. P. Szlyk, W. Seiple, D. J. Laderman, R. Kelsch, K. Ho, and T. McMahon, “Use of bioptic amorphic lenses to expand the visual field in patients with peripheral loss,” *Optometry and Vision Science: Official Publication of the American Academy of Optometry*, vol. 75, no. 7, pp. 518–524, 1998.
- [54] H. Apfelbaum and E. Peli, “Tunnel vision prismatic field expansion: challenges and requirements,” *Translational Vision Science and Technology*, vol. 4, no. 6, pp. 8–8, 2015.
- [55] D. Gottlieb, P. Freeman, and M. Williams, “Clinical research and statistical analysis of a visual field awareness system,” *Journal of the American Optometric Association*, vol. 63, no. 8, pp. 581–588, 1992.
- [56] A. R. Bowers, K. Keeney, and E. Peli, “Randomized crossover clinical trial of real and sham peripheral prism glasses for hemianopia,” *JAMA Ophthalmology*, vol. 132, no. 2, pp. 214–222, 2014.
- [57] J.-H. Jung and E. Peli, “Field expansion for acquired monocular vision using a multiplexing prism,” *Optometry and Vision Science*, vol. 95, no. 9, pp. 814–828, 2018.
- [58] F. Vargas-Martín and E. Peli, “Augmented view for tunnel vision: Device testing by patients in real environments,” *SID Symposium Digest of Technical Papers*, vol. 32, no. 1, pp. 602–605, 2001.

- [59] P. Elango and K. Murugesan, "Cnn based augmented reality using numerical approximation techniques," *International Journal of Signal and Image Processing*, vol. 1, no. 3, pp. 205–210, 2010.
- [60] S. Pundlik, M. Tomasi, and G. Luo, "Evaluation of a portable collision warning device for patients with peripheral vision loss in an obstacle course," *Investigative Ophthalmology and Visual Science*, vol. 56, no. 4, pp. 2571–2579, 2015.
- [61] S. Pundlik, E. Peli, and G. Luo, "Time to collision and collision risk estimation from local scale and motion," in *proceedings of the international symposium on visual computing*. Springer, 2011, pp. 728–737.
- [62] S. Kammoun, G. Parseihian, O. Gutierrez, A. Brilhault, A. Serpa, M. Raynal, B. Oriola, M.-M. Macé, M. Auvray, M. Denis *et al.*, "Navigation and space perception assistance for the visually impaired: The navig project," *Innovation and Research in BioMedical engineering*, vol. 33, no. 2, pp. 182–189, 2012.
- [63] S. Bhatlawande, M. Mahadevappa, J. Mukherjee, M. Biswas, D. Das, and S. Gupta, "Design, development, and clinical evaluation of the electronic mobility cane for vision rehabilitation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 6, pp. 1148–1159, 2014.
- [64] J. Sohl-Dickstein, S. Teng, B. M. Gaub, C. C. Rodgers, C. Li, M. R. DeWeese, and N. S. Harper, "A device for human ultrasonic echolocation," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 6, pp. 1526–1534, 2015.
- [65] G. Balakrishnan, G. Sainarayanan, R. Nagarajan, and S. Yaacob, "Wearable real-time stereo vision for the visually impaired," *Engineering Letters*, vol. 14, no. 2, pp. 1–9, 2007.
- [66] A. Fiannaca, I. Apostolopoulous, and E. Folmer, "Headlock: a wearable navigation aid that helps blind cane users traverse large open spaces," in *proceedings of the 16th international ACM SIGACCESS conference on computers & accessibility*. ACM, 2014, pp. 19–26.



- [67] C. Tsirmpas, A. Rompas, O. Fokou, and D. Koutsouris, “An indoor navigation system for visually impaired and elderly people based on radio frequency identification (rfid),” *Information Sciences*, vol. 320, pp. 288–305, 2015.
- [68] J. Bai, S. Lian, Z. Liu, K. Wang, and D. Liu, “Virtual-blind-road following-based wearable navigation device for blind people,” *IEEE Transactions on Consumer Electronics*, vol. 64, no. 1, pp. 136–143, 2018.
- [69] B. Li, X. Zhang, J. P. Muñoz, J. Xiao, X. Rong, and Y. Tian, “Assisting blind people to avoid obstacles: an wearable obstacle stereo feedback system based on 3d detection,” in *proceedings of the IEEE international conference on robotics and biomimetics*. IEEE, 2015, pp. 2307–2311.
- [70] M.-C. Kang, S.-H. Chae, J.-Y. Sun, J.-W. Yoo, and S.-J. Ko, “A novel obstacle detection method based on deformable grid for the visually impaired,” *IEEE Transactions on Consumer Electronics*, vol. 61, no. 3, pp. 376–383, 2015.
- [71] M.-C. Kang, S.-H. Chae, J.-Y. Sun, S.-H. Lee, and S.-J. Ko, “An enhanced obstacle avoidance method for the visually impaired using deformable grid,” *IEEE Transactions on Consumer Electronics*, vol. 63, no. 2, pp. 169–177, 2017.
- [72] S.-H. Chae, M.-C. Kang, J.-Y. Sun, B.-S. Kim, and S.-J. Ko, “Collision detection method using image segmentation for the visually impaired,” *IEEE Transactions on Consumer Electronics*, vol. 63, no. 4, pp. 392–400, 2017.
- [73] L. Everding, L. Walger, V. S. Ghaderi, and J. Conradt, “A mobility device for the blind with improved vertical resolution using dynamic vision sensors,” in *proceedings of the IEEE 18th international conference on e-Health networking, applications and services (Healthcom)*, 2016, pp. 1–5.
- [74] R. Tapu, B. Mocanu, and T. Zaharia, “Deep-see: Joint object detection, tracking and recognition with application to visually impaired navigational assistance,” *Sensors*, vol. 17, no. 11, p. 2473, 2017.

- [75] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2016, pp. 779–788.
- [76] K. Yang, K. Wang, S. Lin, J. Bai, L. M. Bergasa, and R. Arroyo, “Long-range traversability awareness and low-lying obstacle negotiation with realsense for the visually impaired,” in *proceedings of the international conference on information science and system*. ACM, 2018, pp. 137–141.
- [77] F. Prattico, C. Cera, and F. Petroni, “A new hybrid infrared-ultrasonic electronic travel aids for blind people,” *Sensors and Actuators A: Physical*, vol. 201, pp. 363–370, 2013.
- [78] V. Kaushalya, K. Premarathne, H. Shadir, P. Krithika, and S. Fernando, “Akshi”: Automated help aid for visually impaired people using obstacle detection and gps technology,” *International Journal of Scientific and Research Publications (IJSRP)*, vol. 6, no. 11, 2016.
- [79] J. Bai, S. Lian, Z. Liu, K. Wang, and D. Liu, “Smart guiding glasses for visually impaired people in indoor environment,” *IEEE Transactions on Consumer Electronics*, vol. 63, no. 3, pp. 258–266, 2017.
- [80] S. T. H. Rizvi, M. J. Asif, and H. Ashfaq, “Visual impairment aid using haptic and sound feedback,” in *proceeding of the international conference on communication, computing and digital systems (C-CODE)*. IEEE, 2017, pp. 175–178.
- [81] K. Yang, K. Wang, L. Bergasa, E. Romera, W. Hu, D. Sun, J. Sun, R. Cheng, T. Chen, and E. López, “Unifying terrain awareness for the visually impaired through real-time semantic segmentation,” *Sensors*, vol. 18, no. 5, p. 1506, 2018.
- [82] M. L. Mekhalfi, F. Melgani, A. Zeggada, F. G. De Natale, M. A.-M. Salem, and A. Khamis, “Recovering the sight to blind people in indoor environments with

- smart technologies,” *Expert Systems with Applications*, vol. 46, pp. 129–138, 2016.
- [83] B. Li, P. Munoz, X. Rong, J. Xiao, Y. Tian, and A. Arditi, “Isana: wearable context-aware indoor assistive navigation with obstacle avoidance for the blind,” in *proceedings of the European conference on computer vision (ECCV)*. Springer, 2016, pp. 448–462.
- [84] J. Sosa-García and F. Odone, ““hands on” visual recognition for visually impaired users,” *ACM Transactions on Accessible Computing (TACCESS)*, vol. 10, no. 3, p. 8, 2017.
- [85] S. Bharambe, R. Thakker, H. Patil, and K. Bhurchandi, “Substitute eyes for blind with navigator using android,” in *proceedings of the Texas instruments India educators’ conference*. IEEE, 2013, pp. 38–43.
- [86] A. Aladren, G. López-Nicolás, L. Puig, and J. J. Guerrero, “Navigation assistance for the visually impaired using rgb-d sensor with range expansion,” *IEEE Systems*, vol. 10, no. 3, pp. 922–932, 2014.
- [87] R. K. Katzschnmann, B. Araki, and D. Rus, “Safe local navigation for visually impaired users with a time-of-flight and haptic feedback device,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 3, pp. 583–593, 2018.
- [88] A. Mancini, E. Frontoni, and P. Zingaretti, “Mechatronic system to help visually impaired users during walking and running,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 2, pp. 649–660, 2018.
- [89] S. Meers and K. Ward, “A substitute vision system for providing 3d perception and gps navigation via electro-tactile stimulation,” in *proceedings of the international conference on sensing technology*, 2005, p. 551–556.
- [90] T. Amemiya, “Haptic direction indicator for visually impaired people based on

- pseudo-attraction force,” *International Journal on Human-Computer Interaction*, vol. 1, no. 5, pp. 23–34, 2009.
- [91] T. Amemiya and H. Sugiyama, “Haptic handheld wayfinder with pseudo-attraction force for pedestrians with visual impairments,” in *proceedings of the 11th international SIGACCESS conference on computers and accessibility*. ACM, 2009, pp. 107–114.
- [92] T. Amemiya and H. Sugiyama, “Orienting kinesthetically: A haptic handheld wayfinder for people with visual impairments,” *ACM Transactions on Accessible Computing (TACCESS)*, vol. 3, no. 2, p. 6, 2010.
- [93] S. Sharma, M. Gupta, A. Kumar, M. Tripathi, and M. S. Gaur, “Multiple distance sensors based smart stick for visually impaired people,” in *proceedings of the IEEE 7th annual computing and communication workshop and conference (CCWC)*. IEEE, 2017, pp. 1–5.
- [94] S. Cardin, D. Thalmann, and F. Vexo, “A wearable system for mobility improvement of visually impaired people,” *The Visual Computer*, vol. 23, no. 2, pp. 109–118, 2007.
- [95] V.-N. Hoang, T.-H. Nguyen, T.-L. Le, T.-H. Tran, T.-P. Vuong, and N. Vuillerme, “Obstacle detection and warning system for visually impaired people based on electrode matrix and mobile kinect,” *Vietnam Journal of Computer Science*, vol. 4, no. 2, pp. 71–83, 2017.
- [96] J. Zelek, R. Audette, J. Balthazaar, and C. Dunk, “A stereo-vision system for the visually impaired,” in *University of Guelph*. Citeseer, 1999.
- [97] I. Ulrich and J. Borenstein, “The guidecane-applying mobile robot technologies to assist the visually impaired,” *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, vol. 31, no. 2, pp. 131–136, 2001.
- [98] K. Ito, M. Okamoto, J. Akita, T. Ono, I. Gyobu, T. Takagi, T. Hoshi, and Y. Mishima, “Cyarm: an alternative aid device for blind persons,” in *pro-*

- ceedings of CHI'05 extended abstracts on human factors in computing systems.* ACM, 2005, pp. 1483–1488.
- [99] O. Lahav and D. Mioduser, “Haptic-feedback support for cognitive mapping of unknown spaces by people who are blind,” *International Journal of Human-Computer Studies*, vol. 66, no. 1, pp. 23–35, 2008.
- [100] K. Yatani, N. Banovic, and K. Truong, “Spacesense: representing geographical information to visually impaired people using spatial tactile feedback,” in *proceedings of the SIGCHI conference on human factors in computing systems.* ACM, 2012, pp. 415–424.
- [101] D. Croce, P. Gallo, D. Garlisi, L. Giarрэ, S. Mangione, and I. Tinnirello, “Ar-ianna: A smartphone-based navigation system with human in the loop,” in *proceedings of the 22nd mediterranean conference of control and automation (MED).* IEEE, 2014, pp. 8–13.
- [102] D. Croce, L. Giarрэ, F. G. L. Rosa, E. Montana, and I. Tinnirello, “Enhancing tracking performance in a smartphone-based navigation system for visually impaired people,” in *proceedings of the 24th mediterranean conference on control and automation (MED)*, June 2016, pp. 1355–1360.
- [103] B. Li, J. P. Munoz, X. Rong, Q. Chen, J. Xiao, Y. Tian, A. Ardit, and M. Yousuf, “Vision-based mobile indoor assistive navigation aid for blind people,” *IEEE Transactions on Mobile Computing*, vol. 18, no. 3, pp. 702–714, 2019.
- [104] G. Abowd, A. Dey, P. Brown, N. Davies, M. Smith, and P. Steggles, “Towards a better understanding of context and context-awareness,” in *proceedings of the International symposium on handheld and ubiquitous computing.* Springer, 1999, pp. 304–307.
- [105] M. Markowitz, “Occupational therapy interventions in low vision rehabilitation,” *Canadian Journal of Ophthalmology*, vol. 41, no. 3, pp. 340–347, 2006.

- [106] Y.-H. Ong, M. M. Brown, P. Robinson, G. T. Plant, M. Husain, and A. P. Leff, “Read-right: a “web app” that improves reading speeds in patients with hemianopia,” *Journal of Neurology*, vol. 259, no. 12, pp. 2611–2615, 2012.
- [107] Y.-H. Ong, S. Jacquin-Courtois, N. Gorgoraptis, P. M. Bays, M. Husain, and A. P. Leff, “Eye-search: A web-based therapy that improves visual search in hemianopia,” *Annals of Clinical and Transnational Neurology*, vol. 2, no. 1, pp. 74–78, 2015.
- [108] UCL Institute of Neurology, UCL Multimedia, “Eye-search therapy,” 2012. [Online]. Available: <http://www.eyesearch.ucl.ac.uk>
- [109] UCL Institute of Neurology, “Read-right therapy,” 2012. [Online]. Available: <http://www.readright.ucl.ac.uk>
- [110] N. Bolognini, F. Rasi, M. Coccia, and E. Ladavas, “Visual search improvement in hemianopic patients after audio-visual stimulation,” *Brain*, vol. 128, no. 12, pp. 2830–2842, 2005.
- [111] I. Keller and G. Lefin-Rank, “Improvement of visual search after audiovisual exploration training in hemianopic patients,” *Neurorehabilitation and Neural Repair*, vol. 24, no. 7, pp. 666–673, 2010.
- [112] S. K. Mannan, A. L. Pambakian, and C. Kennard, “Compensatory strategies following visual search training in patients with homonymous hemianopia: an eye movement study,” *Journal of Neurology*, vol. 257, no. 11, pp. 1812–1821, 2010.
- [113] D. Lévy-Bencheton, D. Péliisson, M. Prost, S. Jacquin-Courtois, R. Salemmé, L. Pisella, and C. Tilikete, “The effects of short-lasting anti-saccade training in homonymous hemianopia with and without saccadic adaptation,” *Frontiers in Behavioural Neuroscience*, vol. 9, pp. 332–346, 2016.

- [114] T. Roth, A. Sokolov, A. Messias, P. Roth, M. Weller, and S. Trauzettel-Klosinski, “Comparing explorative saccade and flicker training in hemianopia: a randomized controlled study,” *Neurology*, vol. 72, no. 4, pp. 324–331, 2009.
- [115] A. R. Lane, D. T. Smith, A. Ellison, and T. Schenk, “Visual exploration training is no better than attention training for treating hemianopia,” *Brain*, vol. 133, no. 6, pp. 1717–1728, 2010.
- [116] L. Aimola, A. R. Lane, D. T. Smith, G. Kerkhoff, G. A. Ford, and T. Schenk, “Efficacy and feasibility of home-based training for individuals with homonymous visual field defects,” *Neurorehabilitation and Neural Repair*, vol. 28, no. 3, pp. 207–218, 2014.
- [117] C. Hazelton, A. Pollock, G. Walsh, and M. Brady, “Rehabilitation for visual field loss after stroke: A mixed methods exploration of the effect and feasibility of home-based scanning training,” vol. 10, pp. 165–165, 2015.
- [118] H. M. Livengood and N. A. Baker, “The role of occupational therapy in vision rehabilitation of individuals with glaucoma,” *Disability and Rehabilitation*, vol. 37, no. 13, pp. 1202–1208, 2015.
- [119] W. Lucas Molitor and R. Mayou, “The low vision team: Optometrists’ and ophthalmologists’ perceptions and knowledge of occupational therapy,” *Physical and Occupational Therapy In Geriatrics*, vol. 36, no. 1, pp. 54–71, 2018.
- [120] J. Meyers, *Understanding and Managing Vision Deficits: A Guide for Occupational Therapists*, 2nd ed., 2002.
- [121] G. Kerkhoff, “Neurovisual rehabilitation: recent developments and future directions,” *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 68, no. 6, pp. 691–706, 2000.
- [122] A. Pollock, C. Hazelton, F. J. Rowe, S. Jonuscheit, A. Kernohan, J. Angilley, C. A. Henderson, P. Langhorne, and P. Campbell, “Interventions for visual

- field defects in people with stroke,” *Cochrane Database of Systematic Reviews*, no. 5, 2019.
- [123] A. Sahraie, C. T. Trevethan, M. J. MacLeod, A. D. Murray, J. A. Olson, and L. Weiskrantz, “Increased sensitivity after repeated stimulation of residual spatial channels in blindsight,” *National Academy of Sciences*, vol. 103, no. 40, pp. 14 971–14 976, 2006.
- [124] J. G. Romano, P. Schulz, S. Kenkel, and D. P. Todd, “Visual field changes after a rehabilitation intervention: Vision restoration therapy,” *Journal of the Neurological Sciences*, vol. 273, no. 1-2, pp. 70–74, 2008.
- [125] A. Pollock, C. Hazelton, C. A. Henderson, J. Angilley, B. Dhillon, P. Langhorne, K. Livingstone, F. A. Munro, H. Orr, F. J. Rowe *et al.*, “Interventions for visual field defects in patients with stroke,” *Cochrane Database of Systematic Reviews*, no. 10, 2011.
- [126] “Vision restoration therapy: addressing vision loss after stroke and traumatic brain injury (tbi),” 2019. [Online]. Available: <https://novavision.com>
- [127] F. Schmielau and E. K. Wong, “Recovery of visual fields in brain-lesioned patients by reaction perimetry treatment,” *Journal of Neuroengineering and Rehabilitation*, vol. 4, no. 1, p. 31, 2007.
- [128] R. S. Marshall, M. Chmayssani, K. A. O’Brien, C. Handy, and V. C. Greenstein, “Visual field expansion after visual restoration therapy,” *Clinical Rehabilitation*, vol. 24, no. 11, pp. 1027–1035, 2010.
- [129] C. Gall and B. A. Sabel, “Reading performance after vision rehabilitation of subjects with homonymous visual field defects,” *Physical Medicine and Rehabilitation*, vol. 4, no. 12, pp. 928–935, 2012.
- [130] J. Reinhard, A. Schreiber, U. Schiefer, E. Kasten, B. Sabel, S. Kenkel, R. Vonthein, and S. Trauzettel-Klosinski, “Does visual restitution training change



- absolute homonymous visual field defects? a fundus controlled study,” *British Journal of Ophthalmology*, vol. 89, no. 1, pp. 30–35, 2005.
- [131] P. Jennifer, S. Helen, and R. Yvonne, *Interaction Design: Beyond Human-Computer Interaction*. Wiley, New York, May 2015.
- [132] R. Jafri and M. M. Khan, “User-centered design of a depth data based obstacle detection and avoidance system for the visually impaired,” *Human-centric Computing and Information Sciences*, vol. 8, no. 1, p. 14, May 2018.
- [133] H. Paredes, H. Fernandes, P. Martins, and J. Barroso, “Gathering the users’ needs in the development of assistive technology: A blind navigation system use case,” in *proceedings of the 7th international conference on universal access in human-computer interaction applications and services for quality of life*. Springer, 2013, pp. 79–88.
- [134] A. Benabidvw and M. AlZuhair, “User involvement in the development of indoor navigation system for the visually impaired: A needs-finding study,” in *proceedings of the 3rd international conference on user science and engineering*, 2014, pp. 97–102.
- [135] M. Leo, G. Medioni, M. Trivedi, T. Kanade, and G. M. Farinella, “Computer vision for assistive technologies,” *Computer Vision and Image Understanding*, vol. 154, pp. 1–15, 2017.
- [136] W. Barfield, *Fundamentals of wearable computers and augmented reality*. CRC Press, 2015.
- [137] S. Ong and A. Y. C. Nee, *Virtual and augmented reality applications in manufacturing*. Springer Science & Business Media, 2013.
- [138] A. Al-Ataby, O. Younis, W. Al-Nuaimy, M. Al-Tae, Z. Sharaf, and B. Al-Bander, “Visual augmentation glasses for people with impaired vision,” in *proceedings of the 9th international conference on the developments in esystems engineering (DeSE)*. IEEE, 2016, pp. 24–28.

- [139] Y. Ola, A.-N. Waleed, A.-T. Majid, and A.-A. Ali, “Augmented and virtual reality approaches to help with peripheral vision loss,” in *proceedings of the 14th international multi-conference on systems, signals devices (SSD)*, March 2017, pp. 303–307.
- [140] S. Mitrasinovic, E. Camacho, N. Trivedi, J. Logan, C. Campbell, R. Zilinyi, B. Lieber, E. Bruce, B. Taylor, D. Martineau *et al.*, “Clinical and surgical applications of smart glasses,” *Technology and Health Care*, vol. 23, no. 4, pp. 381–401, 2015.
- [141] “My eye 2.0, Orcam,” 2018. [Online]. Available: <https://www.orcam.com/en/>
- [142] “Microsoft hololens, Microsoft Corporation,” 2018. [Online]. Available: <https://www.microsoft.com/en-gb/hololens>
- [143] “Daqri smart glass,” 2018. [Online]. Available: <https://daqri.com/products/smart-glasses/>
- [144] “Eyetrek insight, Olympus Corporation,” 2018. [Online]. Available: <http://www.getolympus.com/smartglasses?ref=CJs>
- [145] “Glass 2.0, X Company,” 2018. [Online]. Available: <https://www.x.company/glass/>
- [146] S. L. Hicks, I. Wilson, L. Muhammed, J. Worsfold, S. M. Downes, and C. Kennard, “A depth-based head-mounted visual display to aid navigation in partially sighted individuals,” *Plos One*, vol. 8, no. 7, p. e67695, 2013.
- [147] “Smart glasses for people with a visual impairment,” 2019. [Online]. Available: <https://www.oxsight.co.uk/>
- [148] “Moverio BT-200, Epson Seiko Epson Corporation,” 2018. [Online]. Available: <https://www.epson.co.uk>
- [149] P. Gould and R. White, *Mental maps*. Routledge, 2012.

- [150] Y. Zhao, M. Hu, S. Hashash, and S. Azenkot, “Understanding low vision people’s visual perception on commercial augmented reality glasses,” in *proceedings of the 2017 CHI conference on human factors in computing systems*. ACM, 2017, pp. 4170–4181.
- [151] K. Jaya and D. Kruti, “Moving object detection: Review of recent research trends,” in *proceedings of the international conference on pervasive computing (ICPC)*, Jan 2015, pp. 1–5.
- [152] O. Barnich and M. Van Droogenbroeck, “Vibe: A universal background subtraction algorithm for video sequences,” *IEEE Transactions on Image processing*, vol. 20, no. 6, pp. 1709–1724, 2010.
- [153] N. Singla, “Motion detection based on frame difference method,” *International Journal of Information & Computation Technology*, vol. 4, no. 15, pp. 1559–1565, 2014.
- [154] K. A. Joshi and D. G. Thakore, “A survey on moving object detection and tracking in video surveillance system,” *International Journal of Soft Computing and Engineering*, vol. 2, no. 3, pp. 44–48, 2012.
- [155] S. S. Beauchemin and J. L. Barron, “The computation of optical flow,” *ACM Computing Surveys*, vol. 27, no. 3, pp. 433–466, 1995.
- [156] W. Hu, T. Tan, L. Wang, and S. Maybank, “A survey on visual surveillance of object motion and behaviors,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 34, no. 3, pp. 334–352, 2004.
- [157] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, “Performance of optical flow techniques,” *International Journal of Computer Vision*, vol. 12, no. 1, pp. 43–77, 1994.

- [158] M. Kaushal, B. S. Khehra, and A. Sharma, "Soft computing based object detection and tracking approaches: State-of-the-art survey," *Applied Soft Computing*, vol. 70, pp. 423–464, 2018.
- [159] N. Goyette, P. M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "Change detection.net: A new change detection benchmark dataset," in *proceedings of the IEEE computer society conference on computer vision and pattern recognition workshops (CVPR)*, 2012, pp. 1–8.
- [160] G. J. Brostow, J. Shotton, J. Fauqueur, and R. Cipolla, "Segmentation and recognition using structure from motion point clouds," in *proceedings of the European conference on computer vision (ECCV)*, 2008, pp. 44–57.
- [161] G. J. Brostow, J. Fauqueur, and R. Cipolla, "Semantic object classes in video: A high-definition ground truth database," *Pattern Recognition Letters*, vol. 30, no. 2, pp. 88–97, 2008.
- [162] P. Panchal, G. Prajapati, S. Patel, H. Shah, and J. Nasriwala, "A review on object detection and tracking methods," *International Journal for Research in Emerging Science and Technology*, vol. 2, no. 1, pp. 7–12, 2015.
- [163] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.
- [164] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *proceedings of the 7th international joint conference on artificial intelligence*. Morgan Kaufmann Publishers, 1981, pp. 674–679.
- [165] S. Uras, F. Girosi, A. Verri, and V. Torre, "A computational approach to motion perception," *Biological Cybernetics*, vol. 60, no. 2, pp. 79–87, 1988.
- [166] A. Verri and T. Poggio, "Motion field and optical flow: Qualitative properties," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 5, pp. 490–498, 1989.

- 
- [167] H. H. Nagel, "Displacement vectors derived from second-order intensity variations in image sequences," *Computer Vision, Graphics, and Image Processing*, vol. 21, no. 1, pp. 85–117, 1983.
- [168] H. Nagel, "On a constraint equation for the estimation of displacement rates in image sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 1, pp. 13–30, 1989.
- [169] H. H. Nagel and W. Enkelmann, "An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 5, pp. 565–593, 1986.
- [170] P. Anandan, "A computational framework and an algorithm for the measurement of visual motion," *International Journal of Computer Vision*, vol. 2, no. 3, pp. 283–310, 1989.
- [171] F. Glazer, G. Reynolds, and P. Anandan, "Scene matching by hierarchical correlation," Massachusetts university, Amherst department of computer and information science, Tech. Rep., 1983.
- [172] A. Singh, "An estimation-theoretic framework for image-flow computation," in *proceedings of the 3ed international conference on computer vision*. IEEE, 1990, pp. 168–177.
- [173] D. J. Heeger, "Optical flow using spatiotemporal filters," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 279–302, 1988.
- [174] A. M. Waxman and K. Wohn, "Contour evolution, neighborhood deformation, and global image flow: Planar surfaces in motion," *The International journal of Robotics research*, vol. 4, no. 3, pp. 95–108, 1985.
- [175] A. M. Waxman, J. Wu, and F. Bergholm, "Convected activation profiles and the measurement of visual motion," in *proceedings the computer society con-*

- ference on computer vision and pattern recognition (CVPR)*. IEEE, 1988, pp. 717–723.
- [176] Z. Zivkovic, “Improved adaptive gaussian mixture model for background subtraction,” in *proceedings of the 17th international conference on pattern recognition (ICPR)*, vol. 2, Aug 2004, pp. 28–31 Vol.2.
- [177] J. Shi and C. Tomasi, “Good features to track,” in *proceedings of IEEE conference on computer vision and pattern recognition (CVPR)*, Jun 1994, pp. 593–600.
- [178] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [179] iLuvTech. (2016, 3) 4k street view, hongdae, korea. [Online]. Available: <https://youtu.be/qA2W4hLh6Gc>
- [180] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin, “A self-adjusting approach to change detection based on background word consensus,” in *proceedings of the IEEE winter conference on applications of computer vision (WACV)*, 2015, pp. 990–997.
- [181] L. Maddalena and A. Petrosino, “A fuzzy spatial coherence-based approach to background/foreground separation for moving object detection,” *Neural Computing and Applications*, vol. 19, no. 2, pp. pp.179–186, 2010.
- [182] G. Allebosch, F. Deboeverie, P. Veelaert, and W. Philips, “Efic: edge based foreground background segmentation and interior classification for dynamic camera viewpoints,” in *proceedings of the international conference on advanced concepts for intelligent vision systems*. Springer, 2015, pp. 130–141.
- [183] H. Sajid and S.-C. S. Cheung, “Background subtraction for static & moving camera,” in *proceedings of the IEEE international conference on image processing (ICIP)*. IEEE, 2015, pp. 4530–4534.

- [184] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1137–1149, 06 2015.
- [185] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, “Ssd: Single shot multibox detector,” in *proceedings of the European conference on computer vision (ECCV)*. Springer, 2016, pp. 21–37.
- [186] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv:1409.1556*, 2014.
- [187] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” *arXiv:1704.04861*, 2017.
- [188] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, “Microsoft coco: Common objects in context,” in *proceedings of the European conference on computer vision (ECCV)*. Springer, 2014, pp. 740–755.
- [189] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, “The pascal visual object classes challenge: A retrospective,” *International Journal of Computer Vision*, vol. 111, no. 1, pp. 98–136, Jan. 2015.
- [190] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama *et al.*, “Speed/accuracy trade-offs for modern convolutional object detectors,” in *proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2017, pp. 7310–7311.
- [191] J. Dai, Y. Li, K. He, and J. Sun, “R-fcn: Object detection via region-based fully convolutional networks,” in *proceedings of the 30th international conference on neural information processing systems*, 2016, pp. 379–387.

- [192] S.-K. Weng, C.-M. Kuo, and S.-K. Tu, "Video object tracking using adaptive kalman filter," *Journal of Visual Communication and Image Representation*, vol. 17, no. 6, pp. 1190–1208, 2006.
- [193] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, no. 1, pp. 83–97, 1955.
- [194] G. A. Mills-Tettey, A. Stentz, and M. B. Dias, "The dynamic hungarian algorithm for the assignment problem with changing costs," Carnegie Mellon University, Tech. Rep., 2007.
- [195] P. Cika, M. Zukal, Z. Libis, and M. K. Dutta, "Tracking and speed estimation of selected object in video sequence," in *proceedings of the 36th International conference on telecommunications and signal processing (TSP)*. IEEE, 2013, pp. 881–884.
- [196] S. E. Fahlman and C. Lebiere, "Advances in neural information processing systems," D. S. Touretzky, Ed. Morgan Kaufmann Publishers, 1990, ch. The Cascade-correlation Learning Architecture, pp. 524–532.
- [197] S. E. Fahlman, "Faster-learning variations of back-propagation: An empirical study," in *proceeding of the connectionist models summer school*. Morgan Kaufmann, 1988, pp. 38–51.
- [198] S. A. Glantz, B. K. Slinker, and T. B. Neilands, *Primer of applied regression and analysis of variance*. McGraw-Hill New York, 1990, vol. 309.
- [199] A. Patney, M. Salvi, J. Kim, A. Kaplanyan, C. Wyman, N. Benty, D. Luebke, and A. Lefohn, "Towards foveated rendering for gaze-tracked virtual reality," *ACM transactions on graphics (ToG)*, vol. 35, no. 6, p. 179, 2016.
- [200] N. T. Swafford, J. A. Iglesias-Guitian, C. Koniaris, B. Moon, D. Cosker, and K. Mitchell, "User, metric, and computational evaluation of foveated rendering methods," in *proceedings of the ACM symposium on applied perception*. ACM, 2016, pp. 7–14.



- [201] H. W. Hunziker, *Im Auge des Lesers: vom Buchstabieren zur Lesefreude: foveale und periphere Wahrnehmung*. Hunziker Hans-Werner, 2006.
- [202] J. N. Carroll and C. A. Johnson, “Visual field testing: From one medical student to another,” 2013.
- [203] M. S. El-Nasr and S. Yan, “Visual attention in 3d video games,” in *proceedings of the ACM SIGCHI international conference on advances in computer entertainment technology*. ACM, 2006, p. 22.
- [204] A. Cuschieri, “Visual displays and visual perception in minimal access surgery,” in *Seminars in laparoscopic surgery*, vol. 2, no. 3. Sage Publications, 1995, pp. 209–214.
- [205] K. Davids, A. M. Williams, and J. G. Williams, *Visual perception and action in sport*, 1st ed. Routledge, 2005.
- [206] M. Tonnis, V. Broy, and G. Klinker, “A survey of challenges related to the design of 3d user interfaces for car drivers,” in *proceedings of the IEEE symposium on 3D user interfaces (3DUI’06)*, 2006, pp. 127–134.
- [207] “Virtual cable™,” 2011. [Online]. Available: <http://mvs.net/index.html>
- [208] V. Charissis and M. Naef, “Evaluation of prototype automotive head-up display interface: testing driver’s focusing ability through a vr simulation,” in *proceedings of the IEEE symposium on Intelligent Vehicles*, 2007, pp. 560–565.
- [209] A. Sato, I. Kitahara, Y. Kameda, and Y. Ohta, “Visual navigation system on windshield head-up display,” in *proceedings of the 13th world congress on intelligent transort systems*, 2006.
- [210] M. Tonnis, C. Lange, and G. Klinker, “Visual longitudinal and lateral driving assistance in the head-up display of cars,” in *proceedings of the 6th IEEE and ACM international symposium on mixed and augmented reality*. IEEE Computer Society, 2007, pp. 1–4.

- [211] J. Sauerbrey, “Man abbiegeassistent: Ein system zur unfallvermeidung beim rechtsabbiegen von lkw,” in *proceedings of the Tagung aktive sicherheit durch fahrerassistenzsysteme*, 2004.
- [212] S. Henderson and S. Feiner, “Opportunistic tangible user interfaces for augmented reality,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 1, pp. 4–16, 2009.
- [213] S. J. Henderson and S. Feiner, “Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret,” in *proceedings of the 8th IEEE international symposium on mixed and augmented reality*, 2009, pp. 135–144.
- [214] S. Lee and Ö. Akin, “Augmented reality-based computational fieldwork support for equipment operations and maintenance,” *Automation in Construction*, vol. 20, no. 4, pp. 338–352, 2011.

# Appendices

## Appendix A

### Ethical Approval Form



Faculty of Science and Engineering Committee on Research Ethics

22 August 2017

Dear Dr Al-Nuaimy,

I am pleased to inform you that your application for research ethics approval has been approved. Details and conditions of the approval can be found below:

Reference:	1982
Project Title:	Smart Assistive Technology for Peripheral Vision Loss Rehabilitation
Principal Investigator/Supervisor:	Dr Waleed Al-Nuaimy
Co-Investigator(s):	Mrs Ola Younis, Dr Fiona Rowe
Lead Student Investigator:	-
Department:	Electrical Engineering and Electronics
Approval Date:	22/08/2017
Approval Expiry Date:	Five years from the approval date listed above

The application was **APPROVED** subject to the following conditions:

**Conditions**

- All serious adverse events must be reported via the Research Integrity and Ethics Team ([ethics@liverpool.ac.uk](mailto:ethics@liverpool.ac.uk)) within 24 hours of their occurrence.
- If you wish to extend the duration of the study beyond the research ethics approval expiry date listed above, a new application should be submitted.
- If you wish to make an amendment to the research, please create and submit an amendment form using the research ethics system.
- If the named Principal Investigator or Supervisor leaves the employment of the University during the course of this approval, the approval will lapse. Therefore it will be necessary to create and submit an amendment form using the research ethics system.
- It is the responsibility of the Principal Investigator/Supervisor to inform all the investigators of the terms of the approval.

Kind regards,

Faculty of Science and Engineering Committee on Research Ethics

[foseeth@liverpool.ac.uk](mailto:foseeth@liverpool.ac.uk)

0151 795 0649

## Appendix B

### Participant Information Sheet



Committee on Research Ethics

Participant Information Sheet

## Smart Assistive Technology for Peripheral Vision Loss Rehabilitation

Researcher(s): Ola Younis, Waleed Al-Nuaimy, Fiona Rowe

*You are being invited to participate in a research study. Before you decide whether to participate, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and feel free to ask us if you would like more information or if there is anything that you do not understand. Please also feel free to discuss this with your friends, relatives and GP if you wish. We would like to stress that you do not have to accept this invitation and should only agree to take part if you want to.*

*Thank you for reading this.*

### 1. What is the purpose of the study?

This experiment is to have your feedback about a proposed assistive technology that will be implemented on smart glasses to help in peripheral vision loss rehabilitation.

### 2. Why have I been chosen to take part?

You are chosen for this study because you have visual field problems that affect your vision, or because you have a healthy vision but like to give your feedback about the proposed system and its performance.

### 3. Do I have to take part?

This is voluntary task and you are free to withdraw at anytime without explanation and without incurring a disadvantage.

### 4. What will happen if I take part?

1. You will be asked to fill in a questionnaire with two sections:
  - A. General questions about your vision health.
  - B. Questions about your opinion in our proposed assistive technology that will be implemented on smart glasses as a wearable technology to help you in your daily activities.
2. Your feedback and results will be recorded and saved in e-files for future analysis. No personal information will be recorded.
3. Data will be anonymized and used in scientific research only.
4. This experiment will take between 10-15 minutes only.

### 5. Are there any risks in taking part?

There is no risk for the participant to involve in this study.

### 6. Are there any benefits in taking part?

There are no benefits of involving with this experiment. However you may find benefits from helping with research through potential benefits in the future.

**7. What if I am unhappy or if there is a problem?**

"If you are unhappy, or if there is a problem, please feel free to let us know by contacting [Dr Waleed Al-Nuaimy] and we will try to help. If you remain unhappy or have a complaint which you feel you cannot come to us with then you should contact the Research Governance Officer at [ethics@liv.ac.uk](mailto:ethics@liv.ac.uk). When contacting the Research Governance Officer, please provide details of the name or description of the study (so that it can be identified), the researcher(s) involved, and the details of the complaint you wish to make."

**8. Will my participation be kept confidential?**

Yes, no personal information that may identify you (name, address, post code...) will be recorded.  
Data collected before and after performing the experience will be saved in manual (paper) and electronic files with the guarantee of data anonymisation.  
Anonymized data may be stored in laptop computers with password protected folders and files.

**9. What will happen to the results of the study?**

The results will be used to guide our research project. Your answers will help us to clearly identify which objects are harmful for you and what is the suitable type of information for your case.  
Your preferences and answers may be used in research papers in future. Participant will not be identified from published material because data is anonymised

**10. What will happen if I want to stop taking part?**

Participants can withdraw at any time, without explanation. Results up to the period of withdrawal may be used, if you are happy for this to be done. Otherwise you may request that they are destroyed and no further use is made of them. If results are anonymised you should make clear that results may only be withdrawn prior to anonymisation.

**11. Who can I contact if I have further questions?**

Ola Younis,  
[younis@liv.ac.uk](mailto:younis@liv.ac.uk)  
DEPARTMENT OF ELECTRICAL ENGINEERING & ELECTRONICS  
UNIVERSITY OF LIVERPOOL  
Liverpool L69 3GJ, United Kingdom

Dr Waleed Al-Nuaimy,  
[wax@liverpool.ac.uk](mailto:wax@liverpool.ac.uk)  
DEPARTMENT OF ELECTRICAL ENGINEERING & ELECTRONICS  
UNIVERSITY OF LIVERPOOL  
Liverpool L69 3GJ, United Kingdom

Dr Fiona Rowe,  
[rowef@liverpool.ac.uk](mailto:rowef@liverpool.ac.uk)  
Department of Health Service Research/ Psychology Health and Society,  
UNIVERSITY OF LIVERPOOL  
Liverpool L69 3GJ, United Kingdom



## Appendix C

### Participant Consent Form



**Committee on Research Ethics**  
**PARTICIPANT CONSENT FORM**

**Smart Assistive Technology for Peripheral Vision Loss  
 Rehabilitation**

Researcher(s): Ola Younis, Waleed Al-Nuaimy, Fiona Rowe

- |   |   |
|---|---|
| <p>1. I confirm that I have read and have understood the information sheet dated July 2017 for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.</p> <p>2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my rights being affected. In addition, should I not wish to answer any particular question or questions, I am free to decline.</p> <p>3. I understand that, under the Data Protection Act, I can at any time ask for access to the information I provide and I can also request the destruction of that information if I wish.</p> <p>4. I agree to take part in the above study.</p> | <p><b>Please<br/>initial box</b></p> <div style="border: 1px solid black; width: 40px; height: 20px; margin-bottom: 10px;"></div> <div style="border: 1px solid black; width: 40px; height: 20px; margin-bottom: 10px;"></div> <div style="border: 1px solid black; width: 40px; height: 20px; margin-bottom: 10px;"></div> <div style="border: 1px solid black; width: 40px; height: 20px;"></div> |
|---|---|

Participant Name	Date	Signature
------------------	------	-----------

**Principal Investigator:**  
 Name: Waleed Al-Nuaimy  
 Work Address: University of Liverpool, EEE department.  
 Work Telephone: (0)151 794 4512  
 Work Email: wax@liv.ac.uk

**Student Researcher:**  
 Name: Ola Younis  
 Work Address: University of Liverpool,  
 EEE department  
 Work Email: Younis@liv.ac.uk

**Version 1.0**  
**July 2017**

## Appendix D

### Questionnaire 1



UNIVERSITY OF LIVERPOOL  
DEPARTMENT OF ELECTRICAL ENGINEERING AND ELECTRONICS

**Assistive Technology for Visual Field Defects  
Rehabilitation Research Project**

2018

Please take as much time as you need to answer each question. All your answers are confidential.

**INSTRUCTIONS:**

1. In general we would like to have people try to complete these forms on their own. If you find that you need assistance, please feel free to ask the project staff and they will assist you.
2. Please answer every question (unless you are asked to skip questions because they don't apply to you).
3. Answer the questions by circling the appropriate number.
4. If you are unsure of how to answer a question, please give the best answer you can and make a comment in the left margin.
5. Please complete the questionnaire before leaving the centre and give it to a member of the project staff. Do not take it home.
6. If you have any questions, please feel free to ask a member of the project staff, and they will be glad to help you.

**STATEMENT OF CONFIDENTIALITY:**

All information that would permit identification of any person who completed this questionnaire will be regarded as strictly confidential. Such information will be used only for the purposes of this study and will not be disclosed or released for any other purposes without prior consent, except as required by law.

**Research Group Members:**

1. Ola Younis, PhD student.
2. Waleed Al-Nuaimy, main supervisor, senior lecturer.
3. Fiona Rowe, second supervisor, Professor.

### Smart glasses for people with visual field defects

In this project, smart glasses will be used to help people with visual field defects in their navigation and daily activities such as crossing the road and detecting motion around them. The user will be able to wear the smart glasses which contain a video camera and two display units (transparent) and connected to a small computer (could be wired or wireless) to perform some processing.

The goal is to notify the users about any possible threats in their blind area by displaying notifications on the display units at their healthy vision area. This will help the users to (1) avoid obstacles, (2) detect motion (3) detect certain types of objects based on personal preferences.

The system will be customizable based on the visual field test results for each user in order to determine the blind/healthy vision areas before running the smart glasses.

Q1. Using these smart glasses, what type of objects do you think are the most important to be notified about?

- ☐ Stationary objects (obstacles) in your pathway (e.g. chair, couch, car)
- ☐ Moving objects that could cross your path or you may bump into them.

Q2. Using the smart glasses, what type of objects do you prefer to have information about?

*List the most important ones.*

<input type="checkbox"/> Cars	<input type="checkbox"/> Street bollards
<input type="checkbox"/> People	<input type="checkbox"/> Bicycles
<input type="checkbox"/> Chairs	<input type="checkbox"/> Walls
<input type="checkbox"/> Tables	<input type="checkbox"/> Stairs

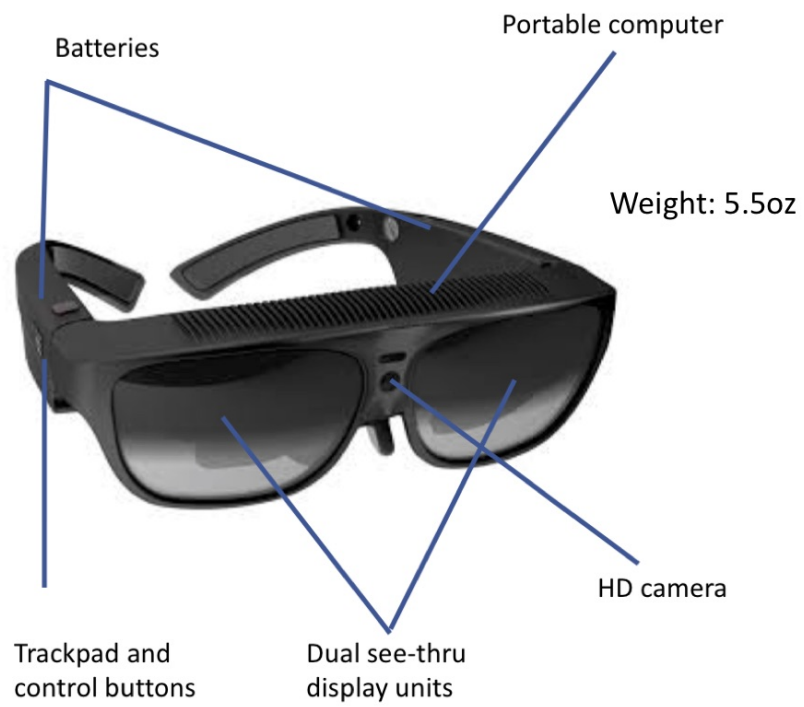
Q3. If the smart glasses detected a possible hazard (expected collision), how early do you prefer to get a warning notification (in seconds)?

Q4. Do you prefer a single level of hazard notification or multiple levels based on the degree of dangerous?

- ☐ One level.
- ☐ Multiple levels.

Q5. How do you prefer to get the notifications?

- ☐ Audio notifications.
- ☐ Visual notification displayed without blocking the normal vision.



## Appendix E

### Questionnaire 2





UNIVERSITY OF LIVERPOOL  
DEPARTMENT OF ELECTRICAL ENGINEERING AND ELECTRONICS

**Assistive Technology for Visual Field Defects  
Rehabilitation Research Project**

Please take as much time as you need to answer each question. All your answers are confidential.

INSTRUCTIONS:

1. In general, we would like to have people try to complete these forms on their own. If you find that you need assistance, please feel free to ask the project staff and they will assist you.
2. Please answer every question (unless you skip questions because they don't apply to you).
3. If you are unsure of how to answer a question, please give the best answer you can and comment on the left margin.
4. Please complete the questionnaire before leaving the centre and give it to a member of the project staff.
5. If you have any questions, please feel free to ask a member of the project staff, and they will be glad to help you.

STATEMENT OF CONFIDENTIALITY:

All information that would permit identification of any person who completed this questionnaire will be regarded as strictly confidential. Such information will be used only for this study and will not be disclosed or released for any other purposes without prior consent.

Research Group Members:

1. Ola Younis, PhD student.
2. Waleed Al-Nuaimy, principal supervisor, Senior lecturer.
3. Fiona Rowe, second supervisor, Professor.

In this project, smart glasses will be used to help people with visual field defects in their navigation and daily activities such as crossing the road and detecting motion around them.

The user will be able to wear the smart glasses which contain a video camera and two display units (transparent) and connected to a small computer (could be wired or wireless) to perform some processing.

The goal is to notify the person about any possible threats (termed as hazards in this project) in their blind area by displaying notifications on the display units within their healthy vision area. This will help the users to (1) avoid obstacles, (2) detect motion (3) identify specific types of objects based on personal preferences.

The system will be customizable based on the visual field test results for each person to determine the blind/healthy vision areas before running the smart glasses.

In this questionnaire, we are investigating people's preferences for developing a smart, wearable assistive technology.

**Section 1:**

1. Do you use portable electronic devices (e.g. kindle, ipad)?

☐ Yes, (specify) .....

☐ No

2. Do you use any navigation aid in your daily life (e.g. white cane)?

☐ Yes, (specify) .....

☐ No

3. Do you wear eye glasses?

☐ Yes

☐ No

4. What type of vision loss do you suffer?

.....

**Section 2**

1. In your navigation, which of the following do you find the most dangerous to detect and avoid? Please rate each from 1 to 5: (1) Least danger, (5) most danger

- ☐ Static hazards outside your pathway, (i.e. not in your way)
- ☐ Static hazards in your pathway
- ☐ Moving objects not in your way (any type)
- ☐ A person moving towards you (or your pathway)
- ☐ Object moving towards you (or your pathway)
- .....

2. How would you prefer your input to the system to be, (i.e. how you add a response to the system?)







- ☐ Speech (*specific predefined words*)
- ☐ Touch
- ☐ Hybrid (*touch and speech*)

3. How do you prefer the feedback of the system to be?

- ☐ Audio (*speech, beeps*)
- ☐ Touch (*vibration*)
- ☐ Visual
- ☐ Hybrid (*specify*).....

For the visual feedback only:

4. How do you prefer the feedback style to be?

- ☐ One shape, different colour (e.g.    )
- ☐ Different shapes, one colour (e.g.    )

5. How many notifications would you prefer to have at one time?

- ☐ One at a time (highest priority only)
- ☐ Multiple at a time (*specify the number*).....

6. If you chose multiple notifications at a time, how do you prefer them to appear?

- ☐ In a sequential way, one after the other
- ☐ All at the same time

7. What information do you prefer to have about the detected hazard?

- ☐ Hazard direction only
- ☐ Hazard type and direction
- ☐ Hazard direction and speed
- ☐ Extra information (explain).....

8. When do you prefer the feedback to appear?

- ☐ In regular time intervals for as long as the hazard is present
- ☐ In an incremental way if the hazard persists
- ☐ Only once when the hazard is detected

9. How do you prefer the feedback to disappear?

- ☐ Automatic after a time interval
- ☐ Manual based on your response (input) to the system

**Section 3**

Please, give us your opinion, impression, ideas and feedback about the proposed system.

We are more than happy to hear from you regarding the design, idea, weight, cost, and any other suggestions for further improvements.