



MONASH University

A Wearable Technology-based System to Negotiate Surface Discontinuities for the Blind and Low Vision People during Navigation

Leong Kuan Yew

Bachelor of Information Technology (Honours) – University Malaysia Sabah, Malaysia

Master of Science in Business Information Technology – Northumbria University,

Newcastle upon Tyne, UK

A thesis submitted for the degree of Doctor of Philosophy at
Monash University in 2019
Faculty of Information Technology

Table of Contents

Table of Contents	i
Copyright Notice	v
Abstract	vi
Declaration	viii
List of Publications	ix
Acknowledgements.....	x
List of Tables	xiii
List of Figures.....	xv
List of Abbreviations.....	xx
Chapter 1: Introduction.....	1
1.1 Overview	1
1.2 Defining the Blind and Low Vision.....	4
1.3 Navigational Challenges Faced by the Blind and Low Vision People	5
1.4 A Brief Understanding of Navigation and Threats along a Pathway.....	7
1.4.1 The Cues	8
1.4.2 The Threats.....	9
1.5 Common Tools for Navigation by the BLVs.....	10
1.5.1 Personal Assistive Tools.....	11
1.5.2 Integrated Assistive Infrastructures	12
1.6 Motivation of the Research.....	14
1.6.1 Factors Affecting Blind Navigation	14
1.6.2 Current Technologies and the Gaps.....	16
1.6.3 Poor Enforcement of Regulations for Accessible Built Environment	23
1.7 Problem Statement.....	26
1.8 Research Objectives and Questions	27
1.8.1 Primary Research Question	28
1.8.2 Subsidiary Research Questions	28
1.9 Research Contributions.....	29
1.10 Thesis Outline	30
Chapter 2: Literature Review.....	34
2.1 Context of Disability – Medical Model versus Social Model	34
2.1.1 Medical Model of Disability.....	34
2.1.2 Social Model of Disability.....	36

2.2	Assistive Technologies for Threats Negotiation.....	38
2.2.1	Laser Cane	38
2.2.2	Ultrasonic Cane	44
2.2.3	Depth Sensing Technologies	46
2.2.4	Computer Vision-based Technologies.....	48
2.2.5	Robotic Solutions	50
2.3	Summary of Major Findings from Literature	56
2.3.1	The Research Gaps	56
2.3.2	What the Research Offers.....	58
2.3.3	Conclusion - Computer Vision and Machine Learning as the Core Approach	60
Chapter 3:	Research Design.....	64
3.1	Overview of the Research Design	64
3.2	Epistemological Position: A Socio-Technologist Approach	66
3.2.1	Background on Design Science Research.....	67
3.2.2	Justification of the Choice of Design Science Methodology.....	68
3.2.3	Scope of the Research	70
3.2.4	Output of Design Science Artifacts from this Research	71
3.3	Process of the Design Science Methodology	73
3.4	Specific Research Techniques	79
3.4.1	Consultation with BLV Service Providers	79
3.4.2	Data Generation	84
3.4.3	System Development	84
3.4.4	System Evaluation	84
3.5	Summary of the Chapter	84
Chapter 4:	Data Generation.....	85
4.1	Instrument Development – the Phase-1 Prototype	85
4.2	Instrument Setup.....	90
4.3	Data Collection.....	93
4.3.1	Crowdsourcing to Identify Potential Locations for Data Collection	93
4.3.2	Data Sampling	93
4.3.3	Measuring the Ground Truth	95
4.3.4	The Data	98
4.4	Data Pre-Processing	100
4.4.1	Cleaning.....	100
4.4.2	Normalization.....	103

4.4.3	PCA Whitening.....	106
4.5	Taxonomy of Surface Discontinuity and Data Labelling.....	108
4.6	Data Augmentation	110
4.7	Summary of the Chapter	112
Chapter 5:	System Development (Phase-2 Prototype)	113
5.1	Overview of Phase-2 Prototype.....	113
5.2	Feature Extraction	115
5.2.1	Disparity Mapping and Depth	115
5.2.2	Technique to Generate the Disparity Map.....	121
5.2.3	Regions of Uniform Intensity	124
5.2.4	Disparity Range	125
5.2.5	Windowing.....	127
5.2.6	Sum of Difference	131
5.3	Assembling a Deep Learning Architecture.....	132
5.3.1	The Proposed Tri-Channel Convolutional Neural Network	135
5.3.2	Architecture of the Proposed CNN	137
5.3.3	Training Algorithm – Backpropagation	141
5.3.4	Model Tuning – an Optimization Approach.....	144
5.3.5	Variable Neighbourhood Search	147
5.4	Summary of the Chapter	149
Chapter 6:	System Evaluation.....	150
6.1	Models Training and Classification Results	150
6.1.1	Model 1 and Its Variants.....	150
6.1.2	Model 2 and Its Variants.....	153
6.1.3	Model 3 and Its Variants.....	155
6.2	Evaluation of the Prototype	158
6.2.1	Classification Accuracy Evaluation.....	159
6.2.2	Algorithmic Efficiency Evaluation	167
6.3	Summary of the Chapter	173
Chapter 7:	Conclusion	174
7.1	Thesis Summary	174
7.2	Answering the Research Questions	176
7.2.1	Subsidiary Research Question 1.....	177
7.2.2	Subsidiary Research Question 2	178
7.2.3	Subsidiary Research Question 3	179

7.2.4	Subsidiary Research Question 4	180
7.3	Impact and Original Contributions	181
7.3.1	Contributions to Practical Application	181
7.3.2	Contributions to Research Methodology	182
7.4	Future Work	183
7.5	Concluding Remarks	187
List of References		189
Appendix		200
Appendix 1: Hardware Specifications of the Prototype.....		200
Appendix 2: Phase-1 Prototype Design Details		203
Appendix 3: Crowdsourcing via Social Media		207
Appendix 4: Variable Neighbourhood Search.....		208

Copyright Notice

© Leong Kuan Yew (2019).

I certify that I have made all reasonable efforts to secure copyright permissions for third-party content included in this thesis and have not knowingly added copyright content to my work without the owner's permission.

Abstract

According to World Health Organization, there is an estimated 285 million blind and low vision (BLV) people worldwide. Apart from difficulties in accessing visual information, one of the greatest challenges to independence for the BLVs is difficulties in self-navigation. Through consultations with several blind and low vision service providers, it is realized that negotiating surface discontinuities is one of the very prominent challenges when navigating an outdoor environment within urban areas. Surface discontinuities are commonly formed by rises and drop-offs along a pathway. They could be a threat to balancing during a walk and perceiving such a threat is a difficult challenge to the BLVs without proper navigational aids or facilities.

The research sets to address the challenge of negotiating surface discontinuities by the BLVs along the urban pathway using technology. The central question of this research is: **“How can a technology assist the blind and low vision people to negotiate surface discontinuities along their navigational pathway?”**

Based on a design science methodology, a lightweight, small and unobtrusive prototype of a wearable technology-based system was proposed. The research design comprises 5 main process steps:

- identification of problem through several participating blind and low vision service providers
- literature analysis and consultation with blind and low vision service providers to conceptualize the design of a prototype
- generation of a novel dataset through the development of phase-1 prototype and several pre-processing steps
- development of the phase-2 prototype by creating a taxonomy from the collected data to train a machine learning model
- evaluation of the accuracy and efficiency of the developed prototype

Tailored for a computer vision task, the prototype was equipped with a tiny stereo camera and an embedded system on a single board computer. With this prototype, a set of video data that exemplifies the issue of surface discontinuity was collected from the field. From the data, a taxonomy of surface discontinuity common to the navigational pathway found at urban environments was created. Through experiments, a purpose-built stacked convolutional neural network was assembled, trained and optimized across a series of configurations and hyperparameters. With the incorporation of the best-trained model, the similar prototype used for data collection was then repurposed such that it is working in real-time to classify the condition of a pathway with simple feedback for the purpose of research study.

Findings from the research confirmed that a wearable technology-based system equipped with a stereo camera running on an embedded system can take advantage of current state-of-the-art computer vision and deep learning approach to detect and classify surface discontinuity in an outdoor urban environment. Through the development of a taxonomy, the research also identified nine classes of surface discontinuity relevant to urban navigational pathways which helped in building a machine learning model. A stacked convolutional neural network was found to be accurate and efficient for classifying the surface discontinuities. With specific hardware used in the implementation of the system, the prototype was evaluated for its accuracy and efficiency. The optimized model achieved a classification accuracy of 96% in field evaluation. In the efficiency evaluation, the prototype was analyzed for its CNN model's algorithmic efficiency, memory usage, speed and power consumption. The model was 50% more efficient with very low memory usage as compared to the original model it was based on. The processing speed was near real-time and the power consumption of the prototype is moderate. The research contributed a methodology and a framework consisting of some algorithms, models and a proof of concept in developing the prototype.

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Signature:

Print Name: Leong Kuan Yew

Date: 28th March 2019

List of Publications

K. Y. Leong, S. Egerton and Carina K.Y. Chan, “A Wearable Technology to Negotiate Surface Discontinuities for the Blind and Low Vision”, *2017 IEEE Life Sciences Conference (LSC)*, Sydney, NSW, 2017, pp. 115-120. DOI: 10.1109/LSC.2017.8268157

K.Y. Leong, “Augmenting the Perception of the Blind and Low Vision to Increase their Levels of Independent Navigation through Wearable Technology” (Oral Presentation), *Image Processing, Image Analysis and Real-Time Imaging Symposium (IPIARTI)* 2018, Selangor, Malaysia.

Publication of dataset at: <https://kyleongsurfacediscontinuitydataset.wordpress.com/>

Acknowledgements

This research owes a great debt of gratitude to three blind and low vision service providers namely Dialogue in the Dark (Malaysia branch), St. Nicholas' Home Penang (Malaysia) and Vision Australia Dandenong (Victoria, Australia). The founder of Dialogue in the Dark Malaysia, Stevens Chan is an active advocate for the blind and low vision people in the country. He was the first person to introduce and bring in the very first guide dog to the local community. It was through continuous consultations with Stevens and his members that the issue of surface discontinuity was then brought up. Appreciation also goes to Peter Khor, a member of Dialogue in the Dark who is also my neighbour, that connected Stevens and his team to me during the early stage of the research. Peter had also actively participated in almost every consultation to offer his constructive opinions to identify the navigational issues which was eventually addressed by the research.

Members of St. Nicholas' Home Penang offered their consultations to the research after knowing about the research's objective to address the issue of surface discontinuity. The executive director, Daniel Soon personally chaired the meeting when I visited their campus for the very first time. With a purpose-built campus to cater for its blind and low vision students and residents, St. Nicholas' Home Penang offered an opportunity for me to observe the accessible environment which benefited the research in many ways. They generously demonstrated several technologically enabled assistive tools they acquired for their members, and they also concurred on the lack of navigational aids and the issue of surface discontinuity.

The research continued to benefit from both Dialogue in the Dark and St. Nicholas' Home Penang in the subsequent sessions of meetings or consultations, to gather feedback on the exact targets of problem through surveys that involved some 3D printed replicas of surface discontinuity. In the following phase of wearable prototype development, the research obtained useful feedback about position setup and mounting configuration of the wearable from these service providers. Thanks to the valuable insights mentioned above, the

research was able to construct a solid wearable design that was implemented in the prototype.

Appreciation also goes to Vision Australia Dandenong. Vision Australia has many offices servicing the blind and low vision people around cities and suburbs in Australia. I visited the Dandenong office to learn more about the surface discontinuity issue and other related matters. Two of their representatives and a certified orientation and mobility (O&M) trainer provided some valuable insights into the issue. They also spent time to inspect the 3D printed replicas of surface discontinuity to offer their opinions on the possible targets that the research would address. At the end of the visit, the O&M trainer demonstrated several types of guide cane for outdoor navigation, and guided me on some proper technique of guide cane usage. The research had greatly benefited from this visit, especially the technique of guide cane handling was applied during data collection to mimic as close as possible the actual situation when a blind and low vision person is navigating a pathway.

I would like to acknowledge the proof-reading services provided by Carol Roche, a professional creative writer in the advertising industry for almost forty years. Recently retired, she now devotes her time to doing social work with refugees and in her community. The proof-reading services rendered by Roche were limited to English language expression, grammar, spelling, text alignment and punctuation checking only.

I would also like to acknowledge Dr Lim Siew Mooi from University Malaysia of Computer Science and Engineering for her assistance in conducting data collection at several urban areas within Klang Valley, Malaysia.

Finally, and most importantly, I would like to express the utmost gratitude to my supervisors Prof Frada Burstein, Dr Simon Egerton and Dr Carina Chan, and the then Associate Dean of Graduate Research Prof Sue McKemmish. I appreciate all the technical guidance provided by both Dr Simon and Dr Carina especially during the early phase of the

research. To Prof Frada, I am especially thankful for her to step in and to take up the role of main supervisor in the midst of my PhD milestone due to relocation of the former main supervisor. Apart from valuable advices on thesis and research related matters, encouragement and motivation from Prof Frada had many times comforted and inspired me to keep up with the challenges of this PhD journey. To Prof Sue, I would like to thank her for all the support given when she was the Associate Dean. There was time when things went complicated and Prof Sue readily looked into the matters, devised solutions and offered assistance. The support given had enabled me to continue my PhD study without much worries. These dedicated individuals as mentioned above, had spent much of their valuable time and efforts to support me along my PhD journey. My gratitude is beyond words.

This research was supported by the Exploratory Research Grant Scheme from the Ministry of Education Malaysia, within the period of February 2015 to January 2016.

List of Tables

<i>Table 1.1: A comparison between several technologies for navigation and wayfinding.</i>	<i>20</i>
<i>Table 2.1: Type of threat or object classification or detection to assist blind navigation from several authors based on depth or computer vision sensing technologies.</i>	<i>57</i>
<i>Table 3.1: Additional details of the proposed research process model illustrated in Figure 3.4.</i>	<i>77</i>
<i>Table 4.1: Distribution of samples from several locations in Petaling Jaya and Kuala Lumpur.</i>	<i>94</i>
<i>Table 4.2: The data was captured using the following configurations.</i>	<i>98</i>
<i>Table 5.1: A disparity mapping technique in two steps used in this research.</i>	<i>122</i>
<i>Table 5.2: Configuration of the best architecture of the proposed tri-channel-single-input CNN.</i>	<i>139</i>
<i>Table 5.3: Items, parameters and ranges used in the model tuning.</i>	<i>147</i>
<i>Table 6.1: Top 10 architectures of model based on AlexNet.</i>	<i>152</i>
<i>Table 6.2: Top 10 architectures of model based on ZF Net.</i>	<i>154</i>
<i>Table 6.3: Selected examples of M3-x model based on VGG Net.</i>	<i>156</i>
<i>Table 6.4: Evaluation samples from several locations around Klang Valley, Selangor, Malaysia.</i>	<i>160</i>
<i>Table 6.5: Number of test units acquired for each class, and the samples used in the analysis with their classification accuracies.</i>	<i>162</i>
<i>Table 6.6: Confusion matrix of the classification from the field evaluation.</i>	<i>162</i>
<i>Table 6.7: The average precision, recall and F1-score of the 9 classes from the field evaluation.</i>	<i>164</i>
<i>Table 6.8: The 2 locations selected for unbiased evaluation in which data were never collected from.</i>	<i>165</i>
<i>Table 6.9: Number of test units acquired for each class, and the samples used in the analysis with their classification accuracies for the unbiased evaluation.</i>	<i>165</i>
<i>Table 6.10: Confusion matrix of the unbiased evaluation.</i>	<i>166</i>
<i>Table 6.11: The precision, recall and F1-score of the 9 classes for the unbiased evaluation.</i>	<i>166</i>
<i>Table 6.12: The optimized versus the original parameters of the proposed CNN.</i>	<i>168</i>

<i>Table 6.13: Average time taken to drain the power over 5 attempts of evaluation based on 3 different cases of application status.....</i>	<i>171</i>
--	------------

List of Figures

<i>Figure 1.1: The author's illustration of surface discontinuities faced by the blind and low vision people, based on several samples found in Malaysia.</i>	2
<i>Figure 1.2: Samples of surface discontinuity taken from several urban areas within Malaysia.</i>	3
<i>Figure 1.3: Classification of visual impairment by the ICD-10 (2010)</i>	4
<i>Figure 1.4: An illustration of a typical tele-assistive system with video camera and GPS receiver, guided by a sighted operator (Hunaiti et al., 2006).</i>	18
<i>Figure 1.5: The breakdown of the main challenges faced by the BLVs, with a focus on the mobility issues. Components in light green are subjects of focus in this research.</i>	27
<i>Figure 1.6: Outline of the research.</i>	33
<i>Figure 2.1: The Medical Model of Disability (adapted from Rieser, R., 2014).</i>	35
<i>Figure 2.2: The Social Model of Disability (adapted from Rieser, R., 2014)</i>	36
<i>Figure 2.3: Four early models of the laser cane, in ascending order: C-2, C-3 and C-4 (Benjamin et al., 1973).</i>	40
<i>Figure 2.4: The C-5 laser cane (Benjamin et al., 1973).</i>	40
<i>Figure 2.5: The protection zone of C-5 laser cane (Benjamin et al., 1973).</i>	41
<i>Figure 2.6: The virtual guide cane by Yuan and Manduchi (2005).</i>	42
<i>Figure 2.7: The assistive walker from Yokota et al. (2013a) and Yokota et al. (2013b).</i>	43
<i>Figure 2.8: The Navbelt, a wearable gadget for obstacles detection by Shoval et al., 2003.</i>	45
<i>Figure 2.9: An illustration of the ultrasonic based GuideCane from Borenstein and Ulrich (1997).</i>	45
<i>Figure 2.10: A depth sensing technology proposed by Takizawa et al. (2012) and Takizawa et al. (2013).</i>	47
<i>Figure 2.11: The RFID based navigation guide robot (right image) from Kulyukin et al. (2006), and the prototype robot chassis (left image) from Yelamarthi et al. (2010).</i>	51
<i>Figure 2.12: The neural networks based robotic guidance by Capi (2012).</i>	52
<i>Figure 2.13: The functional block diagram of Yuanlong and Mincheol (2014)'s guide-dog robotic system.</i>	53

<i>Figure 2.14: A sample of zebra crossing edge detection image from the work of Yuanlong and Mincheol (2014).</i>	54
<i>Figure 2.15: A four-legged guide-dog robot developed by NSK (2011) based on depth sensing technology, capable of obstacles and steps recognition.</i>	55
<i>Figure 3.1: Research design framework adapted from Cecez-Kecmanovic (2011) cited in Weber (2017)</i>	65
<i>Figure 3.2: DSR knowledge contribution framework adapted from Gregor and Hevner (2013).</i>	68
<i>Figure 3.3: Research outputs diagram adapted from Purao (2002).</i>	71
<i>Figure 3.4: Research process model adapted from Mettler’s model with process steps and outputs suggested by Vaishnavi and Kuechler’s mapped on both left and right sides.</i>	76
<i>Figure 3.5: A sample (let’s name it sample A) of steps from three different angles.</i>	81
<i>Figure 3.6: The 3D model and printed replica of Sample A. The 3D printed replica is proportionately scaled down from the actual measurements taken from the sample</i>	81
<i>Figure 3.7: Another sample (let’s call it Sample B) from two different angles.</i>	82
<i>Figure 3.8: The 3D model and printed replica of Sample B.</i>	82
<i>Figure 3.9: Another sample (let’s call it Sample C) from two different angles.</i>	83
<i>Figure 3.10: The 3D model and printed replica of Sample C.</i>	83
<i>Figure 4.1: Structure diagram of the phase-1 prototype.</i>	85
<i>Figure 4.2: Flow chart of the algorithm for the application built for data collection in the field.</i>	89
<i>Figure 4.3: The camera angle was set at 35 degrees facing downward after several experiments.</i>	90
<i>Figure 4.4: The setup and positioning of the prototype.</i>	91
<i>Figure 4.5: Region of interest and distances based on the prototype setup.</i>	92
<i>Figure 4.6: The ROI in red rectangular outline is centered at the lower half position of Sample 95 – Left Image 141, a typical example of down-steps.</i>	92
<i>Figure 4.7: Location map of sample distribution (red dots) around Petaling Jaya and Kuala Lumpur.</i>	94
<i>Figure 4.8: A walkway in front of a bank leading to a car park and road. (Location: Up-Town PJ). The dimensions can be seen labelled in centimeters.</i>	95

Figure 4.9: At one end of a walkway in front of a print shop leading to a road. (Location: Damansara Jaya)	96
Figure 4.10: A pit along a walkway connecting several private services, uncovered and hazardous. (Location: Kota Damansara).....	96
Figure 4.11: Damaged and uneven steps bridging a drainage between the road and walkway along several eateries. (Location: Damansara Jaya)	97
Figure 4.12: Step with drop-offs on both sides bridging a drainage and leading to some shops. (Location: Damansara Jaya).....	97
Figure 4.13: The left and right images of a sample of uneven steps ahead of a walkway.....	99
Figure 4.14: The left and right images of a sample of mix gradient (steps and uncovered drainage). ..	99
Figure 4.15: The left and right images of a sample of uncovered drainage next to a walkway.	99
Figure 4.16: A fix-pattern grey and white noise forming a large area of bended bands in some images.	101
Figure 4.17: An unwanted scene captured at the beginning of a recording.	102
Figure 4.18: Body parts especially the hand or arm could easily become an unwanted object or occlusion in the scene during data collection.....	102
Figure 4.19: Example of a raw image (left) and its normalized version (right) using Z-score. The steps at the centre of the image can be seen clearly in the original version, but are less obvious in the normalized version.....	104
Figure 4.20: Edge detection using Sobel technique on the original image. The edges of the step at the centre of the image are clearly visible as circled in red.....	105
Figure 4.21: As compared to Figure 4.20, the edges of the step at the centre of the image after simple rescaling are not visible. The elevated part of the ramp is also missing.	105
Figure 4.22: Three sample images before PCA whitening.	107
Figure 4.23: Three similar sample images from Figure 4.22 after PCA whitening.....	107
Figure 4.24: The taxonomy of surface discontinuities.....	109
Figure 4.25: Image pair on top is the original version, after the augmentation by horizontally flipping, the bottom image pair was yielded.	111
Figure 5.1: A structure diagram illustrating the proposed system, with the Sensor Module, Processing Unit, User Interface and Output.	114

Figure 5.2: Epipolar rectification of image pair, (a) original image pair, (b) rectified image pair (Gosta and Grgic, 2010).....	116
Figure 5.3: A diagram illustrating the equivalent triangles (top), and the dimensions of the stereo camera (bottom) used in the prototype.....	121
Figure 5.4: A grey scale image pair of Sample 28 (an uncovered drainage).	123
Figure 5.5: A red-cyan composite view of the image pair from Figure 5.4.	123
Figure 5.6: A disparity map from the image pair in Figure 5.4.	124
Figure 5.7: An issue of region with uniform intensity, in which the floor on the walkway can be seen dusted with black dots.	125
Figure 5.8: Another example of region with uniform intensity, in which the floor on the walkway can be seen dusted with black dots.....	125
Figure 5.9: An anaglyph of Sample 97, showing a drop-off at the end of a walkway.	126
Figure 5.10: At disparity range of (0, 16), the drop-off and the road are closely coloured in white.	126
Figure 5.11: At disparity range of (2, 18), the drop-off and road are more distinguishable than before.	127
Figure 5.12: Examples of different window sizes ranging from 5 to 21 units. Their accuracies and blurring effects can be seen here.....	129
Figure 5.13: Example of a disparity map computed from a stereo image pair (from sample 188, image pair of sequence 65).	130
Figure 5.14: A comparison between sum of difference functions SSD and SAD.	131
Figure 5.15: Image with an uncovered drainage ahead (top), its disparity map (middle) and the visualization of its depth profile (bottom), sourced from Leong et al. (2017).	134
Figure 5.16: Architecture of the DBN used in the experiment (Leong et al., 2017).	134
Figure 5.17: The tri-channel fusion of left and right image with disparity map to create a single fused input with three channels like a typical digital coloured image of RGB-channel.....	137
Figure 5.18: An illustration of the full architecture of the proposed tri-channel convolutional neural network for surface discontinuity classification. This architecture has the best accuracy on the testing set.....	140
Figure 6.1: An illustration of the architecture of AlexNet (Krizhevsky et al., 2012b).....	151

<i>Figure 6.2: ZF Net architecture developed by Zeiler and Fergus (2013).</i>	<i>153</i>
<i>Figure 6.3: Adjustable blocks set for the lab test can be used to resemble stairs, rises, drop-offs, curbs and other forms of surface discontinuity.....</i>	<i>159</i>
<i>Figure 6.4: Percentage of power status over time as indicated by a power status application running on the prototype for the three test cases.</i>	<i>172</i>

List of Abbreviations

BLV	Blind and low vision individual
CNN	Convolutional neural network
DBN	Deep belief network
DSR	Design science research
ETA	Electronic travel aid
GIS	Geographic Information System
GPS	Global Positioning System
HCI	Human-computer interaction
ICD-10	the International Classification of Diseases 10 th Revision
O&M	Orientation and mobility
PCA	Principal component analysis
RFID	Radio-frequency identification
ROI	Region of interest
SAD	Sum of absolute differences
SVD	Singular value decomposition
UI	User interface
WHO	the World Health Organization
2D	Two-dimensional
3D	Three-dimensional

Chapter 1: Introduction

The chapter begins with an overview of the research context (Section 1.1), and the next 4 sections (Section 1.2 to 1.5) further define and establish relevant understanding about the subjects of the research. With the background established, the research motivation (Section 1.6) and problem statement (Section 1.7) are then elaborated. The remaining sections describe the objectives and research questions (Section 1.8), the significance and expected contributions (Section 1.9), and lastly the thesis outline (Section 1.10).

1.1 Overview

According to World Health Organization, there is an estimated 285 million blind and low vision (BLV) people worldwide (Pascolini and Mariotti, 2011). Apart from difficulties in accessing visual information (Lennie and Hemel, 2002), one of the greatest challenges to independence for the BLVs is difficulties in self-navigation (Golledge, 1993). To achieve safe and effective navigation, the BLVs need to access global information relevant to orientation and positioning on one hand, and deal with local threats along their pathway on the other hand.

Through consultations with several blind and low vision service providers (namely St. Nicholas' Home Penang, Dialogue in the Dark Malaysia and Vision Australia), it is realized that there is a diverse range of navigational challenges faced by the BLVs. Negotiating surface discontinuities is one of the very prominent challenges when navigating an outdoor environment within urban areas. Surface discontinuities are commonly formed by rises and drop-offs along a pathway (Geruschat and Smith, 2010). The rises and drop-offs change the gradient of the navigational surface. They could be a threat to balancing during a walk, and

perceiving such a threat is a difficult challenge to the BLVs without some proper aids (Kuyk et al., 2004, Goodrich and Ludt, 2002).

If the presence of universal access facilities such as tactile ground surface indicators, handrails along staircases, pedestrian ramps, subways, properly covered drainages and et cetera is a sign of equipping the BLVs with better built environment, it is a little-known reality that even in some modern cities, most of such facilities are often only available around limited public transportation infrastructures, certain well-planned urban landscapes and some government or private properties (Hussein and Mohd. Yaacob, 2013).

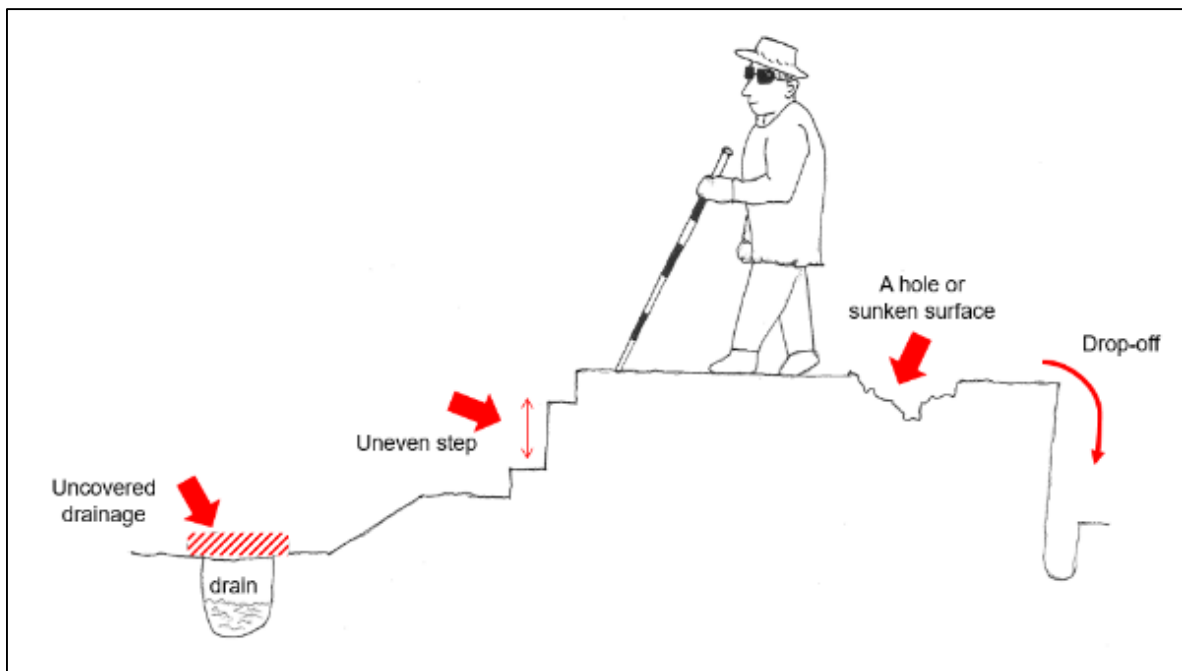


Figure 1.1: The author's illustration of surface discontinuities faced by the blind and low vision people, based on several samples found in Malaysia.

Additionally, traditional navigation aids such as a guide cane might not always be helpful when the built environment does not comply with BLV accessibility design standards. Uneven staircases, steep drop-offs, uncovered drainage, absence of terminal posts between junctions, stairways without proper handrails and other types of surface discontinuity are common hazards to the BLVs due to poor enforcement or implementation of regulations

for accessible design. This issue is especially prominent in low to middle income countries. *Figure 1.1* illustrates some examples of surface discontinuities.

Some surface discontinuities such as staircases, small steps, joints between walkways and curbs are needed for the continuity of the navigation; although some of them might be hazardous due to unregulated circumstances, for example uneven steps with no handrail, uneven heights of rises along a stairway, uncovered drainage and sudden drop-offs. *Figure 1.2* shows some surface discontinuities found at several urban areas within Malaysia. Hazardous conditions are indicated by the red arrows, picture A: steep step, B and C: blended gradients, D and E: uneven steps in between uncovered drainage, F: high altitude drop-off beside staircase without railing.



Figure 1.2: Samples of surface discontinuity taken from several urban areas within Malaysia.

Based on literature, most of the traditional assistive tools or recent technologies on walkway surface threats negotiation have predominantly focused on solving the problem of obstacles (either static objects such as lamp posts and curbs, or dynamic objects such as cars and pedestrians) or stairs detection. However, when dealing with surface discontinuities at an outdoor urban area, there is an even richer taxonomy of surface conditions along the pathway which has not been tackled before. Focusing on the issue of richer taxonomy of surface discontinuity at the outdoor of urban areas, this thesis presents the design and development of a prototype of a wearable technology-based system to assist the BLVs during navigation.

1.2 Defining the Blind and Low Vision

The International Classification of Diseases 10th Revision (ICD-10) classifies the types of visual impairment as shown in *Figure 1.3*.

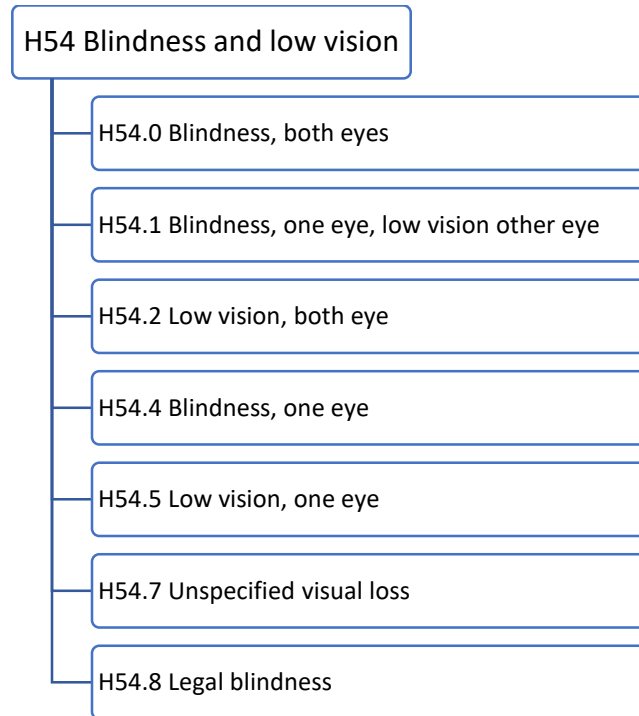


Figure 1.3: Classification of visual impairment by the ICD-10 (2010)

Based on ICD-10, the term “low vision” is inclusive of those with moderate and severe visual impairment, while “blindness” is referred to those with total blindness in either both eyes or a single eye. Some of the individuals might exhibit blindness in one eye and low vision in the other eye, which is a mixture of blindness and low vision. In brief, low vision taken together with blindness represents all visual impairment.

The design and development of a prototype in this research would focus on its applicability for individuals with blindness or low vision in both eyes, as well as individuals with a mixture of both conditions. Thenceforth, the term “blind and low vision” (or BLV) in the context of this research refers to the categories H54.0, H54.1 and H54.2 based on the classification by ICD-10 as listed in *Figure 1.3*.

With an estimated 285 million visually impaired people around the globe, it is undoubtedly important to consider their mobility needs, and offer possible solutions to overcome this challenge. Having defined the blind and low vision categories, next it is also important to take a closer look at the main navigational challenges faced by the BLVs especially when dealing with the physical world in their daily activities.

1.3 Navigational Challenges Faced by the Blind and Low Vision People

To perform their personal and social functions, people must interact with the physical environment. Activities such as accessing services, retrieving information and getting to places are just some of the basic functions people carry out daily. However, societal infrastructure such as walkways, signages, roads and traffic systems, bus stations, hospitals, schools, tertiary facilities, community and sport facilities, government facilities and other ancillary infrastructure are typically designed for sighted people. Sighted people can use their vision to appreciate the physical environment or avoid hazardous encounters. On the other hand, although the BLVs could be quite functional in their daily activities, they often

have to rely on assistive tools, or be dependent on a sighted person to perform some of the mentioned functions.

Two of the greatest challenges to independence for the BLVs are accessibility of visual material or information (Lennie and Hemel, 2002), and difficulties in self-navigation (Golledge, 1993). With advances in technology and cheaper cost of manufacture in recent decades, numerous tools have been created to assist the visually impaired mainly in information accessing. For instance, the Braille translator, Braille printer (or embosser), screen reader, speech synthesizer, optical character recognition system, digital talking book, tactile map, screen magnifier, personal digital assistant, as well as the more recent invention of refreshable Braille display and accessible mobile apps are just some examples of the pervasion of assistive technologies in accessing the world of information. These technologies are getting ubiquitous and inexpensive with each passing year, and thus they are more readily available to the BLVs.

In contrast, similar advances have not been quite so abundant in dealing with self-navigation (Giudice and Legge, 2008) because it is a very hard and complex task. Technologies for blind navigation are sporadic and insufficiently developed. Lack of safe and independent mobility is claimed to be the most significant factor depriving the BLVs of living a normal life (Moore, 2000). The BLV society has traditionally remediated their self-navigation by reducing their movement to unfamiliar places. This practice has further limited their social functions. As posted by the award winning article – *Challenges and Opportunities Facing Visually Impaired Persons* – prepared by a visually impaired author, Beaudin (2011):

‘One hundred years ago, being visually impaired meant being condemned to a life of confinement and institutionalization...’ (para. 1)

The assertion of “being condemned to a life of confinement” indicates the accessibility and mobility challenges experienced by the BLVs in the past. Be it hundred years ago or even

today, this is still one of the biggest challenges to the BLVs (Golledge, 1993, Salminen and Karhula, 2014).

Those with complete blindness or low vision often have problems self-navigating beyond well-known environments and thus confine themselves to only some limited places. Simple activities that sighted people take for granted such as strolling freely in a mall or walking around an office aisle might pose great difficulty to the BLVs without proper facilities or navigation aids.

Based on Beck (2010), even within their own home, the BLVs must learn or remember every detail about the home environment. Large furniture such as sofas and tables must remain in a specific location to prevent injury. As such, if a BLV person lives with others, each member of the house must keep the walkway clear and constantly keep all items or furniture in their original location. Conclusively, moving around a familiar place could be a challenge for the BLVs, let alone navigating some unfamiliar outdoor urban environment.

1.4 A Brief Understanding of Navigation and Threats along a Pathway

Flat and smooth navigational pathways such as streets, roads, pedestrian walkways and aisles can be seen to resemble a two-dimensional plane. However, the real world is not a simple two-dimensional (2D) horizontal plane. The three-dimensional (3D) characteristic becomes more obvious when the vertical component is presented around or along the pathway, e.g. a stairway leading up to another level, or a car blocking the road and so forth. For anyone to navigate in a 3D space, position, orientation and direction are the three basic sets of information a person needs to be aware of (Jeffery et al., 2013, Schiesser, 1986). To utilize this information, it requires an egocentric frame of reference for navigation to work correctly i.e. a frame from the point of view of the person; as opposed to a geocentric frame of reference i.e. a frame from the point of view of a global coordinate system. To specify a navigation, a world-anchored space-defining frame of reference with respect to the

position, orientation and movement direction is needed. Eventually, for the navigation to happen, there also needs to be a system of measurement and coordination that aligns distances and directions in relation to the frame of reference (Jeffery et al., 2013). All the above indicates that navigation is indeed a very complex task even for a sighted person, let alone for a BLV.

1.4.1 The Cues

It is important to point out that while a person navigates through places, one continuously gathers data from the surroundings to update the position, orientation and direction information. The surrounding environment has always been rich with visual data. The primary data comes in the form of landmarks, routes and other ground cues. Landmarks are readily identifiable features of the surrounding landscape and they are normally used as frames of references. Identifying landmarks and recalling one's location in reference to a destination is a fundamental aspect of navigation (Long and Giudice, 2010). Janzen and Jansen (2010) recognized that objects along a route could become crucial landmarks that facilitate successful navigation. Thus, visually impaired individuals unable to see distant iconic landmarks (i.e. buildings and the like) may count on reachable smaller objects along the route. They proceed with these small-scale landmarks one after another in the pursuit of getting to the destination (Fryer et al., 2013).

Besides landmarks, some people are familiar with routes between one location and another, or sometimes between one location and several others. By relying on the known routes, a person can establish new routes by connecting them or, based on their intersections, by navigating to new places where new routes have not yet been learned (McKnight et al., 1993). In addition to landmarks and routes, other commonly used navigation cues are street signboards, terrain or topography, direction information offered by local inhabitants, smell and sound (M. Beatrice Dias et al., 2015).

1.4.2 The Threats

By combining the information from landmarks, routes and other ground cues, navigation becomes possible as one can determine one's position, orientation and directional information. Apart from that, for safe and smooth navigation, one needs to deal with local threats along the pathway. Around social structures, threats can be dynamic (such as pedestrians, animals or pets and moving vehicles) or static (such as lamp posts, rises, drop-offs, partitions, slippery surfaces, drainages and curbs).

Without threats, navigation tasks would be much easier. However, this is not often the case as threats come in a variety of forms and numbers along different pathways, and they are potentially dangerous to navigation. Therefore, people developed strategies to deal with the threats. Based on Joh and Adolph (2006), safe walking requires a person to detect upcoming threats to his / her balance, select strategies for dealing with the threats, and modify these strategies as needed while negotiating the threats.

Most sighted people have little concern about threats negotiation as they unconsciously learn to pick up routes with lesser obstacles (Giudice and Legge, 2008). The BLV individual has to rely on navigation aids such as a white cane or guide dog to exhaustively detect and avoid obstacles (M. Beatrice Dias et al., 2015). Based on Barth and Foulke (1979), some BLV individuals have learnt to utilize reflected sound to gain an understanding of non-visual spatial perception and it enables the traveller to avoid large obstacles before making contact. Typically, an assistive wayfinding system has little to offer in dealing with local threats, as the purpose of most wayfinding systems is to provide information on getting from one point to another, although some of the assistive navigation systems have included obstacles detection and avoidance functions to guide the users through the pathways.

Threats can be hazardous or neutral to a person in navigation. In either case, the standard mechanism is detection and avoidance. Sometimes, there are more complex threats needing further negotiation, for example changes of elevation (or surface discontinuities)

along the pathways. Surface discontinuities are inclusive of rises and drop-offs (Geruschat and Smith, 2010). Changes in the level of navigation surface are common threats to balance, and perceiving these changes is a difficult challenge to BLV pedestrians (Kuyk et al., 2004, Goodrich and Ludt, 2002). Surface discontinuities are considered complex threats because the solution to them might not always be detecting and avoiding, as some surface discontinuities are the only ways to continue with the navigation, as they are integral parts of the navigational pathway.

Based on the nature of a surface, a pedestrian must decide whether to proceed or to take a detour. For example, some surface discontinuities such as staircases, small steps, joints between walkways and curbs are needed for the continuity of the navigation, although some of them might be hazardous due to unregulated circumstances, for example uneven steps with no handrail, uneven heights of rises along a stairway and a sudden drop-off at a supposedly flat surface. As such, the strategies mentioned by Joh and Adolph (2006) would be much more appropriate in dealing with surface discontinuities. Due to the difficulty of detecting surface discontinuities without visual perception, the BLV pedestrians are being trained by the Orientation and Mobility (O&M) specialists to look for reachable features of the environment that are related with elevation changes (Geruschat and Smith, 2010). For example, handrails are associated with stairs and street signs are associated with curbs.

Together with position, orientation and directional information, successfully dealing with potential threats may add a safety aspect to navigation and bring about independent mobility for the BLVs.

1.5 Common Tools for Navigation by the BLVs

Safe and efficient navigation has profound importance to the BLVs, as it might be the prerequisite for getting one to achieve many social functions. To navigate independently in a constantly changing environment, utilizing the surrounding cues and negotiating

potential threats are crucial for them. Besides seeking help from a sighted person, there are several commonly used approaches to assist independent navigation. The following sections briefly discuss some of the commonly used assistive tools from two main distinct ranges:

- (1) the personal assistive tools such as the guide cane, guide dog, electronic travel aids; and
- (2) the integrated assistive infrastructure such as Braille signages, tactile surfaces, accessible pathways and other barrier-free designs.

In most cases, some of the assistive tools are used jointly to achieve a better result.

1.5.1 Personal Assistive Tools

Personal assistive tools refer to personal, ubiquitous, mobile and/or small devices or mechanisms that can be carried or tagged along by the BLVs for aiding them in navigation. The following sections summarize several tools commonly used by the BLVs.

Guide Cane

The guide cane, generally known as the white cane, is the most widely used tool for mobility. The cane is a handy tool that is traditionally used for detecting and identifying obstacles, looking for elevations or drop-offs in the path of travel, or as a symbolic indicator to sighted people that the carrier is a BLV. Even though the guide cane has stood the test of time and remains the most popular choice, it was not used systematically until the twentieth century (Bledsoe, 1997). To properly utilize the guide cane for actual navigation, the user has to go through some training by an Orientation and Mobility (O&M) trainer (Geruschat and Smith, 2010). Typically, the guide cane user has to make contact with the surface, edges, walls or shorelines and follow them to proceed walking. As such, it is acknowledged that the cane is most effective only for proximal cues within an approximation of 1 meter of distance (Barth and Foulke, 1979). In addition to the physical contact made by the cane to the surface along the walkway, some BLVs also rely on the echo cues created by tapping the cane on the walkway or obstacles (Schenkman, 1986).

Guide Dog

The guide dog, also known as seeing-eye dog, performs functions mostly similar to those of the guide cane with additional efficiency due to the dog's capability to lead the user through direct routes between obstacles. Based on a study by Wiggett-Barnard and Steel (2008), a guide dog is able to offer safer, faster and more dynamic mobility aid than a guide cane or residual sight of the person. In the same study, the guide dog is also deemed to give more mobility confidence to the user. An earlier study by Blasch and Stuckey (1995) similarly supported the claim. Further elaboration on the guide dog is given in Section 1.6.1.

Electronic Travel Aids (ETAs)

An ETA is an electronic assistive device comprising sensors, an information processing unit and a communication interface, whereby information is relayed to the user for enhancing independent navigation. ETAs for the BLVs typically operate using audio and haptic signals as the communication medium. As the ETA belongs to the class of assistive tools closely related to the proposed prototype of this research, a detailed literature review about the ETA is given in Chapter 2.

1.5.2 Integrated Assistive Infrastructures

Integrated assistive infrastructures refer to mechanisms installed within the building, landscape, urban environment and other types of public or private property to assist the impaired individuals, to provide universal access to the users.

Accessible Barrier-free Design

An accessible pathway can be regarded as one designed with embedded features in the pathway to facilitate safe and convenient accessibility by disabled people. There are numerous details governing the design of accessible pathways based on the regulations of the country, state or local government. Often, tactile surface is one of the main parts of accessible design. Major accessible designs are centred on the pathways of public or private properties with the inclusion of disabled friendly elements such as sidewalks, curb ramps,

way-finding signages, handrails and terminal posts. Another consideration is given to the construction of pathway surfaces, in which the surface should be smooth, continuous, non-slippery and even. The intersection of two pathways is normally blended at one common level to reduce gradients or drop-offs. The fundamental idea of barrier-free design is to manipulate or alter the built environment such that it is universal to all users.

Tactile Surface

Among the integrated assistive infrastructures, most work has been done in the tactile field (Passini and Proulx, 1988). Tactile surface implementation can offer dozens of applications. It may serve as an indicator for hazardous encounters at places like the entry point of an escalator, descending stairs and the edge of a walkway. It may be an aid to guide the BLVs along a route. It is also specifically devised to inform the BLV pedestrian about a particular change along the walkway such as an arrival at a junction, a terminal point for transportation, or a zebra crossing. Based on Passini and Proulx (1988) from the review of the works of both Sanford (1985) and Steinfeld (1986), they tested various materials and textures to assess their detection threshold. The best result was attained when the tactile cues were combined with a change in the perceived elasticity and reverberation of the material.

Braille Signage

The Braille signage is an informative tactile signboard printed in Braille writing system. Modern cities have commonly included Braille signages around public or commercial infrastructures such as the operating buttons within a lift, room numbers, wash room signages, direction maps at building entrances, street signboards, entrance/exit interfaces and several others. With Braille signs available along the pathway, the BLVs might utilize them as some added information to further assist in navigation.

1.6 Motivation of the Research

Based on Section 1.5, it appears that at present there are choices of navigational aids for the BLVs to help themselves in traversing the physical world independently. However, why is independent navigation still such a challenging and stressful task for the BLVs despite the availability of navigational aids? Some understanding about the factors affecting blind navigation, current assistive technology, and regulations for accessible built environment could offer some justifications to the motivation of this research.

1.6.1 Factors Affecting Blind Navigation

There are several prominent factors affecting the BLVs in acquiring safe and effective navigation. Firstly, the non-visual spatial inference is a very challenging skill to master. With vision (especially stereo vision), the sighted person can acquire distance and direction information to surrounding landmarks and guide one's route. In contrast, the visually impaired person relying on tactile, auditory or olfactory signals may infer less information than vision signals in terms of one's relative motion, environmental layout, distance and direction cues (Strelow, 1985, Thinus-Blanc and Gaunet, 1997). This information discrepancy is especially notable in individuals who lack visual experience in early state according to a study by Loomis et al. (1993).

Secondly, there is a steep learning curve for the BLVs to non-visual navigation. They need to learn to detect obstacles on their pathway, identify curbs and steps, and even avoid hazards. They must learn to interpret their position and orientation relative to the environment and their destination. To acquire these skills, they must go through orientation and mobility (O&M) training by the specialists. These tasks are cognitively much more challenging as compared to visual navigation by sighted people (Rieser et al., 1982, Zimmerman, 2007).

Thirdly, there are shortcomings faced by the conventional navigation aids. Quite often, most of the navigation aids are only able to offer better mobility result when used jointly –

for example, the guide cane when used at an accessible pathway offers safer and more efficient mobility. In other words, these conventional navigation aids exhibit different degrees of dependence on other types of support. Unfortunately, despite the various choices of these aids mentioned, not all of them are readily available to the BLVs.

Besides, some of the aids are known to have some significant limitations. For instance, both the white cane and guide dog have similar limitations in detecting overhanging or above ground level obstacles. In addition, although both aids provide effective proximal cues, they do not communicate much about position and orientation information to the users. In other words, the users are not provided with the heading or facing direction in regards to their environment (Giudice and Legge, 2008). Thus, sometimes a Global Positioning System (GPS) based ETA is used as a complement to the white cane or guide dog to provide additional position and orientation information.

As for the case of the guide dog, to successfully train one into a fully working assistant incurs high cost and is a challenging task (Tomkins et al., 2011). The dog has to go through several temperament tests and kennel behaviour assessments before it is gradually trained into a guide dog. The training takes nearly two years to develop the pup into a responsible guide dog. Eventually the trained guide dog must be matched with a potential BLV handler through some programmes. There is a cost for the maintenance and responsibility for the dog's needs and welfare. At the end, the guide dog will retire from its service after an average of five to six years. In some cases, the guide dog owners exhibit distress due to the end of the partnership (Nicholson et al., 1995).

Moving away from the guide dog, other navigation aids such as the implementation of Braille signages, tactile surfaces, barrier-free accessible designs and other types of alteration to the environment are beyond the control of the BLVs. Their availability is very much dependent on the implementation by the local governments or private developers. These types of navigation aids are part of the built environment and they are more common in

the developed or high-income countries as compared to the underdeveloped or low to middle income countries.

Finally, the adoption of navigation technology is not as pervasive as expected. As pointed out by Giudice and Legge (2008), several issues concerning the adoption are: (1) there has been a lack of both development and acceptance of technologies to aid blind navigation, (2) Williams et al. (2013) revealed that most BLVs are concerned about exposing expensive technological devices in public, (3) more emphasis needs to be given to both the perceptual factors and end-user needs on the developed technologies (Giudice and Legge, 2008), and finally (4) the age factor is another cause for concern in adopting new technologies.

Knowing the multiple factors affecting blind navigation, more tools or technologies were built by further exploring or investigating these factors in the hope of solving some of the underlying issues. Section 1.6.2 provides a summary of these technologies, highlighting and detailing issues being tackled, and shedding light on the technological gap that would motivate this research.

1.6.2 Current Technologies and the Gaps

In order to move around their environment, people perform two main types of physical movement – wayfinding and navigation (Karimi, 2015). Technically these two terms refer to different modes of movement. Having known the shortcomings faced by the conventional mobility aids, various assistive technologies were developed to assist the BLVs. Most BLVs' navigation or wayfinding technologies are tele-assistive, GPS and/or GIS based systems, in which they assist mainly in outdoor mobility needs. Some researchers have introduced position-based localization models to assist navigation or wayfinding within a known indoor environment based on radio-frequency identification (RFID), Bluetooth, pedometer, laser range finder, ultrasonic and several other types of sensor

technologies. The following sections summarize these systems in brief and eventually highlight the gap that motivated this research.

Tele-Assistive, GPS and GIS based Systems

Traditionally, a sighted person is the most reliable companion a BLV individual can count on for the purpose of navigation and wayfinding. Therefore, some technologies were developed to include the sighted guider to arrive at some models of tele-assistive systems. As presented in the works of Hunaiti et al. (2006), Bujacz et al. (2008), Venard et al. (2009) and Baranski et al. (2010) the fundamental idea of most tele-assistive systems is to connect a remote guider (either a sighted operator or a centralized server) to the device carried by the BLV user, with a real-time mechanism to record and transmit the current environmental information, such that the guider would offer a verbal description of the environment as well as directional guidance to the user. Both a GPS receiver and a mobile device with audio communication interface are the common technologies within the mentioned tele-assistive systems.

On the other hand, the work by Venard et al. (2009) known as RAMPE has focused on the objective to provide real-time information to increase mobility and autonomy of the BLV in public transportation. Since the public transport location could partly be an indoor environment (e.g. within a bus or a terminal building), their system has included a central server to retrieve GPS coordinates of the public transport vehicles, and communicate through a Wi-Fi enabled Personal Digital Assistant (PDA) held by the user from the base station. Mata et al. (2011) developed a Bluetooth and GPS based mobile assistive system with a similar purpose as the RAMPE for a Metrobus station. Their system consists of a smartphone, GPS receiver and electronic compass, all of which communicate through Bluetooth. With a menu on the smartphone for the user to browse through by audible interface, the system helps to provide location and orientation information to the user.

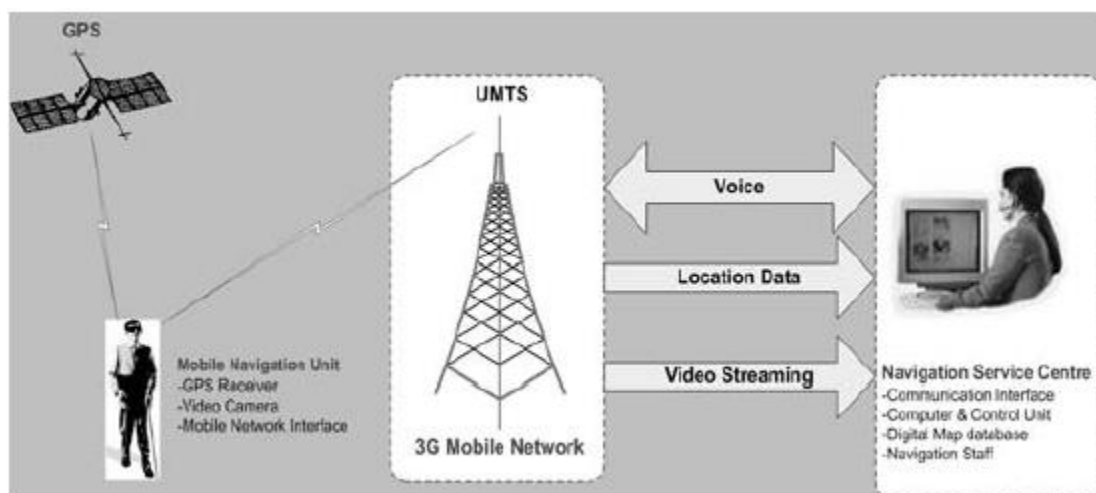


Figure 1.4: An illustration of a typical tele-assistive system with video camera and GPS receiver, guided by a sighted operator (Hunaiti et al., 2006).

Tapping on the advances of the current smartphone technology, Stepnowski et al. (2011) developed an app for point-to-point navigation in urban areas utilizing Android's text-to-speech function. Known as Voice Maps, the app can generate voice messages to continuously monitor the user's direction and position. Sánchez and Torre (2010) developed a similar smartphone-based system that utilizes the text-to-speech and GPS technology from within the smartphone to assist the BLVs in both known and unknown environments. The system provides directional information based on a clockwise metaphor, by assuming that the user is by default facing 12 o'clock, and the turning direction is given based on this orientation – this is indeed a user-friendly element of the system.

A geographic information system (GIS) is used to provide accessibility information in the work of Fernandes et al. (2012). This GIS module is part of a bigger project called Smart Vision that involves several other modules to provide mobility aid to the BLVs. The GIS module, which is the emphasis of this paper, claims to be able to provide proximity information such as a street light or a zebra-crossing in the vicinity of the user. Another simpler GIS solution is presented by Serrão et al. (2015). Their system integrates GIS of a building with computer vision from a smartphone. Visual landmarks such as doors, stairs,

signs and fire extinguishers are used by the computer vision to trace and validate the route for the user's navigation.

Markers and Position-Based Systems

The ability to identify a landmark and map it to one's cognitive map within an environment is crucial for a BLV person. In some solutions, markers are installed at strategic places around the pathway for the assistive tools to detect and trace the navigational route. This type of solution is indeed a position-based system. The markers act as some fixed-point identifiers with known positions relative to the environment. With a relevant device capable of detecting these markers, the visually impaired user will then be guided around the environment. The most widely developed position-based system is based on the RFID technology. Some researchers have proposed visual markers, Bluetooth and several other technologies for similar purpose.

Several applications of RFID can be found in the works of Tatsumi et al. (2007), Liu et al. (2007), Seto (2009), Ganz et al. (2010), Ganz et al. (2012), Alghamdi et al. (2013) and Tsirmpas et al. (2015). In these RFID systems, the transmitter tags are placed at strategic points to act as location identifiers. The user is equipped with the receiver to pick up the signal from the nearest RFID tags such that the location of the user in relative to a known environment can be estimated. This eventually facilitates position-based navigation for the BLV user.

Liu et al. utilized RFID, Bluetooth and the readily available fluorescent light to arrive at an indoor guidance system. They proposed a fluorescent light communication (FLC) model to serve as an indoor global positioning module and developed software to compute the distance from the receiver to each effective fluorescent light. While the RFID is used to serve as a local positioning module, by adding the FLC module, it produces more precise location information.

Hesch and Roumeliotis (2007) presented a position updating guide cane installed with a laser range finder, a pedometer and a 3-axis gyroscope. This multi-sensor tool is capable of performing real-time localization in a known indoor environment. Gallagher et al. (2012) also developed a multi-sensor indoor positioning system on a smartphone platform, in which a Kalman filter is used to fuse the data of all the sensors available on the phone.

Summary

Table 1.1 summarizes the reviewed technologies for navigation and wayfinding to aid the BLVs in dealing with physical mobility.

Table 1.1: A comparison between several technologies for navigation and wayfinding.

Research	Platform/ Main Technologies	Global / local task	Environment	Needs of alteration to the intended environment	Mode of information provided to the user
Hunaiti et al. (2006)	Tele- assistance, GPS, GIS, 3G	Both global and local	Outdoor	Not needed	Verbal guidance from human
Bujacz et al. (2008)	Tele- assistance, GPS	Both global and local	Outdoor	Not needed	Verbal guidance from human
Venard et al. (2009)	GPS, Wi-Fi, PDA	Both global and local	Both indoor and outdoor	Wi-Fi needed	Verbal guidance from server
Baranski et al. (2010)	Tele- assistance, GPS	Both global and local	Outdoor	Not needed	Verbal guidance from human
Mata et al. (2011)	Smartphone, GPS, Bluetooth	Both global and local	Both indoor and outdoor	Bluetooth needed	Audio signal

Research	Platform/ Main Technologies	Global / local task	Environment	Needs of alteration to the intended environment	Mode of information provided to the user
Stepnowski et al. (2011)	Smartphone	Global	Outdoor	Not needed	Computer generated text-to- speech
Sánchez and Torre (2010)	Smartphone	Global	Outdoor	Not needed	Computer generated text-to- speech
Fernandes et al. (2012)	GIS, computer vision	Both global and local	Both outdoor and indoor	Position markers needed	Audio signal
Serrão et al. (2015)	Smartphone, GIS, computer vision	Local	Indoor	Not needed	Audio signal
Liu et al. (2007)	RFID, Bluetooth, FLC	Both global and local	Indoor	RFID tags and Bluetooth needed	Audio signal
Tatsumi et al. (2007)	RFID	Local	Indoor	RFID tags needed	Not available
Seto (2009)	RFID, computer vision (colour recognition)	Local	Indoor	Coloured lines and RFID tags needed	Vibration and audio signal
Ganz et al. (2010) & Ganz et al. (2012)	RFID, Bluetooth	Local	Indoor	RFID tags and Bluetooth needed	Audio signal
Alghamdi et al. (2013)	Long range active RFID	Local	Indoor	RFID tags needed	Audio signal
Tsirmpas et al. (2015)	RFID	Local	Indoor	RFID tags needed	Audio signal
Hesch and Roumeliotis (2007)	laser range finder, pedometer, gyroscope	Local	Both outdoor and indoor	Not needed	Audio signal

Research	Platform/ Main Technologies	Global / local task	Environment	Needs of alteration to the intended environment	Mode of information provided to the user
Gallagher et al. (2012)	Smartphone	Both global and local	Indoor	Not needed	Audio signal

From *Table 1.1*, it appears that the mentioned technologies can support either local, global or both positioning tasks. Some of these technologies are built for the outdoor environment while some are specific to the indoor environment, with a few being utilized for both environments. These technologies provide navigation or wayfinding aids by acquiring global information (i.e. routes, position, orientation and direction) or local ground cues (i.e. landmarks or location markers). From the table, there are more than half of these technologies that require alteration to the intended environment such as installation of Wi-Fi facility, markers or RFID tags. In most cases, this is beyond the control of the BLV users, and the needs of altering the environment could hinder the practicality of these technologies.

Despite the global information and local ground cues, most of these navigation or wayfinding technologies are lacking one crucial component – the local threats negotiation. This missing component is typically addressed by a different range of assistive technologies known specifically for threats negotiation. However, it was later realized that based on literature (Chapter 2), the assistive technologies for threats negotiation are largely developed for obstacles detection, neglecting some prominent forms of local threats negotiation, for instance, the negotiation of various types of surface discontinuity. This research is partly motivated by such a technological gap. To further add on to the research motivation, Section 1.6.3 provides another facet of the issue related to accessible built environment.

1.6.3 Poor Enforcement of Regulations for Accessible Built Environment

According to Imrie and Kumar (1998) and Christensen and Byrne (2014) the built environment is equally important in promoting accessibility for disabled people. Waddington (2009) further emphasized that having an accessible (barrier-free) built environment could promote free movement of goods and services to the disabled community in the European Union. In most developed countries, strict regulations and standards are set to ensure that infrastructures and the built environment are developed accordingly not only for the safety of the public, but also for the accessibility of disabled people. The terms ‘design for all’ or universal design are introduced to some national regulations to refer to the creation of a barrier-free built environment for the accessibility of everyone including disabled people. While in some low to middle income or underdeveloped countries, there are regulations detailing the built environment, quite often, the regulations serve as guidelines, rather than being enforced or implemented.

Accessibility within the Malaysian Context

As Malaysia is now an upper middle income country (2017), the government has started to be more concerned in providing accessible design in the built environment, especially since the country signed the proclamation of ‘Full Participation and Equality of People with Disabilities in the Asia-Pacific Region’ in 1994. Later with the introduction of Act 685 – Person with Disabilities Act in 2008, Malaysia has taken a step further to empower disabled people. In the context of accessibility for the disabled, Part IV Chapter 1 of Act 685 defines several topics such as ‘Access to Public Facilities, Amenities and Services and Buildings’, ‘Access to Public Transport Facilities’, and ‘Access to Recreation, Leisure and Sport’ (2008).

The situation has further improved with Department of Standards Malaysia (DSM) developing several standards in relation to the Person with Disabilities Act 2008. The main outcome from the standardisation is the introduction of three codes of practices (Hussein and Mohd. Yaacob, 2013) on accessibility around built environment for disabled people namely:

- (1) MS1183: Part 8: Code of Practice for Means of Escape for the Person with Disabilities
- (2) MS1184: Code of Practice for Access for Disabled People to Public Buildings
- (3) MS1331: Code of Practice for Access for Disabled People Outside Buildings

With these codes of practices, the Bylaw UBBL 34A (1991) was amended in 1991 to require new buildings to have access for disabled people, and for buildings already constructed or under construction when the bylaw came into effect to have to comply with the requirements mentioned by the codes of practices (with the exception of MS1331) within three years.

To summarize, the Malaysian federal government is aware of the need to provide accessible built environment to disabled people, and hence the bylaw UBBL 34A was amended as a legislated regulation for such a purpose. However, the implementation and enforcement by the local governments are still very lacking. One of the consequences – the BLVs and other users with disabilities are potentially susceptible to the dangers caused by hazardous built environments.

Hazardous Built Environment within Malaysia

The white cane is the primary and the most important tool used by the BLVs to safely self-navigate their way at some known or even unknown environment. A guide dog might be able to offer a safer, faster and more dynamic mobility aid than a white cane (Wiggett-Barnard and Steel, 2008), but this form of aid has yet to be a common practice in the country. The local media have reported several accounts of guide dogs' disputes over the recent years. On 29th May 2014, The Malay Mail reported that the Malaysian Deputy Transport Minister stated that guide dogs could be an inconvenience for the blind and the people at public transportation facilities (Winifred, 2014). On another encounter, when Lashawn, Malaysia's first and only guide dog and the visually impaired owner Stevens Chan were taken to a mall, they were quickly escorted out of the premises (Tan, 2014). The mall

security stated that animals were not allowed unless some sort of permit is given. From the reported cases, it seems that there is still a long way to go before there is acceptance of guide dogs in the country.

As a result, currently the guide cane is the common choice for most BLVs in Malaysia, as it is in low to middle income countries in general. However, the white cane might not always be helpful when the built environment does not comply with BLV accessibility design standards. Uneven steps, steep drop-offs, drainage without proper covers, uneven levels of walkway, rugged surfaces, discontinuous walkways, missing terminal posts between junctions, slippery floors, stairways without proper handrails and unpredictable surface discontinuities along walkways are just some common hazards to the BLVs as a result of poor enforcement and implementation of regulations for accessible design in low to middle income countries.

According to a preliminary survey through consultation with members from Dialogue in the Dark Malaysia and the St. Nicholas' Home Penang, the author arrived at the idea of exploring further into hazardous built environment. These two local NGOs are very active in looking after the welfare of the BLVs. Both organisations were consulted during the preliminary study of the research. It was then identified that amongst the hazardous conditions listed, "surface discontinuity" is a potentially solvable challenge but currently lacks solutions, as compared to some other areas of blind-navigation, wayfinding, or obstacles detection which have received much more attention.

Hazardous surface discontinuities are easily noticeable at places like surrounding areas of public and private service centres, walkways of commodities outlets, around pedestrian zones, recreation parks, surroundings of eatery hubs and several others. The BLVs would need access to these places as they offer services and goods which are crucial for their personal and social functions. A study on a focus group by Imrie and Kumar (1998) revealed that many disabled people feel estranged and oppressed by facets of built environment and

they feel powerless to do anything about it. These issues highlight the need and the significance for developing a solution to assist the BLVs for negotiating surface discontinuities. To conclude, threats negotiation specifically surface discontinuities around the urban areas would be the focus of this research.

1.7 Problem Statement

Based on literature review and consultation with the BLVs, it was found that many of the current technologies or research works are lacking in the following points:

- they know very little about the nature or characteristics of the local threats
- they require the user to wear (or bring) substantial body gears or gadgets
- most solutions for the outdoor environment are focused on solving global problems such as navigation or wayfinding, and are commonly lacking in local threats negotiation, let alone dealing with hazardous surface discontinuities
- most research works tend to focus on obstacles (or objects) detection, and
- most of the technologies for surface discontinuity are limited to only the problem of steps / staircases recognition, while indeed there is a richer taxonomy of surface discontinuity found predominantly in low to middle income countries, and some could be hazardous to blind navigation

Based on the problem statement, a lightweight, small and unobtrusive wearable technology-based system prototype is proposed. The prototype is expected to process the data in real-time to classify the type of surface discontinuity; and it must be able to work in the outdoor environment where most of the surface discontinuity problems were found.

1.8 Research Objectives and Questions

Looking from the perspective of main challenges faced by the BLVs as a bigger picture, the issues focusing on mobility can be broken down into several components as illustrated in *Figure 1.5*. Components in light green are subjects of focus in this research, particularly the surface discontinuities negotiation.

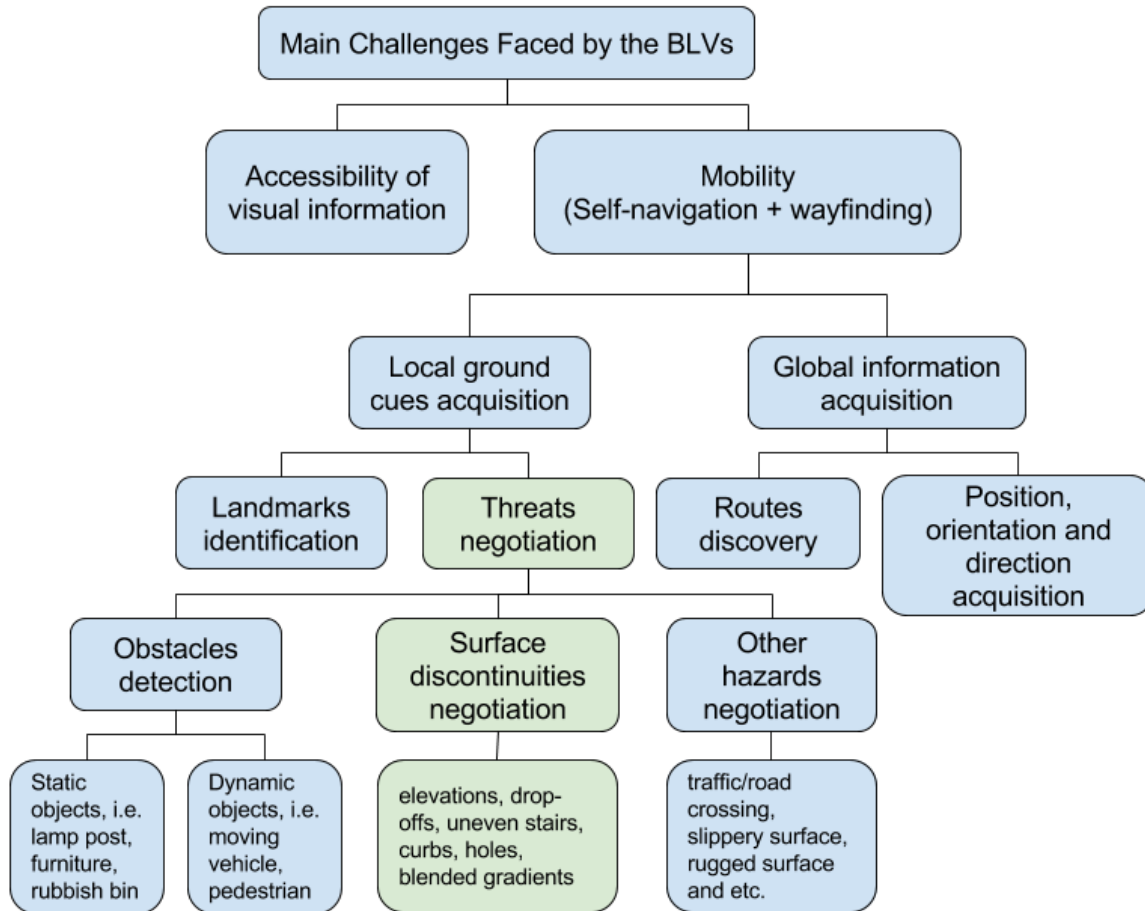


Figure 1.5: The breakdown of the main challenges faced by the BLVs, with a focus on the mobility issues. Components in light green are subjects of focus in this research.

1.8.1 Primary Research Question

Knowing the mobility challenges faced by the BLVs and based on the problem statement discussed in Section 1.7, the following research question was formed:

“How can a technology assist the blind and low vision people to negotiate surface discontinuities along their navigational pathway?”

1.8.2 Subsidiary Research Questions

From the primary research question, the following subsidiary questions were devised to facilitate the research.

1. What tool could be suitable to the sensing of surface discontinuities in an outdoor environment?
2. How can the identified surface discontinuities be classified into a suitable taxonomy to develop a machine learning model as part of the wearable prototype?
3. What are the suitable machine learning models that can be developed to classify the categories within the taxonomy mentioned in (2)?
4. What is the performance (accuracy and efficiency) of the developed prototype?

Based on the above subsidiary research questions, the research would focus on the development of a prototype to classify several types of surface discontinuity through a machine learning model. Relevant dataset would be collected from the field to support the machine learning model training. The development of user interface (UI) or the study of human-computer interaction (HCI) for the prototype would not be part of the research as it could be a very different nature of study. Time constraint is another factor of consideration to exclude the UI or HCI in this research. Having said that, there should be a minimal user interface such that the prototype can be properly operated and tested to answer subsidiary question number 4.

1.9 Research Contributions

This research is the first study into the blind and low vision negotiation of surface continuities in outdoor urban areas, with focus on the richer taxonomy of surface discontinuities that could be hazardous to blind navigation in a middle-income country. The research then proposes an assistive wearable technology-based system prototype to classify the identified surface discontinuities in near real-time.

The following points summarize the contributions of this research:

- *Problem identification*: Through continuous consultation with the BLVs at the early stage of the research, a major problem of surface discontinuities faced by them in navigation was identified. The problem was then classified into a taxonomy that was used to facilitate some machine learning model's development.
- *Dataset generation*: An authentic dataset that could exemplify the identified problem was generated through a systematic procedure that involves several steps. The dataset was collected with a purpose-built prototype. The data were then pre-processed and used to build a meaningful taxonomy to help the development of some machine learning models.
- *Methodological contribution*: One notable part of the research is the method used in data generation. The research first developed a prototype (phase-1) that was used as an instrument for data collection, and it is this similar prototype which was repurposed as the wearable technology-based system (phase-2) to detect and classify surface discontinuities in blind navigation. The phase-1 prototype was repurposed into the phase-2 prototype because it is with the similar hardware settings and sensor (camera) configurations that the surface of a pathway is captured for classification in real-time. Amongst other guidelines within this method are the instrument setup, sampling and pre-processing. Other researchers

can replicate a similar method to generate new dataset which can be added to this research for further study, or to other research projects of a similar nature.

Another contribution from the research is the method applied in developing the machine learning for the prototype. Two main guidelines discussed in the thesis that form the method are firstly the steps of assembling and experimenting with some machine learning algorithms, and secondly the steps of training and optimizing the best model using the algorithms. Again, the same method can be replicated by the research community to produce some similar studies or projects of similar nature.

- *Innovation of machine learning algorithms*: Relevant algorithms from deep learning were assembled and experimented with to solve the classification task, and finally a specific convolutional neural network model was trained and tuned to classify most of the surface discontinuities with high accuracy. The model was also optimized to have lesser parameters as compared to its similar counterparts to achieve higher algorithm efficiency.
- *Proof of concept via the development of a prototype*: The research demonstrates and proves that a wearable technology-based system solely built with stereo computer vision and deep learning could be a potential solution to help the BLVs in negotiating a diverse range of surface discontinuities. The prototype designed with a single tiny sensor installed on a single board computer has created a possibility to achieve a product that meets the BLVs' preferred design for a wearable technology which is lightweight, small and unobtrusive.

1.10 Thesis Outline

This section summarizes the central ideas of each chapter. Figure 1.6 highlights the structure of the research which is based on a design science methodology and it focuses on the design and development of the main artifact (the wearable technology-based system prototype).

Chapter 2 reviews the literature on the main concepts and technologies relevant to the research. It begins by examining the broad context of disability to offer an appropriate perspective for this research and help establish a guideline to review some relevant technologies for blind navigation. The chapter then reviews a range of different assistive technologies for blind navigation based on their sensor types or technology platforms. The chapter summarizes the features of these assistive technologies and identifies their shortcomings. Eventually the chapter concludes on the research gaps and justifies why and how a wearable technology-based system with computer vision as the sole sensing approach can be a potential solution to address the issue of surface discontinuities faced by the BLVs.

Chapter 3 presents the research design based on design science methodology adapted from Vaishnavi and Kuechler (2015) and Mettler et al. (2014). The chapter starts with the epistemological position of the research and follows with justification of the choice of design science methodology. The scope of the research and the output of the design science artifacts are elaborated in this chapter. The process of design science methodology is described. The chapter ends with discussions on the specific research techniques being employed.

Chapter 4 is mainly dedicated to data generation. It describes the data collection instrument with details about its development and setup. The chapter then describes the data collection process which involved crowdsourcing, sampling and ground truth measuring. A section is also provided to describe the collected data. Next, the chapter discusses data pre-processing techniques and justifies some pre-processing choices. The development of taxonomy of surface discontinuity is described. Finally, the chapter describes a data augmentation technique used to increase the volume of the data.

Chapter 5 presents the system development which refers to the phase-2 prototype building. Firstly, the chapter describes the idea of the phase-2 prototype. Next, the feature extraction technique is discussed. The discussion consists of observations and analyses of different configurations or functions used for disparity mapping. The chapter then describes the convolutional neural network being assembled for the classification task in this research.

Chapter 6 describes model training, classification results and evaluation of the developed phase-2 prototype. Three deep learning models and their variants being experimented are elaborated. The chapter then presents the field evaluation being performed on the best model. The evaluation is composed of two aspects – accuracy and efficiency. Details are given about these two aspects of evaluation.

Chapter 7 concludes the research by summarizing the thesis. It also recaps the research questions and provides answers to them based on the findings from the research. The chapter then discusses the impact and original contributions from the research. Future works are suggested before a concluding remark is given.

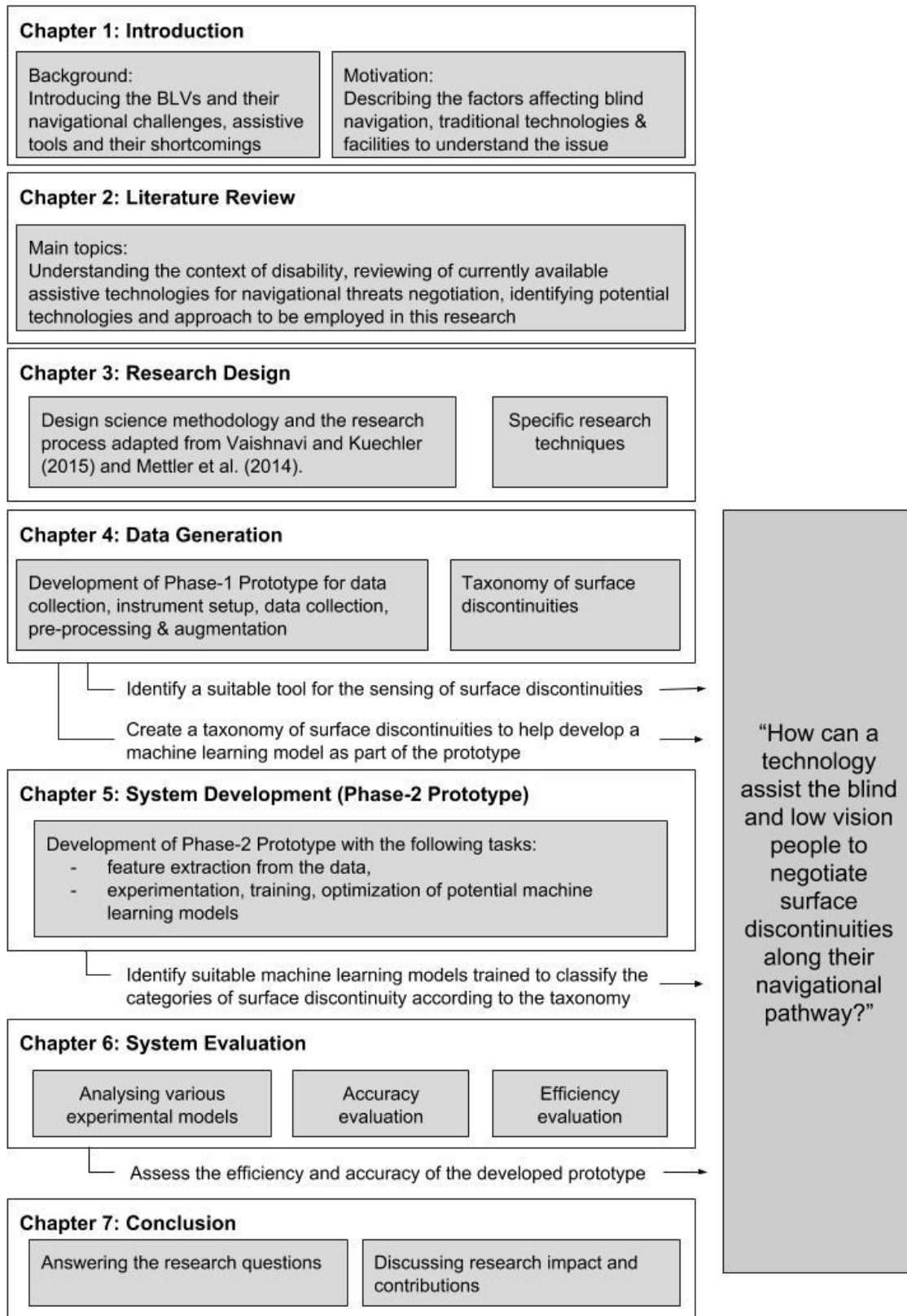


Figure 1.6: Outline of the research.

Chapter 2: Literature Review

This chapter first reviews the context of disability (Section 2.1) to offer an appropriate perspective for the research and based on this perspective, relevant technologies for blind navigation from the literature were then reviewed (Section 2.2). The chapter reviews these technologies and discusses them in five categories (Section 2.2.1 to 2.2.5) based on their major sensor types or technological platforms. Having reviewed the past and recent technologies for blind navigation, the chapter finally summarizes the major findings relevant to this research, before discusses the research gaps and states the possible offerings from this research in Section 2.3.

2.1 Context of Disability – Medical Model versus Social Model

The proposed wearable technology-based system is intended for the usage of the BLVs. These users fall within the category of visual impairment in terms of disability. Thus, before further review into relevant technologies, it would be useful to first consider the context of disability relative to the BLVs. The context will offer an appropriate perspective for this research and help establish a guideline to review some relevant technologies for blind navigation. There are two main models defining the context of disability namely the medical and social models.

2.1.1 Medical Model of Disability

Medical model views the dysfunction of disabled people to participate in society as a direct result of their very intrinsic impairment condition, rather than as the result of other external factors of the society (Read, 2003, Hersh and Johnson, 2008). The model refers to

the International Classification of Impairments, Disabilities and Handicaps by WHO (1980) in defining the three dimensions of condition: (1) impairment, (2) disability, and (3) handicap. All the dimensions take the context of health circumstances within the individual as the parameter for the definitions. In other words, this model defines disability as the residing condition within the individual and it focuses on the individual's impairment(s) as the cause of disadvantages in the society. Under this model, the cure or assistance being offer to the disabled people revolves around clinical treatment on the person such as occupational therapy and rehabilitation, in the attempt to restore the person back to a normal life within the society.

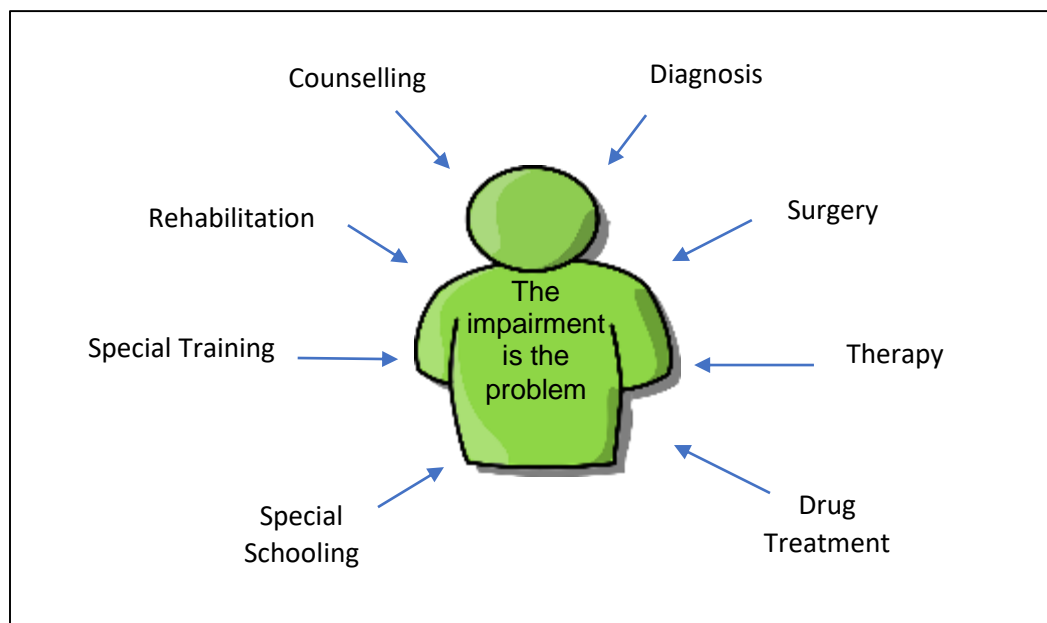


Figure 2.1: The Medical Model of Disability (adapted from Rieser, R., 2014)

Since this research is proposing a technological prototype that does not relate to either clinical treatment, diagnosis or rehabilitation, technologies for BLVs built based on this perspective would be excluded from the literature review in the following sections.

2.1.2 Social Model of Disability

The social model emphasizes the distinction between ‘impairment’ and the ‘disability’ experienced by disabled people. This model views the problems faced by disabled people as mainly caused by the physical, environmental and social barriers extrinsic to the individuals rather than the impairment within them (Swain et al., 2003). ‘Disability’ is thus defined as the loss or limitation of opportunities to take part in the society on an equal level with others due to physical, environmental or social barriers (Hersh and Johnson, 2008, Christensen and Byrne, 2014).

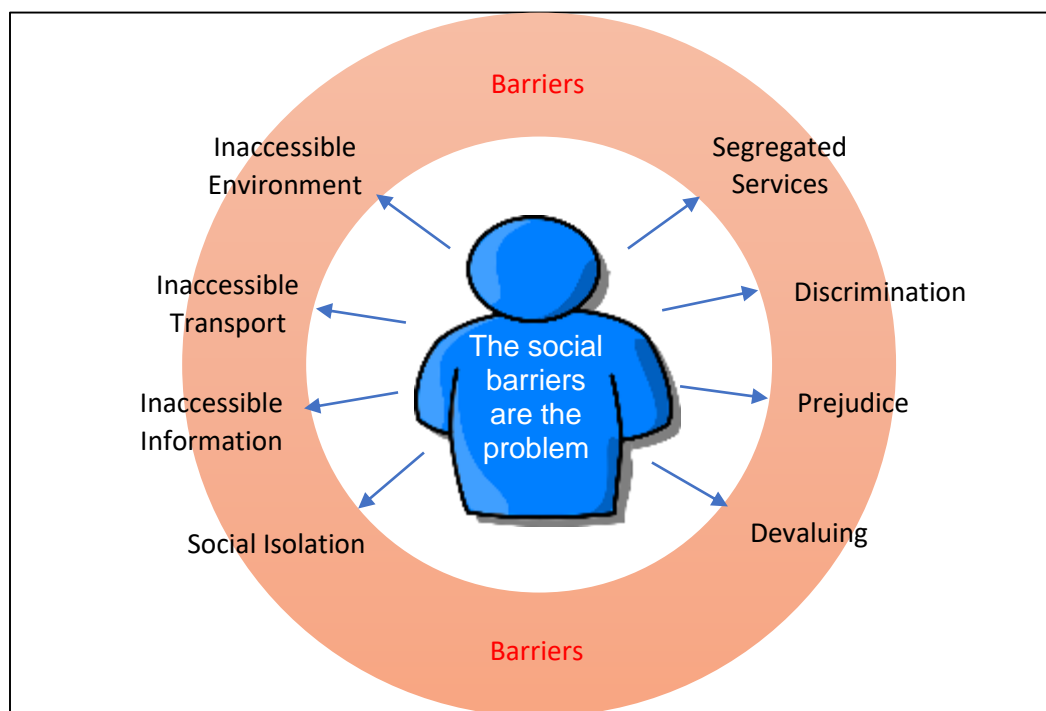


Figure 2.2: The Social Model of Disability (adapted from Rieser, R., 2014)

The lack of consideration in designing a barrier-free society is the main cause preventing disabled people from taking part in most societal functions. Removing the barriers can bring about significant changes to the situation. The new update of the WHO International Classification of Functioning, Disability and Health in 2001 introduced some emphasis on the social model (WHO, 2001). This new version allows for the impact of the environmental and other contextual factors on the functioning of a disabled individual to be compared

using a common metric. In the classification system since 2001, the terms “impairment, disability and handicap” were replaced by “functioning, disability and health”. This new classification system has been demonstrating a broader view with some relevant to social model. Its concern stretches beyond the medical model of disability and allows for the impact of the environmental and other contextual factors on the functioning of a disabled individual to be considered, analysed and recorded

It is through the understanding of the context of disability that some relevant technologies from the literature were chosen for review in the next section. This research would be taking the perspective from the social model of disability in developing a wearable technology-based system prototype. The prototype is aimed at assisting the BLVs in navigation, where the eventual result is to increase the level of independence by removing environmental barriers (in this case negotiating threats). Under the social model context, the purpose of visual impairment assistive technology is to overcome the gap between the needs of the social functions of a BLV and the opportunities offer by the societal infrastructure. The social model also provides another facet of looking at the impairment when the term is distinguished from disability. For the case of BLVs, it is the sensory differences between the people when comparing the blind, the low vision and the sighted; however, everyone should be given equal rights and opportunities in accessing infrastructures within the society such that disability should be reduced to a minimal degree. In view of the social model, it is noticed that most modern cities have given considerable emphasis to the policies of universal accessible design. However, in the less developed or low to middle income countries where the local policies or systems are not favouring the social model, or could take a slower pace to improve, the battle for equal opportunities for the BLVs may rely on some emerging technologies, apart from the traditional tools.

2.2 Assistive Technologies for Threats Negotiation

There are two major classes of problems faced by the BLVs in navigation:

- the problem of wayfinding, and
- the problem of threats negotiation.

The latter is the focus of this research. Most mobility tools are developed to help solve these two classes of problems. Mobility assistive technologies for the BLV navigation are often known as electronic travel aids (ETAs). An ETA is an electronic assistive device comprising sensors, an information processing unit and a communication interface, whereby the information is relayed to the user for enhancing independent navigation. ETAs for the BLVs typically operate using audio and haptic signals as the communication medium.

While performing the navigation or wayfinding task, the BLVs might encounter some local threats as stated in Chapter 1. The conventional guide cane can be a tool for both obstacles and surface discontinuities detection, but with some known limitations. There are some prominent technologies based on the idea of extending the guide cane's function for obstacles detection through some embedded technologies. Other technological variants are built based on mobile devices or robotic platforms. The following sections summarize some of these examples based on the major sensor types implemented within the tools.

2.2.1 *Laser Cane*

One of the most extensive researches since the 60's on extending the guide cane's function is the work by Benjamin et al. (1973), known as the laser cane. When the team worked on the device, it was at the time when laser had just been developed, transistors were just invented, and nickel-cadmium batteries and the first light-emitting diode (LED) had just become available. Benjamin and his team worked on the laser cane based on the guidelines drafted in the Benham (1952)'s report of the time, in which it suggested the following four criteria for a guiding device:

- 1) it must detect obstacles and down-curbs,
- 2) it must be silent and unobtrusive except when giving warning,

- 3) it must be simple to use and that the user should not have to "get dressed up" in it, and
- 4) it should perform range measurement on the principle of optical triangulation used in the Signal Corps Device (a single-channel optical ranging device of that time)

The Benham's report stated as well that the output of the device should be tactile if possible, and if it is auditory, it should not block other aural cues of the users. Working from the first generation of the laser cane named C-2, Benjamin and his team continued to improve on the technology and eventually they produced a series of the laser cane up to the final version, the C-5. Based on their work, the technology involved is quite simple, solely based on optical triangulation using three laser transmitters and three photodiodes as receivers. The upward-looking beam detects obstacles at head height appearing 1 ½ feet to 2 feet beyond the cane tip; the forward-looking beam detects obstacles from the tip of the cane; and lastly the down-looking beam detects drop-offs in front of the user.

The laser sensors extended the distance of detection, providing time enough to evade tree branches, signs, and awnings which the conventional guide cane traveller normally has no way of detecting. The operating range is set by the user, who flips a switch located above the laser housing. Any obstacle detected within the selected range will actuate a stimulator that contacts the index finger when the cane is carried in the usual guide cane manner. The laser canes are shown in Figure 2.3 and Figure 2.4, and the working zone of the C-5 cane is illustrated in Figure 2.5.

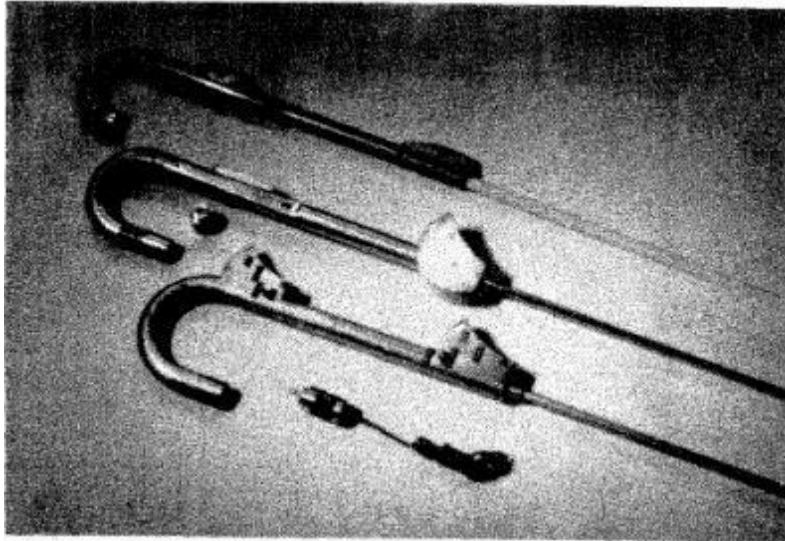


Figure 2.3: Four early models of the laser cane, in ascending order: C-2, C-3 and C-4 (Benjamin et al., 1973).

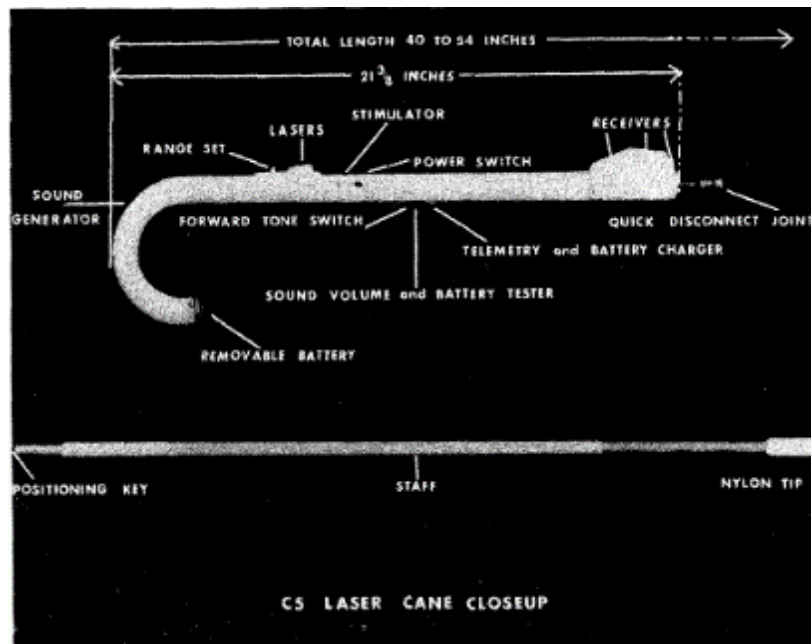


Figure 2.4: The C-5 laser cane (Benjamin et al., 1973).

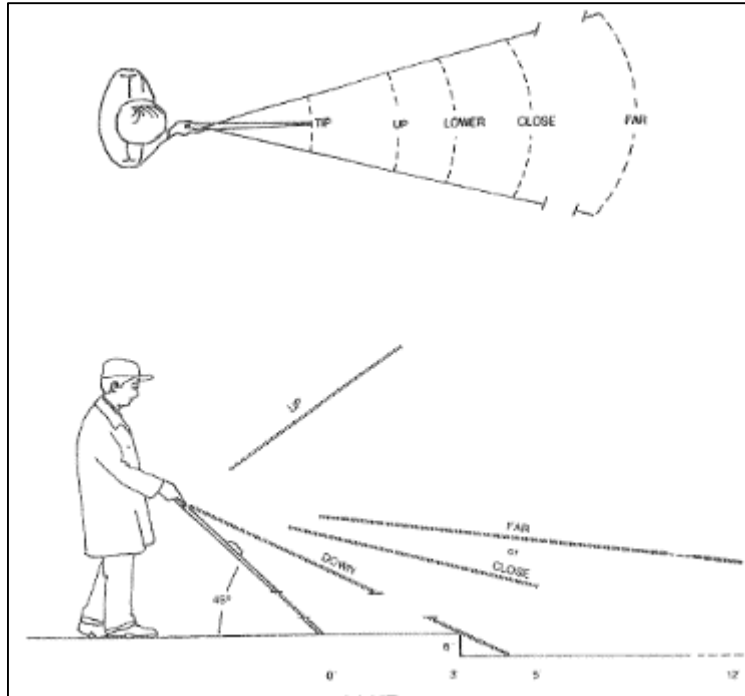


Figure 2.5: The protection zone of C-5 laser cane (Benjamin et al., 1973).

In their first attempt to evaluate the C-2 to C-4 cane models, the team demonstrated the canes to approximately 15 institutions, 100 blind people, and 50 O&M trainers for test and comment. By 1971, the laser cane was ready to be tried for a longer period by some guide cane users. They were followed-up and systematically evaluated. A year later, seven out of the eight selected experienced veteran cane users were still using their canes instead of any other travel aid.

Despite its scale of acceptance by the target group as claimed by the team, there was no publication of empirical study conducted to analyze if the laser cane had actually improved the mobility or safety elements in these users. Since the research by Benjamin et al., a handful of other researchers have worked on laser canes with some added features or improvements. Several recent research efforts in the similar category can be found in the works of Yuan and Manduchi (2005), Hesch and Roumeliotis (2007), Pallejà et al. (2010), Yokota et al. (2013a) and Yokota et al. (2013b).

Yuan and Manduchi developed a virtual guide cane based on laser beam triangulation. The virtual cane, as shown in Figure 2.6 is capable of detecting surface discontinuities such as walls, steps and drop-offs when the user points it to a particular direction as if it was a flashlight. The device can directly detect obstacles with certain heights with range measurement, but not for other surface discontinuity features such as curbs, steps or drop-offs. These features have to be tackled by analyzing the time profile of the measured range. With the base and edge features of the ascending step for example, the recognizable pattern from the time profile can be clearly observed. As such, the team modelled surfaces as piecewise planar, and through tracking the range using suitable models, discontinuities on surfaces can be detected and characterized. To achieve the models, a Jump-Markov process was used to describe the time profile of the range measurement.

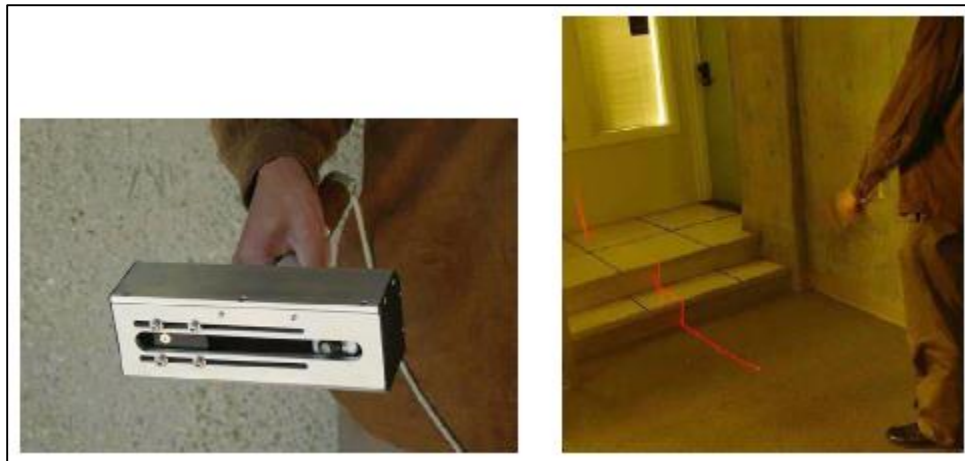


Figure 2.6: The virtual guide cane by Yuan and Manduchi (2005).

Pallejà et al. (2010) proposed a bio-inspired electronic white cane using LIDAR (Light Detection and Ranging) and tri-axial accelerometer to mimic the effect of the whiskers of an animal that tell the animal if something has come into contact with the skin. The device is quite similar to the work by Yuan and Manduchi (2005), but instead of being held by the user, it is mounted over the user's wrist for both information retrievals from the environment as well as to provide tactile stimuli on the user's skin as a means of feedback. The main idea of this tactile mechanism is to prevent the user from using sound input that

might interfere with the individual's daily activities. The device is built to work in two modes – (1) in the floor mode, it uses the intersection between the scanned plane and the floor to define a frontal line, and then it operates to reveal the presence of objects along the path based on the sudden change of the distance to the frontal line; and (2) in the frontal mode, it explores a range of up to 6 meters in front of the user while the user stays still and swings the arm to scan the environment. The validation results showed some high accuracies (between 93 to 100%) of the technology to detect and identify plain floor, up-steps, down-steps and cubic obstacle (100 to 400 mm of heights).



Figure 2.7: The assistive walker from Yokota et al. (2013a) and Yokota et al. (2013b).

In another slightly different work of redefining the guide cane function, Yokota et al. (2013a) started with their assistive walker (Figure 2.7) designed with laser range finder and Omni-wheeled, and it was later followed up with experimental updates (Yokota et al.,

2013b). This walker takes the form of a guide cane, but with Omni-wheels touching the floor surface to passively guide the user. A braking system is implemented to provide grip on the wheels such that the user is informed about approaching obstacles detected by the laser range finder. The advantages of this device are: (1) the system design allows it to move Omni-direction which is most convenient in most environments, and (2) it helps the user to create a mental map of the surroundings.

2.2.2 Ultrasonic Cane

One of the earliest researches on ultrasonic is the UltraCane developed by Sound Foresight (2015c). It works in a similar way as the conventional cane but with embedded front and upward-facing ultrasonic sensors. Another similar technology is known as Miniguide created since 1998 by GDP Research (2015a). Both technologies are capable of detecting drop-offs ahead. Apart from that, the detection of overhangs by these tools is especially useful as the typical guide canes or guide dogs lack this ability. As for UltraCane, it has a distance indicator that informs the user through vibrations felt at the front or rear parts of the holder.

Based on a wearable idea, Navbelt (Shoval et al., 2003) was developed by a group of researchers at the University of Michigan. Navbelt as the name indicates, was a belt embedded with 8 ultrasonic sensors to detect obstacles and provide feedback to the user in audio signals through headphones. When the path is safe without obstacles, the belt will be silent; but when obstacles are detected, the volume of the audio signal increases in proportion to the distance of the obstacles. Through a five-year follow-up and evaluation, the device was not quite accepted by the users as the audio signals were hard to understand and the time needed to process the signals halts the user from walking smoothly.

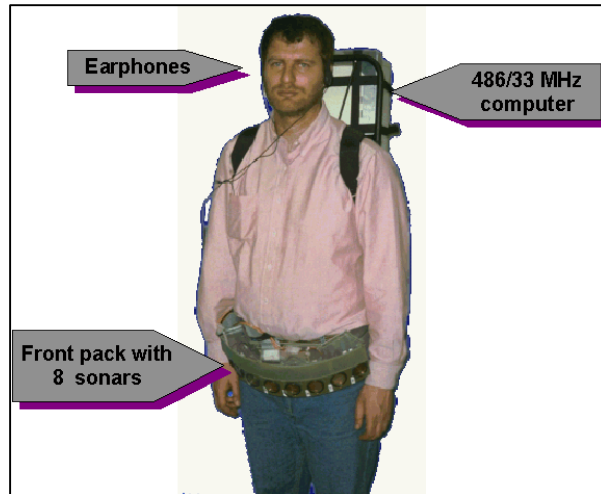


Figure 2.8: The Navbelt, a wearable gadget for obstacles detection by Shoval et al., 2003.

Borenstein and Ulrich (1997) initiated the GuideCane project to partly overcome the problems suffered by the NavBelt, and Borenstein (2001) made some improvement on the project. The GuideCane was developed based on mobile robotic technology using ultrasonic as the main sensors. The tool attempted to reduce conscious effort of the user in obstacles detection and avoidance by acting autonomously.

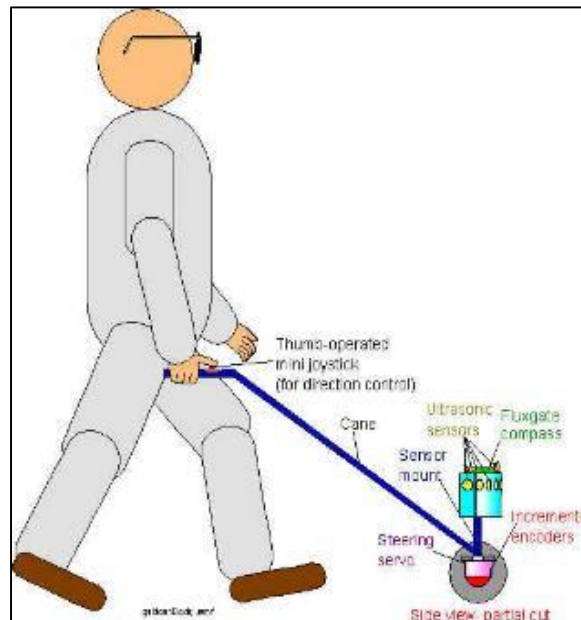


Figure 2.9: An illustration of the ultrasonic based GuideCane from Borenstein and Ulrich (1997).

As moving around without vision can be mentally taxing with the presence of unknown obstacles, the GuideCane was built to overcome this issue by leading the user to a safe path of navigation. Built in the form of a vacuum cleaner, the device is capable of detecting obstacles in 120 degrees of field of view with its 10 ultrasonic sensors. During operation, the user pushes it to move forward and when an obstacle is found, the device's computer determines the correct direction to avoid the obstacle and then steers the user back to the original path of navigation. Odometer, compass and gyroscope technologies were incorporated into the device for extra data to analyze its movement. The major drawback of the device is the limitation of freedom for the user to move around because the user has to follow the decision and movement of the device. This disadvantage could hamper the user from recognizing the surroundings and learning the spatial information.

2.2.3 Depth Sensing Technologies

With the advances in 3-dimensional (3D) depth sensing technology (based on structured light), several more recent works have tapped on this technology to arrive at obstacles detection solutions. Most of the research has attempted to take the form of extending the typical guide cane through depth sensors.

Filipe et al. (2012) presented a depth sensing system for indoor environment. Depth data acquired by the sensor is mapped into a pattern representation and relevant features are extracted from the data to identify possible obstacles along the pathway. The system employed a neural network classifier for the feature extraction. The experiments with 1200 sample images showed that the objects were correctly classified into the representation classes. The real-time application of this tool showed that it is able to provide information about obstacles and stairs for approximately 2 meters before the user.

In the works of Takizawa et al. (2012) and Takizawa et al. (2013), the depth sensing technology is applied for objects recognition and hence the device is capable of informing the user about the presence of a particular class of object. The work indirectly contributes

to threats detection because if the object is classified correctly, the user can decide the next course of movement. The user would instruct the system by pushing one of the keys on the keypad-type controller. Each key can activate a recognition scheme designed to identify the object of a particular class. If the target object is identified in the field of view of the sensor, the tactile device would return vibration feedback to the user. The vibration strength is set to be inversely proportional to the distance between the user and the target object. Based on the vibration feedback from the system, the user could then give a new operation instruction to the system. The prototype was tested for two classes of object – chairs and upward stairs – and recorded precision results of recognition of 0.8 and 1.0 respectively.



Figure 2.10: A depth sensing technology proposed by Takizawa et al. (2012) and Takizawa et al. (2013).

Orita et al. (2013) on the other hand focused on obstacles detection using the measurement of edges or lines from the depth data to identify objects out of the pathway. Due to the nature of the sensor device they had chosen, they had to use two different methods for detecting small and large objects. Kuramochi et al. (2014) proposed a method to recognize elevator doors by applying the RANSAC (RANDOM SAMPLE Consensus) algorithm to depth

data obtained through a depth sensor. They discovered that the proposed method could recognize some types of elevator, but not all types.

Common to all the works on depth sensing technology is that the device works effectively only in an indoor environment. The structured light (commonly based on infrared ray) which is the core element of the depth sensor does not work under certain threshold of illumination in an outdoor environment.

2.2.4 Computer Vision-based Technologies

Several computer vision-based technologies for obstacles detection is based on stereo vision formed by two cameras. By using stereo camera to measure distance, one can estimate the location of an approaching object. With such technology, an obstacle can be identified from the disparity map. This is achieved commonly through some image processing techniques.

With stereo computer vision, Rodriguez et al. (2012) developed a method for obstacles detection to aid blind navigation. A technique with stereo rig calibration and rectification for computing the depth of 3D points was suggested. The team then used the 3D points to resemble a dense of disparity map, with respect to the camera's coordinates to detect obstacles. The team claimed that this method can detect obstacles in both indoor and outdoor environments, and further differentiate between the plane of the walkway and obstacles. The study only presented a limited experiment along a route in Alcala de Henares, Madrid. Apart from planes and standing obstacles, little is reported about other types of surface discontinuity in their study. Despite their focus on planes and standing obstacles, the system was observed to have occasional errors in recognizing a wall as a plane.

Lee et al. (2008) developed a walking guidance system for the visually impaired in unrestricted natural outdoor environments. The system contains several modules for

people detection, text recognition and face recognition. They built the module of obstacle detection using a stereo camera. The team also incorporated a walking guidance module using Differential Global Positioning System for navigation. It is claimed that the system worked well to guide the user in a natural environment.

In the research by Hern et al. (2011), the team used a single camera to detect the localization of indoor stairways. The main approach involves analyzing the edges of a stairway based on planar motion tracking and directional filters. The research, based heavily on some image processing technique, has a very limited focus on detecting indoor stairway and steps only.

Dang et al. (2016) developed a virtual cane based on a camera, a line laser and an IMU for indoor application. The system helps the BLV user to identify the type of obstacle and the distance to it. Although equipped with a visual camera, the system does not rely on the vision data for obstacles detection. Their algorithm works by having the line laser and the camera calibrated to form a line-structured light. Based on the pin-hole model, the system performs a 2D laser point and 3D camera plane matching to estimate depth information of the scene, provided that the position of the system in relation to the user is known. The user would hold the device and swing it horizontally once to scan the environment. The scanned input is analyzed for each image frame captured by the camera. The laser points are then extracted based on their intensities, such that the sets of laser points can be differentiate into basic lines and obstacles. With the differentiation, the obstacles will be further classified into wall, up-stairs, down-stairs, floor and block. When the device is swung, the IMU would provide the orientation of the system using a Kalman filter. The orientation is then used to estimate the pointing angle of the device in the navigation coordinate frame. Their system was tested to accurately recognize both wall and floor (100% accuracy); and for up-stairs, down-stairs and block (96% accuracy). Their tests were performed with the assumption that no occlusion to the objects was recognized. A

limitation to the system is that the laser light is affected by strong illumination, which indirectly reduces the working range.

2.2.5 Robotic Solutions

Recent advances in both hardware and software technologies for robotics could possibly offer a medium to integrate several different sensors to achieve a more robust solution for blind navigation, which is hardly possible in mobile devices or extended guide canes mentioned in Section 2.2.1 to 2.2.4 due to their limited spaces or smaller sizes.

In the work of Kulyukin et al. (2004) and the improved version (2006), the teams used a commercial robotic platform with RFID and laser sensors to assist the BLVs in performing wayfinding tasks in structured indoor environments. In their experiment, the indoor environment intended for the usage of the robot was installed with some passive RFID tags at strategic locations such as permanent landmarks, turning points or doors. Each tag was programmed with a unique ID as a location marker. The robot, apart from navigating the BLV user based on the tags around the building, was capable of detecting obstacles and leading a detour.

Their approach of obstacles detection does not emphasize on interpreting the obstacles; rather the navigation path is treated as either empty space (that defines continuity of the navigation) or non-empty space (that suggests a detour). The navigation of the robot is not egocentric, as long as the environment is instrumented with the RFID tags, the navigation is distributed between the robot and the entire tagged space. The human-robot interface (HRI) is based on audio communication. For the input, they provided the users with audio communication in which a list of standard phrases should be used for telling the robot the intended destination. For the output (or feedback), the users were given a test to choose between speech messages and audio icon (e.g. sound of water bubbles to signify a particular object). The user would follow the robot by holding onto an attached leash. Since the robot was built with orientation-free movement, there were issues of getting lost at certain

events. To counter this disadvantage, the algorithm was added with loss recovery mechanism. In their tests, they observed that the robot can maintain a moderate speed, except at turns and during obstacle avoidance.

A major drawback to the RFID approach is that when a tag is blocked, the robot would miss it and fail to proceed with appropriate behaviour. In addition, the robot would possibly choose a wrong tag at intersections, which could lead to a wrong path or sub-optimized path. For the HRI aspects, they discovered that the speech input is not a viable mode as the robot often faced difficulty recognizing the phrases pronounced by the users.

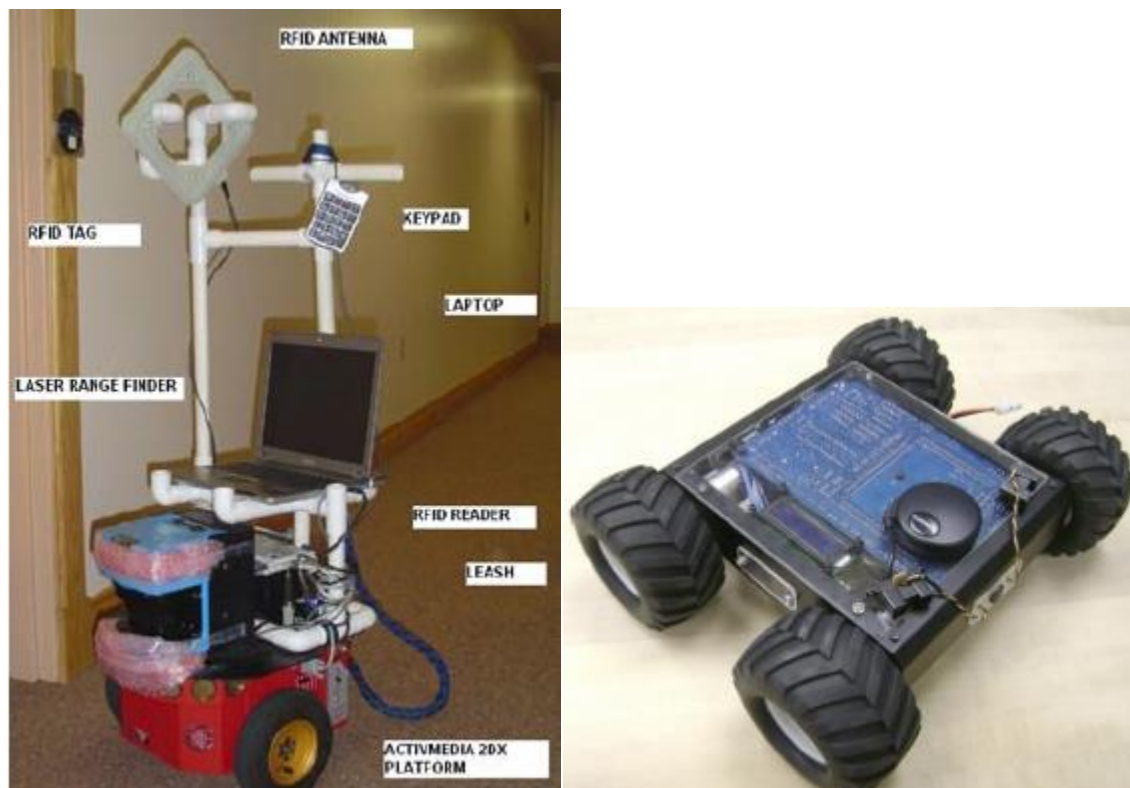


Figure 2.11: The RFID based navigation guide robot (right image) from Kulyukin et al. (2006), and the prototype robot chassis (left image) from Yelamarthi et al. (2010).

Another related work involving RFID technology was proposed by Yelamarthi et al. (2010), in which the team incorporated a GPS technology to create a smart robotic guide. As shown in Figure 2.11, the robot can guide the user to a known destination or add a new route on-

the-go for future usage. The new route creation feature is a distinct ability as compared to several guide robots proposed by others. If a new route is intended, the system will ask the user for route origin and destination. The direction buttons will be active such that the user can use the buttons to move the robot to the destination. At the same time, the system will record the periodic coordinates from RFID while indoor, or from the GPS if outdoor. Equipped as well with an analog compass, and ultrasonic and infrared sensors, it is able to avoid obstacles along the journey and orientate itself back to the destined track. Feedback is provided to the user through a speaker and a vibrating glove. The robot development was still under progress and has yet to be tested thoroughly, despite several pilot experiments focused on destination reaching.

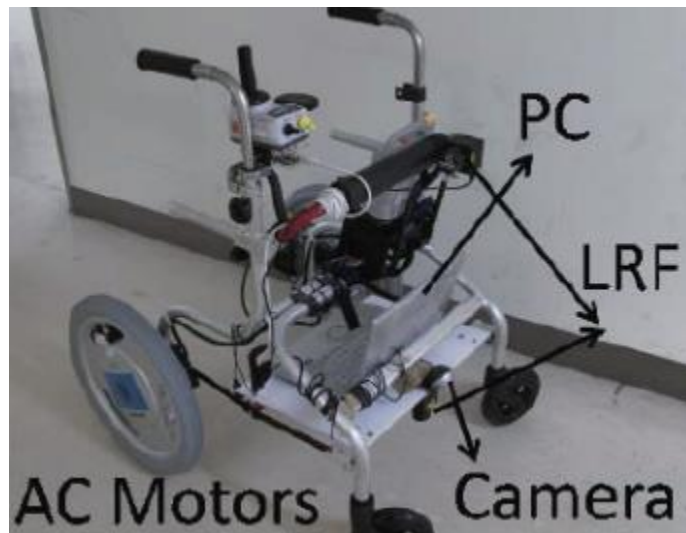


Figure 2.12: The neural networks based robotic guidance by Capi (2012).

Galatas et al. (2011) introduced a robotic guide dog that moves parallel to the direction of the road or corridor based on a visual camera. The image frames captured by the camera are used to estimate the position of the vanishing point on the road or corridor, in which the robot needs to navigate. With the found location of the vanishing point, the robot will be steered by the actuator to the intersection of the point and image centre. Additionally, by using LIDAR as an extra sensor, the robot can perform obstacles detection and avoidance.

Based on two laser range finders and a camera as visual sensor, Capi (2012) developed a neural networks based robotic system that communicates by natural language or beeping signals to guide the visually impaired in public buildings such as hospitals and offices. The main function of the robot is to give information to the user regarding the surroundings through a speaker. The visual sensor is used to detect markers while the laser range finders are used to detect moving persons, steps, stairs and obstacles. They used red and green colours as markers for the doors on the left and right side respectively. When the user selected a specific target (e.g. room within a hospital), the corresponding colour and progressive number of the marker is generated. The robot would then proceed to guide the user to the target by following the black coloured line on the floor. Upon reaching the target, the robot would inform the user that the room is at the left or right side of the user. If obstacles for instance a human gets into the navigation pathway, the robot would stop moving, until the obstacle has left. The robot would then resume the guidance.

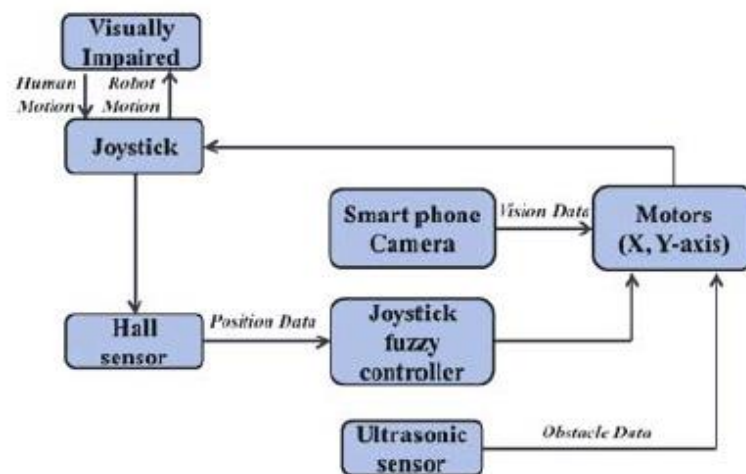


Figure 2.13: The functional block diagram of Yuanlong and Mincheol (2014)'s guide-dog robotic system.

A more recent research by Yuanlong and Mincheol (2014) presented a guide-dog robot system for urban outdoor navigation with a “smart rope” as the main interface for the human-robot interaction. The main concern of the team was the interaction between the user and the robot, in which the “smart rope” (implemented on a hall sensor joystick) is

controlled by a fuzzy logic system able to distinguish between a small involuntary force and the intended movement for navigation. The user would provide different hand movement commands to inform the robot to act e.g. stop or go ahead. With some ultrasonic sensors, the robot is capable of obstacles detection and avoidance. Several distinct features of the robot are (1) traffic lights detection, and (2) zebra crossing detection. Images of either the traffic light or zebra crossing are acquired via a mobile phone's camera, which would be streamed to the main processor using Wi-Fi technology. They employed Hough algorithm, adaptive boosting and template matching methods for the edge detection of the zebra crossing and yielded some promising results.



Figure 2.14: A sample of zebra crossing edge detection image from the work of Yuanlong and Mincheol (2014).

Most of the reviewed robotic solutions were built in the form of a wheeled robot, but a slightly different approach with a four-legged guide-dog robot was developed by Nippon Seikō Kabushiki-kaisha's Emerging Technology Research Center(2011). As shown in Figure 2.15, the robot with a structured light depth sensor is capable of guiding the user to navigate on a smooth floor on its four-legged wheels. Apart from that, the robot is also able to detect unknown stairs and autonomously recognize the shape, altitude and position of the stairs, and then proceed to guide the user up and down each step. The user has to hold on to a handle on the robot for the guidance. Once a step is identified, the robot would stop and

lower its body to inform the user about the elevation changes. The user would then follow the robot's motion to move over the step.



Figure 2.15: A four-legged guide-dog robot developed by NSK (2011) based on depth sensing technology, capable of obstacles and steps recognition.

To offer a more comfortable experience to the user, the handle varies in height and angle to prevent the need for the user to hunch over on stairs. The HCI aspects of the robot is based mainly on computerized speech, which communicates the details of the surroundings to the user including information on obstacles avoidance. Apart from the depth sensor on the robot's head, it has also proximity sensors around the legs to overcome some blind spots unnoticed by the head sensor. To date, the NSK developers have yet to consider the safety aspects of the prototype. For examples, avoidance of falling, recovery, and continuance of climbing if a fall does occur are some critical concerns of the four-legged robot.

2.3 Summary of Major Findings from Literature

2.3.1 *The Research Gaps*

From the review of recent literature in Section 2.2, there are various technologies built in the forms of electronic travel aids (ETA) or robotic platforms for threats negotiation. Most of the mentioned technologies focus on static or dynamic obstacles detection. For instances, Benjamin et al. (1973), Yuan and Manduchi (2005), Hesch and Roumeliotis (2007), Pallejà et al. (2010), Yokota et al. (2013a) and Yokota et al. (2013b) developed their prototypes based on laser sensor for obstacles detection. While these existing works have shown promise in extending the ability of the classic guide cane (i.e. detect farther obstacles or objects above waist level), however, what is missing from these tools is, despite providing the user with indication that an obstacle is detected, they offer very little information about the nature or characteristics of the obstacle. The user must interpret the detected object to proceed with navigation. Additionally, these tools mostly deal with objects above the walkway surface, thus neglecting the surface conditions of the walkway.

Sound Foresight (2015c) and GDP Research (2015a) have marketed their ultrasonic obstacles detection cane with some success, while Shoval et al., 2003, Borenstein and Ulrich (1997) and Borenstein (2001) have made some progress by using ultrasonic technologies. While the precision of the ultrasonic sensor is an issue, the more obvious problem is that most of these tools require the users to abandon their conventional navigation aid such as the guide cane, and this is normally not acceptable by the BLVs. Furthermore, specifically in the case of Shoval et al., 2003, Borenstein and Ulrich (1997) and Borenstein (2001), the user's physical load remains high as they required the user to wear or bring along additional bulky body gear or gadgets.

With recent advancement in structured light depth sensing technologies, Filipe et al. (2012), Takizawa et al. (2012), Takizawa et al. (2013), Orita et al. (2013) and Kuramochi et al. (2014) developed obstacles detection technologies with the extra capability to recognize certain classes of objects. The depth sensing technologies show a promising direction as

they could be the only sensor needed, and they are robust in detecting obstacles. The common drawback of the structured light depth sensing technologies is the limitation of working only indoor. This has limited their applications only to indoor usages, while in most actual cases, the BLVs need to navigate around the outdoor environment within the urban areas.

Several other groups of researchers worked on the similar concept of depth sensing but with computer vision-based technologies (Rodriguez et al., 2012, Lee et al., 2008, Dang et al., 2016). These computer vision approaches have an advantage as they work both indoor and outdoor as compared to the structured light. The information captured by such technologies is also higher in resolution, hence the feedback to the BLVs could be more comprehensible. Other depth sensing or computer vision approaches (from both ETA and robotic platform) have made possible the development of more robust technologies to not only classify the threats, but also to detect (or characterize) some of them. *Table 2.1* summarizes some of the major literature based on these approaches.

Table 2.1: Type of threat or object classification or detection to assist blind navigation from several authors based on depth or computer vision sensing technologies.

Type of threat / object	Author(s)
furniture	Takizawa et al. (2012), Takizawa et al. (2013)
several classes of indoor objects	Filipe et al. (2012)
elevator door	Kuramochi et al. (2014)
traffic light and zebra crossing	Yuanlong and Mincheol (2014)
vanishing points on road and corridor	Galatas et al. (2011)
stairs / steps	Lee et al. (2008), Hern et al. (2011), Jonsson (2011), Tang et al. (2012), Takizawa et al. (2012), Vlaminck et al. (2013), Takizawa et al. (2013), Pérez-Yus et al. (2015), Dang et al. (2016)

From the above works, stairs detection is the most studied subject. However, as detailed in Section 1.6.3, apart from stairs, this research identifies a richer taxonomy of surface discontinuity which has not been tackled before. The situation is especially apparent in underdeveloped (or low to middle income) countries. As for the technologies built on robotic platforms, the idea of having a robot navigating outdoor is fraught with the same problems faced by the BLVs due to surface discontinuities, despite their abilities to carry multiple types of sensors in a single platform.

As a conclusion, the gaps of some of the current technologies can be collectively highlighted in the following points:

- they offer very little about the nature or characteristic of the local threats, especially in surface discontinuities
- they require the users to wear or bring additional bulky body gears or gadgets
- they require the users to abandon their guide canes
- they might not work in an outdoor environment
- robotics solutions are not suitable as assistive tools for outdoor environment
- they focus mostly on obstacles detection above the surface of a walkway
- the technologies for surface discontinuities focus mainly on stairs, while indeed there is a richer taxonomy

2.3.2 What the Research Offers

Through consultation with the BLVs' service providers at the early phase of the research, the technological solution acceptable by the BLVs must be unobtrusive, small and lightweight. They also commented that the guide cane must not be excluded from the adoption of any assistive tools. In addition to the research gaps identified from the literature, the following resolutions are proposed in order to design and develop a prototype:

- (1) the prototype must be able to offer more insights (e.g. the type of surface) about the characteristic of surface discontinuities
- (2) the user must not be required to put on bulky body gears or gadgets
- (3) the prototype must not replace the guide cane, rather it should serve as a complement to the guide cane
- (4) the prototype should be able to work in an outdoor environment
- (5) the prototype should come in a simple wearable form, instead of as a robot
- (6) the research should address a richer taxonomy of surface discontinuity

There are basically two main parts that form the prototype – firstly the sensor, and secondly the system that processes the data from the sensor and classifies the surface discontinuities into their classes. For either the sensor or the system, it must meet (or indirectly help to meet) the proposed resolutions (1) to (6). Typically, more sensors or more types of sensors will enhance the robustness of the system. However, this practice would not be aligned with the proposed resolution (2) if the system is loaded with multiple types of sensors (thus making it bulky) for the sake of robustness. Some sensors may not fulfil resolution (4) because they might not work in an outdoor environment. Sensing approach based solely on computer vision could be a much more suitable choice as it can capture high resolution data (thus making resolution (1) possible) and work well in an outdoor environment, without the need of adding other sensors. There is also a diverse range of small and lightweight camera modules in the market that can be acquired to develop a prototype that is based on computer vision. Resolution (5) to develop a wearable prototype can be achieved if the sensor is small and lightweight. Using a camera as the sensor could also enable a richer range of taxonomy of surface discontinuities to be included to meet resolution (6), provided that a suitable machine learning algorithm could be implemented to correctly classify them. Computer vision is thus a potential choice of sensing approach in this research context.

Since computer vision is the choice for developing the prototype, the data captured from the camera module could be videos or images. The second part of the prototype – the system – must be able to process them in near-real-time and to classify them into their respective classes. It is common that a machine learning model is used to perform such classification and hence it is the core component of such a system. The following section provides additional literature about the state-of-the-art computer vision applications and several machine learning algorithms being employed for each of the applications. The review on these applications would provide a foundation for the research to further design and develop the prototype to address the issue of surface discontinuity as defined earlier.

2.3.3 Conclusion - Computer Vision and Machine Learning as the Core Approach

Several studies based on computer vision to solve navigation problems have utilized machine learning to map the traversability of a pathway to colour histograms. In the works of Dahlkamp et al. (2006), Leib et al. (2005) and Hong et al. (2002), a road-following robot with an RGB camera relies on the immediate ground colour at the front of the robot for navigation. The machine learning algorithm assumes the immediate ground to be traversable and thus, it filters out other colours and searches for a similar colour for the next course of movement. Although this approach seems effective, it is limited by the availability of the inheritance of the similar colour along the pathway.

Hadsell et al. (2008) trained a Deep Belief Network (DBN) to classify complex off-road terrain at a distance captured by a stereo computer vision. Their approach starts with unsupervised training of a DBN with the data to extract features from an input image. The extracted features are then used to train a stochastic gradient descent classifier. The trained network is then used to predict traversability on a pathway for a mobile robot in real-time. Their classifier can differentiate between trees, paths, manmade obstacles and the ground at a distance. The success of their method can be attributed to the availability of large volume of training data, and also due to the use of DBN for better feature extraction.

In another work on computer vision, Ni and Aziz (2016) presented a DBN with Support Vector Machine (SVM) and Histograms of Oriented Gradients (HOG) as a framework to recognize dynamic hand gestures. They split the hand gestures recognition into three processes, i.e. hand detection, hand tracking and hand trajectory recognition. The first two processes work in collaboration with the SVM and HOG for hand detection, and a Mean Shift algorithm for hand tracking. Then the third process uses DBN for hand trajectory recognition and this is a much more challenging task as compared to the first two, partly because the trajectories are accompanied by noise. In the DBN layers, they use Directed Belief Net for the input vectors, Restricted Boltzmann Machine for the hidden (pre-training) layer, and Softmax regression for the output (fine-tuning) layer. Their approach could recognize drawings by hand gesture of a set of numbers and letters with up to 97% to 99% of accuracy. One of the possible factors of their better performance is the capability of the DBN to handle noise as compared to Wang et al. (2011) that they quoted in the comparison.

In the well cited paper of image classification by Krizhevsky et al. (2012a), a deep convolutional neural network (CNN) is trained to classify the 1.3 million high-resolution images in the ImageNet training dataset into thousands of different classes (ImageNet is a large visual database designated for use in visual object recognition research). In their network architecture, they constructed five convolutional layers with a total of 500,000 neurons within. Running on a powerful graphical processing unit, they could speed up the training period with the help of non-saturating neurons. With such a deep learning network, they yielded better prediction on the test set when compared to previous state-of-the-art results.

In another recent work of CNN application, Sun and Qian (2016) attempted to solve Chinese herbal medicine image recognition problems. They used Softmax loss function to optimize the recognition network, and it was later used for retrieval of the similar medicine images. They evaluated their method through a public database of herbal medicine images

that contained cluttered background. The evaluation concluded with improvement with a large margin when compared to the previous state-of-the-art studies.

Apart from the works in Krizhevsky et al. (2012a) and Sun and Qian (2016), several other authors (Cheung, 2012, Ding et al., 2016, Liu et al., 2017, Yeum, 2016) have also applied some variants of CNN to image classification and achieved some accurate predictions.

Tapping on the robustness of CNN, Ren et al. (2015) went ahead with some modifications to a typical CNN to speed up the training for real-time object detection. In their studies, they revealed that most region proposal algorithms could be a bottleneck in network training on object detection. The complexity in object detection as compared to image classification is due to the need of localization of objects in the process. The need of localization creates two main challenges: firstly, numerous candidate locations (of the object) called proposals must be processed; secondly, these candidate locations are rough estimations that need to be further refined. Several variants of region proposal algorithm (which normally take longer time out of the training) are typical solutions to this localization issue. They proposed a fast regional-based convolutional neural network based on the original Fast R-CNN by Girshick (2015), with a Region Proposal Network (RPN). They introduced RPN that enables nearly cost-free computation as compared to other region proposal algorithms when dealing with the localization process prior to object detection. They demonstrated that a CNN can be modified to better suit the needs of object detection, and this work has won them several competitions inclusive of several tracks in ILSVRC (ImageNet Large Scale Visual Recognition Competition) and COCO (Captioning Challenge 2015 - COCO - Common Objects in Context).

In another variant of CNN for object detection, Liu (2017) worked on a real-time solution for autonomous driving. Due to the slow speed in object detection for a typical CNN, they experimented a recent model known as YOLO (Redmon et al., 2016) which has improved speed with the ability to compute regression directly from input images to determine object class scores and positions. Despite its improved speed, YOLO is faced with one issue – it

processes images individually, and this is not favourable for a moving automobile. Having studied the limitation of YOLO, they broadened the scope of YOLO by introducing a memory mapping technique. The memory map takes into consideration inter-frame information and thus enhances YOLO's object detection ability in real-time.

Although the above applications are not directly built for blind navigation, they focus on image classification or object detection via computer vision, and thus their deep learning techniques can be referred to for this research context. Their successes show that deep learning is a potential approach to dealing with object classification challenges in computer vision. In the review "What led computer vision to deep learning?" written by Malik (2017), he concluded that the machine learning community have realized that apart from deep learning's proven performance in computer vision, they also discovered that generally a deeper network performs better. The above insights would suggest deep belief networks and convolutional neural networks as the main deep learning approaches to be experimented by the research.

Through the literature findings, the author has identified the research gaps, and insights from the findings have helped justify the design and development of a prototype - a prototype that is wearable (unobtrusive, small and lightweight), senses via computer vision, and taps on current state-of-the-art deep learning algorithms for the classification of surface discontinuities data.

Chapter 3: Research Design

This chapter describes the methodological approaches to conduct the research based on three levels of abstraction starting from the epistemological position (Section 3.2), research methodology (Section 3.3), to the specific research techniques used (Section 3.4).

3.1 Overview of the Research Design

This chapter adapts the framework for information system research methodology from Cecez-Kecmanovic (2011) cited in Weber (2017) to present the major abstractions of the research design. A research framework consists of a system of concepts, and their definitions to relevant literature, or existing theory that is applicable to a particular study (Bon, 2007). As described by Saunders (2007), the research framework with an epistemological position is important to guide the way for creation of knowledge. The choice of epistemological position also helps to further define the methodology such that the merit of the research can be validated. *Figure 3.1* illustrates the framework from higher to lower level of the three major abstractions for the research design.

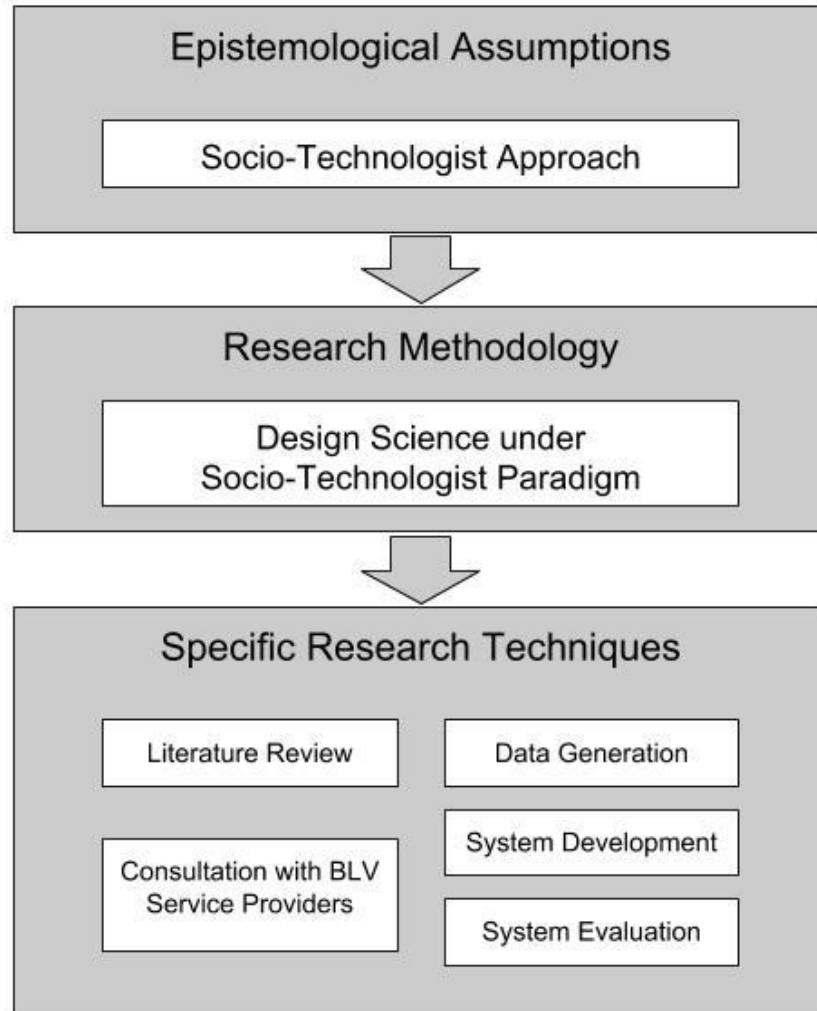


Figure 3.1: Research design framework adapted from Cecez-Kecmanovic (2011) cited in Weber (2017)

Based on Figure 3.1, the first level of the framework addresses the design science epistemological position of the research (more on Section 3.2). The stance on epistemology indicates the view of knowledge in this research, as well as how such knowledge is obtained and synthesized. In this research, the epistemology is positioned on the socio-technologist (which is also known as developmentalist) assumption. The second level of the framework presents the design science research methodology (Section 3.3). The research adapted the process models from Vaishnavi and Kuechler (2015) and Mettler et al. (2014), and combined them to suit the research process. The third level of the framework details the specific research techniques (Section 3.4). These techniques include literature review, consultation

with the BLV service providers, data generation, technology-based system prototype development and evaluation.

3.2 Epistemological Position: A Socio-Technologist Approach

Epistemologically, this research takes the socio-technologist (developmentalist) position. The research would discover knowledge through the process of developing a prototype and all its supportive activities and circumscription. In other words, the research perspective is socio-technologically enabled. It involves the development of a technological artifact (the prototype) that would be impacting a specific group of the society (the BLVs).

The objective of the research is to produce descriptive knowledge through artifact creation in order to answer the primary research question – “how can a technology assist the blind and low vision people to negotiate surface discontinuities along their navigational pathway”. Based on Vaishnavi and Kuechler (2014), a socio-technologist approach would involve the study or measure of artifactual impacts (of the technology) on the composite system (the targeted socio entities). Hence, it is typical that both the artifact and the composite system will be given attention under such an approach. For this research, after assessing the scale of works required as well as the constraints, it placed the development and evaluation (of the performance) of the artifact as the focus. Meanwhile, for the socio part, the research consulted the BLV service providers at the preliminary phases to gather information about surface discontinuity issues and proposed the human-computer interaction (HCI) aspects as future work.

The following sections offer some background on design science research and review the main aspects of the socio-technologist epistemology that are relevant to this research. It is then followed by a section for the justification of the choice of design science methodology.

3.2.1 Background on Design Science Research

Gregg et al. (2001) was the first to term the socio-technologist / developmentalist approach and added it to the epistemological assumptions of design science research that is contrasting positivist and interpretive approaches to research. Epistemologically, in this research, the researcher knows that the information is factual and knows further about that information via the process of development or circumscription. Vaishnavi and Kuechler (2015) compiled a summary of Gregg et al. (2001)'s metaphysical assumptions of the three "ways of knowing" and provided additional insights about these assumptions. Based on the table in Vaishnavi and Kuechler (2015) p. 31, none of the ontology, epistemology or axiology of the paradigms is derivable from any other. Ontological and epistemological perspectives shift in design science research as the projects moves through circumscription cycles described in Vaishnavi and Kuechler (2015) p. 15.

Regarding the contribution of design science research project, Gregor and Hevner (2013) pointed out that the output comes in the form of artifacts (constructs, models, frameworks, architectures, design principles, methods and/or instantiations) and design science theories to address a problem of the research interest. Such research questions are typically problem-solving oriented. *Figure 3.2* shows a knowledge contribution framework adapted from Gregor and Hevner (2013). For this research, the contribution can be categorized as "Adaptation" based on Gregor and Hevner (2013)'s framework (the quadrant highlighted in yellow in *Figure 3.2*). They defined "Adaptation" as a non-trivial or innovative adaptation of knowledge or solutions for new problems. This definition suits the research because the underlying idea of this research is to develop a prototype based on the adaptation of current knowledge to address a newly identified issue (the negotiation of surface discontinuities by BLVs). The outputs from this research are therefore centred around the design and development of the prototype. Artifacts of information system or information technology, as described by Hevner et al. (2004), can be constructs, models, methods or instantiations. The artifacts produced in this research would be further discussed in Section 3.2.4.

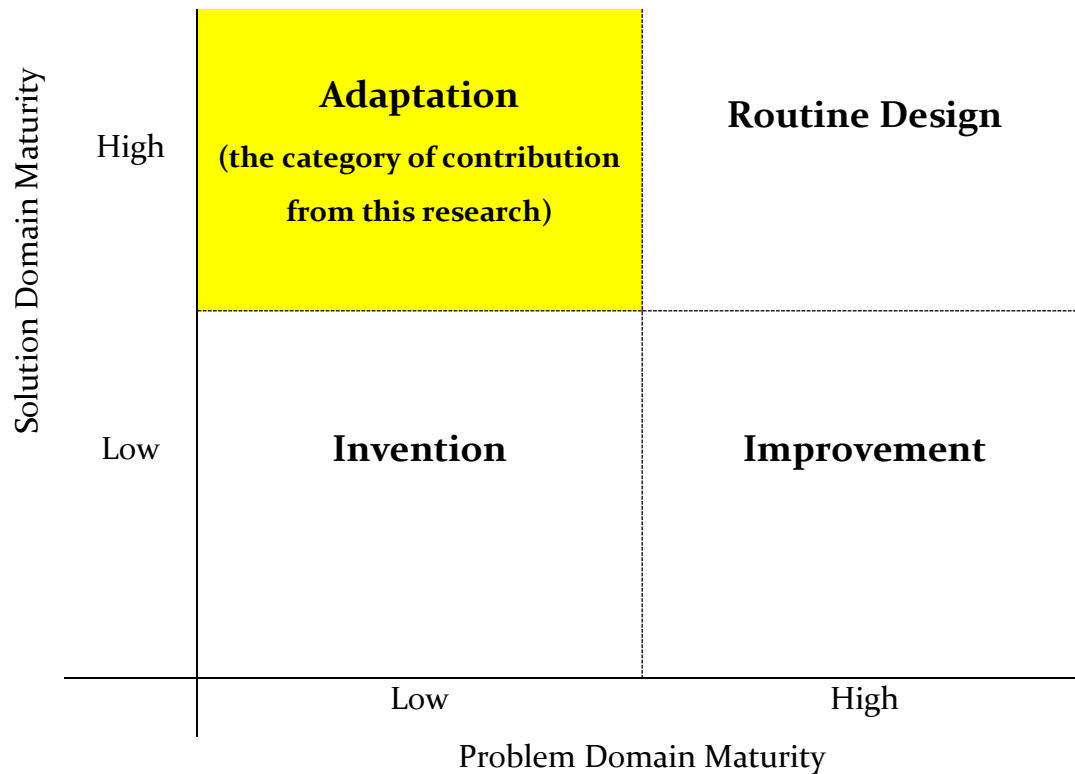


Figure 3.2: DSR knowledge contribution framework adapted from Gregor and Hevner (2013).

3.2.2 Justification of the Choice of Design Science Methodology

This research has chosen design science methodology based on four main justifications:

- the multi-paradigm of the sets of research phenomena,
- the sets of research phenomena are about the “science of the artificial” concerning “how things might be” (Simon, 1996),
- the main objective of developing a technology-based system prototype, and
- the problem-solving orientation of the research.

The multi-paradigm of the sets of research phenomena. In this research, the philosophical perspective of the researcher changes as the research progresses iteratively through the phases (process steps) described by Vaishnavi and Kuechler (2015), p. 15. The research would start with an abduction phase when suggestion of a prototype is proposed

after identification of the research problem. During this phase, the suggestion is formed from some pieces of factual information and it contains some predictions (i.e. the choice of algorithms, sensors and platforms to develop the prototype) set out during this abductive phase. In the following phases of developing and evaluating the prototype, the research would become a deductive process in which iterative progress takes place to build and improve the prototype. The evaluation becomes the basis for a new beginning of the next cycle of the iterative research process. The process continues until termination due to predefined conditions. Through the end of the research process, observations are interpreted. At this point, the research perspective is closer to an interpretive paradigm. The interpretations would become the foundation for new theories that contribute to design science knowledge. Such ontology agrees well with a design science research perspective.

The sets of research phenomena concerning “how things might be”. Knowing through making is the epistemology of the research. Through developing the prototype, knowledge is discovered. This core belief sets the research to ask questions of “how things might be, by design” rather than “how things are” as observed in most intellectual investigations from the fields of natural or social science. Based on the primary research question – “How can a technology assist the blind and low vision people to negotiate surface discontinuities along their navigational pathway?” – the concern is placed on “how a technology might be developed” to address the specific issue faced by the BLVs.

The main objective of developing a technology-based system prototype. The research aims to address the surface discontinuity issue faced by the BLVs through the development of a technology-based system prototype. Data would be generated to establish a taxonomy of surface discontinuity, which would be used to train some machine learning models that form the core functionalities of the system. Evaluation would be performed to understand how well such a prototype can classify the surface discontinuities based on the established taxonomy. Therefore, the prototype as the central artifact, is the

core output of the research. This also indicates that the nature of the research is design. The design science methodology fits well to conduct research of such a nature.

The problem-solving orientation of the research. The research is oriented towards a problem-solving paradigm. This orientation differs greatly with behavioural science paradigm, which is commonly based on cause-and-effect reasoning model to explain and/or predict phenomena instead of problem-solving. In contrary, this research solves a problem through the development of a technology. Design science methodology offers a more relevant approach for this research with problem-solving orientation.

3.2.3 Scope of the Research

The scope of the research is defined in the following points.

- The subject of the research: a prototype of wearable technology-based system
- The knowledge base: literature of relevant technologies and consultation from the BLV service providers at the early phases of the research
- The dataset:
 - Environment type: limited to urban built-environment
 - Location for sampling: Klang Valley, Selangor, Malaysia

The development of the prototype is focused on the assembling of suitable machine learning algorithms that would be used for models training, after acquiring the dataset. Two types of evaluation namely accuracy and efficiency tests would be performed on the prototype to understand its capabilities. The choice of location for data sampling is based on two main reasons – (1) the proposed location has high density of relevant samples, and (2) ease of access.

The development of user interface (UI) or the study of human-computer interaction (HCI) for the prototype would not be part of the research as it could be different in the nature of

study. Time constraint is another factor of consideration to exclude the UI or HCI in this research. Having said that, a minimal user interface would be included such that the prototype can be properly operated and tested for its performance.

3.2.4 Output of Design Science Artifacts from this Research

Design science outputs can be in the form of artifacts (constructs, models, frameworks, architectures, design principles, methods and/or instantiations) and design theories (Vaishnavi and Kuechler, 2015). Research outputs can also be classified into three main levels of abstraction (or generalization) ranging from fully developed design theory to implemented artifacts based on Purao (2002)'s design science knowledge hierarchy.

For this research, the main outputs are design science artifacts. Purao (2002)'s design science knowledge hierarchy diagram is adapted to depict the outputs generated by this research as shown in *Figure 3.3*. There are four main types of design science output produced by the research namely construct, model, method and instantiation.

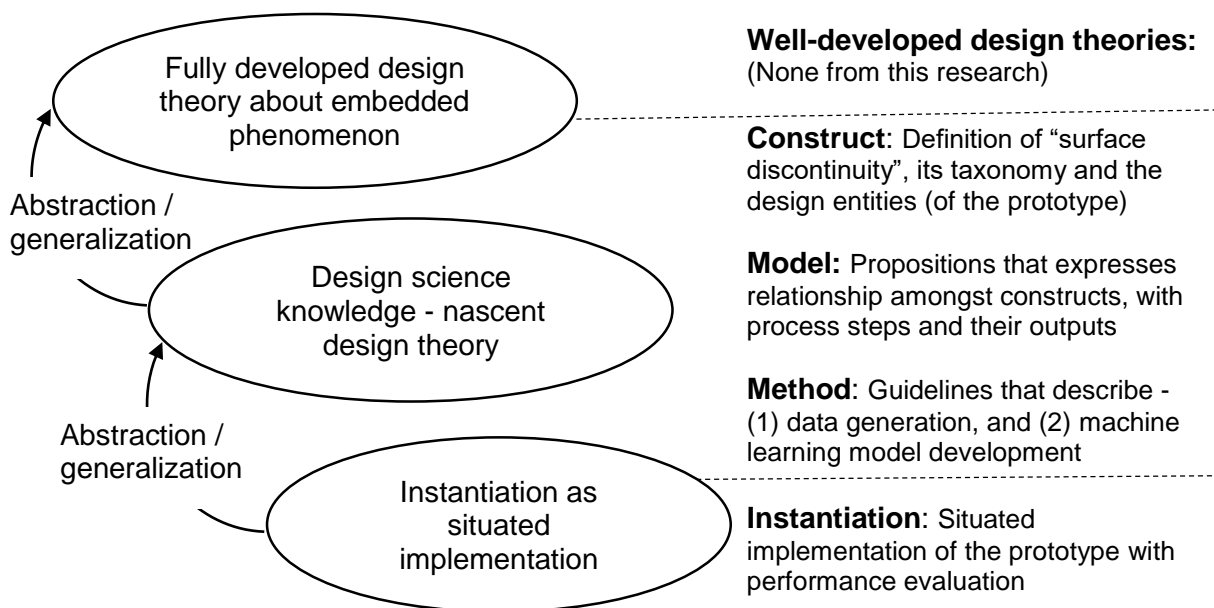


Figure 3.3: Research outputs diagram adapted from Purao (2002).

The design science artifacts produced by this research are described as follows.

Construct. Based on March and Smith (1995), constructs are the conceptual vocabulary of a problem or solution domain. They are languages in which problems and solutions are defined and communicated, defined Hevner et al. (2004). In this research, the definition of “surface discontinuity” pertaining the BLVs navigation issue together with the “taxonomy of surface discontinuities” is a set of constructs which defines the problem domain and conceptualizes the classes of the problem into a taxonomy based on their physical properties. The taxonomy was also developed to exemplify the issue of surface discontinuity sampled from the field, and it guided the generation of the dataset, as well as facilitated models training in a later phase. The design entities of the prototype (as the solution) are also amongst the constructs produced by the research.

Model. A model is a set of propositions or statements expressing relationships among constructs (Vaishnavi and Kuechler, 2015). This research adapted a model from Mettler et al. (2014) and Vaishnavi and Kuechler (2015) that expresses propositions and statements of relationships among the constructs. The model also illustrates the research process steps involved, with additional mapping of the research outputs to each of the steps.

Method. Hevner et al. (2004) defined research method as guidance on how to solve problems. Vaishnavi and Kuechler (2015) summarized methods as “sets of steps used to perform tasks”. In other words, it is “how-to” knowledge in performing a task. The method used to generate dataset in this research contributes a detailed descriptive guideline that includes the development of a data collection instrument (termed as the phase-1 prototype in this research), instrument setup, crowdsourcing for locations with high density of samples, data sampling, ground truth measuring and data pre-processing. The research demonstrated that this method can be replicated to add new data for future study, or for different projects with a similar nature by the research community. The research also produces a method used to assemble and experiment some machine learning algorithms,

and train and optimize the best model from the algorithms which would form the core element of the prototype.

Instantiation. March and Smith (1995) stated that the final output from a design science research effort is an instantiation that “operationalizes constructs, models and methods”. It is seen as the realization of the artifact in an environment. The development (with design and re-design cycles) and evaluation of the prototype at several selected and random locations is an instantiation from this research. The instantiation (which is better phrased as situated implementation) is the most emphasized artifact in this research. It is unlikely that the understanding of “how can a wearable technology be built to assist the BLVs” would ever have occurred in the absence of this working artifact.

3.3 Process of the Design Science Methodology

In this section, a model that describes the research process is provided. This model is a combined adaptation of a general design science research process model described by Vaishnavi and Kuechler (2015) and a specific model of experimentation with design suggested by Mettler et al. (2014).

Vaishnavi and Kuechler adapted a general process of a design science research from a computable design process developed by Takeda et al. (1990). The model elaborates both the knowledge using process and the knowledge building process at several different stages starting from awareness of problem, suggestion, development, evaluation to conclusion. The model also illustrates the outputs (starting from proposal, tentative design, artifact, performance measures to results) corresponding to each process stage and the knowledge flow.

Mettler et al. (2014) proposed the idea of “design experiments” as a design science research synonym for a scientific method that facilitates causal interference while testing the effects

of design alternatives by adhering to the principles of (1) control, (2) randomization and (3) manipulation. Their work refers “control” to the presence of a control group in an experimental setting, apart from the test group. “Randomization” as they put it, is the random assignment of subjects (participants or data sets) to either the test or control group to uniformly distribute the known and unknown confounding factors for some more stable results. Lastly, they defined “manipulation” as a means to evaluate the artifact under differing conditions. These three principles must be carefully addressed in the three basic phases of a design experiment from the pre-experimental phase, actual experimental phase to analysis phase. With these three phases, the role of a researcher differs in terms of planning and conducting the tasks.

The research observed the need of having elements from the two models discussed above because the research emphasizes the instantiation which involves two primary steps: (1) development, and (2) evaluation of the prototype. For these two steps, the researcher is directly involved. The researcher would also play some distinctive roles during the planning versus the conducting phase of these two steps. A model that can reflect the researcher as an important actor with different actions throughout the research process is hence very much practical. Apart from these two emphasized steps, the project is still a design science research that needs to be guided thoroughly, in which other process steps such as awareness of problem, suggestion and conclusion should not be given least attention.

Vaishnavi and Kuechler’s model offers a thorough process of conducting a general design science research, but it lacks description of the involvement of the researcher with distinctive roles in both the planning and conducting phases. Mettler’s model places emphasis on the experimentation with the researcher being involved as both the planner and conductor (executer) in the process, but it lacks a multiplicity guideline (that describes an overall process starting from the awareness of problem to the conclusion step) needed by this research. With a combined adaptation from both models, a process model is proposed as illustrated in *Figure 3.4*. Although the figure illustrates the process flows of the

research specifically, it is difficult to trace the knowledge flows, process steps and outputs explicitly. Thus, some of these relevant descriptions are provide in *Table 3.1* as an additional material.

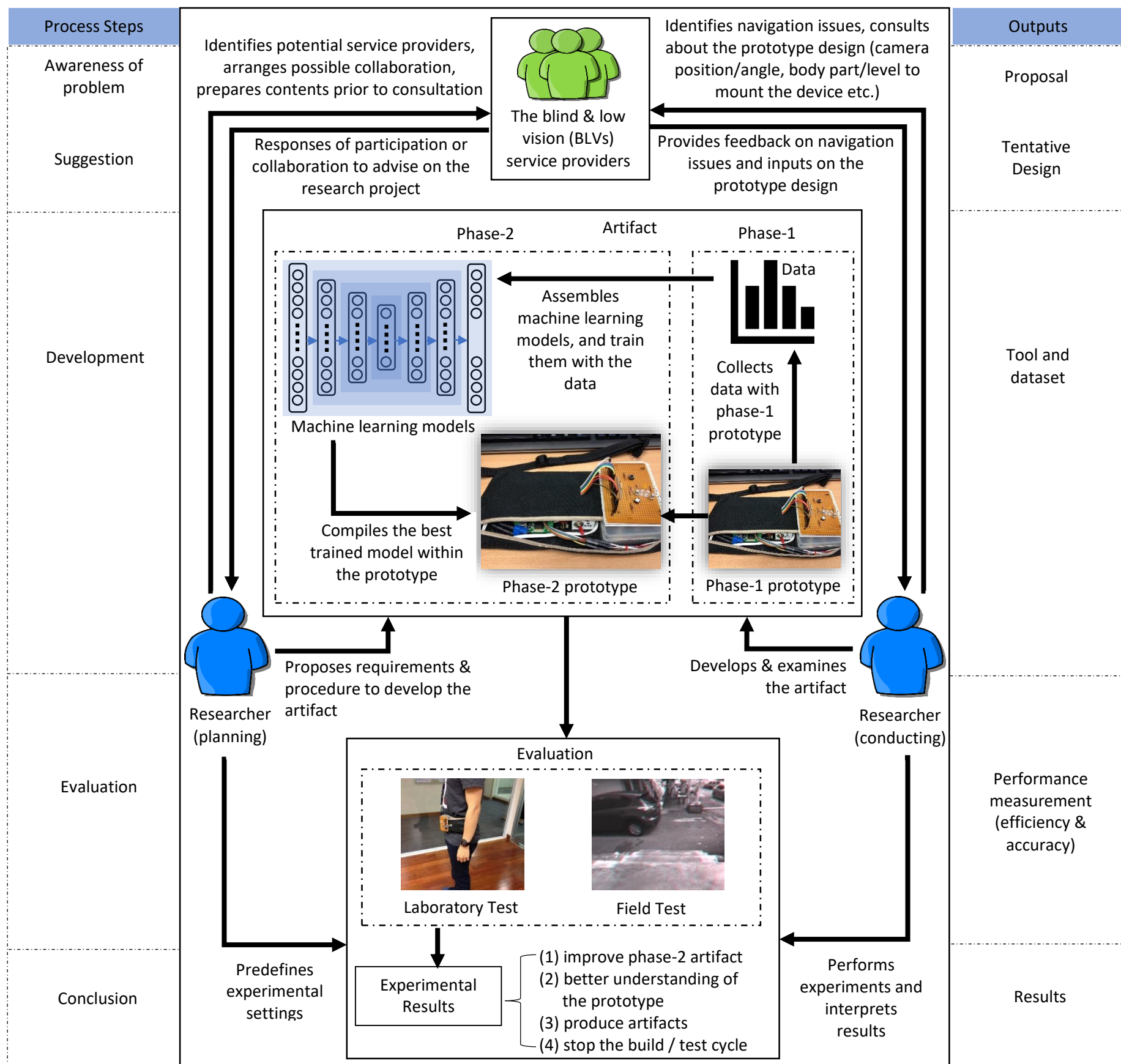


Figure 3.4: Research process model adapted from Mettler's model with process steps and outputs suggested by Vaishnavi and Kuechler's mapped on both left and right sides.

Table 3.1: Additional details of the proposed research process model illustrated in Figure 3.4.

Process Steps	Research Propositions	Resources / actions needed	Outputs
Awareness of problem	The blind and low vision people (BLVs) face challenges when the walkway is hazardous due to surface discontinuity (e.g. uneven down-steps, uncovered drainage, sudden drop-off etc.)	a) Acquire inputs from blind and low vision service providers for problem formulation b) Survey the mentioned problem in the field c) Explore available samples (locations) for possible data collection which will be needed later d) Review of relevant literature	<u>Proposal:</u> A study of possible technological solutions to the problem identified
Suggestion	Design a technological solution to address the identified problem of surface discontinuity.	e) Review current technologies f) Design a practical solution g) Source for available hardware components within allowable budget (i.e. sensors, processors & batteries) based on the design in (f)	<u>Tentative Design:</u> A design of wearable assistive technology-based system prototype
Development	Based on the proposed design, develop a prototype to address the problem of surface discontinuity.	h) Develop the phase-1 prototype meant for data collection by assembling some hardware and software. Along the development, continuously acquire / identify a few parameters from the BLV service providers. These parameters such as the placeable position of the prototype on the user's body, the camera's position/angle, body part/level to mount the device will affect the next phase of the prototype development too. Questions like "should it be placed near the chest, head, arm or embedded within the guide cane" will be asked.	<u>Tool and dataset:</u> <ul style="list-style-type: none"> - phase-1 prototype for data collection - a dataset - a trained machine learning model - phase-2 prototype built with a trained machine learning model and simple user

Table 3.1 (continued): Additional details of the proposed research process model illustrated in Figure 3.4.

Process Steps	Research Propositions	Resources / actions needed	Outputs
		<p>i) Execute data collection from the identified locations in (c) using phase-1 prototype developed in (h)</p> <p>j) Assemble, code and train machine learning models from the collected data in (i). The data will be pre-processed prior to the training.</p> <p>k) Compile the trained machine learning model into the phase-1 prototype and repurpose it for surface discontinuity detection, as opposed to its original purpose which was built only for data collection. Now it is made phase-2 prototype (the hardware components and parameters are mostly unchanged).</p>	interface for surface discontinuity detection
Evaluation	The performance of the prototype can be evaluated through: (1) lab test, (2) field test	<p>l) Predefine the experimental settings and test parameters i.e. the number of tests needed, locations of the field test</p> <p>m) Execution of both lab and field tests</p>	<p><u>Performance Measurement:</u></p> <p>Performance of the prototype in terms of its accuracy and efficiency</p>
Conclusion	The proposed prototype should be able to detect surface discontinuities along the walkway with certain performance and possible limitations.	<p>n) Interpret the results</p> <p>o) Synthesis of the evaluation</p>	<p><u>Results:</u></p> <ul style="list-style-type: none"> - An improved phase-2 artifact - An understanding of the capabilities and limitations of the prototype - Design science artifacts

3.4 Specific Research Techniques

Based on the third level of abstraction from *Figure 3.1*, there are five main specific research techniques involved in this research namely literature review, consultation with BLV service providers, data generation, system (prototype) development and evaluation. Literature review is already provided in Chapter 2, hence it will not be described in this section again. The remaining four techniques are provided in the following sections, with several individual chapters (Chapter 4 to 6) to further explain their detailed implementations.

3.4.1 Consultation with BLV Service Providers

The BLV service providers were consulted at the early phases to identify the problem, understand more about the guide cane as the major aid for navigation, and communicate suggestions of a solution. It is through these consultations that the following insights were gained and benefited the research.

Awareness of Problem. Firstly, the awareness of the problem – surface discontinuities – was identified through several preliminary consultations with members from two BLV service providers namely Dialogue in the Dark Malaysia and St. Nicholas' Home Penang. They are two of the local service providers with very active involvement in looking after the welfare of the BLVs. Both organisations had supported the research by volunteering their members to participate in consultation at some early phases of this research. Apart from these two local service providers, an eventual visit to Vision Australia in Melbourne (a major BLV service provider in Australia with many branches across the country) had also concurred that the issue of surface discontinuities is one of the major threats to blind navigation, even in a developed city like Melbourne. Through the three service providers mentioned, the research had gained insight about the difficulties of a guide cane in detecting certain types of surface discontinuity.

Prototype Design and the Guide Cane. Based on literature review and consultation with the service providers, it was learned that in most cases, the guide cane must not be abandoned by the BLVs even if a new assistive device was to be introduced. Hence, the design of the prototype must include this consideration. An orientation and mobility (O&M) specialist from Vision Australia had demonstrated the formal usage of different types of guide cane for blind navigation to the author. The O&M specialist had also personally conducted a short training for the benefit of the author on the guide cane usage in some urban outdoor environments. This experience and guidance had later helped in the designing and considerations of the wearable prototype, to make sure the prototype is complementing the guide cane, but not replacing it. Other issues during the design phase, i.e. to make sure the prototype does not obstruct the user's guide cane operation, were also considered. This knowledge has eventually shaped the positional setup of the wearable prototype at the chest level of the user, which is deemed to be least obstructive to the user.

Getting Feedback on Surface Discontinuities Data through 3D Printing. One of the most important items before data collection was to understand more precisely the types of surface discontinuity that is deemed hazardous by the BLVs during their outdoor navigation. To achieve this purpose, a better approach of gaining more insights from the BLVs is critical.

Obviously, it is impossible to visually show the BLVs some pictures or video footages of surface discontinuity sampled from the field. Thus, an alternative approach is to let them touch and interpret some 3D printed replicas of the surface discontinuities, where they could form some mental pictures of the samples. During this pre-data collection stage, such communication is important as it would help the research to identify the right targets to be sampled from the field.

The following Figure 3.5 to Figure 3.10 are some examples of targeted samples, ranging from their original images, 3D models, to the 3D printed replicas. All the printed replicas are

proportionately scaled down from the actual measurements taken from the field. During the consultation, these 3D printed replicas were presented to a group of BLV volunteers and they interpreted the details they could touch and sense with their fingers. With their feedback, several types of surface discontinuity were identified for the next stage of data collection.



Figure 3.5: A sample (let's name it sample A) of steps from three different angles.

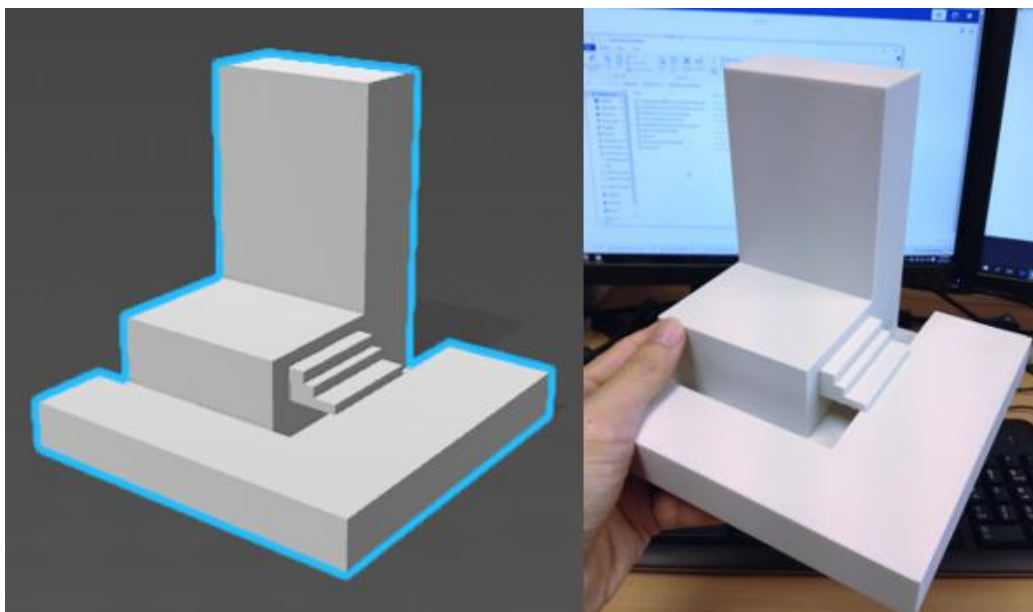


Figure 3.6: The 3D model and printed replica of Sample A. The 3D printed replica is proportionately scaled down from the actual measurements taken from the sample



Figure 3.7: Another sample (let's call it Sample B) from two different angles.

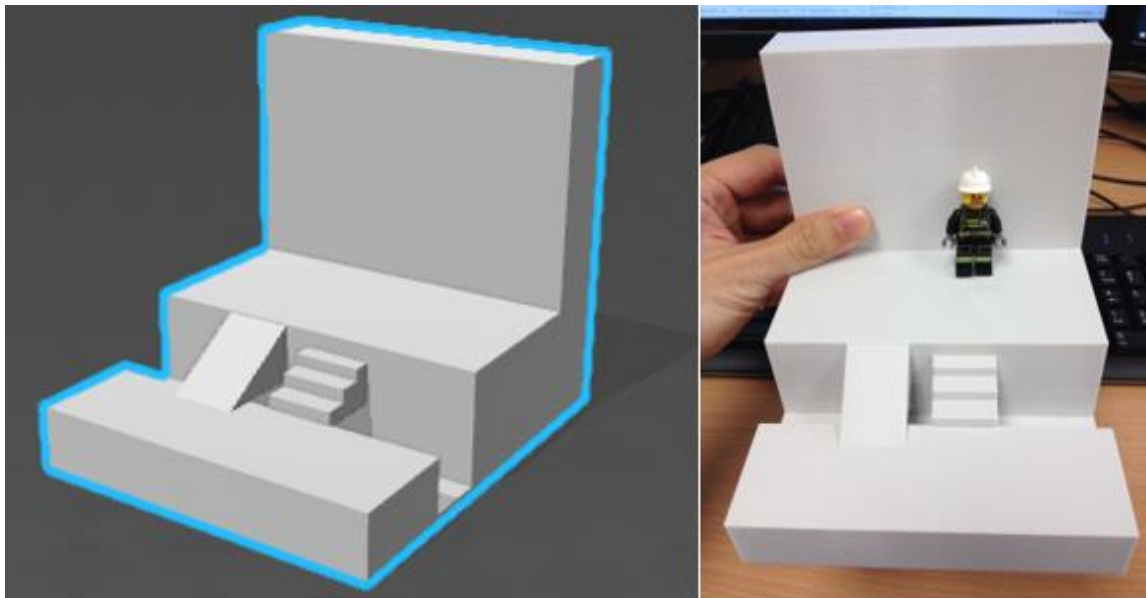


Figure 3.8: The 3D model and printed replica of Sample B.



Figure 3.9: Another sample (let's call it Sample C) from two different angles.

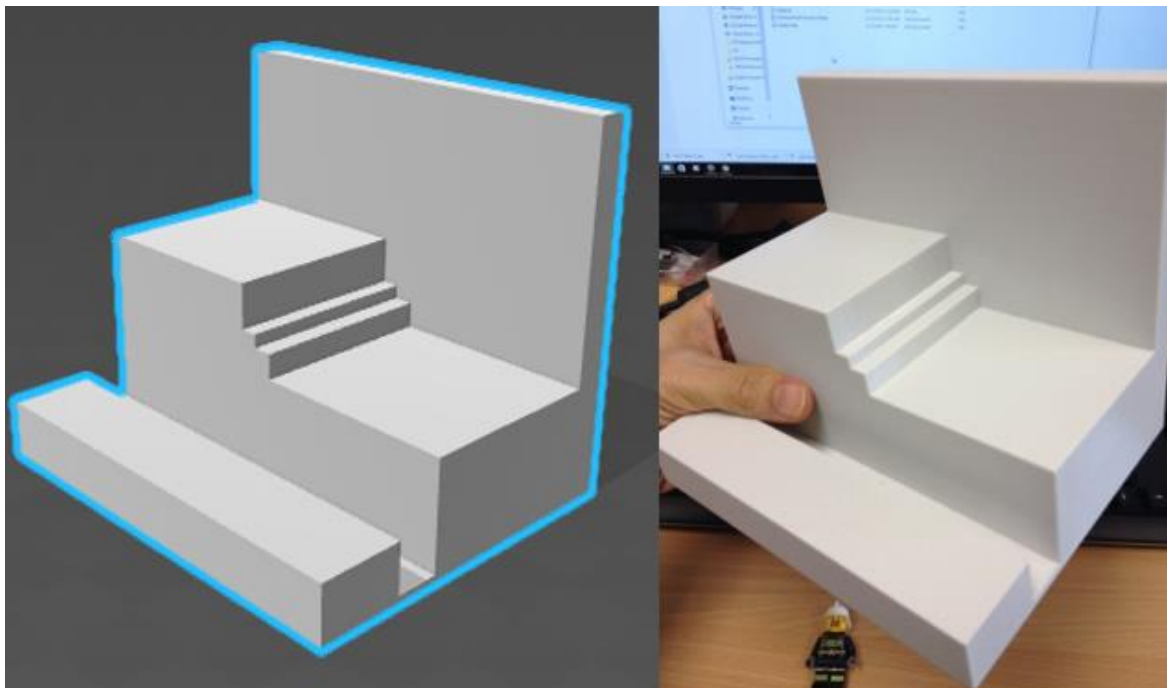


Figure 3.10: The 3D model and printed replica of Sample C.

Based on the feedback, it was found that most of the surface discontinuities that could potentially be hazardous to the BLVs belong to the drop-off types, i.e. uncovered drainage and uneven down-steps. Section 4.5 describes a taxonomy of surface discontinuities built with reference from this feedback.

3.4.2 Data Generation

The data generation technique consists of several parts that are inclusive of development of the data collection instrument (also termed as the phase-1 prototype), setting up and positioning of the instrument, ground truth measuring, data collection and pre-processing. Since data generation is a major contribution from this research, detailed descriptions are given in Chapter 4.

3.4.3 System Development

The system development mainly refers to the repurposing of the phase-1 prototype used for data collection, into the phase-2 prototype meant for classifying the surface discontinuities at field evaluation. This technique involves feature extraction from the generated data, assembling of machine learning algorithms, and training of the machine learning model. This technique is further described in Chapter 5 as it is one of the major works from this research.

3.4.4 System Evaluation

Having built the system (phase-2 prototype), the research sets to measure its performance in the field. Apart from tuning the experimental models, the prototype was evaluated for its accuracy in classifying the test data into their respective taxonomy, and its efficiency in model design, power consumption and speed of classification. Chapter 6 discusses the evaluation technique in detail.

3.5 Summary of the Chapter

This chapter has described the research design based on three abstractions – the socio-technologist epistemology, the design science research methodology and the specific research techniques. Descriptions and justifications for each of the research design considerations were provided, while the details of data generation, system (prototype) development and evaluation will be covered in the following 3 chapters.

Chapter 4: Data Generation

This chapter presents data generation method applied in the research. The method involves several techniques for instrument development (Section 4.1), instrument setup (Section 4.2), data collection (Section 4.3), data pre-processing (Section 4.4) and data augmentation (Section 4.7). One of the major works produced from this research – the development of taxonomy of surface discontinuities – is also presented in this chapter (Section 4.6).

4.1 Instrument Development – the Phase-1 Prototype

The development of a data collection instrument, which is named as the “phase-1 prototype” (to differentiate it from the final prototype) is described here. The prototype was first needed as an instrument to collect the surface discontinuity samples from the field. It is this similar prototype that will be repurposed in a later phase (named as the “phase-2 prototype”) for performance evaluation. *Figure 4.1* shows the structure diagram.

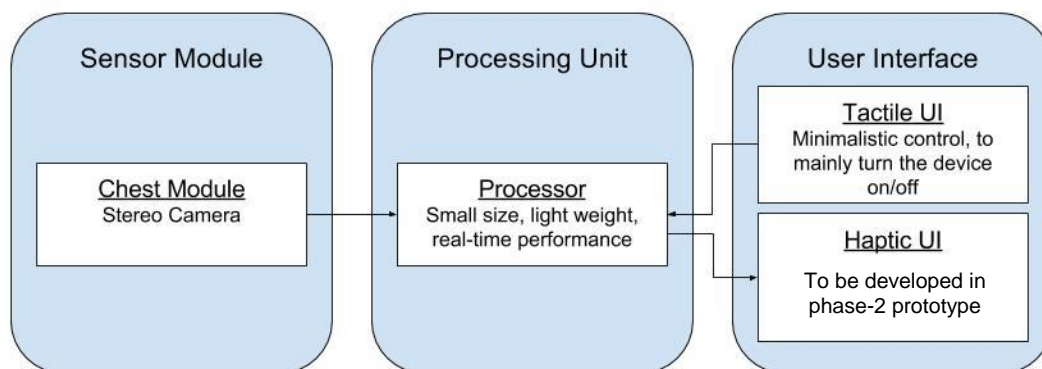


Figure 4.1: Structure diagram of the phase-1 prototype.

As discussed in Chapter 2 that the BLVs would not abandon their guide canes while adopting some new technologies, the similar approach to extend a traditional guide cane with extra functionality as suggested by Benjamin et al. (1973) and several others (Borenstein, 2001, Cang et al., 2014, Dang et al., 2016, Hoyle and Waters, 2008, Kim et al., 2014, Kuramochi et al., 2014, Orita et al., 2013, Pallejà et al., 2010) was first considered in this research. However, it was soon realized that this approach is not suitable due to the following reasons. In the above works that extended the guide cane with extra functionality, the two main sensors used were either laser or ultrasonic. These sensors were literally mounted on the guide canes. During the walk, the sensors will be experiencing additional movement every time when the cane is swung to left and right in a typical maneuver. The additional movement is helpful for the laser and ultrasonic canes to detect obstacles as it widens the sensor's resolution which would result in higher possibility of detecting obstacles along the pathway. This additional movement as mentioned above is not ideal for the stereo camera (the sole sensor) used in this research. Since inputs from the stereo camera are images, it is crucial that the captured images are of good quality for subsequent processing. Any additional movement was observed to have high possibility of causing the captured images to be blurry or defocused. Due to this consideration, the idea of extending the functionality of a guide cane is not suitable here.

The stereo camera must be set at a position that experiences the least effect from user's body movement. Based on Rodriguez et al. (2012) and Pérez-Yus et al. (2015), the chest area of a person would be a strategic location for camera sensor. This area is less affected by the body movement when one is walking in a linear direction. In addition, the typical movement of arms during a walk has little disruption to the view of the camera at this area. Finally, it was decided that the stereo camera is best to be mounted at chest level of the user in the form of a wearable prototype. Descriptions about the camera and instrument setup are provided in Section 4.2.

To operate the sensor and other processing tasks in the form of a wearable prototype, a small computer platform was required. There were two main choices of platform for consideration – (1) off-the-shelf mobile development platform such as Google’s Project Tango (Kastrenakes, 2014) or (2) custom-made platform based on single-board computer. At the time of the prototype development in 2015, Project Tango was one of the most relevant candidate considered amongst the limited off-the-shelf platforms. It has several features that focus on enabling mobile devices to use computer vision to support the development of indoor navigation, 3D mapping, virtual reality applications and etc. Google released two Tango enabled devices to the public. The first was Lenovo’s Phab 2 Pro, and next was the Asus Zenfone AR. Both devices had 16 to 32 mega-pixel camera, depth sensing infrared camera, motion tracking sensor and several other sensors. Since the built-in depth sensor was based on infrared (structured light), it is not suitable for outdoor depth sensing which is the case in this research. Plugging-in of third-party sensors on these devices was difficult to almost impossible due to the lack of support from both the vendors and Project Tango’s official website. Without a workable stereo camera (or depth sensor) for outdoor environment, such platform was withdrawn from consideration for this research.

The selection of platform was then switched to some single-board computers. Single-board computers from several makers such as Raspberry Pi(2018b), Orange Pi(2018a), LattePanda(2019), Odroid(2015b) and several others were compared to identify the most suitable candidate for this research. Based on the operating system compatibility (between the computer and the stereo camera module), hardware specifications (processor and graphical support), form factor and price point, Odroid XU3 was identified to be a better choice for image processing task on a wearable prototype that required high processing power with relatively low power consumption. The stereo camera module acquired for this research was also compatible with Ubuntu, an open-source Linux distribution operating system recommended for Odroid XU3.

The detailed hardware specifications of Odroid XU3 and the stereo camera are given in Appendix 1. To acquire the best quality of data, the stereo camera was set to capture the highest resolution of images that it is capable of (which is 752 x 480 pixels), and the raw images were saved as uncompressed bitmap format (instead of compressed videos). A single-board computer with hardware specifications documented in Appendix 1 is used to implement the phase-1 prototype. An application to launch the camera was developed using C/C++ supported by OpenCV libraries (This is a library of programming functions mainly aimed at real-time computer vision). An algorithm was included to handle the auto-exposure of the camera as the data collection would be done at mostly outdoor environments, where the amount of sunlight could be varied according to time and place. Flow chart in *Figure 4.2* shows the algorithm for the application built for data collection.

The interface of the phase-1 prototype is minimalistic. It has a main power switch and another switch to operate the camera. Several indicators were added to inform the user about: (1) power status, (2) readiness of the system for data capturing, and (3) data storing status during data collection. Details of the development of the circuitry with electronic components (and the camera sensor) connected to the single-board computer are explained and shown in Appendix 2.

To optimize the angle of the camera's view, the camera is mounted on a chest harness which is adjustable in every direction – making it flexible for tuning up to the best angle. At this level, it offers least obstruction and inconvenience to the user. Realizing that the BLVs will be using this device without abandoning their guide cane, the author experimented the camera's position with a cane swinging and tapping to left and right. More descriptions about the instrument setup is given in the next section.

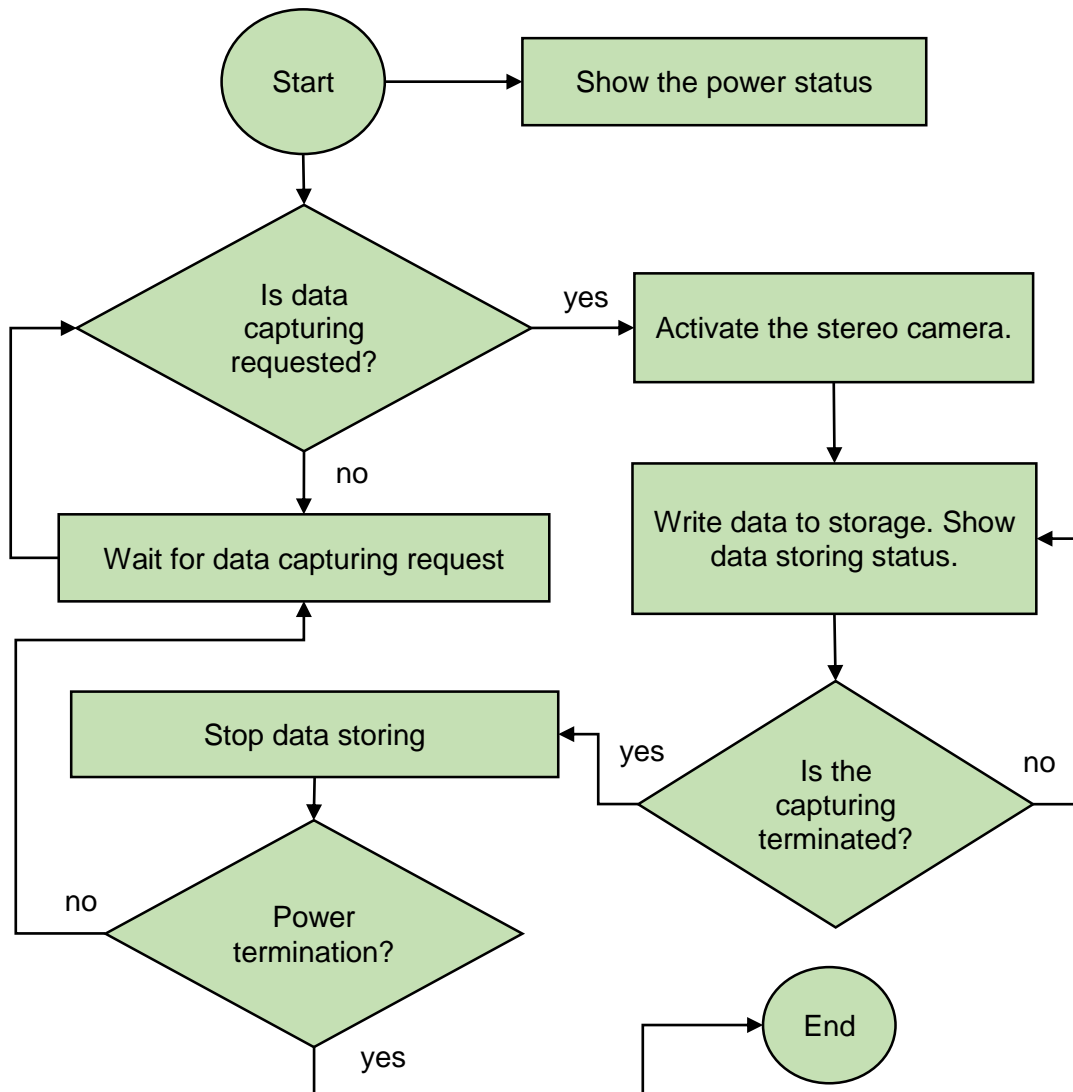


Figure 4.2: Flow chart of the algorithm for the application built for data collection in the field.

At the final stage of development, the phase-1 prototype was tested by five BLV volunteers at their centers for further opinions and comments (mainly to validate that the design is not obstructing their guide cane usage). Only one improvement was suggested – that the BLVs will commonly carry a backpack when going outdoor, thus the belt of the camera mount should be re-designed such that it doesn't require the BLVs to remove their backpack to put it on. A solution was suggested to make it hangable on the user's neck, instead of mounting around the upper body. Other than that, the wearable prototype was positively deemed to be lightweight, unobtrusive, minimalistic and unmonopolizing the

user's attention. While the current design is quite ready for data collection by a sighted person, the above feedback will be included in the future work when the HCI and UI aspects are given full attention.

4.2 Instrument Setup

The research tried to emulate the actual situation of usage as closely as possible, thus the data was captured along with a guide cane swinging and tapping on the surface of the walkway just like the BLVs would do. Based on training from the O&M specialist mentioned in the previous section, proper handling of the guide cane was practised during every data collection session.

To prepare for data collection, the phase-1 prototype must be placed and set at the optimum position and angle (for the camera). One of the major concerns is to aim the camera's angle at the best position such that it can capture all targeted drop-off types with least occlusion from the user's body parts. Rodriguez et al. (2012) and Pérez-Yus et al. (2015) recommended an angle of 45 degrees facing downward, which was tested here. There isn't much issue to capture elevated types of surface discontinuity based on their setting.

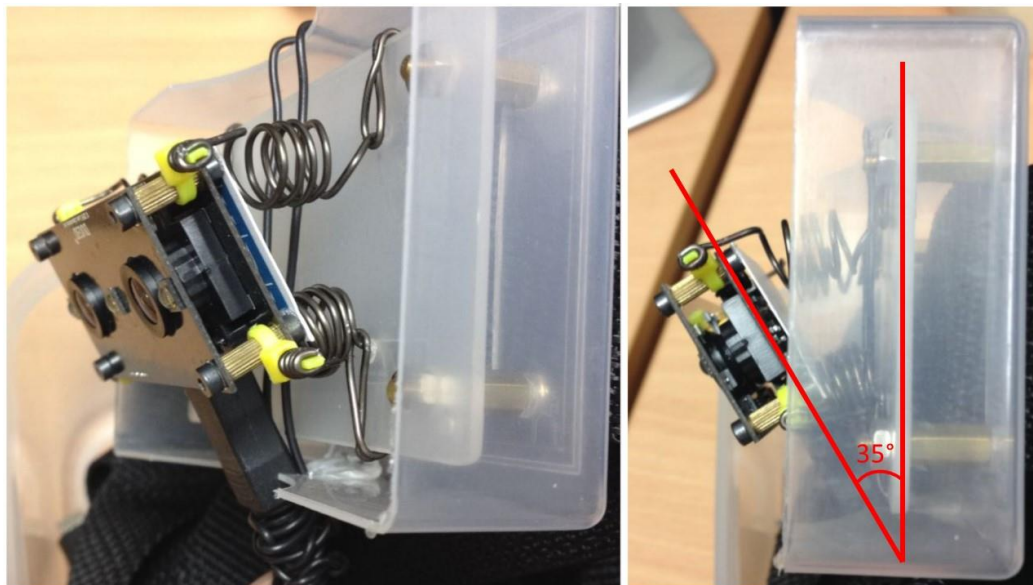


Figure 4.3: The camera angle was set at 35 degrees facing downward after several experiments.

After experimenting with the phase-1 prototype on several locations, the camera was finalized to be set at chest level and point at approximately 35 degrees facing downward (see *Figure 4.3*). At this position and angle, the camera could record most drop-offs without any issues. With 170 degrees of field of view from the camera, it can capture both proximal and distant scenes. However, the guide cane was unnoticed in most of the cases. *Figure 4.4* shows the setup and positioning of the prototype. From the figure, the inserted image A is the stereo camera module is mounted at chest level, and inserted image B is the single board computer is carried inside a waist pouch.

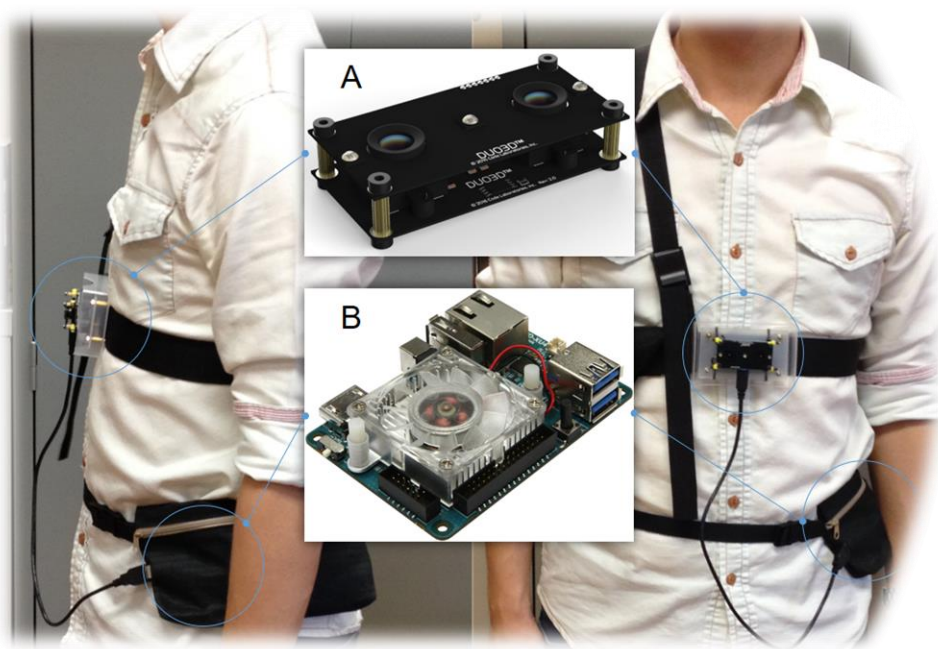


Figure 4.4: The setup and positioning of the prototype.

With this setup, the user is approximately 80 centimeters away from the region of interest (ROI) as illustrated in *Figure 4.5*. This ROI would be translated into the lower half position of an image pair as shown in *Figure 4.6*.

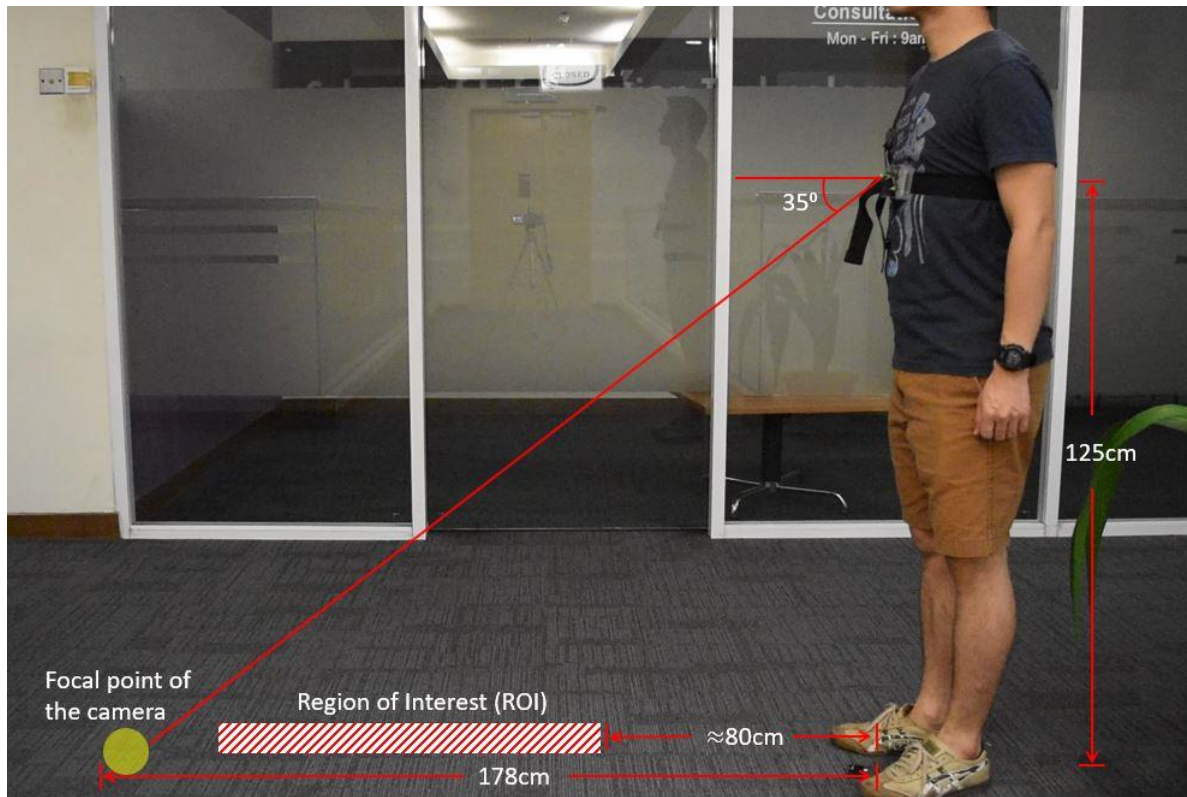


Figure 4.5: Region of interest and distances based on the prototype setup.



Figure 4.6: The ROI in red rectangular outline is centered at the lower half position of Sample 95 – Left Image 141, a typical example of down-steps.

4.3 Data Collection

There were several tasks involved throughout the data collection phase as described in the following sections.

4.3.1 Crowdsourcing to Identify Potential Locations for Data Collection

To identify potential locations with high density of samples around the urban areas, a crowdsourcing technique was applied. Social media pages were set up to advertise the purpose of the research with the intention for the viewers to contribute information of potential locations having high observation of surface discontinuities. The viewers of the social media were encouraged to contribute by snapping photos of surface discontinuities they found along a walkway and share it on the page with caption of the location. By the end of the crowdsourcing activity, a total of 80 locations with different surface discontinuities were contributed by the viewers. Some preliminary observations of these locations were conducted by the research and eventually 10 locations were targeted for actual data collection. Appendix 3 shows the screen shot of a social media page and the advertising details used in this task.

4.3.2 Data Sampling

A total of 225 samples were collected from 10 different locations. The sampling method employed was judgmental sampling, which is a non-probability method based on justification that certain areas could have more samples as compared to others. This is true in most cases because certain urban areas were newly developed and thus new set of building regulations were strictly followed, while certain areas with older buildings or walkways might not have followed proper regulations at the time when they were built. For that reason, the research had to go for more relevant samples at older urban areas, but also included some new areas too as not all developers followed the regulations due to slack enforcement. The following *Figure 4.7* indicates the location of data collected, and *Table 4.1* shows the size of samples taken from each of the locations.

Table 4.1: Distribution of samples from several locations in Petaling Jaya and Kuala Lumpur.

Location	Sample Size
Damansara Perdana	25
Kota Damansara	25
Bandar Utama	25
Up-Town	30
Damansara Jaya	20
Bandar Sunway – PJS 7	20
Bandar Sunway – PJS 9	15
Bandar Sunway – PJS 10	15
Brickfields	25
Pudu	25
Total	225

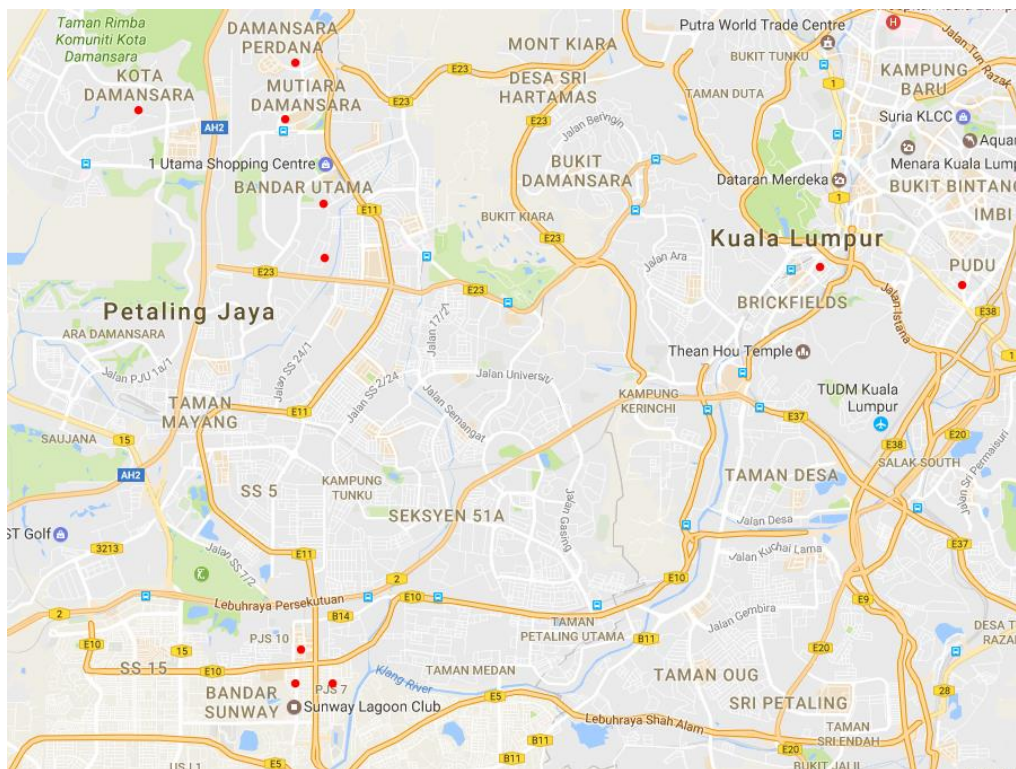


Figure 4.7: Location map of sample distribution (red dots) around Petaling Jaya and Kuala Lumpur.

Next section presents the measuring of the ground truth from these locations where data were sampled.

4.3.3 Measuring the Ground Truth

During data collection, apart from using the phase-1 prototype to capture stereo image sequences from the identified locations, actual ground truth (the dimensions of the sample) was also measured. A common measuring tape was used for this purpose. The dimensions taken are widths, heights or depths of the samples. A simple process of drafting the shape of the sample and labelling the dimensions was used. For example, some of the photos of the samples and the drafted versions with their corresponding dimensions are shown in *Figure 4.8* to *Figure 4.12*.

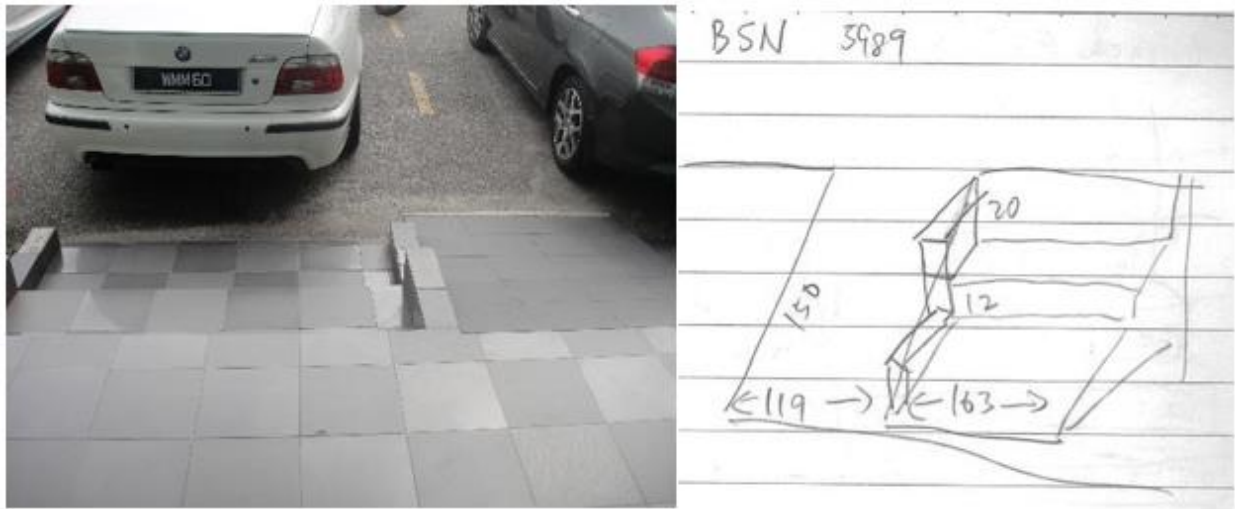


Figure 4.8: A walkway in front of a bank leading to a car park and road. (Location: Up-Town PJ). The dimensions can be seen labelled in centimeters.



Figure 4.9: At one end of a walkway in front of a print shop leading to a road. (Location: Damansara Jaya)



Figure 4.10: A pit along a walkway connecting several private services, uncovered and hazardous. (Location: Kota Damansara)

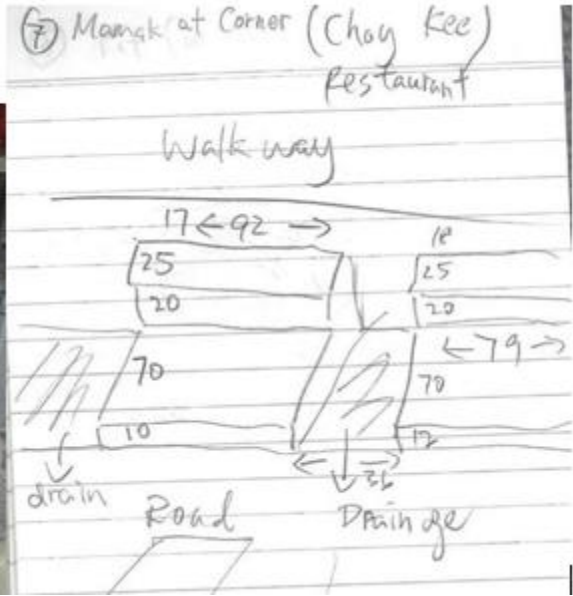


Figure 4.11: Damaged and uneven steps bridging a drainage between the road and walkway along several eateries. (Location: Damansara Jaya)

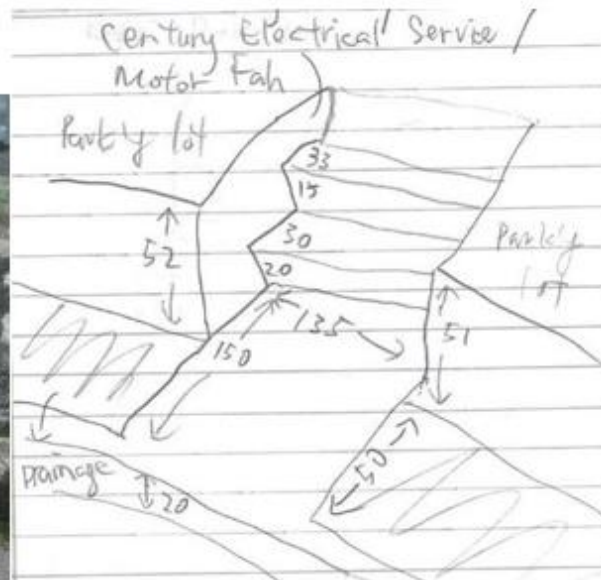


Figure 4.12: Step with drop-offs on both sides bridging a drainage and leading to some shops. (Location: Damansara Jaya)

The ground truths as shown in the above figures were used to validate data annotation task in a later phase.

4.3.4 The Dataset

The dataset contains a total of 225 samples of image sequences of surface conditions from the 10 locations mentioned in Section 4.3.2. These data are short navigations by a sighted person along the walkways. Each sample is a pair of left and right images (hence stereo vision) sequence of a walkway connected to an incoming surface discontinuity. A sample could contain averagely 135 image pairs based on the duration of the recording at the field (the recording was set to 30 frames per second). To maintain the best quality of the images, the camera was set to capture and save the raw images without any compression (in bitmap file format). The image resolution was set to the highest possible as allowed by the camera module, which is 752 x 480 pixels in width and height. The colour mode of the camera is monochrome which is the only mode provided by the camera module. A frame rate of 30 frames per second was used in the recording. *Table 4.2* summarizes the configurations of the camera module that was used to capture the data.

Table 4.2: The data was captured using the following configurations.

Specification	Value
Image file format	Bitmap (BMP)
Image resolution	752 x 480 pixels
Colour mode	Monochrome
Frame rate	30 fps
Camera exposure value	Automatic (by algorithm)
Pixel format	8-bit integer (0 - 255)

The data collection was performed over the period of November 2016 to March 2017. They were collected during sunny days under direct sunlight or indirect sunlight depending on the locations. The recording tasks were performed under natural (uncontrolled) environment hence some data might contain pedestrians or vehicles. *Figure 4.13* to *Figure 4.15* shows some examples of the collected data.



Figure 4.13: The left and right images of a sample of uneven steps ahead of a walkway.



Figure 4.14: The left and right images of a sample of mix gradient (steps and uncovered drainage).



Figure 4.15: The left and right images of a sample of uncovered drainage next to a walkway.

These data need further pre-processing before they can be used for the following phase of model training. Next section discusses the data pre-processing techniques.

4.4 Data Pre-Processing

According to Han et al. (2012), the collected data have quality if they satisfy the requirement of the intended use. The quality as referred to can be defined in three aspects: accuracy, completeness, and consistency. To achieve quality data, some pre-processing tasks needed to be performed on the raw data. Since the data are purely images, some relevant pre-processing tasks such as cleaning, normalization and whitening were carried out as described in the following sections.

4.4.1 *Cleaning*

The raw data tends to be incomplete, noisy and inconsistent. Data cleaning (or cleansing) is an attempt to overcome these three issues by filling in missing values, filtering out noise while looking for outliers, and correcting inconsistencies in the data (Han et al., 2012). For the image data recorded, incompleteness and inconsistency were not so much the issues. Sorting out noise was then the major cleaning task.

In statistics, noise is a random error (or variance) in a measured variable. Statistical description techniques such as boxplots or scatterplots, and visualization methods can be used to identify outliers which normally represent noise. However, for the case of this research with digital image data, noise is random variation of brightness or colour in the image. Such image noise is normally produced by the camera sensor and/or circuitry of the device, and hence known as electronic noise.

With careful inspection of the raw data, it was found that the digital noise in the images was very little, and the only issue was predominantly a fix-pattern noise as shown in *Figure 4.16*. The occurrence of this noise was random, and it typically had a banded grey and white banding covering some large areas of the images. The noise was removed upon manual inspection.

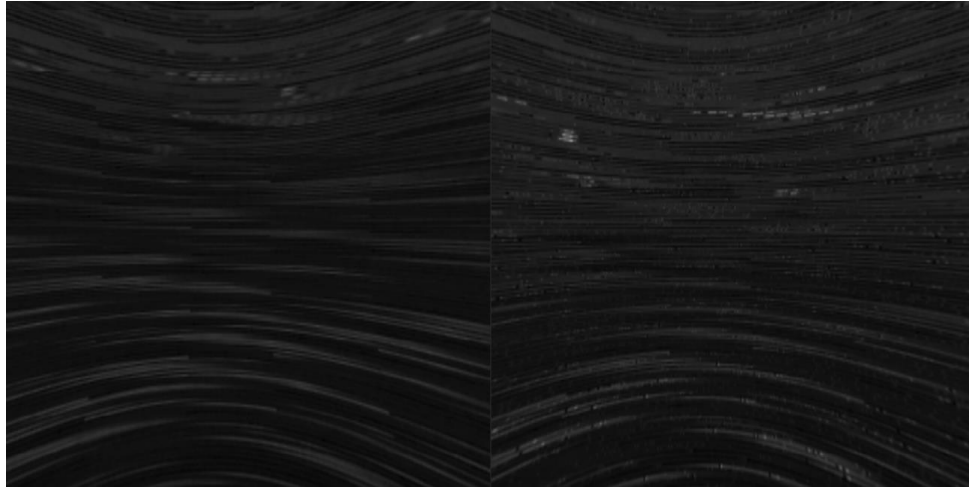


Figure 4.16: A fix-pattern grey and white noise forming a large area of banded bands in some images.

Apart from the digital noise, there were also large amounts of irrelevant areas within the image data. These irrelevant areas were mainly image sequences captured at the very beginning and/or ending of the recording. They are irrelevant because at the beginning of the recording, sometimes the camera was not pointed at the correct direction while the person operating it tried to adjust it to the intended position (see *Figure 4.17*). In other cases, there would be some unwanted elements recorded accidentally (see *Figure 4.18*). At the end of the recording, one might accidentally turn to another direction before the device had stopped its operation, thus capturing some unwanted scenes. These irrelevant parts were all removed and backed-up into another storage upon manual inspection.



Figure 4.17: An unwanted scene captured at the beginning of a recording.



Figure 4.18: Body parts especially the hand or arm could easily become an unwanted object or occlusion in the scene during data collection.

For example, the camera was supposed to record the walkway, but in *Figure 4.17*, it captured mainly the vehicles nearby. In another example (*Figure 4.18*), large irrelevant region was recorded and it must be removed.

4.4.2 Normalization

Once the cleaning task was over, the following task was normalization. The typical methods for image normalization are simple rescaling, per-example mean subtraction and feature standardization (Ng et al., 2013). In simple rescaling, the original data will be rescaled along each data dimension such that the resulting data vectors lie in the range of zero to one $[0, 1]$, or negative one to one $[-1, 1]$. In per-example mean subtraction as the name suggests, the mean value for each example will be subtracted. Feature standardization would independently set each dimension of the data to have zero mean and unit variance.

The choice of normalization method depends on the data type as well as the subsequent steps. The recorded data are grey scale images, and such images have stationarity property – a condition where statistical properties in an image such as mean, variance and autocorrelation could be theoretically constant over pixels. In other words, it is quite meaningless to estimate a separate mean and variance for each pixel as the statistical properties in one part of the image should be almost the same as any other part.

In the subsequent step, Principal Component Analysis (PCA) whitening technique (JOLLIFFE, 2002) was applied to get the data ready for training with deep learning approach. For PCA to work well, it is required that (1) the feature vectors have almost zero mean, and (2) the different feature vectors have similar variances to each other. For the naturally grey images collected in this research, the above condition in (2) is already satisfied without any types of variance normalization. Leaving out variance normalization, mean normalization would be the only choice. Moreover, although feature standardization is good at restoring balance to the components of the feature when there are components with greater influence over most of the other components, this is not a point to consider with the image data here.

Based on the above justifications, per-example mean subtraction was applied as the normalization method. This would remove the mean component from each sample

independently, causing the data to be zero-centred and thus making them ready for the next stage of PCA whitening.

The per-example mean subtraction was performed based on an equation from probability theory, the Z-score (*Equation 4-1*). The data would be normalized such that the mean is 0 and the variance is 1. From the equation, x is the pixel value of the raw image (unnormalized), μ is the mean and σ is the standard deviation.

$$Z = \frac{x - \mu}{\sigma} \quad \text{Equation 4-1}$$

Applying the equation, the pixel values are processed in the following expression:

normalized image = (raw image – mean of raw image) / standard deviation of raw image

However, the results might not be ideal for the subsequent use to perform disparity mapping from the stereo image pairs. Pixels that are supposed to have richer intensity appear to be darkened or whitened in the same tone. Edge detection didn't work well for most images in these normalized versions as compared to the original grey scale images. For example, *Figure 4.19* shows both the original (left) and the normalized (right) images for a similar scene. The step at the middle of the scene is not clear in the normalized version. In both *Figure 4.20* and *Figure 4.21*, a subsequent edge detection applied on both versions revealed that the step is not visible in the normalized image, but not for the original one.

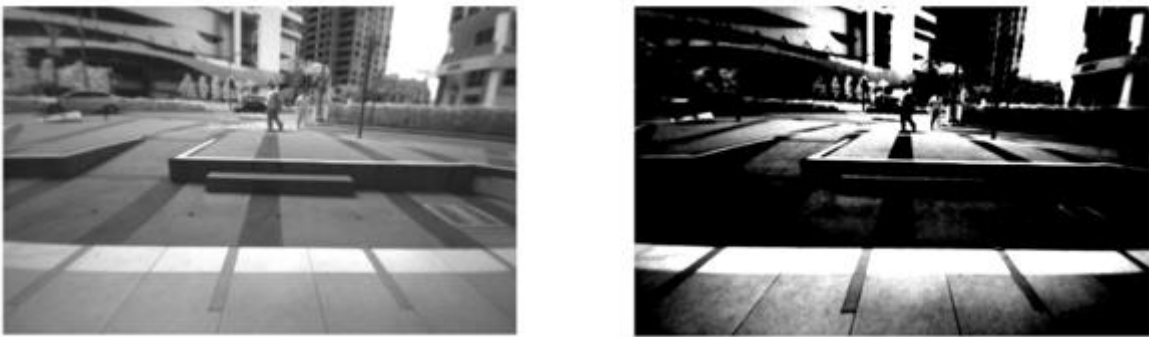


Figure 4.19: Example of a raw image (left) and its normalized version (right) using Z-score. The steps at the centre of the image can be seen clearly in the original version, but are less obvious in the normalized version.

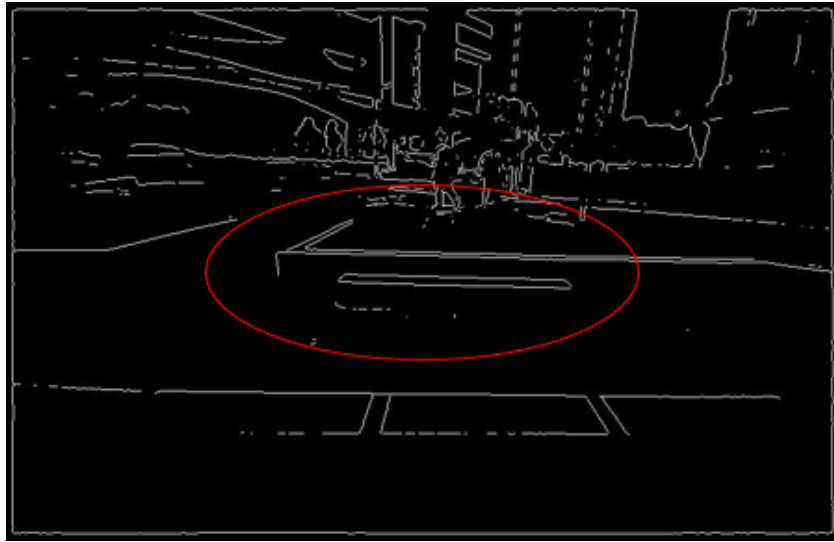


Figure 4.20: Edge detection using Sobel technique on the original image. The edges of the step at the centre of the image are clearly visible as circled in red.

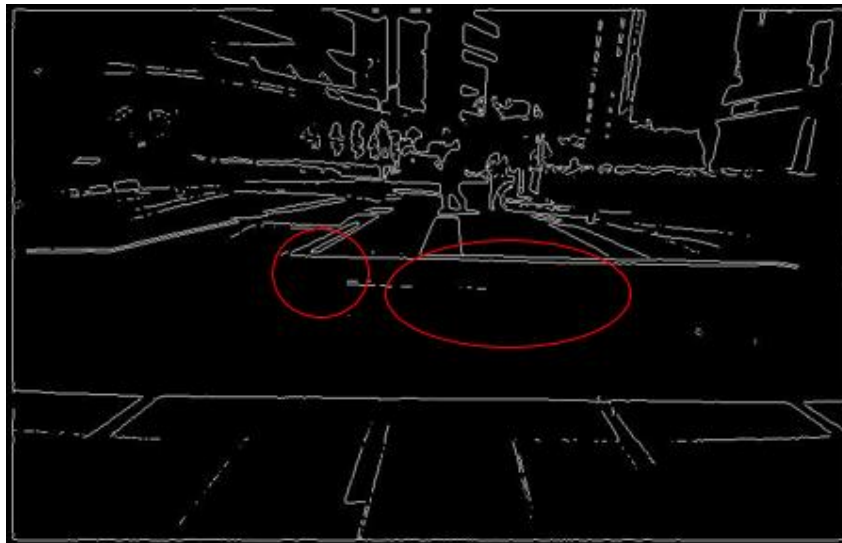


Figure 4.21: As compared to Figure 4.20, the edges of the step at the centre of the image after simple rescaling are not visible. The elevated part of the ramp is also missing.

At this point, it appears that the above normalization might reduce possibly useful features from the original image. In image classification or regression, it is understood that a higher pixel intensity could mean that a feature might be useful. If the mean is subtracted, the operation could result in features with original value of zero having a negative value. This

will result in a higher magnitude for the features that are probably less significant, making them more important. The research cautiously considered the consequence of such a change on the classification task and abandoned the above normalization technique. Instead, PCA whitening which is explained in the next section was applied.

4.4.3 PCA Whitening

Principal Component Analysis (PCA) (JOLLIFFE, 2002) is a widely used method for reducing the number of dimensions in the feature vectors of a data. It is a linear mapping technique to compress the data by exploiting correlations between the dimensions. A whitening process (or transformation) is a linear transformation that converts the feature vectors based on a covariance matrix into a new set of vectors whose are uncorrelated and have a variance of value 1 for each of them. As such, it is generally a beneficial practice to pre-process the data with whitening, since it de-correlates the data and makes them easier to model. However, this point is arguable as Koivunen and Kostinski (1999) pointed out that its benefit is dependent on the data and its subsequent processes. The main benefit of applying PCA whitening to images is that it could enable a convolutional neural network to get better classification results, because it would have eliminated most of the noise in the images.

In the first step of PCA whitening, Singular Value Decomposition (SVD) (Weisstein) was applied to compute the eigenvector decomposition from the zero-centred data. SVD is a matrix analysis that decomposes a high dimensional matrix to a low dimensional representation. SVD makes it easy to eliminate the less important components of that representation and produce any desired number of dimensions from the elimination.

The image resolution is 752 x 480 pixels, and this would eventually produce a dimension of 360960. For a trial, the author first performed SVD on 10 samples (the process had taken some time due to the high dimension) from the data and analysed the singular values. It is

crucial to preserve as much information as possible from the singular values. A calculation on total energy from the singular values was performed to retain the components that contribute to 90% of the information. From the first few trials, it was observed that 292380 out of the 360960 components preserve about 90% of the information. Based on the select 292380 singular values, the other low singular values were zeroed out. Now with the remaining eigenvectors, the rotated version of the data (let's call it $xRotate$) was computed by the product of inversed eigenvectors to the original pixel matrix. Finally, the PCA whitened data, $whitenedX$ was computed with the following equation, in which epsilon ϵ is a small constant to prevent division by zero:

$$whitenedX = diag(1/\sqrt{(diag(singular\ values) + \epsilon)}) \times xRotate \quad \text{Equation 4-2}$$

Figure 4.22 shows three patches of image before the whitening, and Figure 4.23 is the result after their whitening process. After observing better edge detection results from the whitened images, this technique was applied to all the remaining data.



Figure 4.22: Three sample images before PCA whitening.



Figure 4.23: Three similar sample images from Figure 4.22 after PCA whitening.

4.5 Taxonomy of Surface Discontinuity and Data Labelling

To prepare the data for supervised training, the data were then labelled manually. Before the labelling process could take place, a taxonomy of the data – surface discontinuities – must be developed. This taxonomy is important to provide a guideline for the labelling task, which would then facilitate the machine learning training.

The purpose of the research is to develop a prototype that assists the BLVs in negotiating surface discontinuity, and for that, the surface of a pathway must be differentiated between a “continuity” (indicating a smooth surface) and “discontinuity” (i.e. indicating a drop-off). For the discontinuity type, the prototype must be able to tell if it is a down-step, drop-off, down-ramp, uncovered drainage or blended gradient (when two or more types of surface discontinuity meet at one point along the supposed navigational pathway, it is known as a blended gradient in this research). Up-steps and up-ramps (or collectively known as rises) were also included in the model training, although they are not hazardous to blind navigation because these properties are easily handled by the BLV’s guide cane, and there were quite some successful classification or object detection studies done on them (Lee et al., 2008, Hern et al., 2011, Pérez-Yus et al., 2015, Vlaminc et al., 2013, Tang et al., 2012, Takizawa et al., 2012, Takizawa et al., 2013, Dang et al., 2016, Jonsson, 2011). Having said that, the system still needs to be able to recognize rises along a navigation task, such that the classification does not fall into false positive.

According to observation of the physical attributes of the collected data, there are 9 distinctively differentiable classes of surface conditions. In addition to the physical attributes, the Building Regulations from Document K entitled “Protection from Falling, Collision and Impact” (2013), England and the Uniform Building (Amendment) By-Laws (1991), Malaysia were also referred to, to guide the taxonomy development. *Figure 4.24* shows the taxonomy with sample images along with their corresponding class label.



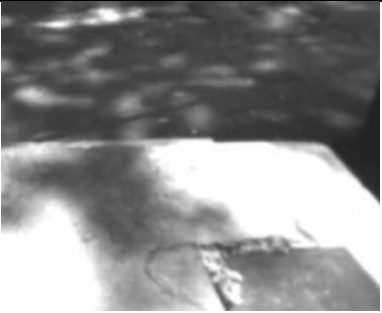


Image Sample	Common Name / Attribute	Class Label
	Continuity – Smooth walkway	SW
	Discontinuity – Down-steps	D1
	Discontinuity – Drop-off	D2
	Discontinuity – Down-ramp	D3
	Discontinuity – Uncovered drainage	D4

Figure 4.24: The taxonomy of surface discontinuities.

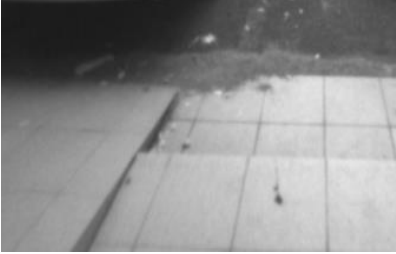



Image Sample	Common Name / Attribute	Class Label
	Discontinuity – Blended gradient	D5
	Discontinuity – Rise	R1
	Discontinuity – Up-steps	R2
	Discontinuity – Up-ramp	R3

Figure 4.24 (continued): The taxonomy of surface discontinuities.

4.6 Data Augmentation

Since deep learning and other neural networks need to be trained on large volume of data, a technique known as data augmentation can be used to increase the data volume (Frühwirth-Schnatter, 1994, Ding et al., 2016). Simard et al. (2003), Cireşan et al. (2011) and Cireşan et al. (2012) are several research works that have used data augmentation to increase the data volume prior to deep neural networks training. They applied the typical

transformation techniques such as horizontal flipping, random cropping, shifting and jittering in their augmentation. Another work by Krizhevsky et al. (2012b) proposed a purpose-built technique involving PCA to alter the intensities of RGB channels in their training images.



Figure 4.25: Image pair on top is the original version, after the augmentation by horizontally flipping, the bottom image pair was yielded.

Through data collection from the field, a total of 225 samples were recorded for this research. Each sample is a collection of image sequences of approximately 60 to 150 sets depending on the duration of the records (the camera module was set to 30 frames per second during recording). This means that a recording with 5 seconds of duration would yield 150 images. From the 225 samples, a total of 30375 images were yielded (after pre-processing). To double up this amount, a data augmentation technique was applied by simply flipping the images horizontally. This has given the research a final amount of 60750 images – a number large enough for the next phase of machine learning training.

4.7 Summary of the Chapter

The chapter presented the data generation process which involved several techniques ranging from the instrument development, crowdsourcing for locations for data collection, data sampling, ground truth measurement, description of the collected data and pre-processing. The development of the taxonomy of surface discontinuity was also presented, in which it would be an important guideline for data labelling process, and this would facilitate the next phase of machine learning model training. Finally, the chapter ended with a technique used for data augmentation to increase the amount of data.

Chapter 5: System Development (Phase-2 Prototype)

This chapter is dedicated to the development of the system, which is also named as the phase-2 prototype. It starts with an overview of the phase-2 prototype (Section 5.1) and follows with some detailed descriptions of feature extraction techniques used (Section 5.2). Next, the chapter presents model training method with elaborations on several supportive techniques.

5.1 Overview of Phase-2 Prototype

“System Development” refers to the development of the phase-2 prototype. In the earlier section, it was described that the phase-1 prototype was used only for data collection. Over here, the phase-2 prototype is a wearable technology-based system that can be used in the field by a researcher to evaluate the performance of the trained machine learning model. From the previously built phase-1 prototype, several modifications were introduced to transform it into the phase-2 prototype. The modifications include the following five main items:

1. Addition of a trained machine learning model into the prototype
2. Reprogramming of the application to apply model mentioned in (1) to classify the images captured by the stereo camera
3. Reworking on the user interface to control the turning on and off the prototype, which would then activate the real-time classification of images being captured
4. Addition of user interface to produce simple output signals representing different types of the classes of surface discontinuity, after the classifier predicted their classes

5. Storing of the classification results into a text file for evaluation

Figure 5.1 shows the structure diagram of the phase-2 prototype. It consists of four main components - the Sensor Module, Processing Unit, User Interface and Output. The Sensor Module has a stereo camera as the only sensor, and it provides the main input (sequences of image pairs) to the prototype. The Processing Unit is responsible for real-time disparity mapping of the image pair, before conducting classification on it. Having the class identified, it would send a feedback signal to the User Interface and store the class type into a file together with a timestamp. The User Interface also allows the user to switch on the prototype for usage or turn it off for termination of the application before shutting-down the prototype.

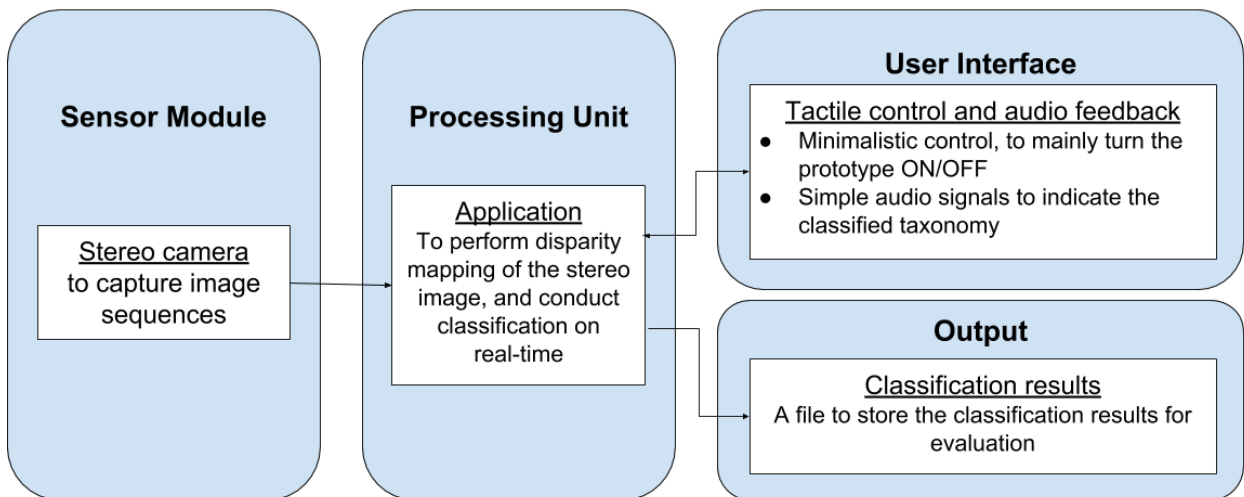


Figure 5.1: A structure diagram illustrating the proposed system, with the Sensor Module, Processing Unit, User Interface and Output.

Apart from transforming the phase-1 prototype into the phase-2 version as described above, two more main tasks were involved in this phase namely feature extraction and model training. The following sections will discuss some concepts involved and steps taken in these two main tasks.

5.2 Feature Extraction

With the data generated, there is one main feature that must be extracted before the next step of training the data. This feature is the disparity map of the image pairs. During data collection, the phase-1 prototype with the help of some algorithms had processed the lens undistortion and Epipolar rectification of the image pairs on real-time before saving them to the storage. This has simplified the feature extraction process.

5.2.1 *Disparity Mapping and Depth*

One of the challenges in computing stereo vision is image pair matching or mapping (Gosta and Grgic, 2010). With two input images from two cameras, a matching method is needed to identify the corresponding points from both images and generate a disparity map. Disparity refers to the difference in location of an object in correspondence to the image pair viewed by two cameras. Additional information such as the depth of the object can be obtained through the disparity.

The image pair matching comprises of search and comparison tasks for the corresponding points of two images. To achieve the matching, Gosta and Grgic (2010) explained that if the camera geometry is known, the two dimensional search for corresponding points could be simplified as single dimensional search. This would involve the rectification of the image called Epipolar rectification (Zhang, 1998). *Figure 5.2* shows an example of Epipolar rectification of an image pair.

Despite the simplification of the matching task, there are still several challenging problems in stereo vision. Firstly, a common phenomenon is the lack of consistency between the image pair due to variation of intensity and colour from the two different points of view. Secondly, electronic components of the camera module often produce noise that might affect the image acquisition. Another issue is the difficulty in uniquely matching two points due to the presence of constant luminance in large regions. In such a region, more than one corresponding point could be detected. Finally, the more serious problem is that by

nature some pixels in the left image do not have corresponding pixels in the right image. The typical reason for this problem is that some scene parts are visible to one camera but not the other due to occlusion of obstacle affecting one camera solely (Brown et al., 2003). The first two issues mentioned above were already resolved during the data pre-processing phase, thus in this phase of disparity mapping, the last two issues were the main concern.

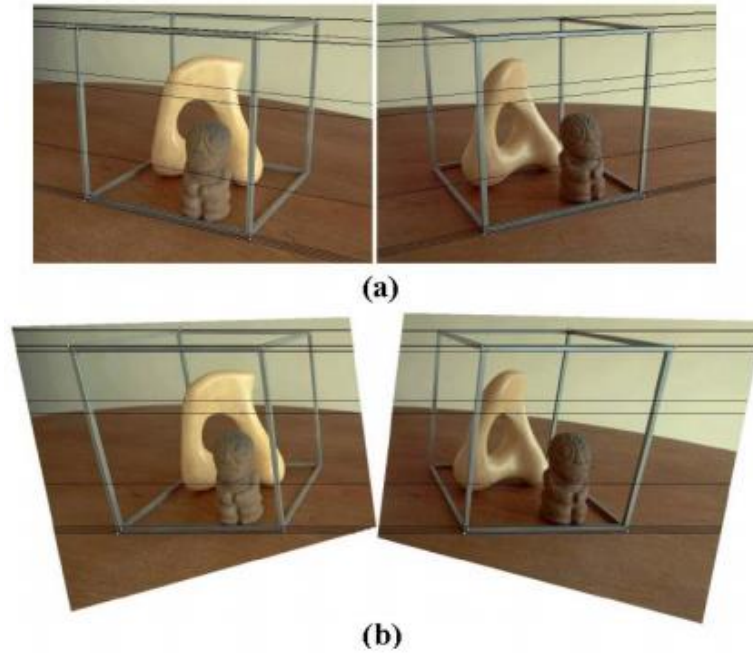


Figure 5.2: Epipolar rectification of image pair, (a) original image pair, (b) rectified image pair (Gosta and Grgic, 2010).

Due to the nature of the matching issues, it takes quite some work in finding a better solution to it. The matching algorithms can be classified into two classes: the feature-based methods and the intensity-based methods (which can be further divided into local and global approaches). Feature-based methods first detect features such as edges and corners of the image pair, and then proceed to matching these features (Grimson, 1985, Hoff and Ahuja, 1989). One of the feature-based methods relies on segmentation techniques for feature extraction (Klaus et al., 2006, Wei-Qun and Xian-Ming, 2010, Lin and Tomasi, 2004).

Local intensity-based methods rely on pixels within a selected window to compute correlation or sum squares differences between the image pair. The disparity that produces the best match is then set as the disparity of the pixels (Kanade and Okutomi, 1994, Zitnick and Kanade, 2000). Although typically the local intensity-based methods can produce dense disparity maps, they are troubled by noise as the disparities are estimated from local information only. Global intensity-based methods perform extraction of a surface from a three-dimensional volume consisting of $u - v - d$ as the disparity space. In a pixel (u, v) , the probability of the disparity, d can be obtained from the value of any voxel in the disparity space. With this definition, the disparity map can be constructed with the satisfaction of two constraints: (1) a smooth surface, (2) it passes through high probabilities voxel (Yang et al., 1993).

Since the matching task is a class of search and compare problems, several authors have proposed optimization approaches to the solution, especially with the objective to perform a global search for some better results. Gong and Yang (2002) and Wenlong Fu et al. (2012) proposed a genetic-based algorithm (GA) as a global solver to the edge mapping problem. However, as mentioned in a previous section, GA as a form of stochastic search typically has quite a few operators with parameters to be optimized, and thus it requires higher computational effort. For example, in the proposed approaches used by Gong and Yang (2002) and Han et al. (2001), the segmented regions are coded as rectangular chromosome blocks. Since the disparity block is also rectangular, which would greatly increase the length of their chromosome, and eventually increase the complexity, Zhang et al. (2008) proposed a PSO-based stereo mapping method which can perform global optimization but with lesser computational complexity when compared to the GA methods. They produced a partial disparity map using dynamic programming and sum-of-absolute-differences (SAD), and then followed by PSO for the segmented region to be searched and matched optimally.

Based on the above literature, the simplest form of disparity mapping is basically composed of two steps: (1) compute contrast using edge detection, and (2) compute disparity. Thus, before further decision is made on the disparity computation, a technique for edge detection must be first decided. Edges are boundaries between different textures. In this research context, for an example, the edge could be the boundary between a floor plane and the rising plane, indicating an elevation of the surface, hence a surface discontinuity.

There are several methods for edge detection – two common categories are: (1) computing of the first-order derivatives in an image (known as the gradient method), and (2) computing of the second-order derivatives using the Laplacian method (Petrou, 1999, Bovik, 2005). Technically, the gradient method detects the edges by identifying the maximum and minimum in the first-order derivatives of the images, while the Laplacian method detects zero-crossing in the second derivatives of the images to locate edges. Several traditional image filters for edge detection such as the Roberts, Prewitt and Sobel are based on the gradient method.

Another popular variant developed by Canny (1986) using a multi-stage algorithm with Gaussian derivative is capable of a wider range of edge detection within an image. The Canny method is considered to be a complete and well-defined process of edge detection with good localization ability (Setayesh et al., 2011). One typical problem with the Canny method is the double edges construction in locations with high frequencies of information (Kang and Wang, 2007), while other variants of Gaussian-based method suffer from displacement, removed and false edges (Basu, 2002), as well as not performing well at corners and curves (Sharifi et al., 2002).

As recommended by Umbaugh (2005), an edge detection algorithm should ideally be able to recognize the continuous contours of the object boundaries. However, this task can be challenging due to several types of noise in particularly both the Gaussian and impulsive noises. The Gaussian noise arises during the image acquisition, possibly caused by poor

illumination, high temperature and transmission within the electronic circuit. The impulsive noise is caused by bit errors during data transmission or converter errors during analog to digital conversion.

In order to overcome the noise issue and improve accuracy, there are several statistical approaches to deal with noise. Lim (2006) proposed a robust rank-order (RRO) detector and benchmarked it against other statistical methods namely the t-detector and Wilcoxon detector. These methods use statistical tests (t-test, Wilcoxon test and RRO test) to classify the pixel as edge or non-edge by analyzing the intensity of the neighbours of a pixel. When comparing these statistical methods to other edge detection methods, the former ones are more capable of operating on a large area of an image. Since statistical methods are data-driven, they are not able to recognize edge intensities which are typically required for techniques in edge thinning and linking. This often leads to thick and sometimes broken edges by the statistical methods.

Some other methods have been proposed based on soft computing approaches, for instances, fuzzy logic methods (Haq et al., 2015, Li and Li, 2012, Patel and More, 2013), neural networks (Stevens, 2015, Meftah et al., 2010, Dingran Lu et al., 2011) and evolutionary methods (Fu et al., 2011, Gong and Yang, 2002, Wenlong Fu et al., 2012). These soft computing methods have distinctively achieved improvement on accuracy over the traditional methods, but the known drawbacks common to these methods are longer convergence rate and higher time complexity.

Edge detection can be viewed as a search task needing a global optimization solution too. To solve it as global optimization, another relatively recent computational method for edge detection is the Particle Swarm Optimization (PSO) (Alipoor et al., 2010, Setayesh et al., 2011). According to Hassan et al. (2004), the advantage of PSO over most evolutionary methods such as GA-based methods is that PSO is more computationally efficient (having lesser number of operators or function evaluations) than the GA, although both of them

are population-based approaches. In the work of Setayesh et al. (2011), a constrained PSO-based edge detector is proposed and compared against the Canny, RRO and another PSO methods. Their results showed that although the Canny methods can perform well in most images without or with little noise, their PSO method outperformed the other methods for detecting images corrupted by Gaussian and/or impulsive noises. One notable outcome from Setayesh et al. (2011) is the comparison of time taken by the different methods. Their PSO method was recorded to take a longer time (50 – 60 seconds) than the Canny (2 – 3 seconds) and RRO (3 seconds).

To justify the choice of an edge detection method, it is important to refer again to the aim of the research. The research needs to process the data in real-time, such that the BLVs could receive timely feedback to negotiate the threats. In other words, the disparity mapping process needs an edge detection technique with better accuracy and time efficiency (lower computational complexity). Based on review from the literature, it appears that the Canny and Sobel filters generally perform well with lower time complexity. This suits the need of the disparity mapping here.

With the decision on edge detection techniques decided, the next stage of computing the disparity based on a block matching technique (Thaher and Hussein, 2014) was applied. A block matching technique consists of several steps to solve the stereo vision correspondence problem. Every block from the left image is matched to the right image by shifting the left block over the search space for pixels in the right image. Based on the SAD method, Thaher and Hussein (2014) found that the error rate would increase when the window size of the matching block is not set to the optimum. They have performed several systematic tests to identify the best window size for their SAD method, and this would be referred to for the research to identify the best size.

5.2.2 Technique to Generate the Disparity Map

The stereo camera used in the prototype has a baseline (distance between two camera lenses), B of 30.02 mm (≈ 0.03 m). As illustrated in *Figure 5.3*, disparity (denoted as $x - x'$) is expressed as:

$$\text{disparity} = x - x' = Bf/Z \quad \text{Equation 5-1}$$

where x and x' are the distances between points in an image plane corresponding to the point on actual scene and their camera centres. f is the focal length of the camera, which is 2.0 to 2.1 mm. In other words, *Equation 5-1* suggests that the depth of a point in a scene is inversely proportional to the displacement of corresponding image points and their camera centres. Since values of B and f are known, the depth of all pixels in the image can be derived.

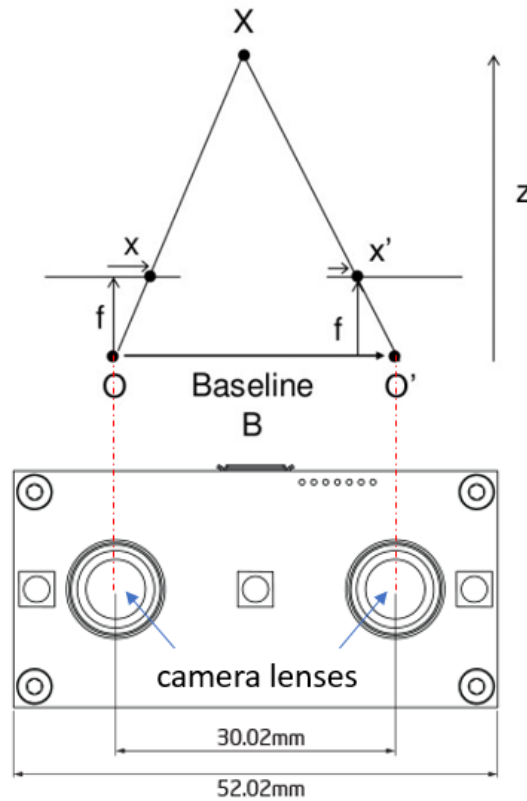


Figure 5.3: A diagram illustrating the equivalent triangles (top), and the dimensions of the stereo camera (bottom) used in the prototype.

There are fast and accurate algorithms to perform the disparity mapping based on the above theory. In this research, the lens undistortion and Epipolar rectification of the images had been processed during data collection on real-time before saving them to the storage. Therefore, a two-step disparity mapping technique was used (see *Table 5.1*). Within the two steps, some alterations on the algorithms were added to observe their accuracies, before a final decision was made on the implementation.

Table 5.1: A disparity mapping technique in two steps used in this research.

Step	Variation
1. Compute contrast based on edge detection filter.	Edge detection filter: Sobel / Canny
2. Compute disparity for each of the pixels using block matching with sum of absolute differences (SAD).	Window sizes of 5, 7, 9, 15 and 21 were tested; several disparity ranges were tested

In *Figure 5.4* the outcome of image pair of Sample 28 is matched into a composite view of red-cyan stereo anaglyph as shown in *Figure 5.5*. In the following *Figure 5.6*, the depth of the similar sample is represented using a colour gradient from black (further object) to white (proximal object). It can be seen in the figure that the drop-off part (which is the drainage) is clearly darkened as compared to its surrounding walkway or vehicles. Its 3-dimensional values of width, height and depth are stored in a 3-D array, ready to be used for the next phase of training.



Figure 5.4: A grey scale image pair of Sample 28 (an uncovered drainage).



Figure 5.5: A red-cyan composite view of the image pair from Figure 5.4.



Figure 5.6: A disparity map from the image pair in Figure 5.4.

Based on Figure 5.6, the depth is represented using a colour gradient from black (distant object) to white (proximal object). Its 3-dimensional values of width, height and depth are stored in a 3-D array. More observations are discussed in the following sections with variations applied on the disparity mapping steps.

5.2.3 Regions of Uniform Intensity

One of the biggest issues in most of the mapping outcomes is the occurrence of black dots on large regions with the same intensity. It appears that when a large region has neighbouring pixels with uniform intensity, the mapping ended up with dusts of black dots. This issue can be seen in the following two cases of Sample 120 and 127, with the flat floor on both walkways (Figure 5.7 and Figure 5.8). It is suspected that the intensity-based error minimization technique used in the process could have picked up very small differences and assigned some arbitrary disparity to that region. Since this is an issue relevant to intensity, it was left for the machine learning to handle because most error minimization mechanisms in deep learning should have no problem overcoming it.

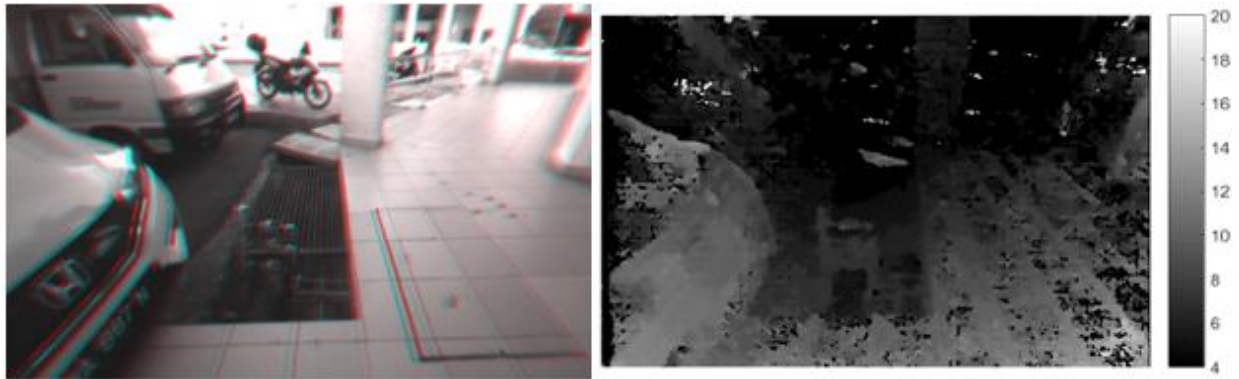


Figure 5.7: An issue of region with uniform intensity, in which the floor on the walkway can be seen dusted with black dots.

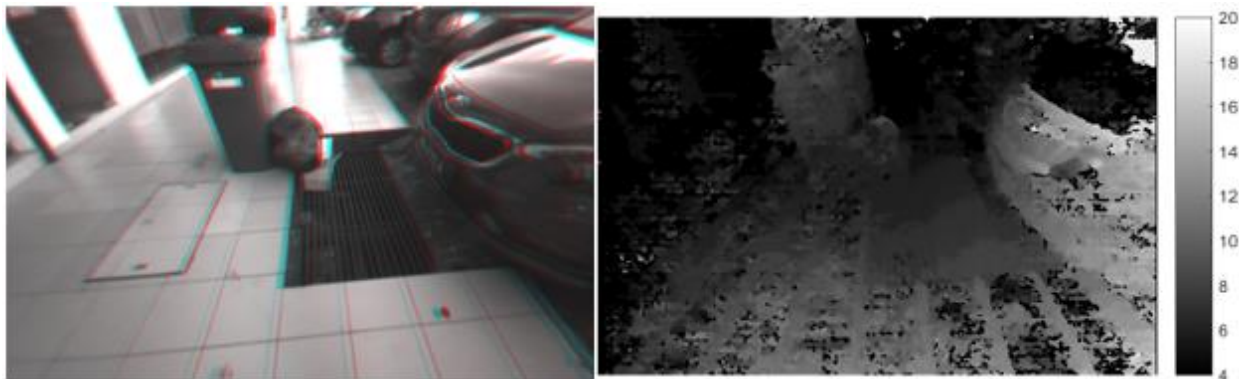


Figure 5.8: Another example of region with uniform intensity, in which the floor on the walkway can be seen dusted with black dots.

5.2.4 Disparity Range

In the second step of the disparity mapping technique, one of the adjustable parameters is the disparity range. Disparity range is the distance between the two cameras and the distance between the cameras and the point of the actual scene. It was observed that this range must be optimized or else the mapping algorithm would get confused at the region where there is not too much of intensity variation. The result of the error can be observed through the black and white depth map of Sample 97 (Figure 5.9) that when it is set too

low at $(0, 16)$, for example in *Figure 5.10*, the drop-off is not clearly differentiable from the road ahead – both are coloured close to white.



Figure 5.9: An anaglyph of Sample 97, showing a drop-off at the end of a walkway.



Figure 5.10: At disparity range of $(0, 16)$, the drop-off and the road are closely coloured in white.







Figure 5.11: At disparity range of (2, 18), the drop-off and road are more distinguishable than before.

As for the map in *Figure 5.11* in which the range is set slightly higher to (2, 18), both objects are more distinguishable. At the end of the experiment, it was found that a range of (4, 20) is mostly acceptable for the data used in this research. Hence, the disparity range is chosen to be (4, 20).

5.2.5 Windowing

Window size has some significant effect on the disparity mapping. An optimal size would make the map clearer than a randomly picked one, which could blur the map. In several observations, when the window size was gradually increased, it helped to overcome noise better. However, when the window was increased beyond the optimal size, the depth maps began getting blurred, and thus reduced the accuracy (see **Error! Reference source not found.**). It was found that a window size of 9 units could be optimal for most of the collected stereo images, although there were some exceptions which are not significant. The research settled at a window size of 9 units. *Figure 5.13* shows a typical example of the disparity mapping using this window size.

	<p>The red-cyan composite anaglyph of Sample 143, a drop-off next to an uneven step.</p>
	<p>Window size: 5 Map with much noise, but quite precise</p>
	<p>Window size: 7 Map with lesser noise</p>
	<p>Window size: 9 Map with even lesser noise, but blurring effect is acceptable in most cases</p>


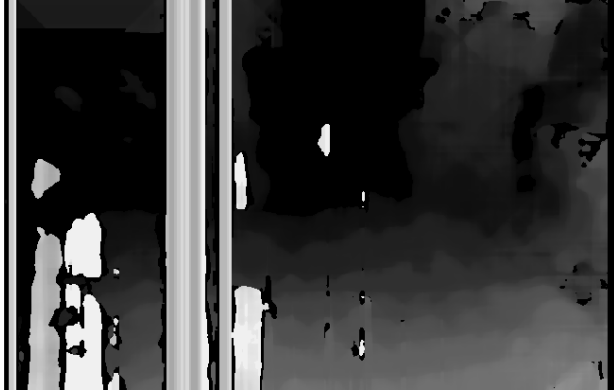
	<p>Window size: 15</p> <p>Map with very little noise, but most of the edges are blurred</p>
	<p>Window size: 21</p> <p>Inaccurate mapping in all regions</p>

Figure 5.12: Examples of different window sizes ranging from 5 to 21 units. Their accuracies and blurring effects can be seen here.



Figure	Description																																																																																																																																																
	This is the left-image of a typical drop-off.																																																																																																																																																
	The disparity map of the sample as visualized in grayscale.																																																																																																																																																
<table><tr><td></td><td>21</td><td>22</td><td>23</td><td>24</td><td>25</td><td>26</td><td>27</td><td></td></tr><tr><td>21</td><td>6.4375</td><td>8</td><td>8</td><td>8</td><td>8.0625</td><td>8.1250</td><td>8.1250</td><td></td></tr><tr><td>22</td><td>6.5000</td><td>8</td><td>8.1250</td><td>8.1250</td><td>8.1875</td><td>8.2500</td><td>8.3125</td><td></td></tr><tr><td>23</td><td>6.6875</td><td>8.1250</td><td>8.2500</td><td>8.3125</td><td>8.3750</td><td>8.4375</td><td>8.6250</td><td></td></tr><tr><td>24</td><td>6.8125</td><td>8.1875</td><td>8.3125</td><td>8.3750</td><td>8.4375</td><td>8.6250</td><td>8.8125</td><td></td></tr><tr><td>25</td><td>6.8750</td><td>8.1875</td><td>8.3125</td><td>8.3750</td><td>8.6250</td><td>8.8125</td><td>8.9375</td><td></td></tr><tr><td>26</td><td>6.8750</td><td>8.1250</td><td>8.2500</td><td>8.3750</td><td>8.6250</td><td>8.9375</td><td>9</td><td></td></tr><tr><td>27</td><td>6.6875</td><td>8.1250</td><td>8.1875</td><td>8.3125</td><td>8.6250</td><td>8.9375</td><td>9.0625</td><td></td></tr><tr><td>28</td><td>6.5000</td><td>8.1250</td><td>8.1875</td><td>8.3125</td><td>8.5625</td><td>8.9375</td><td>9.0625</td><td></td></tr><tr><td>29</td><td>6.4375</td><td>7.9375</td><td>8.1250</td><td>8.1875</td><td>8.3750</td><td>8.6875</td><td>8.9375</td><td></td></tr><tr><td>30</td><td>6.3750</td><td>7.8125</td><td>7.9375</td><td>8</td><td>8.1875</td><td>8.3750</td><td>8.6875</td><td></td></tr><tr><td>31</td><td>6.3125</td><td>7.7500</td><td>7.8125</td><td>7.9375</td><td>8</td><td>8.1875</td><td>8.4375</td><td></td></tr><tr><td>32</td><td>6.3125</td><td>7.6250</td><td>7.7500</td><td>7.8125</td><td>7.9375</td><td>8.1250</td><td>8.3125</td><td></td></tr><tr><td>33</td><td>6.3125</td><td>7.2500</td><td>7.6250</td><td>7.7500</td><td>7.8750</td><td>8.0625</td><td>8.3125</td><td></td></tr><tr><td>34</td><td>6.3125</td><td>7.1875</td><td>7.4375</td><td>7.6250</td><td>7.8125</td><td>8</td><td>8.3125</td><td></td></tr><tr><td>35</td><td>6.3125</td><td>7.1875</td><td>7.2500</td><td>7.5000</td><td>7.7500</td><td>8</td><td>8.3750</td><td></td></tr></table>		21	22	23	24	25	26	27		21	6.4375	8	8	8	8.0625	8.1250	8.1250		22	6.5000	8	8.1250	8.1250	8.1875	8.2500	8.3125		23	6.6875	8.1250	8.2500	8.3125	8.3750	8.4375	8.6250		24	6.8125	8.1875	8.3125	8.3750	8.4375	8.6250	8.8125		25	6.8750	8.1875	8.3125	8.3750	8.6250	8.8125	8.9375		26	6.8750	8.1250	8.2500	8.3750	8.6250	8.9375	9		27	6.6875	8.1250	8.1875	8.3125	8.6250	8.9375	9.0625		28	6.5000	8.1250	8.1875	8.3125	8.5625	8.9375	9.0625		29	6.4375	7.9375	8.1250	8.1875	8.3750	8.6875	8.9375		30	6.3750	7.8125	7.9375	8	8.1875	8.3750	8.6875		31	6.3125	7.7500	7.8125	7.9375	8	8.1875	8.4375		32	6.3125	7.6250	7.7500	7.8125	7.9375	8.1250	8.3125		33	6.3125	7.2500	7.6250	7.7500	7.8750	8.0625	8.3125		34	6.3125	7.1875	7.4375	7.6250	7.8125	8	8.3125		35	6.3125	7.1875	7.2500	7.5000	7.7500	8	8.3750		A small fraction of the disparity map in a vector of numerical values, at row pixel 21 to 27, and column pixel 21 to 35.
	21	22	23	24	25	26	27																																																																																																																																										
21	6.4375	8	8	8	8.0625	8.1250	8.1250																																																																																																																																										
22	6.5000	8	8.1250	8.1250	8.1875	8.2500	8.3125																																																																																																																																										
23	6.6875	8.1250	8.2500	8.3125	8.3750	8.4375	8.6250																																																																																																																																										
24	6.8125	8.1875	8.3125	8.3750	8.4375	8.6250	8.8125																																																																																																																																										
25	6.8750	8.1875	8.3125	8.3750	8.6250	8.8125	8.9375																																																																																																																																										
26	6.8750	8.1250	8.2500	8.3750	8.6250	8.9375	9																																																																																																																																										
27	6.6875	8.1250	8.1875	8.3125	8.6250	8.9375	9.0625																																																																																																																																										
28	6.5000	8.1250	8.1875	8.3125	8.5625	8.9375	9.0625																																																																																																																																										
29	6.4375	7.9375	8.1250	8.1875	8.3750	8.6875	8.9375																																																																																																																																										
30	6.3750	7.8125	7.9375	8	8.1875	8.3750	8.6875																																																																																																																																										
31	6.3125	7.7500	7.8125	7.9375	8	8.1875	8.4375																																																																																																																																										
32	6.3125	7.6250	7.7500	7.8125	7.9375	8.1250	8.3125																																																																																																																																										
33	6.3125	7.2500	7.6250	7.7500	7.8750	8.0625	8.3125																																																																																																																																										
34	6.3125	7.1875	7.4375	7.6250	7.8125	8	8.3125																																																																																																																																										
35	6.3125	7.1875	7.2500	7.5000	7.7500	8	8.3750																																																																																																																																										

Figure 5.13: Example of a disparity map computed from a stereo image pair (from sample 188, image pair of sequence 65).

5.2.6 Sum of Difference

The two commonly used sum of differences functions are Sum of Absolute Difference (SAD) and Sum of Squared Difference (SSD). According to the mathematical interpretation of both the functions, SSD is more sensitive than SAD in getting the difference of pixels because it squares the difference and thus magnifies it for comparison. However, based on some experiments, SSD was observed to be more susceptible to noise as compared to SAD (see *Figure 5.14* for comparison on Sample 224). Small black dots are more obvious in the SSD map, while they are not present in the SAD map. Zhang et al. (2008) and Thaher and Hussein (2014) have recommended SAD in their works which are relevant to this research. For the above reasons, SAD is chosen as part of the disparity mapping steps.


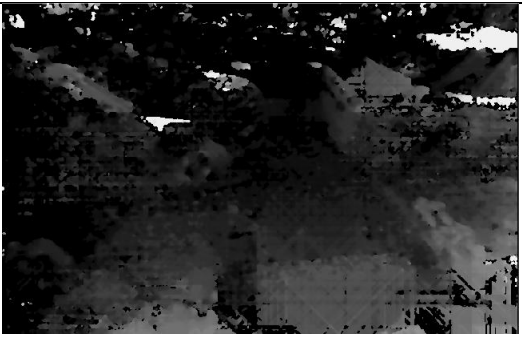

	Red-cyan anaglyph of Sample 224, two uneven steps bridging the walkway and road between an uncovered drainage.
	Mapping done with SSD, more black dots can be seen.
	Mapping done with SAD, less black dots on the map.

Figure 5.14: A comparison between sum of difference functions SSD and SAD.

5.3 Assembling a Deep Learning Architecture

After a disparity map is extracted from the data, the map can now be used to train a machine learning model to classify the classes of surface discontinuity based on the developed taxonomy. Based on literature reviewed in Chapter 2, two potential deep learning approaches for the task are deep belief network (DBN) and convolutional neural network (CNN). As the task involves training a model using images as data (the disparity map is a type of image), CNN is the current state-of-the-art technique used widely to train such a model. However, before a CNN was experimented, the more easily train DBN was first tested.

Deep belief network. From the disparity map, the author proposed a simple technique inspired from a 2-dimensional image intensity histogram to profiling the type of surface from the data (Leong et al., 2017). It is presumed that along the same row of the scene, the disparity should be identical for a smooth surface in the row, unless there is a drop-off (if the disparity is greater than the average of the same row) or elevation (if the disparity is lesser than the average of the same row). This is called the depth profile. The depth profiles were computed as the sum of disparity along the horizontal pixels over the vertical pixels. The depth profiles were then labelled into two categories – continuity (smooth surface) and discontinuity. *Figure 5.15* shows the visualization of a depth profile of a pathway leading to an uncovered drainage.

The dimension of this profile is scalable, and the maximum size being the height of the original image at 480. Various scales of the profile were tested, and the highest dimension was observed to give the best classification result. With further analyses, it was found that most of the significant surface conditions were presented at the lower parts of the images. Hence, the experiment proceeded by reducing the dimension of the image pair to retain only the $\frac{1}{2}$ lower region. The depth profiles were recreated based on the reduction. The new map was about half of the size of the original image. At this point, the dimension of

the original depth map was reduced from 360960 (computed from the height of 752 multiplies the width of 480) to as low as 240.

The proposed depth profiles were then classified using a DBN. A typical DBN is built as a generative learning model. Its implementation is simple with its two training parts of an unsupervised pre-training and a supervised fine-tuning. In this experiment, a Generative Restricted Boltzmann Machine was used for its pre-training, and a backpropagation for its fine-tuning. The architecture of the network is illustrated in *Figure 5.16* with H_1 to H_4 as the hidden layers. The process of the DBN adopted from Hinton and Salakhutdinov (2006) is described as follows:

1. Train the first layer (visible input layer) as an GRBM that models the input vector $x = H_0$.
2. Use H_0 to obtain a representation of the input for H_1 , the first hidden layer. Gibbs sampling is used to compute the joint probability of $p(H_1|H_0)$ for the representation.
3. Train H_1 as an GRBM by taking the representation from H_0 as training samples.
4. Repeat Step 2 and 3 for H_2 to H_4 , each time propagating forward the samples.
5. Fine-tuning this deep architecture (all the layers H_1 to H_4) with a backpropagation supervised training to learn the two classes of “continuity” or “discontinuity”.

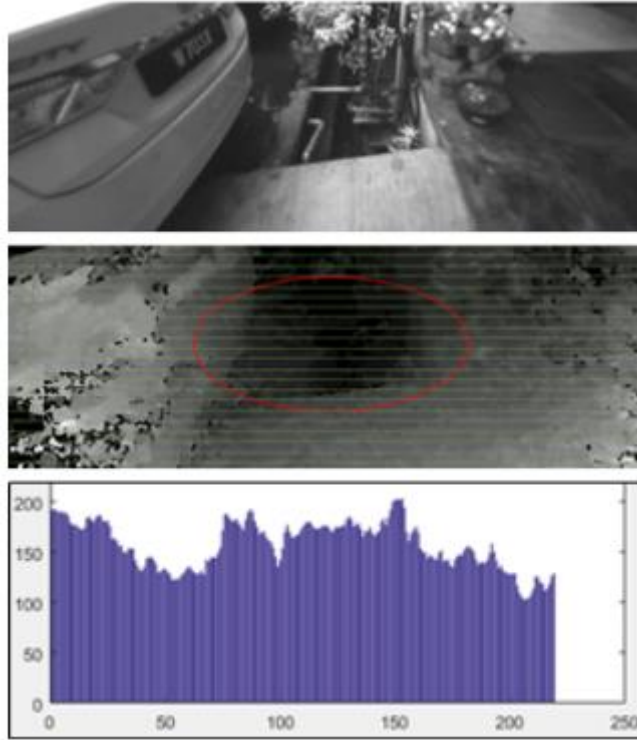


Figure 5.15: Image with an uncovered drainage ahead (top), its disparity map (middle) and the visualization of its depth profile (bottom), sourced from Leong et al. (2017).

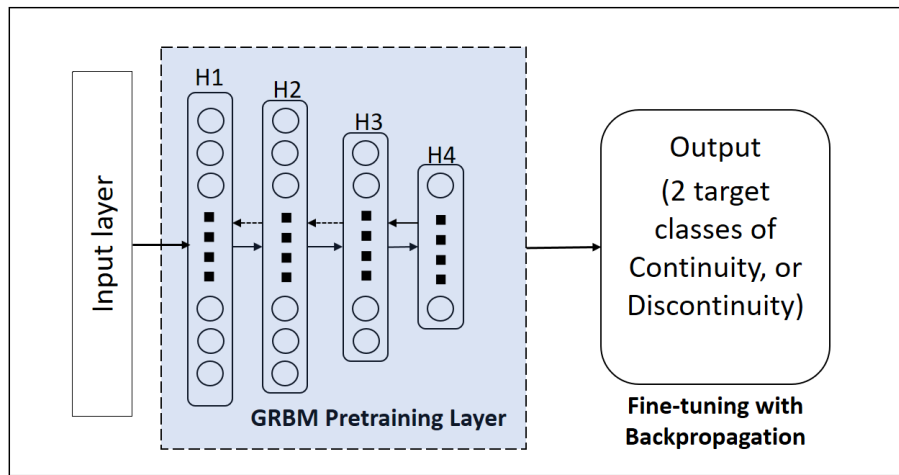


Figure 5.16: Architecture of the DBN used in the experiment (Leong et al., 2017).

The best trained DBN model achieved accuracy of 86% in classifying the continuity and discontinuity classes. No further improvements were observed even with more experiments. The model was not able to classify the surface discontinuity taxonomy, which

is the main objective of the deep learning model training. Due to this reason, the research proceeded to train some CNN models which are typically good at computer vision tasks.

Convolutional neural network. CNN was originally devised by LeCun et al. (1990) for handwritten digit recognition which ended up with great accuracy. Since then, CNNs or the variants of CNN have been successful used in various computer vision, image identification, object detection and image segmentation tasks (Cheung, 2012, Krizhevsky et al., 2012b, Girshick, 2015, Ren et al., 2015, Sun and Qian, 2016, Liu, 2017). CNN is known to have precise accuracy in large scale image classification.

The architecture of a CNN is typically composed of three types of layers – convolutional layer, pooling layer and fully-connected layer. Convolutional layer is unique for a CNN, in which it could have several layers and it is used to extract features of the dataset by convolving image regions with multiple filters. With more such layers, the CNN understands an image from the extracted features progressively. As for the pooling layer, it reduces the size of output maps from the convolutional layer and it has the nature of preventing overfitting by regularizing the network. These two layers make sure the numbers of neurons, parameters and connections of the network are much fewer, thus making a CNN more efficient than a typical backpropagation neural network of similar size and layers. Lastly, the fully-connected layer focuses on the classification task, commonly with a backpropagation algorithm.

5.3.1 The Proposed Tri-Channel Convolutional Neural Network

At an earlier stage, the experiment started with CNN of a single channel input solely using the disparity map for model training. From this experiment, the best model was only able to achieve about 92.6% accuracy of classification on the test set. With further experiment, a tri-channel-single-input CNN which also takes the image pair as input (both right and left image), apart from the disparity map was proposed. The motivation of experimenting

a tri-channel-single-input CNN is very straight forward – although the disparity map holds useful details proportional to depth, the original raw image pair could top up additional details about the surface textures, edges, shapes and any other useful features which can be further exploited by the convolutional layers.

Figure 5.17 illustrates the fusion of both left and right image with disparity map to create a tri-channel input like a typical digital coloured image of RGB-channel. The input fusion approach was inspired by Liu and Liu (2017) in which they combined three different images at pixel-level to be used for object detection. In this implementation, since the image pair and its corresponding disparity map have similar pixel size of 600 x 200 at the region of interest, the author concatenated the image pair and its corresponding disparity map into a three-dimensional array of 600 x 200 x 3 for each unit of input. This is called the fused input. It is important to note that the author did not modify any pixel values of the original images and the disparity map, by concatenation, which means that they were just placed into independent channels of a single fused input. This fused input was then fed into a CNN for model training.

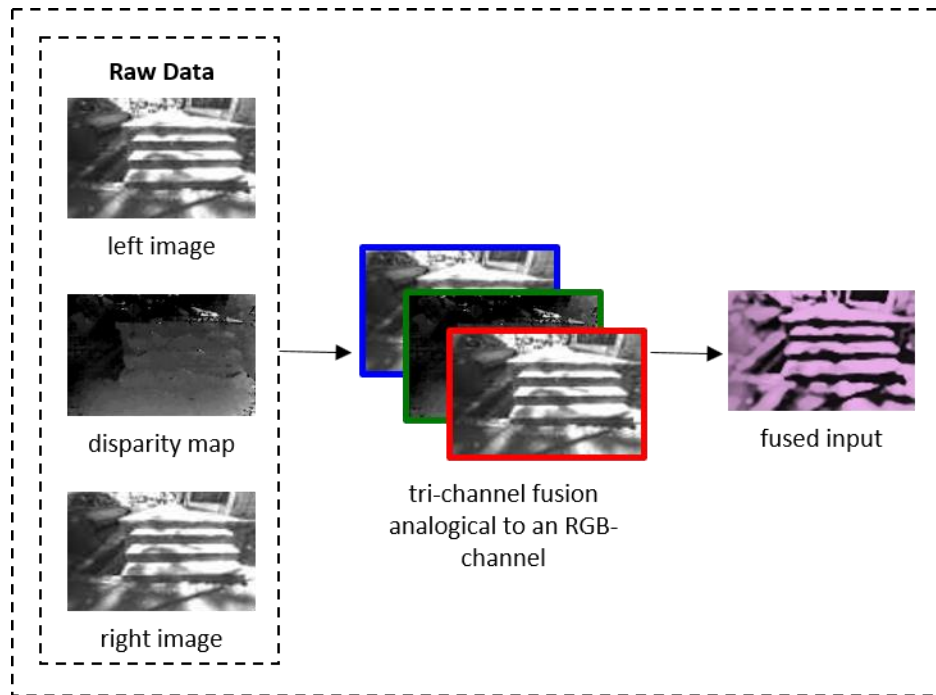


Figure 5.17: The tri-channel fusion of left and right image with disparity map to create a single fused input with three channels like a typical digital coloured image of RGB-channel.

The fused input as shown in the figure has little visualization purpose, yet the actual benefit of such fused input is two-fold in its implementation:

- (1) it allowed for a direct implementation of a single input (but multi-channel) CNN using most of the readily available deep learning framework or APIs; and
- (2) it enabled multiple sources to be used as a single input, and in this case, there are exactly three sources which were fused together.

5.3.2 Architecture of the Proposed CNN

In the development process, the research went through some selection and optimization to reach some potential CNN models. As a start, the experiment was guided by the relatively simple AlexNet's (Krizhevsky et al., 2012b) architecture, but soon it was realized that AlexNet was not the best for this dataset (AlexNet is made up of 5 convolutional layers, each followed by a max-pooling layer, dropout layer, and 3 fully-connected layers). The exploration for a better model continued with references to several other successful deep

learning models such as ZF Net (Zeiler and Fergus, 2013) and VGG Net (Simonyan and Zisserman, 2015). The best model's architecture is somehow different from the above, but it shares some fundamental qualities of VGG Net which will be described in the following section. VGG Net was the winner of ILSVRC 2014 with error rate as low as 7.3%. It has simple but deep 16-layer CNN that strictly uses 3×3 filter with stride and pad of 1, along with maximum pooling layers with stride of 2. Another uniqueness of VGG Net is that it has several convolutional layers stacked up back to back.

The full architecture of the proposed tri-channel-single-input CNN with the best classification accuracy is presented in *Figure 5.18*. The first convolutional layer filters $600 \times 200 \times 3$ input with 32 kernels of size 3×3 and stride of 1×1 pixel. A typical CNN would follow by a pooling layer after the convolutional layer, but for this architecture, it is followed by another convolutional layer instead of a pooling layer. Such design would be observed more often in the following layers, and it shares some similarities to the VGG Net – simple and deep.

The feature map resulting from the first convolution is then convolved by another layer with 32 kernels of size 3×3 and stride of 1×1 pixel. A max-pooling layer follows with a size of 3×3 and stride of 2×2 . Now, it repeats again a double convolutional layer both with 64 kernels of size 3×3 and stride 1×1 pixel. The resulting feature map after the fourth convolutional layer then goes through a max-pooling layer with size of 3×3 and stride of 2×2 . The next layers and the overall architecture are simplified in *Table 5.2* for better comprehension.

Table 5.2: Configuration of the best architecture of the proposed tri-channel-single-input CNN.

Layer	Configuration
Input	600 x 200 x 3
Convolutional layer 1 & 2	Filter size: 3 x 3, 32 kernels, stride: 1 x 1
Max-pooling layer 1	Filter size: 3 x 3, stride: 2 x 2
Convolutional layer 3 & 4	Filter size: 3 x 3, 64 kernels, stride: 1 x 1
Max-pooling layer 2	Filter size: 3 x 3, stride: 2 x 2
Convolutional layer 5, 6 & 7	Filter size: 3 x 3, 128 kernels, stride: 1 x 1
Max-pooling layer 3	Filter size: 3 x 3, stride: 4 x 4
Fully-connected layer 1 (flatten)	12 x 37 x 128 = 56832 neurons, drop-out: 0.1
Fully-connected layer 2 & 3	4096 neurons
SoftMax output layer	9 categorical classes

The tuning and optimization processes used to arrive at this architecture and their respective results are described in Chapter 6.

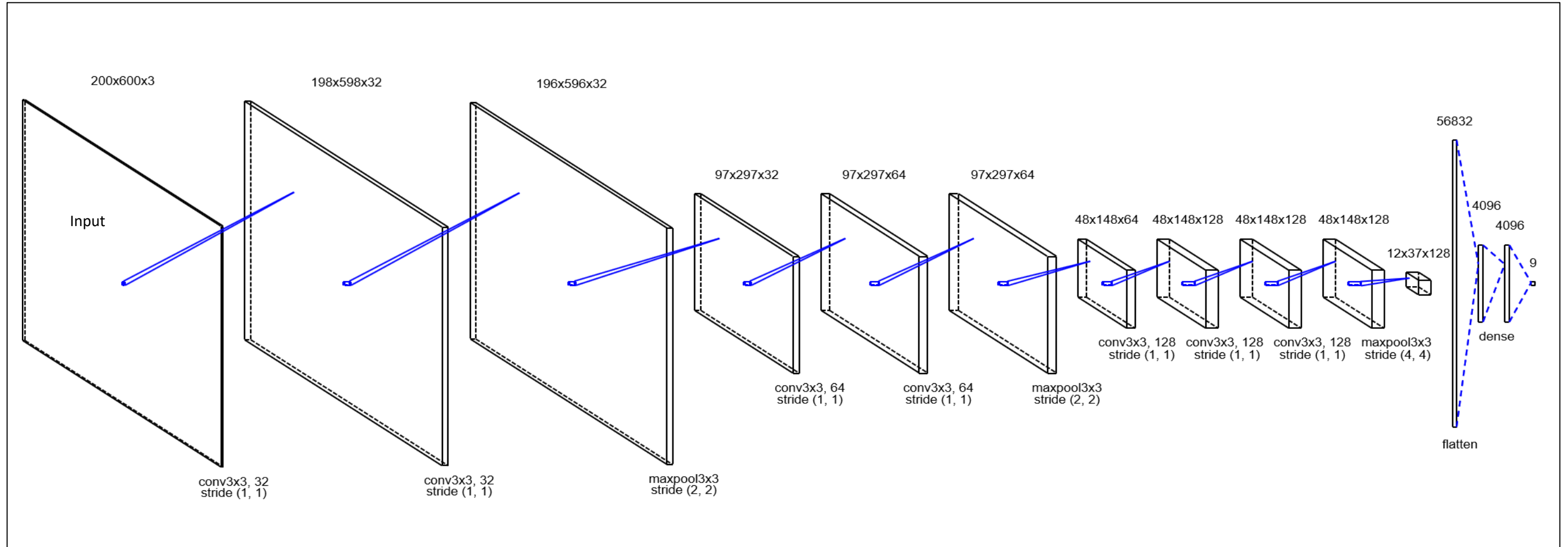


Figure 5.18: An illustration of the full architecture of the proposed tri-channel convolutional neural network for surface discontinuity classification. This architecture has the best accuracy on the testing set.

5.3.3 Training Algorithm – Backpropagation

The fundamental technique of CNN model training is adjusting its filter values or weights through an operation called backpropagation (Goodfellow et al., 2016). For a typical neural network as well as a CNN, the algorithm of a backpropagation can be described in four distinct but related steps – (1) the forward-pass, (2) the loss function, (3) the backward-pass, and (4) the parameters update (this basically refers to weights). In the following paragraphs, the similar backpropagation process used in this tri-channel-single-input CNN is discussed.

In the forward-pass, the operation takes the input of 600 x 200 x 3 array of numbers and passes it through the next layer, and this is continued through the whole network. At the convolutional layer, it uses trainable kernels to filter the result of the previous layer to the output known as the feature map, via an activation function. The operation is denoted in Equation 5-2.

$$x_j^l = f \left(\sum_{i \in M_j} k_{ij}^l x_i^{l-1} + b_j^l \right) \quad \text{Equation 5-2}$$

The equation denotes l layers with any number of neurons in each layer, where M_j is the set of input (or feature maps), b is the bias added to each output map, k is the kernel in which k_{ij}^l means the weight of row i and column j in each kernel at layer l . f is a ReLU activation function.

While for the pooling layer, the operation is sometimes to subsample or summarize the outputs of neighboring neurons of a feature map by a kernel map of:

$$x_j^l = f(\beta_j^l D(X_j^{l-1}) + b_j^l) \quad \text{Equation 5-3}$$

where β is a multiplicative bias, b is an additive bias and $D()$ is a subsampling function. In the network architecture, a max-pooling subsampling function is used instead of average pooling. According to Cireřan et al. (2012), max-pooling is shown to be better than average pooling in extracting important features such as edges of an object from the feature map.

The fully-connected layers are treated similarly as the typical hidden layers of a multilayer Perceptron.

The activation function used in the proposed CNN is Leaky ReLU (rectified linear unit) as defined in Equation 5-4. This is the same activation that is used in the convolutional and fully-connected layers. A standard ReLU has one caveat known as “the dead ReLU” in which it happens to output zero when the input it takes is negative. This would hinder backpropagation because the gradients will be zero after a negative value has been entered to the activation function. On a larger scale, a dead ReLU will output the same value which is zero for all activation. The choice of Leaky ReLU allows a small and positive value to the gradient when the unit is not active, thus it avoids generating a large scale of zeros that will corrupt the entire network.

$$f(x) = \begin{cases} x, & \text{if } x > 0 \\ 0.01x, & \text{otherwise} \end{cases} \quad \text{Equation 5-4}$$

The activation function of output layer is SoftMax function, devised for a multi-class classification. The function is given as:

$$\text{class output, } a_j = \frac{e^{Z_j}}{\sum_{k=1}^K e^{Z_k}} \text{ for } j = 1, \dots, K \quad \text{Equation 5-5}$$

where Z represents K-dimensional vector with a range of (0, 1], and e is an exponential function. The K-dimensional vector is used to represent a categorical distribution. In the proposed CNN, K is simply 9 as it has nine different possible outcomes based on the taxonomy of surface discontinuity developed earlier.

At this point of forward-passing, the algorithm needs to update parameters (or weights) to minimize the difference between the predicted output and the actual output. Before the update is possible, the network needs to have an idea of the difference. This is where the loss function comes into play. The difference can be computed in terms of loss. There are several commonly used loss functions. The author reviewed and analyzed classification errors, mean squared errors and cross entropy loss functions, and found the latter works well for the task, and it is defined as:

$$Loss_i = - \sum_i y'_i \log(y_i) \quad \text{Equation 5-6}$$

where y'_i is the actual output (or the labelled class) of i^{th} training instance and y_i is the prediction result of the classification for the similar i^{th} training instance. Loss is minimized during training using a gradient descent technique.

There are various gradient descent methods, and experiments on several variants of them have found mini-batch gradient descent to be the most suitable. However, the learning momentum for mini-batch gradient descent needs some optimization. There are a few widely used modern optimizers for this task and the research experimented Adam (Adaptive Moment Estimation) (Kingma and Ba, 2015) and its variants. Although AdaMax (Ruder, 2016), a variant of Adam, had sometimes recorded some better results in the training set, results from Adam are more consistent across all experiments with the test set.

The ultimate purpose of model training is to get the predicted output as close as possible to the actual output (the label). To get there, the network is minimizing the loss, which also means that the algorithm has to identify which weights are most directly contributing to the error (loss) of the network. This is when backward-pass is performed through the network. The task of a backward-pass is to determine which weights contributed to the loss, and to adjust the weights such that the loss can be minimized. This task is traditionally performed as a derivative of:

$$w_j = w_i - r \left(\frac{dL}{dW} \right) \quad \text{Equation 5-7}$$

where w_j is the adjusted weight at a particular layer, w_i is the initial weight, L is the error or loss, and r is the learning rate. Traditionally the learning rate has to be predefined before training takes place. The author did not define the learning rate in this network since it is using a modern optimizer – Adam, and it allows for an adaptive learning rate throughout the training. Despite Adam's adaptive learning, there are two controls that were applied to its rate in the training process. Firstly, the initial learning rate was set to a relatively low value of 0.0003. Secondly, the decay rates of Adam were set to some boundaries to guide

the adaptive learning momentum. With the weights updated through this process, the four-step backpropagation algorithm as mentioned earlier is finally completed.

5.3.4 Model Tuning – an Optimization Approach

Large numbers of experiments were conducted to identify some of the best models to classify the training set of the generated data. The experimental architectures of the models varied by their sizes of kernel filter (which also resulted in the different sizes of feature maps), formations of convolutional layers and fully-connected layers. With each different network architecture, the hyperparameters such as some variants of ReLU activation function, loss function and network optimizer were also analyzed. A technique to tune or optimize the models in the effort to achieve a predefined target was employed. Before the model tuning, a target of achievement was set. In the earlier experiment with a deep belief network using a proposed depth profile (Leong et al., 2017) for model training, it was able to achieve a classification accuracy of 86%. With this accuracy as a baseline, it is believed that CNN could perform even better according to several success cases from the literature. Hence, an accuracy of 95% for the following experiments was targeted.

To train the best model that can help achieve the target is a challenging task. Firstly, the qualities that contribute to a good model in the context of this research must be defined. The qualities that should be examined for this research are both efficiency and accuracy. In a cloud-based solution, it is typical that the computing power is not given much concern, while the accuracy will be very much emphasized. However, in an embedding system such as in this prototype, the cost of the processing power is a great part of the trade-off. The challenge for this model is how to increase efficiency (reduction in the processing cost), which is measured by processing power and memory requirement, with as little impact to accuracy as possible.

Secondly, an appropriate approach to improving the model's performance is needed. This approach should be as simple as possible such that the technique employed should not be another open research problem itself. Based on recommendation from Goodfellow et al. (2016), the performance of a deep learning model could be improved through:

- (1) data
- (2) right choice of model (or algorithm)
- (3) model (or algorithm) tuning
- (4) ensembles

For improvement with data, it was discussed earlier that the author sampled the sequences at every 3 frames to optimize the difference between each frame of the image sequences. Data augmentation was performed to eventually obtain more data. To ensure the right choice of model, the author has based it on literature, selecting several relevant CNN models with proven performance as a start. For improvement with ensembles, it is generally known that ensembles increase the complexity of the computation. Hence, it will not be the first choice of consideration as the author will have to deal with its deployment to the prototype with limited processing power and memory. Finally, the improvement choice is left with model tuning which can be done by repeating the model training process with different network configurations or hyperparameters to identify the best performer.

To better manage the model tuning, a search or optimization algorithm can be utilized. The tuning can be viewed as a combinatorial and global optimization problem. The objective is to maximize the classification accuracy while minimizing the model size or complexity. Typically, the following items could be altered by the tuning algorithm to get the best model.

- (1) Number of layers
- (2) Dimensions of convolution kernel filter in each layer
- (3) Number of strides
- (4) Connectivity between layers

- (5) Number of neurons in each fully-connected layer
- (6) Variant of ReLU activation
- (7) Type of loss function
- (8) Type of momentum optimizer

The optimization was executed in two steps. Step one was to train the network to get the best classification rate. Step two was to vary the parameters of items (1) to (8) above using a search or optimization algorithm, and then the cycle was repeated until the best model was found. Based on the theoretical analysis and review described in Section 5.3.4, the optimization for items (6) and (7) was eliminated. Item (4) - the connectivity between layers – was not part of the tuning as well, since some well-developed architectures were referred, the connectivity or organization of those models was preserved. By doing so, the author had reduced the scale of the optimization task. For the remaining items on the list above, the author implemented the optimization within certain scopes or groupings as shown in *Table 5.3*. There are some specifications in tuning the parameters for each different model which will be described in Chapter 6, *Table 5.3* is just a general summary about the tuning.

Based on the number of parameters to be tuned, a combinatorial and global optimization algorithm was needed. Exhaustive search is not possible for this task as it will have to explore an approximation of 245,760 discrete possibilities for each model. One such model will take a few days to weeks to complete a single training, thus making an exhaustive search impractical for the experiment. Heuristic (or meta-heuristic) search may be a better choice. It could produce an optimized solution in a reasonable time frame faster than the exhaustive counterparts. Although there is no guarantee that the solution is the best of all, the trade-off is acceptable if it could systematically guide the model tuning to reach the target accuracy within a constrained time frame. Having pointed out that heuristic search could be the choice, caution should be practiced as quite many such algorithms are themselves yet another optimization problem with large pre-defined parameters needed

before they can be employed. Most of the population-based heuristic or meta-heuristic optimizations (such as genetic algorithm or particle swarm optimization) were skipped, and some single-solution optimizations (such as simulated annealing) were also avoided because apart from the complexity to implement them for a neural network design, they required some heavy parameter setting prior to their usage.

Table 5.3: Items, parameters and ranges used in the model tuning.

Item to be tuned	Sub-item (group)	Parameter / Range
Number of layers	Convolution	3 – 10
	Fully-connected	2 – 3
Dimensions of convolution kernel filter in each layer	width x height	3 x 3, 5 x 5, 7x 7, 9 x 9, 11 x 11
	depth	2^n , $n = [4, 5, 6, 7]$
Number of strides	Convolutional layer	1 x 1, 2 x 2
	Pooling layer	1 x 1, 2 x 2, 3 x 3, 4 x 4
Number of neurons in each fully-connected layer	size	2×2^n , $n = [8, 9, 10, 11]$
	dropout rate, d	None, or $d = [0.1, 0.2, 0.3, 0.4, 0.5]$
Type of momentum optimizer for mini-batch gradient descent	-	Adam, AdaMax

5.3.5 Variable Neighbourhood Search

One of the meta-heuristic search algorithms stands out from the rest in terms of its simplicity to optimize a network architecture. This is the variable neighbourhood search (VNS) developed by Mladenović and Hansen (1997). Another additional point is that it requires little to no parameter setting. It exploits systematically the change in the neighbourhood, both in the descent to local minima, and in the escape from the valleys which trap the potential solutions (the algorithm of VNS is given in Appendix 4). Araújo et

al. (2017) proposed a VNS method for CNN architecture design and tested it on the famous CIFAR-10 image dataset. CIFAR-10 dataset (Canadian Institute for Advanced Research) is a collection of images that are commonly used to train machine learning and computer vision algorithms. It is one of the most widely used datasets for machine learning research. The dataset contains 60,000 32x32 colour images in 10 different classes. Their optimized model has improved the validation loss by 15% as compared to their benchmarked models. The model tuning adopted the similar method proposed by Araújo et al. but restricted the optimization to only tuning the parameters mentioned in *Table 5.3*. Since a discrete set of parameters to be optimized has already been defined, one may consider the combinatorial cum global optimization problem as:

$$\min f(x) \quad \text{Equation 5-8}$$

subject to:

$$x \in S \quad \text{Equation 5-9}$$

where $f(x)$ is the cost function to be minimized and S is the set of feasible solutions. The author has defined $f(x)_i = -\sum_i x'_i \log(x_i)$ in Equation 5-6, and minimizing $f(x)$ is the same as maximizing the predictive power of the model, hence the optimization. For VNS to work on the CNN model tuning, the solution representation can be denoted in a structure array:

$$S_m = \text{struct}(p_1[m_1 \times n_1], p_2[m_2 \times n_2], \dots, p_i[m_i \times n_i]) \quad \text{Equation 5-10}$$

where S is the solution, m is the identifier of the solution represented by an integer, p is the parameter with size of $m \times n$, and there could be up to i number of parameters to be optimized. For example, based on the parameters from *Table 5.3* there is a total of 9 sets of parameters with different discrete variables to be optimized. Thus *Equation 5-10* can be now partially written as:

$$S_1 = \text{struct}(p_1[1 \times 8], p_2[1 \times 2], p_3[2 \times 5] \dots) \quad \text{Equation 5-11}$$

for a solution S_1 , with a possibility to select a set of discrete values from each of the parameters of $p_1 = \{3,4,5,6,7,8,9,10\}$, $p_2 = \{2,3\}$, $p_3 = \{[3 \times 3], [5 \times 5], [7 \times 7], [9 \times 9], [11 \times 11]\}$ and so on. The structure array is a convenient way to provide the VNS algorithm with

the index of each variable, and with these indices, corresponding values can be retrieved such that they are then taken into the solution space representing a specific solution S_m .

5.4 Summary of the Chapter

This chapter presented the major techniques involved in the system development. The system structure was described, and the major components of the system were elaborated. Two main techniques needed for building a classifier model – feature extraction and assembling of a convolutional neural network were provided with detailed descriptions of considerations and steps involved. Several decisions were made along the development and justifications of certain choices were discussed.

Chapter 6: System Evaluation

The chapter summarizes model training which was conducted based on three different widely cited CNN models and their variants. From the training, model with the best classification accuracy was implemented for field evaluation. Next, the chapter presents the evaluation in two aspects – classification accuracy evaluation and efficiency evaluation – and provides further analyses and conclusions for these evaluations.

6.1 Models Training and Classification Results

To train the best model using the data, some experiments with variation on the CNN were carried out. From the data augmentation, a total of 60750 images were obtained. Since the data were stored as image sequences just like any uncompressed videos, the author sampled the sequences at every 3 frames to optimize the difference between each frame. With this sampling, eventually 20250 images were used to train or test the models. 80% of the dataset was randomly selected for training, and 20% of the dataset was used for testing. 10% of the training set was also randomly selected as a validation set during model training.

6.1.1 *Model 1 and Its Variants*

The first model was designed based on AlexNet's architecture (Krizhevsky et al., 2012b). Although in the past few years, several CNN models have outperformed AlexNet in some similar image classification tasks, the author would still like to take it as a starting point because this model was the first one to perform so exceptionally well on a historically

challenging ImageNet dataset, and most of the techniques developed for the model are still widely used today.

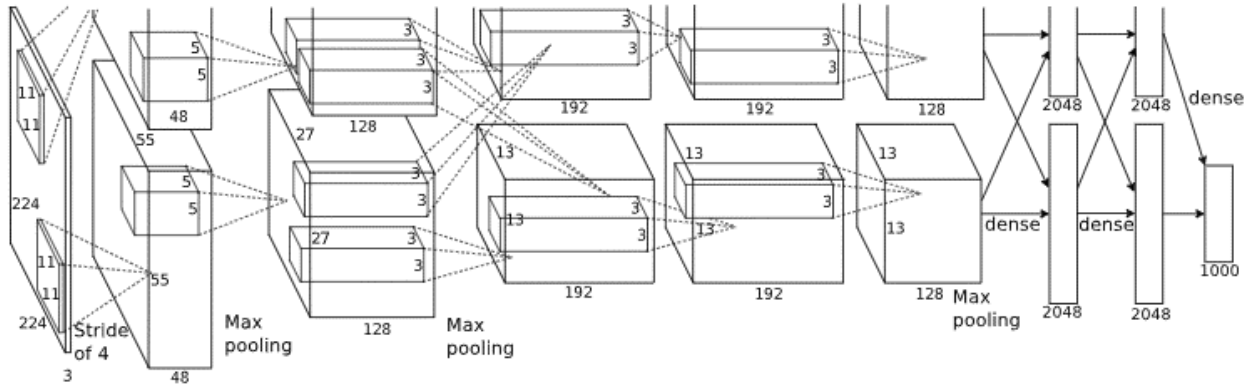


Figure 6.1: An illustration of the architecture of AlexNet (Krizhevsky et al., 2012b).

AlexNet has a relatively simple layout as compared to most modern architectures. The network contains eight learned layers inclusive of five convolutional layers and three fully-connected layers. The fully-connected output layer is fed to a 1000-way SoftMax activation function to produce a distribution of the 1000 class labels (based on Equation 5-5, $K = 1000$). One notable part of AlexNet is the two-streamed network after the input layer, which are eventually met at the first fully-connected layer. Both streams have identical architecture as described above. Along the two streams, the second max-pooling layer on both streams output features to each other in the following convolutional layer. At the fully-connected layers, outputs from both streams are shared across each other, before they eventually meet at the single output layer. AlexNet also implemented some overlapping pooling layers at the first and the last feature maps, which was claimed to reduce overfitting.

In the experiment, the model tuning was initialized with exactly the same network architecture and other hyperparameters. Leaky ReLUs were used after each convolutional or fully-connected layer, which is slightly different from AlexNet that uses ReLU followed by a response-normalization layer. The author named all models based on AlexNet as M1-x, in which x indicates the variants that had been experimented, i.e. M1-2 is the second of such model in the experiment. Table 6.1 summarizes some best models from the

experiment. Based on the table, “11x11,96” represents the size of filter kernels which is 11 x 11, with 96 kernels. Under the column for “Fully-connected”, “4096, 0.5” indicates the fully-connected layer has 4096 connected neurons and a dropout rate of 0.5. All models have an overlapping Max-pooling layer at the first and the last feature maps. The tabulated classification accuracies are based on the test set.

Table 6.1: Top 10 architectures of model based on AlexNet.

Model	Convolutional layers							Fully-connected		Acc. (%)
	C1	C2	C3	C4	C5	C6	C7	FC1	FC2	
M1-1	11x11, 96 Stride 4x4	5x5, 256 Stride 2x2	3x3, 384 Stride 2x2	3x3, 384 Stride 2x2	3x3, 256 Stride 2x2	-	-	4096, 0.5	4096, 0.5	85.8
M1-4	11x11, 96 Stride 4x4	5x5, 256 Stride 2x2	3x3, 384 Stride 2x2	3x3, 384 Stride 2x2	3x3, 256 Stride 2x2	-	-	4096, 0.5	2048, 0	85.2
M1-6	11x11, 64 Stride 4x4	5x5, 128 Stride 2x2	3x3, 256 Stride 2x2	3x3, 256 Stride 2x2	3x3, 256 Stride 2x2	-	-	2048, 0.3	2048, 0	76.2
M1-18	11x11, 64 Stride 4x4	5x5, 256 Stride 2x2	3x3, 512 Stride 2x2	3x3, 512 Stride 2x2	3x3, 256 Stride 2x2	-	-	4096, 0.5	4096, 0.2	77.3
M1-22	9x9, 128 Stride 4x4	5x5, 256 Stride 2x2	3x3, 256 Stride 2x2	3x3, 256 Stride 2x2	3x3, 256 Stride 2x2	-	-	4096, 0.5	2048, 0.1	76.1
M1-24	9x9, 128 Stride 4x4	5x5, 256 Stride 2x2	3x3, 384 Stride 2x2	3x3, 384 Stride 2x2	3x3, 384 Stride 2x2	-	-	4096, 0.5	2048, 0	78.4
M1-54	9x9, 64 Stride 4x4	5x5, 128 Stride 2x2	3x3, 256 Stride 2x2	3x3, 256 Stride 2x2	3x3, 256 Stride 2x2	3x3, 128 Stride 2x2	-	4096, 0.5	4096, 0.2	80.6
M1-56	9x9, 96 Stride 4x4	5x5, 256 Stride 2x2	3x3, 384 Stride 2x2	3x3, 384 Stride 2x2	3x3, 256 Stride 2x2	3x3, 256 Stride 2x2	-	4096, 0.5	4096, 0.1	86.7
M1-71	7x7, 64 Stride 4x4	3x3, 256 Stride 2x2	3x3, 384 Stride 2x2	3x3, 384 Stride 2x2	3x3, 256 Stride 2x2	3x3, 256 Stride 2x2	-	4096, 0.5	2048, 0.2	88.6
M1-87	7x7, 64 Stride 4x4	5x5, 64 Stride 2x2	3x3, 128 Stride 2x2	3x3, 128 Stride 2x2	3x3, 128 Stride 1x1	3x3, 256 Stride 1x1	3x3, 256 Stride 1x1	2048, 0.4	2048, 0.1	86.1

Model M1-1 is closely similar to AlexNet (except that it uses a different activation function), and this well-cited model has a top 10 place in the list when compared to its variants optimized by VNS. It is observed that some models (M1-71 and M1-87) with smaller convolving size and stride contributed to better accuracy. This finding concurred with the following experiments in which ZF Net – a fine-tuned version of AlexNet by Zeiler and Fergus (2013) was referred to. From the top models, it is also observed that the fully-connected layer with averagely 0.5 dropout rate has performed well on the test set. This indicates higher dropout rates helped the models generalize better.

6.1.2 Model 2 and Its Variants

The research continued to experiment with another famous model which outperformed AlexNet in ILSVRC 2013 competition. This is the ZF Net (Zeiler and Fergus, 2013). ZF Net was trained on 1.3 million images only as compared to AlexNet of 15 million images, but it had beaten the latter in the similar competition. ZF Net has a very similar architecture to AlexNet, except for a few modifications (see *Figure 6.2*). First it uses smaller filter size of 7×7 instead of 11×11 , and a smaller stride of 3 for all layers. Secondly, it has only a single branch of network layers as opposed to AlexNet, while it follows the exact network organization and connectivity.

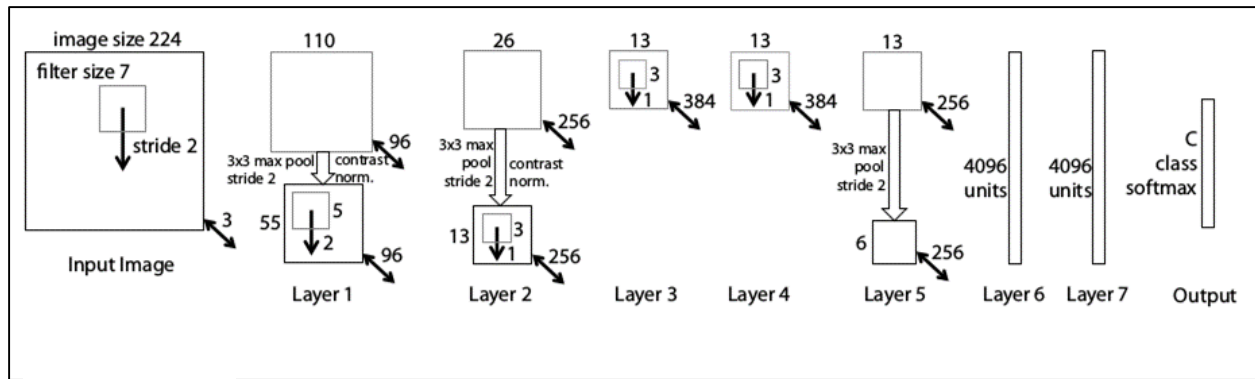


Figure 6.2: ZF Net architecture developed by Zeiler and Fergus (2013).

In the experiment, the model tuning was initialized with exactly the same network architecture and other hyperparameters of ZF Net. Again, leaky ReLU was used after each convolutional or fully-connected layer. The author named all models based on ZF Net as M2-x, in which x indicates the variants that had been experimented. *Table 6.2* summarizes some best models from the experiment.

Table 6.2: Top 10 architectures of model based on ZF Net.

Model	Convolutional layers							Fully-connected		Acc. (%)
	C1	C2	C3	C4	C5	C6	C7	FC1	FC2	
M2-1	7x7, 96 Stride 2x2	5x5, 256 Stride 2x2	3x3, 384 Stride 2x2	3x3, 384 Stride 2x2	3x3, 256 Stride 2x2	-	-	4096, 0.5	4096, 0.5	87.1
M2-39	7x7, 64 Stride 2x2	3x3, 64 Stride 2x2	3x3, 64 Stride 2x2	3x3, 128 Stride 2x2	3x3, 128 Stride 2x2	-	-	2048, 0.3	2048, 0.2	86.9
M2-65	7x7, 64 Stride 3x3	5x5, 128 Stride 2x2	3x3, 256 Stride 2x2	3x3, 256 Stride 2x2	3x3, 64 Stride 2x2	-	-	2048, 0.2	1024, 0.2	76.3
M2-71	7x7, 64 Stride 2x2	5x5, 128 Stride 2x2	3x3, 256 Stride 2x2	3x3, 256 Stride 2x2	3x3, 512 Stride 2x2	-	-	4096, 0.5	4096, 0.2	86.1
M2-78	7x7, 96 Stride 2x2	5x5, 128 Stride 2x2	3x3, 256 Stride 2x2	3x3, 256 Stride 2x2	3x3, 512 Stride 2x2	-	-	4096, 0.5	4096, 0.5	87.0
M2-85	7x7, 128 Stride 2x2	5x5, 256 Stride 2x2	3x3, 256 Stride 2x2	3x3, 256 Stride 2x2	3x3, 512 Stride 2x2	-	-	4096, 0.5	4096, 0.3	87.7
M2-95	5x5, 64 Stride 2x2	3x3, 128 Stride 2x2	3x3, 128 Stride 2x2	3x3, 256 Stride 2x2	3x3, 256 Stride 2x2	3x3, 256 Stride 2x2	3x3, 256 Stride 1x1	4096, 0.5	4096, 0.4	88.6
M2-96	5x5, 128 Stride 2x2	5x5, 256 Stride 2x2	3x3, 384 Stride 2x2	3x3, 384 Stride 2x2	3x3, 256 Stride 2x2	3x3, 256 Stride 2x2	3x3, 256 Stride 1x1	4096, 0.5	4096, 0.4	90.2
M2-97	5x5, 96 Stride 2x2	5x5, 128 Stride 2x2	3x3, 128 Stride 2x2	3x3, 384 Stride 2x2	3x3, 256 Stride 2x2	3x3, 256 Stride 2x2	3x3, 256 Stride 1x1	4096, 0.5	2048, 0.4	88.8
M2-99	5x5, 128 Stride 2x2	5x5, 256 Stride 2x2	3x3, 384 Stride 2x2	3x3, 384 Stride 2x2	3x3, 256 Stride 2x2	3x3, 256 Stride 1x1	3x3, 256 Stride 1x1	2048, 0.4	2048, 0.5	88.6

A similar version of ZF Net is implemented as M2-1, and it is indeed a pruned-down version of AlexNet with a tiny improvement when compared to its predecessor (an accuracy of 87.1% versus 86.8%). With smaller filter size (7 x 7) for convolving the input, it helped to retain more original pixel information and hence the improvement. A trend about the optimization was observed, in which a smaller filter size to convolve the input gave better results. Possible explanation for this observation is: the original implementation of 11 x 11 filter could have skipped a lot of useful information from the pixel, especially from the first input layer. Another important finding that could be inferred from M2-95 to M2-99 is when the network gets deeper (with a growing number of filters as well), it seems to improve the generalization, hence achieving better accuracy for the test set. Model M2-96 is one such example with accuracy of 90.2%, the highest ever recorded in this experiment.

6.1.3 Model 3 and Its Variants

With VGG Net (Simonyan and Zisserman, 2015) achieving an even lower error rate in ImageNet classification, it outperformed both AlexNet and ZF Net which were developed earlier (although it wasn't the winner of ILSVRC 2014). Based on the previous experiment on ZF Net in which a smaller filter at the first layer and deeper network produced better result, this finding offers a hint that VGG Net's architecture is worth exploring. VGG Net is basically just simple and deep. The section will provide more explanation for this model because this is where some of the best models come from.

Simonyan and Zisserman (2015)'s VGG Net has several different versions ranging from 11 to 19 layers of network, with the 16-layer being the best performer. VGG Net uses strictly 3×3 kernel filters with stride of 1 for all convolutional layers, hence the simplicity of it. One notable feature of the architecture of the network is that it has two to three convolutional layers stacked together back to back without any pooling layer between them. This formation has created a deeper network as compared to the previous models trained. The author named all models based on VGG Net as M₃-x, in which x indicates the variants that had been experimented. *Table 6.3* summarizes some selected models from the experiment.

The tuning was initialized with exactly the same configurations and hyperparameters of VGG Net as can be seen in Model M₃-1 in the table. It is observed that many variants from this model achieved some good classification accuracies above 90%. The tuning allowed variable neighbourhood search to optimize up to some seven stacked convolutional layers, but all the models beyond 5 stacked layers were seen saturated in gaining further improvement (these examples are M₃-67 and M₃-82 shown in *Table 6.3*). Based on *Table 6.3*, there is a Max-pooling layer with 2×2 stride between two stacked layers, and before the first fully-connected layer. 3×3 kernel filter with stride of 1×1 is used for all convolutional layers. "K-64, K-64, Pool- 2×2 " represents a stacked of two convolutional layers with 64 kernel filters followed by a 2×2 sized Max-pooling layer.

Table 6.3: Selected examples of M_3 -x model based on VGG Net.

Model	Convolutional layers							Fully-connected		Acc. (%)
	Stacked 1	Stacked 2	Stacked 3	Stacked 4	Stacked 5	Stacked 6	Stacked 7	FC1	FC2	
M3-1	K-64 K-64 Pool-2x2	K-128 K-128 Pool-2x2	K-256 K-256 K-256 Pool-2x2	K-512 K-512 K-512 Pool-2x2	K-512 K-512 K-512 Pool-2x2	-	-	4096, 0.5	4096, 0.5	95.8
M3-45	K-32 K-32 Pool-3x3	K-64 K-64 Pool-3x3	K-128 K-128 K-128 Pool-3x3	-	-	-	-	4096, 0.5	4096, 0.2	96.7
M3-5	K-64 K-64 Pool-2x2	K-128 K-128 Pool-2x2	K-256 K-256 K-256 Pool-2x2	K-512 K-512 K-512 Pool-2x2	K-512 K-512 K-512 Pool-2x2	-	-	4096, 0.5	4096, 0.4	91.0
M3-11	K-64 K-64 Pool-2x2	K-128 K-128 Pool-2x2	K-256 K-256 K-256 Pool-2x2	K-512 K-512 K-512 Pool-2x2	K-512 K-512 K-512 Pool-2x2	-	-	4096, 0.4	2048, 0.4	90.2
M3-20	K-64 K-64 Pool-2x2	K-128 K-128 Pool-2x2	K-256 K-256 K-256 Pool-2x2	K-512 K-512 K-512 Pool-2x2		-	-	4096, 0.5	4096, 0.3	93.7
M3-34	K-32 K-32 Pool-2x2	K-128 K-128 Pool-2x2	K-256 K-256 K-256 Pool-2x2	K-512 K-512 K-512 Pool-2x2	-	-	-	4096, 0.5	4096, 0.5	87.1
M3-41	K-32 K-32 Pool-3x3	K-64 K-64 Pool-3x3	K-128 K-128 K-128 Pool-3x3	-	-	-	-	4096, 0.5	2048, 0.3	94.6
M3-44	K-32 K-32 Pool-3x3	K-64 K-64 K-64 Pool-3x3	K-128 K-128 K-128 Pool-3x3	-	-	-	-	4096, 0.4	2048, 0.2	94.8
M3-67	K-64 K-64 Pool-2x2	K-128 K-128 Pool-2x2	K-256 K-256 K-256 Pool-2x2	K-512 K-512 K-512 Pool-2x2	K-512 K-512 K-512 Pool-2x2	K-512 K-512 K-512 Pool-2x2	-	4096, 0.5	4096, 0.5	96.5
M3-82	K-64 K-64 Pool-2x2	K-128 K-128 Pool-2x2	K-256 K-256 K-256 Pool-2x2	K-512 K-512 K-512 Pool-2x2	K-512 K-512 K-512 Pool-2x2	K-512 K-512 K-512 Pool-2x2	K-512 K-512 K-512 Pool-2x2	4096, 0.5	4096, 0.4	96.3

According to Simonyan and Zisserman, two convolutional layers with 3×3 filter size when stacked together will produce an effective receptive field of 5×5 . This resulted in a larger filter while at the same time maintained the benefits of smaller filters that could retain more pixel information from the input or the consecutive feature maps. The tuning results agreed with this statement. In a similar way, the stack of three convolutional layers with 3×3 filter size provides a receptive field of 7×7 , just like the first convolutional layer of ZF Net, but with the extra benefit of keeping more details as explained. Another unique

structure of this model is that the number of filters doubles after each Max-pooling layer. The shrinking spatial dimensions of the feature maps could be reinforced by the growing depth, and thus possibly contribute to some good performance as observed from the experiment.

From all the experiments, it was observed that a smaller convolutional filter at the input layer gives better classification. With the stacked convolutional layers in the network, there was a great jump in the classification accuracy. The error of classification was mainly contributed by the labelled class of D5 (discontinuity – blended gradient). It could be that this class – a blended gradient – consists of some very complex and physically different samples as compared to some less complex classes such as a smooth walkway or a down-stairs with obvious identical perceptual features.

From the above experiments, M3-45 was found to be the best model. Although there are other popular deep learning networks such as GoogLeNet and Microsoft ResNet developed for similar image classification (or object detection) tasks, the research did not pursue further experiments on these networks based on the following justifications. Firstly, the model training and tuning have not only reached, but exceeded the pre-defined target of 95% accuracy in classification. Secondly, it will be very costly to explore several other well-cited deep learning networks due to their very deep layers and network complexities. For examples, GoogLeNet is a very complex network with deep layers (about 100 layers), while Microsoft ResNet could be relatively simple, yet it has a very deep network with 152 layers; hence making them not some ideal models to be further explored in the experiment. A shallower model suits the prototype as this research needs to consider the lower memory resource and computing power of the mobile platform.

6.2 Evaluation of the Prototype

It was set at the phase of research design (Chapter 3) that the system will be evaluated in two aspects: (1) accuracy, and (2) efficiency of the prototype. Accuracy refers to the precision of the model to classify the surface discontinuity as defined in the taxonomy. Efficiency is the ratio of the useful work performed by the prototype to the total resource taken in. These two aspects are largely the result of the trained classification model. Firstly, the model was trained with the training set and evaluated with the testing set to be able to classify the taxonomy as accurately as described in Section 6.1. However, the accuracy of the classification must also be evaluated in the field to validate whether the model has yielded some similar results as indicated by the accuracy performance on the testing set.

Secondly, the efficiency of the prototype could be affected by both the software design and the hardware implementation. As this research has not placed any scope on the study of hardware, the only aspect of efficiency should be concerning the software design. For the case of this prototype involving a deep learning model, the efficiency of software design is largely affected by the model. The model consists of network configuration and weights, apart from the algorithm developed to execute it. Based on the definition of “efficiency as the ratio of the useful work performed by the prototype to the resources taken in”, both terms “useful work” and “resources” must be defined. Based on the research context, “useful work” is referred to the task of executing the model to perform classification, while main “resources” related to the above task are memory and computation. The convolutional layers are computation-intensive, while the fully connected layers are memory-intensive (Sun et al., 2017). These circumstances have brought many challenges to the implementation of CNN on embedded platforms. Both memory and computation loads would affect the power consumption of the prototype, and this is an important aspect to be evaluated especially for the prototype being a wearable device.

6.2.1 Classification Accuracy Evaluation

Prior to the accuracy evaluation, some lab tests were conducted. The lab tests were needed to serve as a precaution before the research ventures out to the field for further evaluation of the prototype. The lab tests had eliminated some possible unforeseen circumstances before actual field evaluation was conducted. Utilizing the landscape around the university campus as a convenient built environment, the author tested the readiness of the phase-2 prototype. A set of purpose-built adjustable blocks for the lab testing was prepared. The set of blocks as shown in *Figure 6.3* can be adjusted to form or resemble stairs, rises, drop-offs, curbs and other forms of surface discontinuity. Several test cases were conducted, and some issues related to the stereo camera's parameter setting (exposure rate and illumination control) were found. After the issues were rectified, the prototype was then tested for classification without much problem. It was then ready for the next activity – field evaluation.



Figure 6.3: Adjustable blocks set for the lab test can be used to resemble stairs, rises, drop-offs, curbs and other forms of surface discontinuity.

Field Evaluation in Locations where Data were Collected

Field evaluation was performed by the author (a sighted person) at five locations where the surface discontinuity samples were collected during the data collection phase. As there were nine different classes to examine, these five locations will provide the nine different surface discontinuities for evaluation (smooth surface is the other one class that is available at most of these locations).

Table 6.4: Evaluation samples from several locations around Klang Valley, Selangor, Malaysia.

Location	Class Label	Unit of test conducted
Damansara Perdana	SW	32
	D ₁	11
	D ₃	3
	D ₄	4
	R ₂	11
	R ₃	3
Bandar Utama	SW	12
	D ₂	4
	D ₄	4
	R ₁	4
Up-Town 1	SW	12
	D ₃	3
	D ₄	4
	D ₅	2
	R ₃	3
Up-Town 2	SW	7
	D ₄	3
	D ₅	4
Damansara Jaya	SW	22
	D ₁	4
	D ₂	3
	D ₃	3
	D ₅	2
	R ₁	3
	R ₂	4
	R ₃	3
Total Unit tested		170

Table 6.4 shows the units of test conducted, with their class labels and definitions (SW – smooth walkway, D1 – down-steps, D2 – drop-off, D3 – down-ramp, D4 – uncovered drainage, D5 – blended gradient, R1 – rise, R2 – up-steps, R3 – up-ramp).

A minimal feedback mechanism to aid the author (a sighted person) in the evaluation was added to the prototype. It was a simple audio feedback using a buzzer to produce different tones to signify the classes detected during the field evaluation. The prototype was the only instrument used for the recording of the video (stored as stereo image sequences) as well as the real-time classifications of the surface condition (stored as comma separated values with time-stamp corresponding to the image sequence). The recorded data were then transcribed in the lab to calculate the accuracy of the classification.

A unit of test was counted based on a single walk on a pathway with the occurrence of a specific class. For example, a navigation along a walkway leading to a drop-off has one-unit test for smooth walkway (class label = SW), and one-unit test for drop-off (class label = D2). If the tester walks from the other direction at the similar walkway, then the record will have a unit test for smooth surface before another one for rise (class label = R1). By this definition, more SW class was observed than the other classes. This is true even for an actual navigation. This imbalance also justifies that the evaluation will be more sensible if each class has similar sample sizes to compute their classification accuracy (as shown in *Table 6.5*). The video recording was set to 30 frames per second. Since the data are stored as image sequences just like any uncompressed videos, the experiment sampled the sequences at every 3 frames to optimize the difference between each frame for the analysis. Each sample was recorded three times, and the author selected one of the best for analysis. Although the main objective of this research is to classify surface discontinuity, it is also equally important for the prototype to know if it is sensing a smooth walkway.

Table 6.5: Number of test units acquired for each class, and the samples used in the analysis with their classification accuracies.

Class Label	SW	D ₁	D ₂	D ₃	D ₄	D ₅	R ₁	R ₂	R ₃
Units of tests conducted	85	15	7	9	15	8	7	15	9
Total Images sampled for the analysis	315 from each class								
Classification Accuracy (%)	99.4	98.4	99.0	98.1	92.4	86.9	98.4	98.7	96.8

The evaluation result reflects that most classes were correctly classified with respective accuracies as specified in the table; however there were two exceptions – the class D₄ (uncovered drainage) and D₅ (blended gradient). The low accuracy for D₅ was expected as the trained model had shown some high confusion rate. The overall evaluation accuracy is 96.4% which is very closed to the result of earlier test set (which was 96.7%). A very small number of smooth surfaces were classified as a down-ramp; this agrees to the test set earlier, in which these two classes could be perceptually similar. Both up-ramp and down-ramp have a minor tendency to be classified as smooth surface. Classes of steps were mostly well-classified. The rest of the misclassification cases can be seen from the confusion matrix in Table 6.6.

Table 6.6: Confusion matrix of the classification from the field evaluation.

Predicted class Actual class	SW	D ₁	D ₂	D ₃	D ₄	D ₅	R ₁	R ₂	R ₃
SW	313	0	0	2	0	0	0	0	0
D ₁	2	310	2	0	1	0	0	0	0
D ₂	0	3	312	0	0	0	0	0	0
D ₃	3	1	1	309	1	0	0	0	0
D ₄	0	8	6	0	291	10	0	0	0
D ₅	0	11	7	7	13	274	0	3	0
R ₁	0	0	0	0	0	0	310	4	1
R ₂	0	0	0	0	0	0	3	311	1
R ₃	0	0	0	0	0	0	5	6	305

Apart from accuracy measure, the precision, recall and F1-score were also calculated. As the 9 classes were not recorded with similar number of samples, accuracy alone would not be a good performance measure. There is a potential imbalance classification issue. Precision is the number of “true positives” divided by the number of “true positives” and “false positives”. In short, it is the number of positive predictions divided by the total number of positive class values predicted. Precision expresses the proportion of the data points the classification model says was relevant were actually relevant. Recall is the number of “true positives” divided by the number of “true positives” and the number of “false negatives”. It is the number of positive predictions divided by the number of positive class values in the data. Recall is the ability of a model to find all the relevant cases within a dataset. The F1-score is the harmonic mean of precision and recall. It is used to convey the balance between precision and recall. The F1-score is defined as:

$$F1_score = 2 \times \frac{precision \times recall}{precision + recall} \quad \text{Equation 6-1}$$

From the above definitions of precision, recall and F1-score, it can be simplified that the precision reflects the exactness while the recall conveys the completeness of the classification model. F1-score offers a hint of the balance between the precision and recall. *Table 6.7* shows the average precision, recall and F1-score of the model based on the field evaluation. One notable part is class D5 (blended gradient). While class D5 was measured with low accuracy of 86.9%, its precision is relatively low with a rate of 0.87 but it has a higher rate of recall of 0.96. This indicates that the model’s ability to find all the relevant cases within the evaluated data for class D5 is high, although its ability to “identify only” the relevant data for this class is slightly low. Overall, the model showed balance performance across all the classes based on the relatively high F1-scores.

Table 6.7: The average precision, recall and F1-score of the 9 classes from the field evaluation.

Class	Precision	Recall	F1-score
SW	0.99	0.98	0.98
D1	0.98	0.93	0.95
D2	0.99	0.95	0.97
D3	0.98	0.97	0.97
D4	0.92	0.95	0.93
D5	0.87	0.96	0.91
R1	0.98	0.97	0.97
R2	0.98	0.96	0.97
R3	0.96	0.99	0.97

Evaluation on Random Locations

The results in Table 6.5 to Table 6.7 could be biased because the locations involved are places where data were collected to train the model. In order to perform unbiased evaluation, two new locations were selected based on the presence of most of the classes. These two locations are SS2 and SS17 – two urban areas within the Klang Valley, Selangor, Malaysia. Data were never collected from these two locations for model training. Table 6.8 summarizes the number of unit tests conducted for different classes available at the locations.

The similar procedures used for field evaluation described in the earlier section were repeated here. The evaluation was again performed by a sighted person. The recorded data were then transcribed in the lab to calculate the accuracy, precision, recall and F1-score of the classification. Table 6.9 shows the overall classification accuracy of this evaluation. It appears that the accuracy rates have slightly dropped for most classes as compared to the results in Table 6.5, except for classes D4, R1, R2 and R3 where the accuracy rates have not changed much. The average classification accuracy observed was 94.6%.

Table 6.8: The 2 locations selected for unbiased evaluation in which data were never collected from.

Location	Class	Units of tests conducted
SS ₂	So	28
	D ₁	9
	D ₃	3
	D ₄	4
	R ₂	9
	R ₃	3
SS ₁₇	So	18
	D ₁	4
	D ₂	2
	D ₃	2
	D ₅	2
	R ₁	3
	R ₂	3
	R ₃	2
Total Unit tested		92

Table 6.9: Number of test units acquired for each class, and the samples used in the analysis with their classification accuracies for the unbiased evaluation.

Class Label	SW	D ₁	D ₂	D ₃	D ₄	D ₅	R ₁	R ₂	R ₃
Units of tests conducted	46	13	2	5	4	2	3	12	5
Total Images sampled for the analysis	300 from each class								
Classification Accuracy (%)	92.0	97.5	97.0	97.2	92.4	83.0	98.2	98.6	96.1

Table 6.10: Confusion matrix of the unbiased evaluation.

Predicted class \ Actual class	SW	D1	D2	D3	D4	D5	R1	R2	R3
SW	276	0	0	1	1	0	6	6	10
D1	4	292	3	0	1	0	0	0	0
D2	0	5	291	1	2	0	1	0	0
D3	3	1	1	291	1	0	0	0	0
D4	0	4	2	0	291	3	0	0	0
D5	0	12	8	9	14	249	2	6	0
R1	0	0	0	0	0	1	295	3	1
R2	0	0	0	0	0	1	2	296	1
R3	4	0	0	0	0	0	5	3	288

Table 6.11: The precision, recall and F1-score of the 9 classes for the unbiased evaluation.

Class	Precision	Recall	F1-score
SW	0.92	0.96	0.94
D1	0.97	0.93	0.95
D2	0.97	0.95	0.96
D3	0.97	0.96	0.96
D4	0.97	0.94	0.95
D5	0.83	0.98	0.90
R1	0.98	0.95	0.96
R2	0.98	0.94	0.96
R3	0.96	0.96	0.96

D4 is uncovered drainage and it was observed that this class has a very standard physical attribute across the locations, and this could be the reason the model performed quite similarly over them. The similar observations were also applicable to the classes R1 (rises), R2 (up-steps) and R3 (up-ramps) which appeared to share closer physical attributes. Again, class D5 (blended gradient) was observed to record an even lower accuracy rate as before.

The physical attributes of class D5 often vary from one to another at different locations, and it could be challenging for the model to learn some hidden patterns that can better represent this class. The model achieved at least 0.90 or higher of F1-score for all classes (*Table 6.11*), indicating a balance of exactness versus completeness in the classification evaluation.

6.2.2 Algorithmic Efficiency Evaluation

Efficiency in the context of this prototype is the ratio of the task to perform classification, to the resources taken in. As hardware is not a subject of study in this evaluation, the main aspects of efficiency lie within the software engineering or in other words, the algorithmic efficiency. The major part of a deep learning algorithm's efficiency is contributed by the network configuration and its weights, which are also termed parameters in most literature. These parameters were tuned and optimized during the development phase of the phase-2 prototype. Hence the efficiency of the algorithm can be analyzed based on the parameters being employed. Other indirect analysis of the efficiency is through inspecting the power consumption of the prototype when the algorithm is coded into the application. Both the model's algorithmic and power consumption analyses are presented in the following sections.

CNN Model Analysis

At the point of deployment for classification task, the only operation involved in the CNN is the forward-pass. The number of computations during the forward-pass for the optimized model depends on the configuration and parameters summarized in *Table 6.12*. Through model tuning and optimization, the final version of the model was based on the VGG Net. When this model is compared to the original version, it has about 50% reduction of parameters indicating a 50% leverage of algorithmic efficiency. *Table 6.12* also shows the comparison of the optimized versus the original parameters of a VGG Net being employed as the final model in the prototype for field evaluation. The 50% reduction of parameters indicates that the algorithm is two times more efficient as compared to the case if it uses

the original VGG Net. The removal of convolutional layers 8 to 13 also indicates that the algorithm required lesser resources to perform the same task.

Table 6.12: The optimized versus the original parameters of the proposed CNN.

Network Layer	Number of Weights (before tuning) in Million, M	Number of weights retained (%)
Conv. 1 or 2	7.58M	50.0
Conv. 3 or 4	3.69M	49.9
Conv. 5, 6 or 7	1.82M	49.5
Conv. 8, 9 or 10	0.28M	-
Conv. 11, 12 or 13	0.056M	-
Fully-connected 1	0.010M	560.0
Fully-connected 2	0.0041M	100.0
Fully-connected 3	0.0041M	100.0
Total (in Million, M)	13.44M	6.6M

Based on *Table 6.12*, the stacked convolutional layer 1 and 2 is the most memory intensive. It is worth calculating the memory requirement here to perform further analysis. The equations to calculate output pixels for a convolutional layer are defined as:

$$output\ width = \frac{W - F_W + 2P}{S_W} + 1 \quad \text{Equation 6-2}$$

and

$$output\ height = \frac{H - F_H + 2P}{S_H} + 1 \quad \text{Equation 6-3}$$

where W and H are the width and height of the image, F is the filter size, P is the padding and S is the stride being applied to the convolutional layer.

The input image has an ROI of 600 x 200 pixels, filter size of 3x3, zero padding and 1 stride. Applying both *Equation 6-2* and *Equation 6-3* the output width and height are:

$$\text{output width} = \left(\frac{600 - 3 + 2(0)}{1} \right) + 1 = 598$$

and

$$\text{output height} = \left(\frac{200 - 3 + 2(0)}{1} \right) + 1 = 198$$

Based on the above calculations, the output size is then $598 \times 198 = 118,404$ pixels. Since the pixel values is a grey scale, they could range from 0 to 255. Thus the 256 pixel values are integers, and each of them takes one byte of storage in the prototype. With 118,404 pixels, the memory required is about 118 kilo bytes only. This memory would be released once the process moves on to the next layer and so forth. From the analysis, the memory available on ODROID XU-3 (the single-board computer used for the prototype, see Appendix 1) is more than enough to handle this requirement as it has 2 giga bytes of CPU memory. In terms of memory efficiency, the most memory intensive layer has utilized merely 0.000059 percent of the total CPU memory.

Apart from the memory space usage, another measure of the model's efficiency is speed – how long does the algorithm take to complete a task. To observe the speed of the model, the analysis was based on the time needed by the application (in which the model's algorithm was coded in C/C++ language) to complete a classification task for a single frame of image. The measurement was performed by implementing the first anchor point in the beginning of the code when an image frame is grabbed, and the second anchor point at the end of the code before the next image frame. By calculating the time taken to execute the code from the first anchor point to the second anchor point for a duration of 1 minute for 30 observations, the speed of the application to complete a task was averaged to 10.7 frames

per second. This is a reasonable speed for actual usage of the prototype, although the frame rate for a video has to be set to around 20 to 30 frames per second to be smooth enough in a typical movie (which is not the purpose of the prototype). The speed of 10.7 frames per second is analogous to the sampling of 1 over every 3 frames for the case of a 30 frames per second movie. Thus, with the application's speed of 10.7 frames per second, the prototype could still yield near real-time processing of the classification task.

Power Consumption Analysis based on a Fixed Hardware Setting

The proposed prototype is developed to be a wearable assistive device for the BLVs. As a wearable device, the most convenient way of supplying power to it is via batteries. This is a similar scenario to almost all other smart phones or mobile devices. Due to the importance of batteries as the source of power for such devices, there are numerous studies on the energy efficiency of applications running on them (Son et al., 2014, Tawalbeh et al., 2016, Li et al., 2014, Fowdur et al., 2016). In a study by Rashid et al. (2015) on the energy consumption of algorithms implemented on an ARM based mobile device, they analyzed different sorting algorithms coded with different languages and observed that these algorithms and languages exhibit different energy consumption significantly. They concluded that algorithms with less operation cycles coded in lower level languages are the most power efficient. This research has used a similar ARM based processor and the application to run the classification is developed with C/C++ language, a language ranked as highly energy efficient next to a few other languages according to Rashid et al. (2015).

The power consumption analysis is conducted for three different test cases: (1) not running the application, only the operating system is running, (2) not running the application, only the operating system is running with camera module activated, and (3) running the application. The application refers to the small program developed to launch the stereo camera module and run the CNN model to perform the classification task as described in Section 5.1. Activating the camera module means turning the camera on without any data

processing from the camera. The test cases were observed over a period from the turning on of the prototype, until the battery's power is drained empty. Each one of the test cases was conducted 5 times to obtain their average results. To avoid possible inconsistency from different batteries, the evaluations were performed with a similar battery with specifications as described in *Figure A2.4*, Appendix 2.

Table 6.13 shows the average time taken by each test case to drain the battery's power. Test case (1) provides a baseline of the battery usage by the operating system when neither the camera module nor the application was involved. Test case (2) provides an insight of the power consumption by the camera module. Eventually test case (3) shows the power consumption by the application, which is the major part of the prototype system.

Table 6.13: Average time taken to drain the power over 5 attempts of evaluation based on 3 different cases of application status.

Application status Evaluation	Case 1: Not running, only turning on the operating system	Case 2: Not running, only turning on the operating system with the camera module activated	Case 3: Running
Attempt 1	249 minutes	208 minutes	133 minutes
Attempt 2	251 minutes	205 minutes	136 minutes
Attempt 3	246 minutes	201 minutes	128 minutes
Attempt 4	248 minutes	213 minutes	125 minutes
Attempt 5	244 minutes	210 minutes	132 minutes
Average time taken to drain the power empty	247.6 minutes	207.4 minutes	130.8 minutes

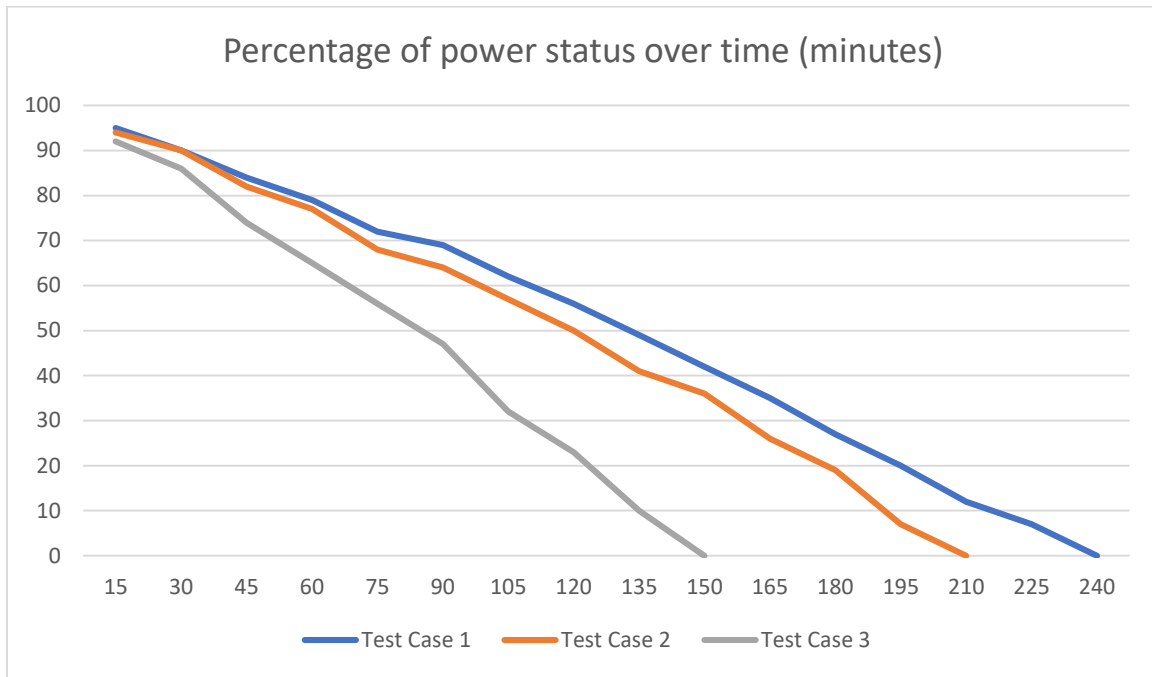


Figure 6.4: Percentage of power status over time as indicated by a power status application running on the prototype for the three test cases.

It was observed that when the application is not running on the prototype, the power consumption by the operating system alone can last for more than 4 hours. The activation of the camera module would increase the power consumption by about 16%, and it took approximately 3 hours and 27 minutes to drain the battery's power. When the application is running, the power lasted only for about 2 hours. From the baseline provided by test case (1), it appears that the application has increased the power consumption by about 35%, after factoring in the power consumption by the camera module based on test case (2). Based on the chart in *Figure 6.4*, the application exhibited near-linear power consumption over time. In addition to the near-linear power consumption, when the power is drained to 50% or below, the power draining appears to slightly speed up for all test cases.

Apart from direct power consumption by the application and the camera, it was observed that the indirect power consumption needed for cooling purpose could contribute to the quick draining of the battery's power. Most of the time when the application was not

running, the computer's operating system conserved the power by not running the cooling fan. However, the frequency of the cooling fan's activation was much higher when the application was running. This cooling activity have resulted in the indirect power consumption that shorten the power draining time.

6.3 Summary of the Chapter

This chapter first summarizes the model training which was based on three different widely cited CNN models and their variants. From the model training, model M3-45 with the best classification accuracy was implemented for field evaluation. Next, the chapter presents the evaluation in two aspects – classification accuracy evaluation and efficiency evaluation. From the accuracy evaluation in the field, firstly the result obtained from locations where data were collected shows a high accuracy of 96.4%. Secondly, the result obtained from random locations where data were never collected shows a lower accuracy of 94.6%. Finally, two aspects of efficiency were analysed – the model's algorithmic efficiency and power consumption efficiency. Analysis on the model's algorithmic efficiency shows that model M3-45 has approximately 50% less parameters as compared to the original model it was based on. Further analysis indicates that the model required very low CPU memory with a near real-time speed in processing the task. Next, the power consumption analysis provides an insight of the application's power usage under different conditions. The analysis shows that the application, being a major part of the prototype system, could increase the power consumption for about 35% as compared to the case when it is not running.

Chapter 7: Conclusion

This chapter concludes the thesis by providing a summary of the research, an elaboration of the answers to the research questions, a description of the research impact and contributions, and a glimpse into possible future work. The chapter then ends with some concluding remarks.

7.1 Thesis Summary

There is an estimated 285 million BLVs worldwide according to WHO, and this is about a ratio of 1 to 25 against the estimated world population of 7.7 billion. With one BLV in every twenty-five persons as estimated, it is undoubtedly crucial to offer more universal accessible environments to connect the BLVs to societal facilities. However, despite the significant size of the BLV population, even in some advanced cities, such safe environments for the BLVs are not sufficiently available, let alone in less developed places.

At the beginning of the research, it was pre-supposed using a mobile robot as a guide dog substitute and targeting home environments. However, the author quickly realised after some consultations with BLVs in different countries (Australia and Malaysia) that the problem of navigating around the home is minimal at best and in Malaysia, the idea of a robot navigating outdoors is fraught with the same problems faced by the BLVs. The research focus soon shifted to wearable technologies and navigational pathways in outdoor environments.

Motivated by the issue faced by the BLVs and tapping on the applicability of current technology, this research worked on the development of a wearable technology-based system prototype to help the BLVs in navigating their pathway. It addressed specifically the challenge of negotiating surface discontinuities typically found within urban areas.

The research made both practical and methodological contributions in a few ways. Through continuous consultations with the BLVs which also involved using 3D printing to gather better insights about the issue, the research came to aware of the problem, and eventually identified a richer taxonomy of surface discontinuities which was never tackled before. Based on the findings, the research generated a novel dataset that could exemplify the problem and used it to train some deep learning models to classify the identified surface discontinuities. The trained model was then built into a wearable technology-based system prototype. With the prototype evaluated for its performance, the research demonstrated that an assistive technology based solely on stereo computer vision and deep learning can be a potential solution to help the BLVs in negotiating the mentioned problem.

The following paragraphs summarize the works completed and results achieved from this research. The first stage of the research involved problem awareness through consultation with the BLV service providers. It was identified that during blind navigation, the BLVs are often faced with challenges of negotiating surface discontinuities within some urbans.

Based on input from the BLV service providers and literature, the second stage of the research proposed and designed a lightweight, small and unobtrusive wearable technology-based system prototype to address the issue.

The third stage was set for data generation. Tailored for a computer vision task, the phase-1 prototype was built with a tiny stereo camera and a single board computer. With this prototype, a set of video data that exemplifies the issue of surface discontinuity was collected from the field. A taxonomy of surface discontinuity was also built based on the

physical attributes of the data and some definitions. These data were then used to train a machine learning model to classify the condition of the pathway.

In the following stage, the phase-2 prototype was developed. Firstly, a purpose-built stacked convolutional neural network was assembled, experimented, trained and optimized across a series of configurations and hyperparameter settings. Next, with the incorporation of the best trained model, the phase-1 prototype is then repurposed as a phase-2 prototype such that it is working to classify the condition of a pathway with some simple feedback to the researcher.

Finally, the performance of the prototype to classify surface condition was evaluated in terms of accuracy and efficiency in the field. The evaluation observed high accuracy from the prototype in classifying the taxonomy. Analyses of the CNN model and power consumption were conducted to evaluate the efficiency of the prototype.

7.2 Answering the Research Questions

Based on the completed works, it is important to recap the research questions and provide answers to these questions based on the results or findings. The research has answered the primary research question:

“How can a technology assist the blind and low vision people to negotiate surface discontinuities along their navigational pathway?”

In the following sections, the answers to the subsidiary questions which collectively address the primary research question, are elaborated. However, before the following questions are addressed, it is imperative to point out that to fully address the primary research question, the author must work on several major areas of research which are inclusive of problem identification (from the blind and low vision people), dataset generation, prototype development and evaluation. In the prototype development and evaluation, the works

addressed only the efficiency and accuracy of the machine learning models trained to classify the types of surface discontinuity along a pathway based on the developed taxonomy. Another aspect within this task – the development of UI and its HCI aspects for the prototype was minimal just to aid the sighted researcher to perform evaluation. It was justified earlier that this aspect itself can be another major research area, hence it was proposed as a future work due to time and resource constraints.

7.2.1 Subsidiary Research Question 1

What tool could be suitable to the sensing of surface discontinuities in an outdoor environment?

During the early stage of considering a suitable tool for sensing the surface discontinuities in some outdoor environments, the research had maintained that the prototype should use as few sensors as possible. This consideration was to ensure the prototype design is aligned with the requirements of being a lightweight and small prototype by the BLVs.

Next, the research had identified from the literature some key approaches of computer vision and machine learning that are closely relevant to surface discontinuity problem. Using structured light depth sensing technologies, Filipe et al. (2012), Takizawa et al. (2012), Takizawa et al. (2013), Orita et al. (2013) and Kuramochi et al. (2014) developed obstacles detection technologies with extra capability to recognize some classes of objects. These technologies with depth sensing capability are moving in a promising direction as they could be small and robust in detecting objects. However, a common drawback of the structured light depth sensing technologies is the limitation of working only indoor. Several other groups of researchers worked on the similar concept of depth sensing but with computer vision-based technologies (Rodriguez et al., 2012, Lee et al., 2008, Dang et al., 2016). These computer vision approaches have the advantage of working both indoor

and outdoor as compared to the structured light, although both techniques produce depth data.

Knowledge provided by the literature, in addition to some preliminary tests, later became important insights in determining that stereo computer vision sensing technique is the most suitable choice to capture the identified surface discontinuity data for the research.

7.2.2 *Subsidiary Research Question 2*

How can the identified surface discontinuities be classified into a suitable taxonomy to help develop a machine learning model as part of the wearable prototype?

Based on surveys from the BLV service providers and observations in the field, some preliminary samples that can exemplify the surface discontinuity problem were collected at various urban areas. Through literature review and surveys from the BLVs (some 3D printings of the collected samples were presented to the BLVs during the consultations), the navigational styles practiced by a BLV with trained orientation and mobility skills were understood. Based on these findings, some assumptions about blind navigation were made and these contributed to how the prototype could approach the problem. Some of these assumptions are:

- a) Walls are not a problem to blind navigation using a guide cane, and thus they are not part of surface discontinuity in this research context.
- b) At most of the venues (aisle or corridor of shops or service centers) in which the data were collected, an orientation and mobility trained BLV will search for walls (of a building if outdoor) to get oriented before heading off with the guide cane to negotiate the surface condition. Thus, the data were collected following the direction of the walkways, but not off the direction (not off-road with different angles facing the surface discontinuity).

- c) At places without walls, the BLVs will try to reach the curbs or tiles along the walkway using their guide cane to get oriented before moving further.
- d) The average walking speed of the BLVs is relatively slow, thus the recording of the data was conducted at a slower speed as compared to a typical sighted person.

According to the assumptions, more than 200 samples of relevant data were then collected from the field using the phase-1 prototype (purposed-built for data collection). With the data ready, a taxonomy of classes as shown in *Figure 4.24* was created. The taxonomy development was based on the physical attributes of the samples and guided by some building regulations. Eight classes of distinguished surface discontinuities and one class of surface continuity were identified. The developed taxonomy was then used in the labelling process of the collected data. With the data labelled, some supervised deep learning models were experimented. Such deep learning models were the fundamentals to the phase-2 prototype design.

7.2.3 Subsidiary Research Question 3

What are the suitable machine learning models that can be developed to classify the categories within the taxonomy mentioned in subsidiary research question 2?

The research first identified some potential deep learning models from the literature related to computer vision. These models are state-of-the-art convolutional neural networks relevant to image classification or object detection tasks. It was then followed by testing and training to identify some of the best neural network models for the task. These models were then fine-tuned with an optimization technique. The optimized model as mentioned in Section 6.1.3 was then taken as the best model to be employed to classify the categories within the taxonomy mentioned in (7.1.2).

7.2.4 *Subsidiary Research Question 4*

What is the performance (accuracy and efficiency) of the developed prototype?

It was learned that the best model with an architecture and parameters as mentioned in Section 6.1.3 could classify the categories within the developed taxonomy up to an accuracy of 96.7% for the test set, and 96.4% for the field evaluation. At the field evaluation, the model was good at classifying smooth walkway, down-steps, drop-off, down-ramp, rise, up-steps and up-ramp, all with more than 96% accuracy. Uncovered drainage which is a highly hazardous surface discontinuity had a slightly lower classification accuracy of 92.4%. The prototype performed moderately in classifying blended gradient which is a relatively less commonly found surface discontinuity with only 86.9% accuracy.

In the efficiency evaluation, the prototype was analyzed for its CNN model's algorithmic efficiency, memory usage, speed and power consumption. The optimized CNN model has 50% more efficient configuration and parameters, with very low memory usage as compared to the original model it was based on. Based on the analysis, the most memory intensive layer utilized merely 0.000059 percent of the total CPU memory. The processing speed was near real-time at 10.7 frames per second averagely. The power consumption of the prototype is two times higher when the application is running, and it was observed to exhibit near-linear power consumption over time.

Overall, the prototype can help in classifying surface discontinuities into the developed taxonomy as exemplified by the dataset collected. However, the prototype was developed only for built environment within urban areas where the demand for blind navigation is more significant due to the necessities of daily activities or businesses. Surface discontinuities in natural landscape or terrain are not relevant to this research, hence they are not included in the evaluation.

7.3 Impact and Original Contributions

This research is one of the few studies of blind and low vision negotiation within some urban areas consisting of a rich range of surface discontinuities that could be hazardous to BLV pedestrians. Having identified the issue from consultation with the BLV service providers, the research then proposed an assistive tool in the form of a wearable technology-based system prototype to classify the surface discontinuities. The following sections summarize the contributions of this research.

7.3.1 Contributions to Practical Application

Problem identification: Through continuous consultation with the BLVs, a major problem of surface discontinuities faced by them in navigation was identified. It was discovered that a rich range of surface discontinuities which was never studied before in the reviewed literature, is a major issue in blind navigation.

Dataset generation: A novel dataset that could exemplify the identified problem was generated through a systematic procedure that involved several steps described in Chapter 4. The dataset was originally collected with the help of a purposed-built phase-1 prototype. It was then pre-processed and labelled into several classes based on a taxonomy developed for this research. This labelled dataset will be a valuable contribution even for future or other studies.

Framework, algorithms, models and proof of concept prototype: The research developed a framework that utilizes the surface discontinuity taxonomy to assemble some machine learning algorithms and train some classifier models. The fine-tuned and optimized model is a major contribution from the research, and it was a core component used to develop the wearable technology-based system prototype. Such a prototype designed with a single sensor working on a single board computer is a proof of concept to achieve a device that meets the BLVs' preferred design for an assistive device that is lightweight, small and

unobtrusive. The research demonstrated and showed that a wearable assistive technology-based system built solely using stereo computer vision and machine learning guided by the framework can be a potential solution to help the BLVs in negotiating a diverse range of surface discontinuities.

7.3.2 Contributions to Research Methodology

Methodological contributions from the research can be chronologically described from the stage of problem identification, data generation, to prototype development.

Method of surveying the problem: After some consultations with the BLVs to learn about the problem of surface discontinuities, several techniques were applied to further understand it. Firstly, the details of the problem were further studied by pre-sampling at various locations within the urban areas. Several surface discontinuity samples were first captured by camera and their ground truths were sketched. These images and ground truths were then translated into 3D models using computer software, before they were 3D printed. The 3D printed replicas of the samples were presented to the BLVs to finally identify surface discontinuity types that could cause issues in blind navigation. It was then discovered that a rich range of surface discontinuities which was never studied before in the reviewed literature, is a major issue in blind navigation. Overall, the method used to identify the target samples is a valuable contribution.

Method of data generation: One notable contribution from the research is the method used in data generation. The research first developed a prototype (phase-1) that was used as an instrument for data collection, and it is this similar prototype which was repurposed as the wearable technology-based system (phase-2) to detect and classify surface discontinuities in blind navigation. The phase-1 prototype was transformed into the phase-2 prototype because it is with the similar hardware settings and sensor (camera) configurations that the

surface of a pathway is captured for classification in near real-time. Several techniques involved in this data generation method are the instrument setup, data sampling, data pre-processing, data augmentation and labelling as detailed in the thesis. Other researchers can replicate this similar method to generate new dataset which can be added to this research for further study, or to other research of a similar nature.

Method of innovating the machine learning algorithms: Relevant algorithms from deep learning were assembled and experimented to solve the classification task, and finally a stacked convolutional neural network model was trained and tuned to classify most of the surface discontinuities with high accuracy. The techniques and procedures involved in training and tuning the model were detailed and documented in this thesis. These similar techniques and procedures can be replicated or referred to in other research with similar nature.

7.4 Future Work

In developing a prototype to assist the BLVs, the next steps of this research are to develop the user interface (UI) and study its human-computer interaction (HCI). To proceed with these steps, it would involve some surveys and further research. A primary research question could focus on the feedback mechanism from the prototype to communicate the detected surface discontinuity to the users. Based on earlier consultations with the BLVs, it was found that most of them would not want to have their hearing (auditory sense) occupied by any assistive devices if possible. Some of them tend to prefer haptic feedback instead. Thus, the understanding about the feedback and communication mechanism is another major study.

Without vision, the BLVs must count on the other remaining senses to perceive the environment. The auditory sense perceives information in sequential encoding, as compared to visual information that is normally simultaneous. In order to form a bigger

picture about a particular information through hearing, the cognitive load on memory is greater than the spatial information organized simultaneously by visual interpretation (Hersh, 2008). The olfactory sense may not be a good choice for navigation as humans are not able to localize the source of smell and benefit from it. Gustation is also not quite possible to be a practical choice. This eventually leaves auditory and haptic as the two main senses that the BLVs can rely on to complement the vision sense in perceiving the spatial environment. Hence, in future work, a mechanism of communicating through either the auditory or haptic perception, or a multimodal device utilizing both senses through multisensory perception can be considered.

Some assistive devices produce only one single type of signal, for instance, audio signal. This type of single modality devices based on audio normally represent the environment information in sequential encoding. Sequential signals may be easy to encode (Thinus-Blanc and Gaunet, 1997), but they might not be easy or natural to interpret and comprehend (Hersh, 2008). Especially in dealing with environmental information for navigation, the signals have to be able to represent position, orientation and direction information. Therefore, the signals produced should be able to represent visual-spatial information alike for fast and easy interpretation. A multimodal device may have some advantages in this issue.

In a multimodal technology, multisensory perception is required. There could be a possibility that the different senses affect or influence each other. For the case of a navigational device for the BLVs, the technology will typically be producing audio and tactile signals in order to represent the visual-spatial information. There are some known effects about multisensory perception such as the McGurk effect and the ventriloquism effect. The McGurk effect is an auditory illusion produced by a visual stimulus (Magnotti and Beauchamp, 2015). For instance, in lip reading, the visual information seen by an observer might change the way he/she hears the sound. In the ventriloquism effect, the perceived location of a sound is shifted to a visual stimulus located at another location

(Callan et al., 2015). An example for this effect is the talking puppet, in which the listener has an impression that the puppet is talking because the movement of the puppet's lips are matched to the handler's voice. Such effects must be handled carefully if future designs of the feedback mechanism involving multisensory perception of the BLVs.

On the other hand, it is notable that in multisensory perception, one sense typically dominates the other(s) (Soto-Faraco et al., 2004). Perception will be dominated by the sense that is capable of providing the best information in terms of precision and appropriateness about the perceived stimulus (Klemen and Chambers, 2011). For example, in sighted people, vision is the most dominant sense under normal circumstances. Since BLVs mostly rely on auditory and haptic senses for environmental stimuli, the researchers must devise some approach to identify which one of them would be the more dominant sense. Each BLV individual could have a different preference. The answer to this question is important to the design of any multimodal assistive technologies.

To inspect the multisensory effect of auditory and haptic senses, both motions involving the two senses have to happen simultaneously. In this case, according to Soto-Faraco et al. (2004), either modality is able to dominate the other, however, tactile motion has a stronger impact on audio motion perception than vice versa. When a person manipulates an object that produces both sounds and causes some haptic sensations to facilitate the recognition of the object, there could be some binding relation between audio and haptic perception. Kassuba et al. (2013) conducted a study to identify the neural correlates of audio-haptic binding features in a group of volunteers that handled an object. The results show that multisensory perception interacts at different hierarchical levels for audio-haptic object processing. They noticed that object-specific multimodal interactions are highest in the left fusiform gyrus (a part of the temporal lobe in the human brain), indicating a higher order of convergence zone for conceptual object knowledge. In another study concerning multi-parallel haptic perception by Dupin et al. (2015), the combined signals from static tactile surface and kinesthetic (moving) surface were examined if they are coherently

perceived through two separate hands of a person. The results revealed that both discrete and continuous tactile and kinesthetic signals are perceived as a combined information as if they came from the same hand. The similar effects of interaction are not limited to moving haptic perception. This finding eventually led to the suggestion that the brain simplifies complex coordinate tasks by remapping sensory inputs from haptic perception. The above-mentioned multisensory effects have to be taken into account in future work on the prototype, if multimodality is a possible choice.

Apart from the UI design and HCI considerations, other technical improvements to the prototype can also be further explored. For instance, although there are known issues of uneven down-stairs which might be more threatening than the properly built version in blind navigation, not a large amount of such sample is available currently for model training. Hence, the current model is only able to be trained on a standard down-stairs, but it can't differentiate between a properly built one and an unevenly built one. An upgrade of the sensor (camera) to a higher sensitivity might help too, as future hardware might be getting less costly and more robust.

Finally, the problems encountered during field evaluation could offer some important hints for future work too. The following points summarize some of the main issues faced, in which they could expose some important concerns in future work.

- Weather: rainy days prevented some of the evaluation activities at certain locations without shelter (the prototype was not rainproof) – this issue will lead to further consideration to build a more robust hardware for the end-product.
- Problem of camera's mounting angle: The supposed angle should be 35 degrees facing downward, but at certain points during data collection, the camera mount was tilted unnoticed to be higher and hence some samples were out of the frame. It was then realized that the mounting gear was not able to keep the camera strictly at

the supposed angle for long. Again, this issue will lead to further consideration to build a more robust hardware for the end-product.

The future work should look into the above issues and suggest viable design or solution, for instance, some rain-proof casing and better mounting feature. However, the HCI considerations and UI design as mentioned earlier could affect the casing and other design features too. These designs are inter-connected, and they should be studied thoroughly and carefully in future work.

7.5 Concluding Remarks

The BLVs need to constantly negotiate surface discontinuities along the walkways during self-navigation. Although the guide cane is a primary assistive tool for navigation, it has known limitations in dealing with certain types of surface discontinuity. Other technologies were introduced to augment the guide cane's ability, but limited works have been done in dealing with a richer range of surface discontinuities. The foremost motivation of the research is to design and develop a prototype to help the BLVs in negotiating such an issue of surface discontinuities.

Through a design science research methodology, the research has:

- identified specifically the issue of surface discontinuities,
- generated a set of data to exemplify the issue,
- designed and developed both the phase-1 and phase-2 prototypes, and
- evaluated the phase-2 prototype.

Via the prototype, the research has shown that a wearable assistive technology-based system built predominantly with a lightweight single-board computer, a tiny stereo camera and a fine-tuned machine learning model can be a potential solution to the mentioned issue.

The multi-faceted navigation challenge of the BLVs is still largely an unsolved problem. However, every single contribution from the research community could narrow the scope of this challenge. Based on this very belief, this research has contributed its portion by tapping on some recent state-of-the-art technologies. Had Louis Braille not spent years of effort in perfecting the Braille system, reading would still be a tedious chore for the BLVs today. The BLVs of the past might have lesser choices of assistive tool, but it is hoped that the BLVs in the near future could be well-equipped with some practical tools thanks to the advance in technologies.

As aspired from the last line of Beaudin's article(2011): "the challenges of yesterday are the opportunities of today" (para. 8), technologies have enabled the transformation of challenges from the past into opportunities of today. Certainly, the proposed wearable technology-based system could be a timely solution in reducing the gaps between the mobility needs and the limitations of traditional navigation aids for the BLVs.

List of References

- _____. 1991. Uniform Building (Amendment) By-Laws 1991. *UBBL 34A*. National Council of Local Government, Malaysia.
- _____. 2008. Act 685 - Person with Disabilities Act 2008, Laws of Malaysia. JKM, Government of Malaysia.
- _____. 2013. Protection from Falling, Collision and Impact, The Building Regulations 2010. *Approved Document K*. England: Her Majesty's Government (England).
- _____. 2015a. *GDP Research: The Miniguide mobility aid* [Online]. Available: http://www.gdp-research.com.au/miniguide_1.htm [Accessed 10 November 2015].
- _____. 2015b. ODROID HardKernel. Available from: <https://www.hardkernel.com/> [Accessed 5 July 2015].
- _____. 2015c. *UltraCaneputting the world at your fingertips* [Online]. Available: <http://www.ultracane.com/soundforesighttechnologyltd> [Accessed 10 November 2015].
- _____. 2017. Country Income Groups (World Bank Classification), Country and Lending Groups, The World Bank Group.
- _____. 2018a. Orange Pi. Available from: <http://www.orangepi.org/> [Accessed 21 May 2019].
- _____. 2018b. Raspberry Pi. Available from: <https://www.raspberrypi.org/> [Accessed 24 March 2019].
- _____. 2019. LattePanda – A Windows 10 Computer with integrated Arduino. Available from: <https://www.lattepanda.com/> [Accessed 20 May 2019].
- ALGHAMDI, S., VAN SCHYNDEL, R. & KHALIL, I. 2013. Accurate positioning using long range active RFID technology to assist visually impaired people. *Journal of Network and Computer Applications*, 41, 135-147.
- ALIPOOR, M., IMANDOOST, S. & HADDADNIA, J. Designing edge detection filters using Particle Swarm Optimization. *Electrical Engineering (ICEE), 2010 18th Iranian Conference on*, 11-13 May 2010 2010. 548-552.
- ARAÚJO, T., ARESTA, G., ALMADA-LOBO, B., MENDONÇA, A. M. & CAMPILHO, A. Improving Convolutional Neural Network Design via Variable Neighborhood Search. In: KARRAY, F., CAMPILHO, A. & CHERIET, F., eds. *Image Analysis and Recognition*, 2017// 2017 Cham. Springer International Publishing, 371-379.
- BARANSKI, P., POLANCZYK, M. & STRUMILLO, P. 2010. A remote guidance system for the blind. *The 12th IEEE International Conference on e-Health Networking, Applications and Services*.
- BARTH, J. L. & FOULKE, E. 1979. Preview: A neglected variable in orientation and mobility. *Journal of Visual Impairment and Blindness*, 73, 41-48.
- BASU, M. 2002. Gaussian-based edge-detection methods-a survey. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 32, 252-260.
- BEAUDIN, L. 2011. *Challenges and Opportunities Facing Visually Impaired Persons* [Online]. Alliance for Equality of Blind Canadians. Available: <http://www.blindcanadians.ca/publications/cbm/10/challenges-and-opportunities-facing-visually-impaired-persons> [Accessed 18 June 2015].
- BECK, K. 2010. Challenges that Blind People Face. Available from: <http://www.livestrong.com/article/241936-challenges-that-blind-people-face/> [Accessed 18 June 2015].
- BENHAM, T. A. 1952. Evaluation of Signal Corps Sensory Aid for the Blind, AN/PVQ-2 (XE-2). Report on VA Contract V1001M-1900.

- BENJAMIN, J. M., JR., NAZIR, A. A. & ADAMO, F. S. A Laser Cane for the Blind. Biomedical Symposium, 1973 San Diego.
- BLASCH, B. B. & STUCKEY, K. A. 1995. Accessibility and Mobility of Persons Who Are Visually Impaired: A Historical Analysis. *Journal of Visual Impairment & Blindness*, v89, p417-22.
- BLEDSON, C. W. 1997. Originators of orientation and mobility training, In B.B. Blasch, W.R. Wiener and R.L. Welsh (Eds.) *Foundations of orientation and mobility*, 2nd. Edn., pp. 580-623. New York: AFB Press.
- BON, J. V. 2007. *IT Service Management best practices*, pp. 351, Van Haren Publishing.
- BORENSTEIN, J. 2001. The GuideCane-applying mobile robot technologies to assist the visually impaired. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 31, 131-136.
- BORENSTEIN, J. & ULRICH, I. The GuideCane-a computerized travel aid for the active guidance of blind pedestrians. *Robotics and Automation*, 1997. Proceedings., 1997 IEEE International Conference on, 20-25 Apr 1997 1997. 1283-1288 vol.2.
- BOVIK, A. C. 2005. *Handbook of image and video processing*, Boston, MA, Amsterdam, Academic Press.
- BROWN, M. Z., BURSCHKA, D. & HAGER, G. D. 2003. Advances in computational stereo. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25, 993-1008.
- BUJACZ, M., BARAŃSKI, P., MORAŃSKI, M., STRUMIŁŁO, P. & MATERKA, A. Remote mobility and navigation aid for the visually disabled. 7th ICDVRAT with ArtAbilitation, 2008 Maia, Portugal. ICDVRAT/University of Reading, 263 - 270.
- CALLAN, A., CALLAN, D. & ANDO, H. 2015. An fMRI Study of the Ventriloquism Effect. *Cerebral cortex (New York, N.Y. : 1991)*, 25, 4248.
- CANG, Y., SOONHAC, H. & XIANGFEI, Q. A Co-Robotic Cane for blind navigation. 2014 IEEE International Conference on Systems, Man and Cybernetics (SMC), 5-8 Oct. 2014 2014. 1082-1087.
- CANNY, J. 1986. A Computational Approach to Edge Detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-8, 679-698.
- CAP, G. Development of a new robotic system for assisting and guiding visually impaired people. IEEE International Conference on Robotics and Biomimetics, December 11-14 2012 Guangzhou, China. 229-234.
- CECEZ-KECMANOVIC, D. 2011. On methods, methodologies and how they matter. *19th European Conference on Information Systems*. Helsinki, Finland.
- CHEUNG, B. 2012. Convolutional Neural Networks Applied to Human Face Classification.
- CHRISTENSEN, K. M. & BYRNE, B. C. 2014. The Built Environment and Community Integration. *Journal of Disability Policy Studies*, 25, 186-195.
- CIREŞAN, D., MEIER, U. & SCHMIDHUBER, J. 2012. Multi-column Deep Neural Networks for Image Classification. *Arxiv preprint arXiv:1202.2745*, 9.
- CIREŞAN, D. C., MEIER, U., MASCI, J., GAMBARDILLA, L. M. & GAMBARDILLA, J. 2011. High-performance neural networks for visual object classification. *Arxiv preprint arXiv:1102.0183*, 9.
- DAHLKAMP, H., KAEHLER, A., STAVENS, D., THRUN, S. & BRADSKI, G. 2006. Self-supervised monocular road detection in desert terrain. *Robotics: Science and Systems (RSS)*. The MIT Press.
- DANG, Q., CHEE, Y., PHAM, D. & SUH, Y. 2016. A Virtual Blind Cane Using a Line Laser-Based Vision System and an Inertial Measurement Unit. *Sensors*, 16, 95.

- DING, J., CHEN, B., LIU, H. & HUANG, M. 2016. Convolutional Neural Network With Data Augmentation for SAR Target Recognition. *IEEE Geoscience and Remote Sensing Letters*, 13, 364-368.
- DINGRAN LU, Q., XIAO-HUA YU, Q., XIAOMIN JIN, Q., BIN LI, Q., CHEN, Q. & JIANHUA ZHU, Q. 2011. Neural network based edge detection for automated medical diagnosis.
- DUPIN, L., HAYWARD, V. & WEXLER, M. 2015. Direct coupling of haptic signals between hands. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 619.
- FERNANDES, H., CONCEIÇÃO, N., PAREDES, H., PEREIRA, A., ARAÚJO, P. & BARROSO, J. 2012. Providing accessibility to blind people using GIS. *Universal Access in the Information Society*, 11, 399-407.
- FILIPE, V., FERNANDES, F., FERNANDES, H., SOUSA, A., PAREDES, H. & BARROSO, J. 2012. Blind Navigation Support System based on Microsoft Kinect. *Procedia Computer Science*, 14, 94-101.
- FOWDUR, T. P., HURBUNGS, V. & BEEHARRY, Y. Statistical analysis of energy consumption of mobile phones for web-based applications in Mauritius. 2016 International Conference on Computer Communication and Informatics (ICCCI), 7-9 Jan. 2016. 1-8.
- FRÜHWIRTH-SCHNATTER, S. 1994. Data Augmentation and Dynamic Linear Models. *Journal of Time Series Analysis*, 15, 183-202.
- FRYER, L., FREEMAN, J. & PRING, L. 2013. What verbal orientation information do blind and partially sighted people need to find their way around? A study of everyday navigation strategies in people with impaired vision. *The British Journal of Visual Impairment*, 31, 123-138.
- FU, W., JOHNSTON, M. & ZHANG, M. 2011. Genetic programming for edge detection: A global approach.
- GALATAS, G., MCMURROUGH, C., MARIOTTINI, G. L. & MAKEDON, F. 2011. eyeDog: an assistive-guide robot for the visually impaired. *Proceedings of the 4th International Conference on Pervasive Technologies Related to Assistive Environments*. Heraklion, Crete, Greece: ACM.
- GALLAGHER, T., WISE, E., BINGHAO LI, A. G., DEMPSTER, C., RIZOS, E. & RAMSEY-STEWART, E. 2012. Indoor positioning system based on sensor fusion for the Blind and Visually Impaired. *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. Sydney, NSW.
- GANZ, A., GANDHI, S. R., WILSON, C. & MULLETT, G. 2010. INSIGHT: RFID and Bluetooth enabled automated space for the blind and visually impaired. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*.
- GANZ, A., SCHAFER, J., GANDHI, S., PULEO, E., WILSON, C. & ROBERTSON, M. 2012. PERCEPT Indoor Navigation System for the Blind and Visually Impaired: Architecture and Experimentation. *International Journal of Telemedicine and Applications*, 2012, 12.
- GERUSCHAT, D. R. & SMITH, A. J. 2010. Low Vision for Orientation and Mobility. In: WIENER, W. R., WELSH, R. L. & BLASCH, B. B. (eds.) *Foundations of Orientation and Mobility (3rd edition): History and Theory*. New York: American Foundation for the Blind.
- GIRSHICK, R. 2015. Fast R-CNN. *IEEE International Conference on Computer Vision (ICCV)*.
- GIUDICE, N. A. & LEGGE, G. E. 2008. Blind Navigation and the Role of Technology. In: HELAL, A., MOKHTARI, M. & ABDULRAZAK, B. (eds.) *Engineering Handbook of Smart Technology for Aging, Disability and Independence*. John Wiley & Sons.
- GOLLEDGE, R. G. 1993. Geography and the Disabled: A Survey with Special Reference to Vision Impaired and Blind Populations. *Transactions of the Institute of British Geographers*, 18, 63-85.

- GONG, M. & YANG, Y.-H. 2002. Genetic-Based Stereo Algorithm and Disparity Map Evaluation. *International Journal of Computer Vision*, 47, 63-77.
- GOODFELLOW, I., BENGIO, Y. & COURVILLE, A. 2016. *Deep Learning*, MIT Press.
- GOODRICH, L. G. & LUDT, L. R. 2002. Distance Vision Recognition Assessment in Low Vision Mobility. *Optometry and Vision Science*, 79, 275-275.
- GOSTA, M. & GRGIC, M. 2010. Accomplishments and Challenges of Computer Stereo Vision. *International Symposium ELMAR*. Zadar, Croatia.
- GREGG, D., KULKARNI, U. & VINZE, A. 2001. Understanding the Philosophical Underpinnings of Software Engineering Research in Information Systems. *Information Systems Frontiers*, 3, 169 - 183.
- GREGOR, S. & HEVNER, A. 2013. Positioning and Presenting Design Science Research for Maximum Impact. *Management Information Systems Quarterly*, 37, 337 - 355.
- GRIMSON, W. E. L. 1985. Computational Experiments with a Feature Based Stereo Algorithm. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-7, 17-34.
- HADSELL, R., ERKAN, A., SERMANET, P., SCOFFIER, M., MULLER, U. & YANN, L. Deep belief net learning in a long-range vision system for autonomous off-road driving. 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, 22-26 Sept. 2008 2008. 628-633.
- HAN, J., KAMBER, M. & PEI, J. 2012. *Data Mining Concepts and Techniques*, USA, Morgan Kaufmann.
- HAN, K.-P., SONG, K.-W., CHUNG, E.-Y., CHO, S.-J. & HA, Y.-H. 2001. Stereo matching using genetic algorithm with adaptive chromosomes. *Pattern Recognition*, 34, 1729-1740.
- HAQ, I., ANWAR, S., SHAH, K., KHAN, M. T. & SHAH, S. A. 2015. Fuzzy Logic Based Edge Detection in Smooth and Noisy Clinical Images. *PloS one*, 10, e0138712.
- HASSAN, R., COHANIM, B. & WEEK, D., OLIVIER 2004. A Comparison of Particle Swarm Optimization and the Genetic Algorithm. US: American Institute of Aeronautics and Astronautics.
- HERN, NDEZ, D. C., CERES, KIM, T. & JO, K.-H. 2011. Stairway detection based on single camera by motion stereo. In: MEHROTRA, K. G., MOHAN, C. K., OH, J. C., VARSHNEY, P. K. & ALI, M. (eds.) *IEA/AIE'11*. Springer-Verlag.
- HERSH, M. A. 2008. Perception, the Eye and Assistive Technology Issues. In: HERSH, M. A. & JOHNSON, M. A. (eds.) *Assistive Technology for Visually Impaired and Blind People*. UK: Springer London.
- HERSH, M. A. & JOHNSON, M. A. 2008. Disability and Assistive Technology Systems. In: HERSH, M. A. & JOHNSON, M. A. (eds.) *Assistive Technology for Visually Impaired and Blind People*. Springer London.
- HESCH, J. A. & ROUMELIOTIS, S. I. An Indoor Localization Aid for the Visually Impaired. Robotics and Automation, 2007 IEEE International Conference on, 10-14 April 2007 2007. 3545-3551.
- HEVNER, A., MARCH, S., PARK, J. & RAM, S. 2004. Design Science in Information Systems Research. *MIS Quarterly*, 28, 75-105.
- HINTON, G. E. & SALAKHUTDINOV, R. R. 2006. Reducing the Dimensionality of Data with Neural Networks. *Science*, 313, 504-507.
- HOFF, W. & AHUJA, N. 1989. Surfaces from stereo: integrating feature matching, disparity estimation, and contour detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11, 121-136.

- HONG, T., CHANG, T., RASMUSSEN, C. & SHNEIER, M. Road detection and tracking for autonomous mobile robots. *Proc. SPIE 4715, Unmanned Ground Vehicle Technology*, 2002. 311.
- HOYLE, B. & WATERS, D. 2008. Mobility AT: The Batcane (UltraCane). In: HERSH, M. A. & JOHNSON, M. A. (eds.) *Assistive Technology for Visually Impaired and Blind People*. UK: Springer London.
- HUNAITI, Z., GARAJ, V. & BALACHANDRAN, W. 2006. A Remote Vision Guidance System for Visually Impaired Pedestrians. *J. Navigation*, 59, 497-504.
- HUSSEIN, H. & MOHD. YAACOB, N. 2013. Malaysian Perspective on the Development of Accessible Design. *Asian Journal of Environment-Behaviour Studies*, 3, 101 - 116.
- ICD-10 2010. International Statistical Classification of Diseases and Related Health Problems. *H54 Blindness and low vision*. Malta: World Health Organization.
- IMRIE, R. & KUMAR, M. 1998. Focusing on Disability and Access in the Built Environment. *Disability & Society*, 13, 357-374.
- JANZEN, G. & JANSEN, C. 2010. A neural wayfinding mechanism adjusts for ambiguous landmark information. *NeuroImage*, 52, 364-370.
- JEFFERY, K. J., JOVALEKIC, A., VERRIOTIS, M. & HAYMAN, R. 2013. Navigating in a three-dimensional world. *Behavioral and Brain Sciences*, 36, 523-587.
- JOH, A. S. & ADOLPH, K. E. 2006. Learning From Falling. *Child Development*, 77, 89-102.
- JOLLIFFE, I. T. 2002. *Principal Component Analysis*, New York, Springer-Verlag New York, Inc.
- JONSSON, R. 2011. NSK develops four-legged robot "guide dog". <http://www.gizmag.com/nsk-four-legged-robot-guide-dog/20559/>: Gizmag - Robotics.
- KANADE, T. & OKUTOMI, M. 1994. A stereo matching algorithm with an adaptive window: theory and experiment. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 16, 920-932.
- KANG, C.-C. & WANG, W.-J. 2007. A novel edge detection method based on the maximizing objective function. *Pattern Recognition*, 40, 609-618.
- KARIMI, H. A. 2015. *Indoor Wayfinding and Navigation*, CRC Press.
- KASSUBA, T., MENZ, M. M., RÖDER, B. & SIEBNER, H. R. 2013. Multisensory Interactions between Auditory and Haptic Object Recognition. *Cerebral Cortex*, 23, 1097-1107.
- KASTRENAKES, J. 2014. *Google announces Project Tango, a smartphone that can map the world around it* [Online]. The Verge. Available: <https://www.theverge.com/2014/2/20/5430784/project-tango-google-prototype-smartphone-announced> [Accessed 13 Oct 2015].
- KIM, D., KIM, K. & LEE, S. 2014. Stereo camera based virtual cane system with identifiable distance tactile feedback for the blind. *Sensors (Basel, Switzerland)*, 14, 10412.
- KINGMA, D. P. & BA, J. 2015. Adam: A Method for Stochastic Optimization. *3rd International Conference for Learning Representations*. San Diego.
- KLAUS, A., SORMANN, M. & KARNER, K. 2006. Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure.
- KLEMEN, J. & CHAMBERS, C. D. 2011. Current perspectives and methods in studying neural mechanisms of multisensory interactions. *Neuroscience and Biobehavioral Reviews*, 36, 111-133.
- KOIVUNEN, A. C. & KOSTINSKI, A. B. 1999. The Feasibility of Data Whitening to Improve Performance of Weather Radar. *Journal of Applied Meteorology*, 38, 741 - 749.
- KRIZHEVSKY, A., SUTSKEVER, I. & HINTON, G. E. 2012a. ImageNet Classification with Deep Convolutional Neural Networks. *Neural Information Processing Systems*

- KRIZHEVSKY, A., SUTSKEVER, I. & HINTON, G. E. 2012b. ImageNet Classification with Deep Convolutional Neural Networks. In: WEINBERGER, F. P. A. C. J. C. B. A. L. B. A. K. Q. (ed.) *Advances in Neural Information Processing Systems 25 (NIPS 2012)*. Curran Associates, Inc.
- KULYUKIN, V., GHARPURE, C., NICHOLSON, J. & OSBORNE, G. 2006. Robot-assisted Wayfinding for the Visually Impaired in Structured Indoor Environments. *Autonomous Robots*, 21, 29-41.
- KULYUKIN, V., GHARPURE, C., NICHOLSON, J. & PAVITHRAN, S. RFID in Robot-Assisted Indoor Navigation for the Visually Impaired. IEEE/RSJ International Conference on Intelligent Robots and Systems, September 28 -October 2 2004 Sendai, Japan.
- KURAMOCHI, Y., TAKIZAWA, H., AOYAGI, M., EZAKI, N. & SHINJI, M. Recognition of elevators with the Kinect cane system for the visually impaired. System Integration (SII), 2014 IEEE/SICE International Symposium on, 13-15 Dec. 2014 2014. 128-131.
- KUYK, T., ELLIOTT, J. L., WESLEY, J., SCILLEY, K., MCINTOSH, E., MITCHELL, S. & OWSLEY, C. 2004. Mobility function in older veterans improves after blind rehabilitation. *Journal of rehabilitation research and development*, 41, 337.
- LECUN, Y., BOSER, B., DENKER, J. S., HENDERSON, D., HOWARD, R. E., HUBBARD, W. & JACKEL, L. D. 1990. Handwritten digit recognition with a back-propagation network. In: TOURETZKY, D. (ed.) *Advances in Neural Information Processing Systems (NIPS)*.
- LEE, S.-W., KANG, S. & LEE, S.-W. 2008. A WALKING GUIDANCE SYSTEM FOR THE VISUALLY IMPAIRED. *International Journal of Pattern Recognition and Artificial Intelligence*, 22, 1171-1186.
- LEIB, D., LOOKINGBILL, A. & THRUN, S. 2005. Adaptive road following using self-supervised learning and reverse optical flow. *Robotics: Science and Systems (RSS)*.
- LENNIE, P. & HEMEL, S. B. V. 2002. *Visual impairments determining eligibility for social security benefits*, Washington, D.C., Washington, D.C. : National Academy Press.
- LEONG, K. Y., EGERTON, S. & CHAN, C. K. Y. A wearable technology to negotiate surface discontinuities for the blind and low vision. 2017 IEEE Life Sciences Conference (LSC), 13-15 Dec. 2017 2017. 115-120.
- LI, D., HAO, S., GUI, J. & HALFOND, W. G. J. An Empirical Study of the Energy Consumption of Android Applications. 2014 IEEE International Conference on Software Maintenance and Evolution, 29 Sept.-3 Oct. 2014 2014. 121-130.
- LI, G. & LI, X. 2012. Research on the Gaussian image Edge Detection Module based on Fuzzy Logic Parameters. *International Journal of Advancements in Computing Technology*, 4, 484-490.
- LIM, D. H. 2006. Robust edge detection in noisy images. *Computational Statistics and Data Analysis*, 50, 803-812.
- LIN, M. H. & TOMASI, C. 2004. Surfaces with occlusions from layered stereo. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26, 1073-1078.
- LIU, G. 2017. Real-Time Object Detection for Autonomous Driving Based on Deep Learning. In: RAHNEMOONFAR, M., BELKHOUCHE, M. & LI, L. (eds.). ProQuest Dissertations Publishing.
- LIU, S. & LIU, Z. 2017. Multi-Channel CNN-based Object Detection for Enhanced Situation Awareness. *arXiv.org*, arXiv:1712.00075v1 [cs.CV].
- LIU, X., JIANG, J., ZHANG, K., LONG, E., CUI, J., ZHU, M., AN, Y., ZHANG, J., LIU, Z., LIN, Z., LI, X., CHEN, J., CAO, Q., LI, J., WU, X., WANG, D. & LIN, H. 2017. Localization and diagnosis framework for pediatric cataracts based on slit-lamp images using deep features of a convolutional neural network.(Research Article)(Report). *PLoS ONE*, 12, e0168606.
- LIU, X., MAKINO, H., KOBAYASHI, S. & MAEDA, Y. 2007. Design of an indoor self-positioning system for the visually impaired--simulation with RFID and Bluetooth in a visible light

- communication system. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, 2007, 1655.
- LONG, R. G. & GIUDICE, N. A. 2010. Establishing and Maintaining Orientation for Mobility. In: WIENER, W. R., WELSH, R. L. & BLASCH, B. B. (eds.) *Foundations of Orientation and Mobility (3rd edition): History and Theory*. New York: American Foundation for the Blind.
- LOOMIS, J. M., KLATZKY, R. L., GOLLEDGE, R. G., CICINELLI, J. G., PELLEGRINO, J. W. & FRY, P. A. 1993. Nonvisual navigation by blind and sighted: assessment of path integration ability. *Journal of experimental psychology. General*, 122, 73.
- M. BEATRICE DIAS, ERMINE A. TEVES, GEORGE J. ZIMMERMAN, HEND K. GEDAWY, SARAH M. BELOUSOV & DIAS, M. B. 2015. Indoor Navigation Challenges for Visually Impaired People. *Indoor Wayfinding and Navigation*. CRC Press.
- MAGNOTTI, J. & BEAUCHAMP, M. 2015. The noisy encoding of disparity model of the McGurk effect. *Psychonomic Bulletin & Review*, 22, 701-709.
- MALIK, J. 2017. Technical Perspective: What led computer vision to deep learning? *Communications of the ACM*, 60, 82-83.
- MARCH, S. & SMITH, G. 1995. Design and Natural Science Research on Information Technology. *Decision Support Systems*, 15, 251 - 266.
- MATA, F., JARAMILLO, A. & CLARAMUNT, C. A mobile navigation and orientation system for blind users in a metrobus environment. 10th International Conference on Web and Wireless Geographical Information Systems (W2GIS), 2011 Kyoto, Japan. 94 - 108.
- MCKNIGHT, C., DILLON, A. & RICHARDSON, J. 1993. Space - the final chapter: Or why physical representations are not semantic intentions. In C. McKnight, A. Dillon, & J. Richardson (Eds.). *Hypertext: A psychological perspective*. New York: Ellis Horwood.
- MEFTAH, B., LEZORAY, O. & BENYETTOU, A. 2010. Segmentation and Edge Detection Based on Spiking Neural Network Model. *Neural Processing Letters*, 32, 131-146.
- METTLER, T., EURICH, M. & WINTER, R. 2014. On the Use of Experiments in Design Science Research: A Proposition of an Evaluation Framework. *Communications of the Association for Information Systems*, 34, 222 - 241.
- MLADENOVIC, N. & HANSEN, P. 1997. Variable neighborhood search. *Computers and Operations Research*, 24, 1097 - 1100.
- MOORE, N. 2000. The Information Needs of Visually Impaired People: A Review of Research. *The Royal National Institute of Blind People (RNIB)*, London.
- NG, A., NGIAM, J., FOO, C. Y., MAI, Y. & SUEN, C. 2013. *Data Preprocessing* [Online]. Available: http://ufldl.stanford.edu/wiki/index.php/Data_Preprocessing#Per-example_mean_subtraction [Accessed 1 April 2017].
- NI, L. & AZIZ, M. A. A. A robust deep belief network-based approach for recognizing dynamic hand gestures. 2016 13th International Bhurban Conference on Applied Sciences and Technology (IBCAST), 12-16 Jan. 2016 2016. 199-205.
- NICHOLSON, J., KEMP-WHEELER, S. & GRIFFITHS, D. 1995. Distress Arising from the end of a Guide Dog Partnership. *Anthrozoös: A multidisciplinary journal of the interactions of people and animals*, 8, 100-110.
- ORITA, K., TAKIZAWA, H., AOYAGI, M., EZAKI, N. & SHINJI, M. Obstacle Detection by the Kinect Cane System for the Visually Impaired. System Integration (SII), 2013 IEEE/SICE International Symposium on, 15-17 Dec. 2013 2013. 115-118.
- PALLEJÄ, T., TRESANCHEZ, M., TEIXIDÓ, M. & PALACIN, J. 2010. Bioinspired Electronic White Cane Implementation Based on a LIDAR, a Tri-Axial Accelerometer and a Tactile Belt. *Sensors*, 10, 11322.

- PASCOLINI, D. & MARIOTTI, S. P. 2011. Global estimates of visual impairment: 2010. *British Journal Ophthalmology Online*, 10.1136/bjophthalmol-2011-300539v1.
- PASSINI, R. & PROULX, G. 1988. Wayfinding without Vision. An Experiment with Congenitally Totally Blind People. *Environment & Behavior*, 20, 227-252.
- PATEL, D. K. & MORE, S. A. 2013. Edge detection technique by fuzzy logic and Cellular Learning Automata using fuzzy image processing.
- PÉREZ-YUS, A., LÓPEZ-NICOLÁS, G. & GUERRERO, J. 2015. Detection and Modelling of Staircases Using a Wearable Depth Sensor. In: AGAPITO, L., BRONSTEIN, M. M. & ROTHER, C. (eds.) *Computer Vision - ECCV 2014 Workshops*. Springer International Publishing.
- PETROU, M. 1999. *Image processing : the fundamentals*, Chichester, Wiley.
- PURAO, S. 2002. Design Research in the Technology of Information Systems: Truth or Dare. In: GSU DEPARTMENT OF CIS, A., GA (ed.).
- RASHID, M., ARDITO, L. & TORCHIANO, M. Energy Consumption Analysis of Algorithms Implementations. 2015 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), 22-23 Oct. 2015 2015. 1-4.
- READ, J. 2003. Controversial Issues in a Disabling Society, John Swain, Sally French and Colin Cameron, Buckingham, Open University Press, 2003, pp. v 198, ISBN 0 335 20904 1, 17.99. *The British Journal of Social Work*, 33, 837-839.
- REDMON, J., DIVVALA, S., GIRSHICK, R. & FARHADI, A. You only look once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. 779-788.
- REN, S., HE, K., GIRSHICK, R. & SUN, J. 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Neural Information Processing Systems (NIPS)*.
- RIESER, J. J., GUTH, D. A. & HILL, E. W. 1982. Mental processes mediating independent travel: Implications for orientation and mobility. *Journal of Visual Impairment and Blindness*, 76, 213-218.
- RIESER, R. 2014. Models of Disability. *Center for Accessible Resources - Our Philosophy* [Online]. [Accessed 11 Nov 2015].
- RODRIGUEZ, A., YEBES, J., ALCANTARILLA, P., BERGASA, L., ALMAZAN, J. & CELA, A. 2012. Assisting the Visually Impaired: Obstacle Detection and Warning System by Acoustic Feedback. *Sensors*, 12, 17476-17496.
- RUDER, S. 2016. An overview of gradient descent optimization algorithms. [Online]. Available: http://www.gdp-research.com.au/minig_1.htm. [Accessed 16th January 2017].
- SALMINEN, A. L. & KARHULA, M. E. 2014. Young persons with visual impairment: challenges of participation. *Scand J Occup Ther*, 21, 267-76.
- SÁNCHEZ, J. & TORRE, N. D. L. 2010. Autonomous navigation through the city for the blind. *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility*. Orlando, Florida, USA: ACM.
- SANFORD, A. J. 1985. Designing for Orientation and Safety. *International Conference on Building Use and Safety Technology*.
- SAUNDERS, M. 2007. *Research methods for business students*, New York, New York: Financial Times/Prentice Hall.
- SCHENKMAN, B. 1986. Identification of ground materials with the aid of tapping sounds and vibrations of long canes for the blind. *Ergonomics*, 29, 985-998.
- SCHIESSER, E. R. 1986. *Principles of navigation*, New York, Wiley.

- SERRÃO, M., SHAHRABADI, S., MORENO, M., JOSÉ, J., RODRIGUES, J. & BUF, J. 2015. Computer vision and GIS for the navigation of blind persons in buildings. *International Journal*, 14, 67-80.
- SETAYESH, M., ZHANG, M. & JOHNSTON, M. Edge detection using constrained discrete particle swarm optimisation in noisy images. *Evolutionary Computation (CEC)*, 2011 IEEE Congress on, 5-8 June 2011. 246-253.
- SETO, F. T. 2009. A navigation system for the visually impaired using colored navigation lines and RFID tags. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*.
- SHARIFI, M., FATHY, M. & MAHMOUDI, M. T. 2002. A classified and comparative study of edge detection algorithms. USA.
- SHOVAL, S., ULRICH, I. & BORENSTEIN, J. 2003. NavBelt and the Guide-Cane (obstacle-avoidance systems for the blind and visually impaired). *Robotics & Automation Magazine, IEEE*, 10, 9-20.
- SIMARD, P. Y., STEINKRAUS, D. & PLATT, J. C. Best practices for convolutional neural networks applied to visual document analysis. *Proceedings of the Seventh International Conference on Document Analysis and Recognition*, 2003. 958 - 962.
- SIMON, H. A. 1996. *The sciences of the artificial*, Cambridge, MA, Cambridge, MA : MIT Press.
- SIMONYAN, K. & ZISSERMAN, A. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv.org*, arXiv:1409.1556v6 [cs.CV].
- SON, D. O., CHOI, H. J., KIM, J. M. & KIM, C. H. Analysis on the Power Efficiency of Mobile Systems Varying Device Parameters. 2014 International Conference on IT Convergence and Security (ICITCS), 28-30 Oct. 2014. 1-3.
- SOTO-FARACO, S., SPENCE, C. & KINGSTONE, A. 2004. Congruency effects between auditory and tactile motion: Extending the phenomenon of cross-modal dynamic capture. *Cognitive, Affective, & Behavioral Neuroscience*, 4, 208-217.
- STEINFELD, E. E. A. 1986. Detectable Tactile Surface Treatments. *The Environmental Design Research Association (EDRA) Conference*.
- STEPNOWSKI, A., KAMIŃSKI, Ł. & DEMKOWICZ, J. Voice maps: the system for navigation of blind in urban area. 10th International Conference on Applied Computer and Applied Computational Science (ACACOS), 2011 Venice, Italy. 201 - 206.
- STEVENS, C. F. 2015. Novel neural circuit mechanism for visual edge detection. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 875.
- STRELOW, E. R. 1985. What is needed for a theory of mobility: Direct perception and cognitive maps—lessons from the blind. *Psychological Review*, 92, 226-248.
- SUN, F., WANG, C., GONG, L., XU, C., ZHANG, Y., LU, Y., LI, X. & ZHOU, X. A Power-Efficient Accelerator for Convolutional Neural Networks. 2017 IEEE International Conference on Cluster Computing (CLUSTER), 5-8 Sept. 2017. 631-632.
- SUN, X. & QIAN, H. 2016. Chinese Herbal Medicine Image Recognition and Retrieval by Convolutional Neural Network. *PloS one*, 11, e0156327.
- SWAIN, J., GRIFFITHS, C. & HEYMAN, B. 2003. Towards a Social Model Approach to Counselling Disabled Clients. *British Journal of Guidance & Counselling*, 31, 137-52.
- TAKEDA, H., VEERKAMP, P., TOMIYAMA, T. & YOSHIKAWA, H. 1990. Modeling Design Processes. *AI Magazine*, Winter: 37 - 48.
- TAKIZAWA, H., YAMAGUCHI, S., AOYAGI, M., EZAKI, N. & MIZUNO, S. 2012. Kinect cane: An assistive system for the visually impaired based on three-dimensional object recognition. *IEEE/SICE International Symposium on System Integration (SII)*.

- TAKIZAWA, H., YAMAGUCHI, S., AOYAGI, M., EZAKI, N. & MIZUNO, S. 2013. Kinect cane: Object recognition aids for the visually impaired. *The 6th International Conference on Human System Interaction (HSI)*.
- TAN, K. C. 2014. Man's best friend not loved by everyone. *The Malay Mail Online*.
- TANG, T., LUI, W. L. D. & LI, W. 2012. Plane-based detection of staircases using inverse depth.
- TATSUMI, H., MURAI, Y. & MIYAKAWA, M. 2007. RFID for aiding the visually impaired recognize surroundings. *IEEE International Conference on Systems, Man and Cybernetics (ISIC)*.
- TAWALBEH, M., EARDLEY, A. & TAWALBEH, L. A. 2016. Studying the Energy Consumption in Mobile Devices. *Procedia Computer Science*, 94, 183-189.
- THAHER, R. H. & HUSSEIN, Z. K. 2014. Stereo Vision Distance Estimation Employing SAD with Canny Edge Detector. *International Journal of Computer Applications*, 107, 38-43.
- THINUS-BLANC, C. & GAUNET, F. 1997. Representation of space in blind persons: vision as a spatial sense? *Psychological bulletin*, 121, 20.
- TOMKINS, L. M., THOMSON, P. C. & MCGREEVY, P. D. 2011. Behavioral and physiological predictors of guide dog success. *Journal of Veterinary Behavior: Clinical Applications and Research*, 6, 178-187.
- TSIRMPAS, C., ROMPAS, A., FOKOU, O. & KOUTSOURIS, D. 2015. An indoor navigation system for visually impaired and elderly people based on Radio Frequency Identification (RFID). *Information Sciences*, 320, 288.
- UMBAUGH, S. E. 2005. *Computer imaging : digital image analysis and processing*, Boca Raton, FL, Boca Raton, FL : Taylor & Francis.
- VAISHNAVI, V. & KUECHLER, W. 2014. Design Science Research in Information Systems. Available: <http://www.desrist.org/design-research-in-information-systems/> [Accessed 12 October 2017].
- VAISHNAVI, V. & KUECHLER, W. 2015. *Design science research methods and patterns : innovating information and communication technology*, Boca Raton, Florida : CRC Press.
- VENARD, O., BAUDOIN, G. & UZAN, G. 2009. Field experimentation of the RAMPE interactive auditive information system for the mobility of blind people in public transport : Final evaluation. *9th International Conference on Intelligent Transport Systems Telecommunications*.
- VLAMINCK, M., JOVANOVIĆ, L., VAN HESE, P., GOOSSENS, B., PHILIPS, W. & PIZURICA, A. Obstacle detection for pedestrians with a visual impairment based on 3D imaging. 3D Imaging (IC3D), 2013 International Conference on, 3-5 Dec. 2013. 1-7.
- WADDINGTON, L. 2009. A Disabled Market: Free Movement of Goods and Services in the EU and Disability Accessibility. *European Law Journal*, 15, 575-598.
- WANG, X., XIA, M. & CAI, H. 2011. Hidden-markov-models-based dynamic hand gesture recognition. *Mathematical Problems in Engineering*.
- WEBER, R. 2017. Design-science Research. In: WILLIAMSON, K. & JOHANSON, G. (eds.) *Research Methods: Information, Systems, and Contexts*. 2nd ed. United Kingdom: Elsevier Ltd.
- WEI-QUN, L. & XIAN-MING, C. 2010. A fast stereo matching using image segmentation for high quality dense disparity maps.
- WEISSTEIN, E. W. "Singular Value Decomposition." From MathWorld--A Wolfram Web Resource. [Online]. Available: <http://mathworld.wolfram.com/SingularValueDecomposition.html>. [Accessed 25th April 2018].

- WENLONG FU, M., JOHNSTON, M. & MENGJIE ZHANG, M. 2012. Soft edge maps from edge detectors evolved by genetic programming.
- WHO 1980. International Classification of Impairments, Disabilities and Handicaps.
- WHO 2001. International Classification of Functioning, Disability and Health (ICF).
- WIGGETT-BARNARD, C. & STEEL, H. 2008. The experience of owning a guide dog. *Disability & Rehabilitation*, 2008, Vol.30(14), p.1014-1026, 30, 1014-1026.
- WILLIAMS, M. A., HURST, A. & KANE, S. K. 2013. "Pray before you step out": describing personal and situational blind navigation behaviors. *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*. Bellevue, Washington: ACM.
- WINIFRED, A. 2014. Use of guide dogs must follow city by-laws, says deputy minister. *The Malay Mail Online*.
- YANG, Y., YUILLE, A. & LU, J. 1993. Local, global, and multilevel stereo matching.
- YELAMARTHI, K., HAAS, D., NIELSEN, D. & MOTHERSELL, S. 2010. RFID and GPS Integrated Navigation System for the Visually Impaired. *53rd IEEE International Midwest Symposium on Circuits and Systems*. IEEE Conference Publications.
- YEUM, C. 2016. Computer vision-based structural assessment exploiting large volumes of images. In: DYKE, S. J., BENES, B., PIZLO, Z., PUJOL, S., RAMIREZ, J. & WACHS, J. (eds.). ProQuest Dissertations Publishing.
- YOKOTA, S., HASHIMOTO, H., CHUGO, D. & KAWABATA, K. The assistive walker using hand haptics - Report of the experiment. Industrial Electronics Society, IECON 2013 - 39th Annual Conference of the IEEE, 10-13 Nov. 2013 2013a. 8329-8334.
- YOKOTA, S., NAGAI, K., MORISHITA, K., MORI, M., CHUGO, D. & HASHIMOTO, H. The assistive walker using hand haptics: The design of the prototype. Human System Interaction (HSI), 2013 The 6th International Conference on, 6-8 June 2013 2013b. 214-218.
- YUAN, D. & MANDUCHI, R. 2005. Dynamic environment exploration using a virtual white cane. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. USA.
- YUANLONG, W. & MINCHEOL, L. A guide-dog robot system research for the visually impaired. Industrial Technology (ICIT), 2014 IEEE International Conference on, Feb. 26 2014-March 1 2014 2014. 800-805.
- ZEILER, M. D. & FERGUS, R. 2013. Visualizing and Understanding Convolutional Networks. *arXiv.org*, arXiv:1311.2901v3 [cs.CV].
- ZHANG, H., ZHAO, C., TANG, Z. & YANG, J. 2008. Particle swarm based stereo algorithm and disparity map evaluation.
- ZHANG, Z. 1998. Determining the Epipolar Geometry and its Uncertainty: A Review. *International Journal of Computer Vision*, 27, 161-195.
- ZIMMERMAN, G. J. 2007. "Mental Processes Mediating Independent Travel: Implications for Orientation and Mobility." (This Mattered to Me) (Book review). *Journal of Visual Impairment & Blindness*, 101, 119.
- ZITNICK, C. L. & KANADE, T. 2000. A cooperative algorithm for stereo matching and occlusion detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22, 675-684.

Appendix

Appendix 1: Hardware Specifications of the Prototype

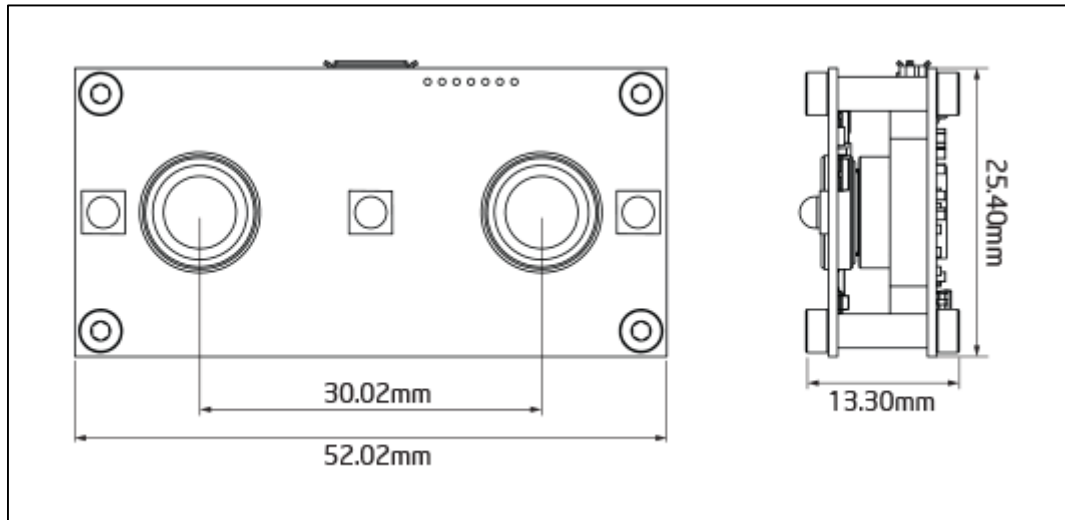


Figure A1.1: Dimensions of the Stereo Camera
(source: <https://duo3d.com/product/duo-minilx-lv1#tab=specs>)

Table A1.1: Specifications of the Stereo Camera (based on details from the manufacture, Code Laboratory Inc.)

Model	DUO-MINILX-LV1 (DUO MLX R2)
Baseline	30.0 mm
Frame Rates	0.1 – 3000+ FPS
Stereo Resolutions	Configurable Binning / Windowing: <ul style="list-style-type: none"> - 45 FPS @ 752x480 - 49 FPS @ 640x480 - 98 FPS @ 640x240 - 192 FPS @ 640x120 - 86 FPS @ 320x480 - 168 FPS @ 320x240 - 320 FPS @ 320x120
Pixel Size	6.0 x 6.0 μm
Shutter Speed	0.3 μsec ~ 10 sec
Lens Mount	Standard M8 x P0.5
Field of View	170° Wide Angle Lens Low Distortion < 3%
Focal Length	2.0mm - 2.1mm
Filters	850-870nm Narrow Band-Pass
Illumination	Fully Programmable LED Array 3 Independently controlled 3.4W 850nm IR LEDs 170° light cone

Illumination Control	Individual brightness sequence programmable in 256 linear steps
Motion Sensing	100Hz Sampling Rate Six Degree of Freedom (DoF) Accelerometer/Gyroscope IMU/Temperature
Colour Modes	Monochrome (S/N Ratio > 54dB Linear)
Control Functions	Exposure/Shutter/Brightness
Scanning Modes	Progressive Scan/Global Shutter
Power Consumption	~2.5 Watt @ +5V DC from USB
Interface	480 Mbps - USB 2.0 Interface (Micro USB)
Weight	12.5g
Operating Systems	DUO OS -Custom Linux Kernel, Linux, Mac OS or Windows
Operating Temperature	0° to 40-50° C (32° to 104-122° F)
Storage Temperature	-20° to 45° C (-4° to 113° F)
Relative Humidity	0% to 90% non-condensing

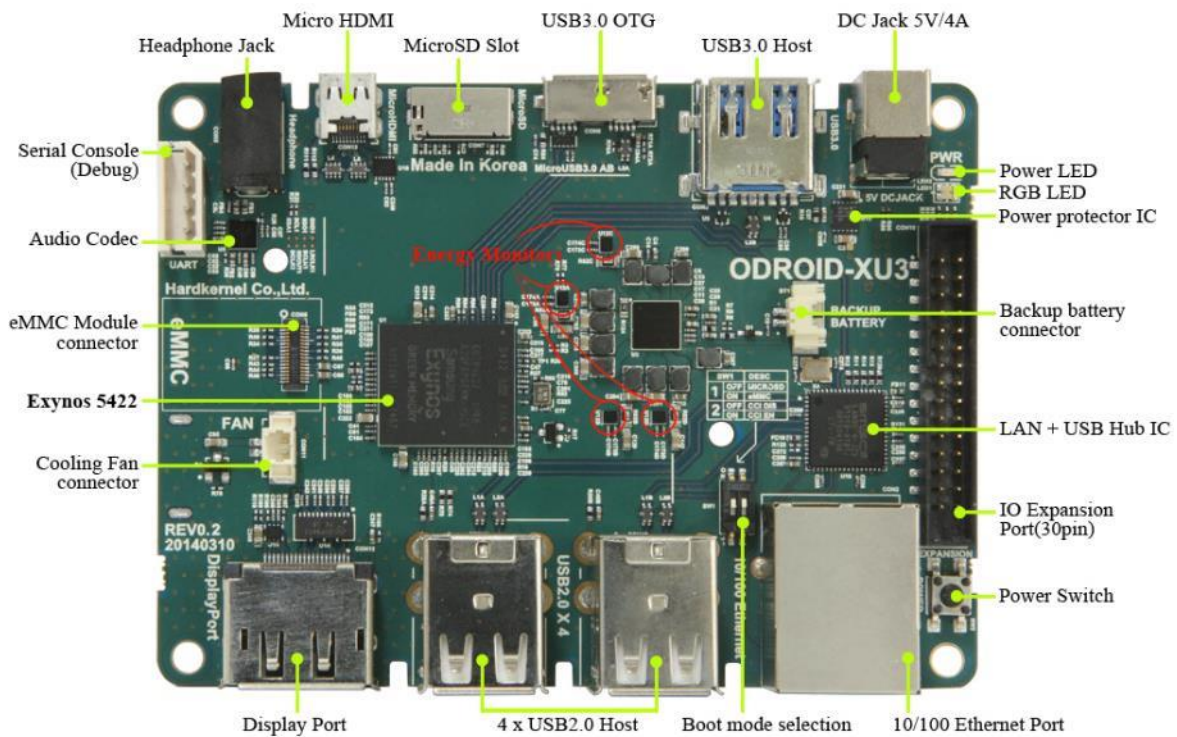


Figure A1.2: Board detail of Odroid XU₃ from the manufacture, Odroid HardKernel Co., Ltd.

Table A1.2: Specifications of Odroid XU3

CPU	Samsung Exynos-5422 : Cortex™-A15 2GHz and Cortex™-A7 big.LITTLE processor with 2GByte LPDDR3 RAM
GPU	Arm Mali-T628 (MP6)
eMMC 5.0 module	16GB/32GB : Sandisk iNAND Extreme 64GB : Toshiba eMMC
LAN/USB Hub	LAN9514 4-port Hi-Speed USB 2.0 hub and 10/100 Ethernet controllers from SMSC/Microchip
USB Load Switch	NCP380 Protection IC for USB power supply from OnSemi.
Audio Codec	MAX98090 is a full-featured and high-performance audio CODEC from Maxim
Power protector	NCP372 Over-voltage, Over-current, Reverse-voltage protection IC from OnSemi.
LED indicator	Tri-color RGB LED to display the status of operating system
HDMI connector	Standard Micro-HDMI, supports up to 1920 x 1080 resolution
DisplayPort connector	Standard DisplayPort, supports up to 3840 x 2160 resolution
IO Ports	USB 3.0 Host x 1, USB 2.0 Host x 4, USB 3.0 OTG x 1, PWM for Cooler Ethernet RJ-45, Headphone Jack, 30 Pin : GPIO/IRQ/SPI/ADC
Storage slot	Micro-SD slot, eMMC 5.0 module connector
DC Input	5V / 4A input, plug specification is inner diameter 2.1mm and outer diameter 5.5mm
Energy Monitor	4 separated current sensors to measure the power consumption of Big CPU, Little CPU, GPU and DRAM in real time

Appendix 2: Phase-1 Prototype Design Details

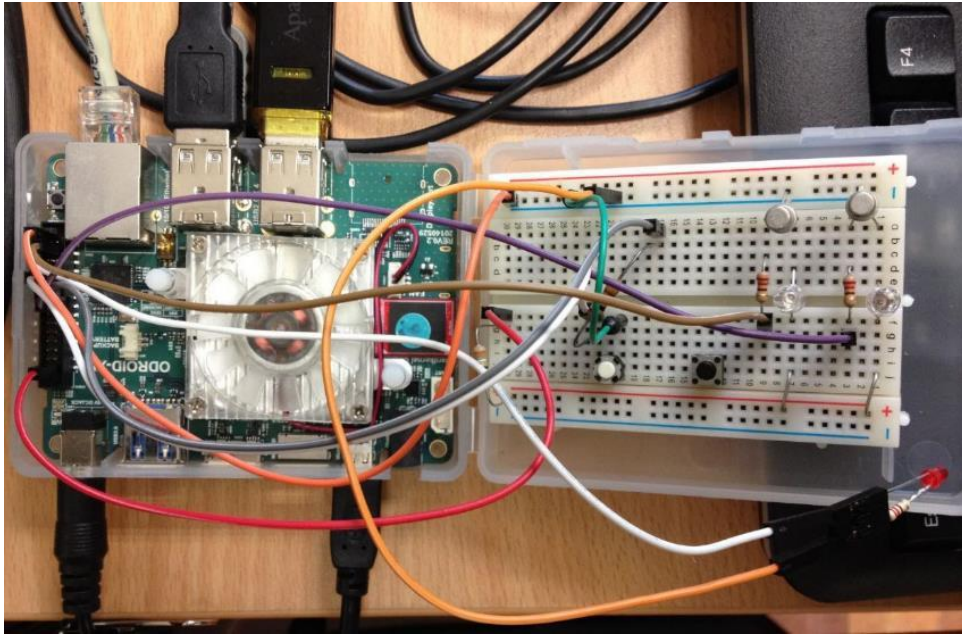


Figure A2.1: ODROID XU3 board (left) and a breadboard (right) during the initial phase of designing the circuitry. The circuit was designed mainly to operate data capturing during this phase.

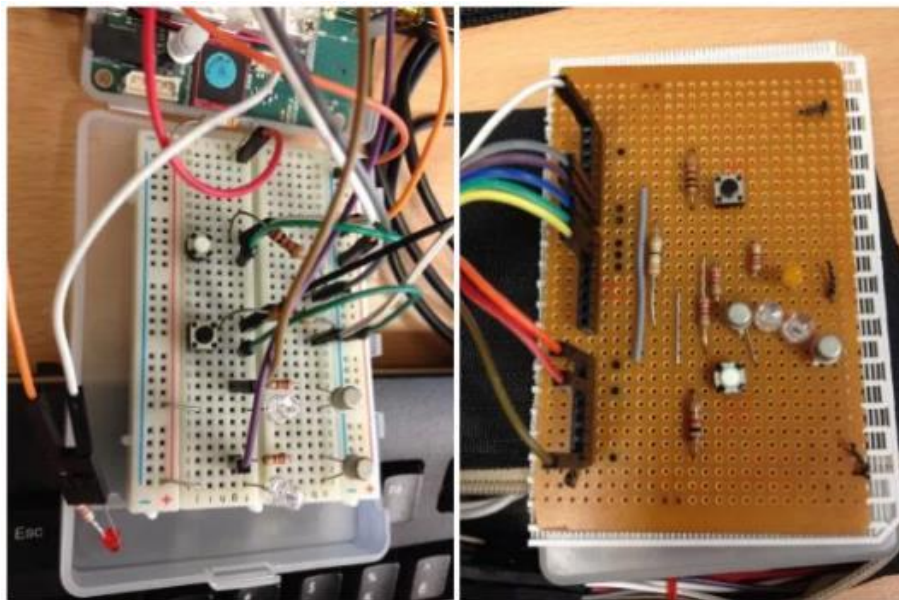


Figure A2.2: Once finalized, the workable circuit on the left photo was moved to a Veroboard as shown on the right photo to make it handy for outdoor usage. 3 LEDs were added as indicators during data collection to indicate: (1) power status, (2) readiness of the system for data capturing, and (3) data storing status.

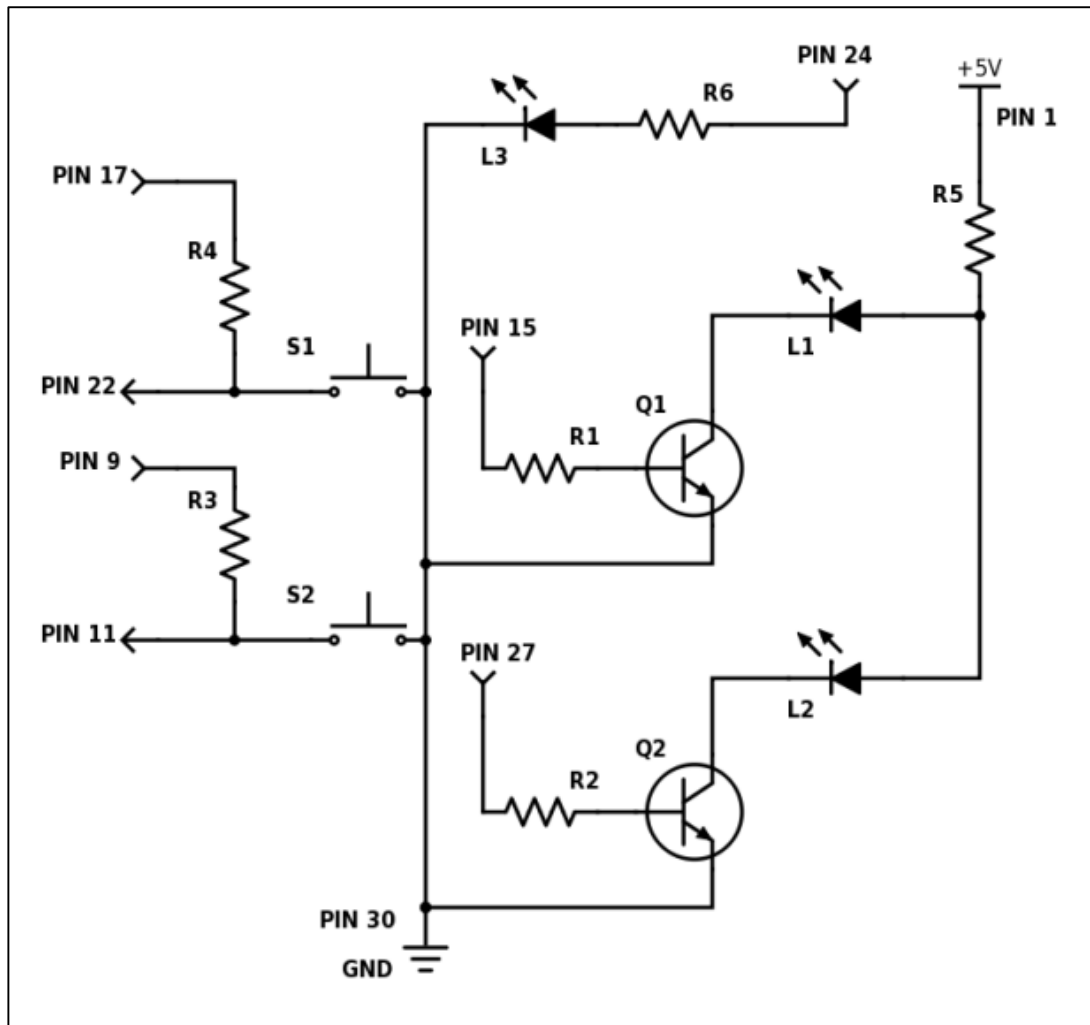


Figure A2.3: The schematic diagram of the circuit to turn the camera program and data storing function ON/OFF by accessing the GPIO on ODROID XU3. LEDs were included as indicator.

More details about the GPIO pins can be found in Appendix 1.

List of electronic components used in the circuit as labelled in Figure A2.3:

- Q1, Q2: two N2222A transistors
- L1, L2: two LEDs (3 – 5 volts)
- L3: one LED (1.5 volts)
- S1, S2: two button switches (momentary press contact)
- R1, R2, R6: three 220 Ω resistors
- R3, R4: two 1 kilo Ω resistors
- R5: one 390 Ω resistor



Figure A2.4: The lithium power source. Since ODROID XU3 is running on 5 volts, a voltage regulator (the green board) is added to convert the 11.1 volts from the lithium battery. Battery specifications: 11.1V, 2300 mAh, 30 C, Li-Polymer battery.

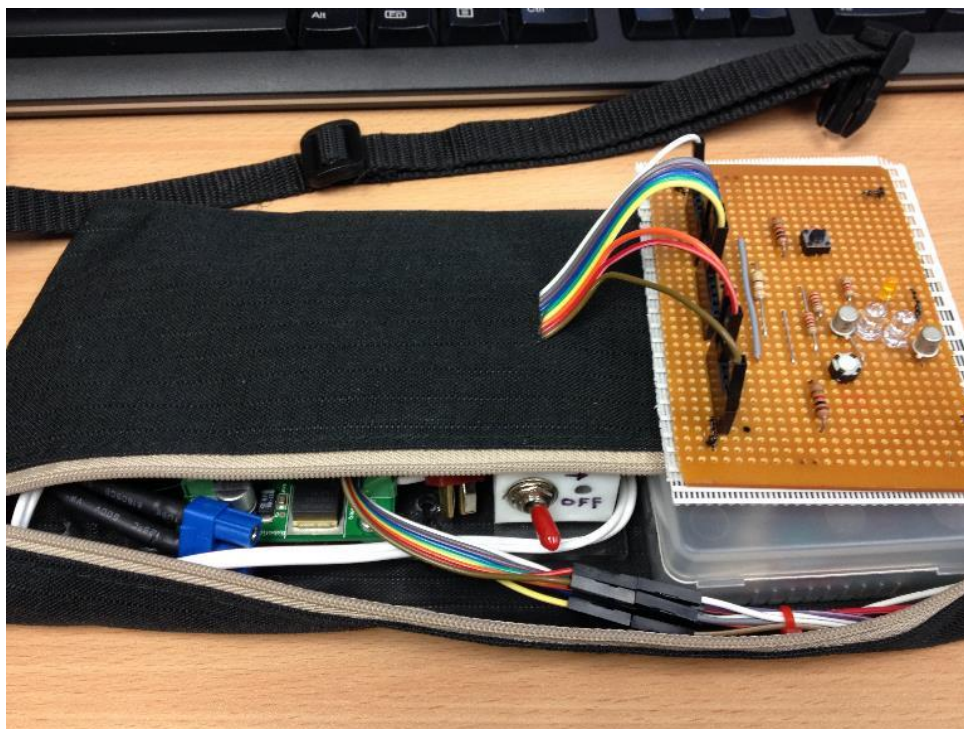


Figure A2.5: All in a waist pouch – putting together the lithium power source, the ODROID board (can be seen hidden in a plastic box inside the pouch) and the circuit board ready for data collection.



Figure A2.6: Wearing the pouch at waist level of a user.

Appendix 3: Crowdsourcing via Social Media



A3.1: Screen shots from crowdsourcing via social media to locate potential samples within urban areas in Malaysia, prior to actual data collection. The viewers of the social media were encouraged to contribute by snapping photos of surface discontinuities they found along a walkway and share it on the page with caption of the location.

Appendix 4: Variable Neighbourhood Search

The full algorithm based on Mladenović and Hansen (1997) is split into three parts. There are some variants to these algorithms since its introduction, but the following are some original works from Mladenović and Hansen, which were referred to in this research.

Algorithm 1: Best improvement (highest descent) heuristic

Function BestImprovement(x)

```
1: repeat
2:    $x' \leftarrow x$ 
3:    $x \leftarrow \operatorname{argmin}_{y \in N(x)} \{f(y)\}$ 
4: until ( $f(x) \geq f(x')$ )
5: return x
```

Algorithm 2: – Neighbourhood change

Function NeighbourhoodChange (x, x', k)

```
1: if  $f(x') < f(x)$  then
2:    $x \leftarrow x'$  // Make a move
3:    $k \leftarrow 1$  // Initial neighbourhood
4: else
5:    $k \leftarrow k+1$  // Next neighbourhood
```

Algorithm 3: Basic Variable Neighbourhood Search

Function VNS (x, kmax, tmax);

```
1: repeat
2:   k ← 1;
3:   repeat
4:     x' ← Shake(x, k); // randomly generate a neighbourhood solution
5:     x'' ← BestImprovement(x') // Local search ;
6:     x ← NeighbourhoodChange(x, x'', k) // Change neighbourhood ;
7:   until k = k_max ;
8:   t ← CpuTime() // if termination is based on CPU time
9: until t > t_max ;
```