

Summer 2014

Multimodal perception of histological images for persons blind or visually impaired

Ting Zhang
Purdue University

Follow this and additional works at: https://docs.lib.purdue.edu/open_access_theses



Part of the [Industrial Engineering Commons](#)

Recommended Citation

Zhang, Ting, "Multimodal perception of histological images for persons blind or visually impaired" (2014). *Open Access Theses*. 715.
https://docs.lib.purdue.edu/open_access_theses/715

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

**PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance**

This is to certify that the thesis/dissertation prepared

By Ting Zhang

Entitled

Multimodal Perception of Histological Images for Persons Blind or Visually Impaired

For the degree of Master of Science in Industrial Engineering



Is approved by the final examining committee:

Bradley S. Duerstock

Juan P. Wachs

Vincent Duffy

To the best of my knowledge and as understood by the student in the *Thesis/Dissertation Agreement, Publication Delay, and Certification/Disclaimer (Graduate School Form 32)*, this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Bradley S. Duerstock

Approved by Major Professor(s):

Juan P. Wachs

Approved by: Steven Landry

07/29/2014

Head of the Department Graduate Program

Date

MULTIMODAL PERCEPTION OF HISTOLOGICAL IMAGES FOR PERSONS
BLIND OR VISUALLY IMPAIRED

A Thesis

Submitted to the Faculty

of

Purdue University

by

Ting Zhang

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science in Industrial Engineering

August 2014

Purdue University

West Lafayette, Indiana

ACKNOWLEDGMENTS

Foremost, I would like to express my sincere gratitude to my advisors, Prof. Bradley S. Duerstock and Prof. Juan P. Wachs, for their continuous support of my graduate study and research. Their encouragement, patience, motivation and immense knowledge helped me in all the time of research and writing of this thesis. I could not have imagined having better advisors and mentors for my graduate study.

Besides my advisors, I would like to thank the rest of my thesis committee: Prof. Vincent Duffy. In the first year I started my graduate study, Prof. Vincent Duffy led me in understanding what research is and how to do research.

This thesis is supported by the National Institute for Health Director's Pathfinder Award to Promote Diversity in the Scientific Workforce (1DP4GM096842-01). The author would also thank the State of Indiana through support of the Center for Paralysis Research at Purdue University.

My sincere thanks also goes to all my lab mates, Yu-Ting Li, Hairong Jiang, Dr. Greg J. Williams and Mithun G. Jacob, for their insightful comments and stimulating discussions. Also, many thanks to my parents, Yong-Sheng Zhang and Jing Wang and my boyfriend, Ying-Long Chen for their company and support spiritually throughout my life.

Last but not the least, I would like to thank my friends, Chen-Xing Niu, A-Zhu Liu and Si-Ming Xu, for all their encouragement and support.

TABLE OF CONTENTS

	Page
LIST OF TABLES	vi
LIST OF FIGURES	vii
LIST OF ABBREVIATIONS.....	ix
ABSTRACT	x
CHAPTER 1. INTRODUCTION.....	1
1.1 Overview	1
1.2 Research Problem.....	4
1.2.1 Research Questions	5
1.3 Contribution	5
1.4 Thesis Structure.....	6
CHAPTER 2. LITERATURE REVIEW	7
2.1 Sensory Substitution.....	7
2.1.1 Tactile Sensory Substitution	8
2.1.1.1 Tactile-visual substitution.....	11
2.1.2 Auditory Sensory Substitution	13
2.1.2.1 Auditory-visual substitution	14
2.2 Image Processing.....	15
2.2.1 Color to Grayscale.....	15
2.2.2 Edge Detection.....	17
2.2.3 Texture Analysis	19
2.2.3.1 Statistical Texture Analysis Approach	21
2.2.3.2 Structural Texture Analysis Approach	22
2.3 Multimodal Sensory Interpretation	23

	Page
2.4	Bayesian Network 26
2.5	Linear Assignment Problem..... 28
CHAPTER 3.	METHODOLOGY 32
3.1	Image Feature Bayesian Network 33
3.1.1	Primary Feature Extraction34
3.1.1.1	Intensity 34
3.1.1.2	Texture..... 35
3.1.1.3	Shape 37
3.1.1.4	Color 39
3.1.2	Peripheral Feature Extraction.....39
3.1.2.1	Expert-based Modeling..... 40
3.1.2.2	Probability Calculation 41
3.1.2.3	Bayesian Network Optimization 42
3.2	Modality Assignment Problem..... 43
3.2.1	Problem Definition.....44
3.2.2	Cost Weighing.....45
3.2.3	Linear Assignment Algorithm.....45
CHAPTER 4.	EXPERIMENTS AND RESULTS..... 48
4.1	Experiments..... 48
4.1.1	Experiment 1: Finding the Rank of Modalities.....49
4.1.2	Experiment 2: Comparing with print-out tactile paper51
4.2	Results 53
4.2.1	Experiment 1: Finding the Rank of Modalities.....53
4.2.1.1	Intensity 54
4.2.1.2	Texture..... 56
4.2.1.3	Shape 58
4.2.1.4	Color 60
4.2.1.5	Cost Matrix 62
4.2.2	Experiment 2: Comparing with print-out tactile paper63

	Page
CHAPTER 5. CONCLUSIONS AND FUTURE WORK.....	66
5.1 Possible Changes to Bayesian Network	67
5.2 Expanding the Modality Assignment Problem	67
5.3 Considerations for Future Experiments.....	69
5.4 Possible Improvements for Human-computer Interaction	69
BIBLIOGRAPHY	72
APPENDICES	
Appendix A Consent Form	86
Appendix B Experiment Procedures	89
VITA	91
PUBLICATIONS	93

LIST OF TABLES

Table	Page
Table 2.1 Summary of electrotactile sensation thresholds and pain/sensation current ratios	10
Table 3.1 Definition of Discrete States for Each Node	40
Table 3.2 Candidate Modalities for Each Feature	44
Table 4.1 Summary of Test Images and Tasks for Each Primary Feature	50
Table 4.2 Test Images and Tasks for Experiment 2.....	52

LIST OF FIGURES

Figure	Page
Figure 1.1 A blind subject navigating a blood smear image using a haptic device with a stylus grip and perceiving blood smear image through multiple modalities in real-time...	4
Figure 2.1 Vision substitution system with vibration stimulators.	8
Figure 2.2 Example of reading paper material using Optacon.	11
Figure 2.3 Tactile-vision substitution through the tongue.	13
Figure 2.4 Mobile TVSS system with the use of smartphones.	13
Figure 2.5 Examples of color image to grayscale conversion. (a) Original image; (b) Converted grayscale image; (c) Red channel; (d) Green channel; (e) Blue channel.	17
Figure 2.6 Results of different edge detection algorithms.	18
Figure 2.7 Vegetation communities derived from automated segmentation based on image texture analysis.	20
Figure 2.8 Artificial image texture examples.	23
Figure 2.9 HOMERE multimodal system for virtual environment navigation for persons blind or visually impaired.	25
Figure 2.10 A simple Bayesian network.	26
Figure 2.11 Linear assignment problem bi-graph.	29
Figure 3.1 System Architecture.	33

Figure	Page
Figure 3.2 Intensity computed from color information. (a) Color image (b) Grayscale image of (a).....	35
Figure 3.3 Different textures compared between red blood cells (b) and white blood cells (c) in a blood smear image (a).	36
Figure 3.4 Shape difference between normal red blood cells and sickle cells.	38
Figure 3.5 Edge detection results for 4 tested blood smear images. (a) ~ (d): Original images; (e) ~ (h) corresponding detected edges.	38
Figure 3.6 Candidate Bayesian structures generated by typical user.....	41
Figure 3.7 Optimal Bayesian structure.	43
Figure 4.1 Tactile paper using specialized thermal capsule paper.....	53
Figure 4.2 Response time and error rate for feature Intensity.	56
Figure 4.3 Response time and error rate for feature Texture.....	58
Figure 4.4 Response time and error rate for feature Shape.....	60
Figure 4.5 Response time and error rate for feature Color	61
Figure 4.6 Optimal matching of modalities and primary features	62
Figure 4.7 Response time and error rate for all tasks in experiment 2.	64
Figure 5.1 Force Dimension 7 DOF haptic device with a gripper end-effector.	71

LIST OF ABBREVIATIONS

NHIS	National Health Interview Survey
BVI	Blind or Visually Impaired
AT	Assistive Technologies
LM	Light Microscopy
NSF	National Science Foundation
HCI	Human-computer Interfaces
TVSS	Tactile-vision Sensory Substitution
LAP	Linear Assignment Problem
QAP	Quadratic Assignment Problem
BN	Bayesian Network
DAG	Directed Acyclic Graph
GA	Genetic Algorithms
QAP	Quadratic Assignment Problem

ABSTRACT

Zhang, Ting. M.S.I.E., Purdue University, August 2014. Multimodal Perception of Histological Images for Persons Blind or Visually Impaired. Major Professors: Bradley S. Duerstock and Juan P. Wachs.

Currently there is no suitable substitute technology to enable blind or visually impaired (BVI) people to interpret visual scientific data commonly generated during lab experimentation in real time, such as performing light microscopy, spectrometry, and observing chemical reactions. This reliance upon visual interpretation of scientific data certainly impedes students and scientists that are BVI from advancing in careers in medicine, biology, chemistry, and other scientific fields. To address this challenge, a real-time multimodal image perception system is developed to transform standard laboratory blood smear images for persons with BVI to perceive, employing a combination of auditory, haptic, and vibrotactile feedbacks. These sensory feedbacks are used to convey visual information through alternative perceptual channels, thus creating a palette of multimodal, sensorial information. A Bayesian network is developed to characterize images through two groups of features of interest: primary and peripheral features. Causal relation links were established between these two groups of features. Then, a method was conceived for optimal matching between primary features and sensory modalities. Experimental results confirmed this real-time approach of higher accuracy in recognizing and analyzing objects within images compared to tactile images.

CHAPTER 1. INTRODUCTION

1.1 Overview

From the 2011 National Health Interview Survey (NHIS) Preliminary Report (American Foundation for the Blind, n.d.), there are estimated 21.2 million adults in the United States, more than 10% of all adult Americans have impaired sight. Over 59,000 children (through age 21) in the United States are enrolled in elementary through high schools. Among 6,607,800 are working-age blind or visually impaired (BVI) persons. Of these individuals 64% stated that they did not finish high school and only 16% received high school diplomas. The BVI population who earned Bachelor's or higher degrees was much less, only 374,400 or 5.7% of those aged 21 to 64. The lack of proper and effective assistive technologies (AT) is a major roadblock for individuals that are BVI wanting to actively participate in science and advanced research activities (W. Yu, Reid, & Brewster, 2002). A major challenge for them is to perceive and understand scientific visual data acquired during wet lab experimentation, such as viewing live specimens through a stereo microscope or histological samples through light microscopy (LM) (Bradley S. Duerstock, Lisa Hillard, & Deana McDonagh, 2014). According to Science and Engineering Indicator 2014 published by the National Science Foundation (NSF), no more than 1% of blind or visually impaired people are involved in advanced science and engineering research and receive doctoral degrees (National Science Board, 2014).

Images have always been a direct way to convey information, like the adage says “A picture is worth a thousand words”. Although this may not be true to all the cases, there is an increasing trend that images, as well as diagrams, charts and scientific data have been applied in diverse situations replacing word description to assist people to understand the content. This trend has also led to the recent growth of the visual analytics discipline (Wong & Thomas, 2004). However, the popularizations of proper assistive technologies are not sufficient for BVI students and scientists to easily interpret images. More than 70% of textbooks consists of diagrams without word description (Burch & Pawluk, 2011). Braille, audio books and screen readers are common assistive technologies applied to help blind students reading word material, while tactile papers are utilized to show images. Tactile graphics work similar as Braille in that surfaces are slightly raised to highlight important features of an image. Although computer-aided tactile graphics printing systems have alleviated the load for people who manually create the tactile graphics (Takagi, 2009), the information that tactile graphics can convey is much less than what visual perception provides. With the popularization and cost reduction of 3D printers, increasing interest in their use to generate tactile graphics has surged. Now, 3D models can be created for 2D images by mapping pixel intensity to plate height (Greg J. Williams et al., 2014). By utilizing 3D printing technology, more information, like intensity, pattern and relative relationship, can be revealed to the visually impaired. However, it is still time consuming for a 3D printer to print out a tactile plate (from 5 to 7 hours depending on image’s size and resolution). This cannot be a viable real-time solution.

Real-time methods leveraging from hearing and tactual sensoria have been studied as well. However, by using current single-modality human-computer interfaces (HCI), only limited visual information can be accessed. Tactile-vision sensory substitution (TVSS) technologies, such as Tongue electrotactile array (P. Bach-y-Rita, Kaczmarek, Tyler, & Garcia-Lara, 1998), and dynamic tactile pictures (Heller, 2002), have been demonstrated capable of conveying visual information (Paul Bach-y-Rita & W. Kercel, 2003) of spatial phenomenology (Ward & Meijer, 2010). Nevertheless, the low resolutions of somatosensory display arrays have been reiteratively reported as a major limitation to convey complex image information. Auditory-vision sensory substitution has also been studied in image perception (Capelle, Trullemans, Arno, & Veraart, 1998; De Volder et al., 2001) as a potential solution. Trained early blind participants showed increased performance in localization and object recognition (Arno, Capelle, Wanet-Defalque, Catalan-Ahumada, & Veraart, 1999) through auditor-vision sensory substitution. Auditory-vision sensory substitution involves the memorization of different audio forms and training is required to map from different audio stimulus to visual cues. In addition, it has been shown that the focus on auditory feedback may decrease the subjects' ability to get information from the environment (Meers & Ward, 2005).

The current gap for this problem is that existing solutions cannot help convey to blind persons the richness, complexity and amount of visual data readily understood by persons without disabilities. In this study, a real-time multimodal image perception approach (see Figure 1.1) is investigated that offers feedback through multiple sensory channels, including auditory, haptics and vibrotactile. Through the integration of multiple sensorial substitutions, participants supported using this studied platform showed higher analytic

performance than when using the standard interface based on tactile sensory feedback only.

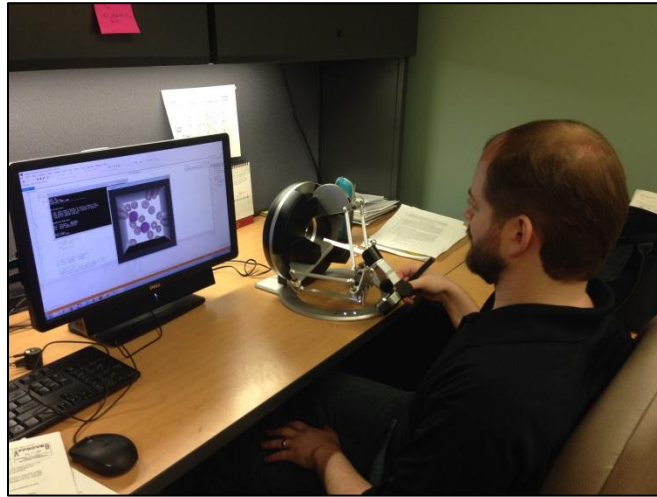


Figure 1.1 A blind subject navigating a blood smear image using a haptic device with a stylus grip and perceiving blood smear image through multiple modalities in real-time.

1.2 Research Problem

Representing visual images for visually impaired people is not a new problem; however, the solutions suggested so far have shown to represent only a small fraction of the original visual content. Most of current methods are only based on one modality to convey visual information. For instance, perception of tactile graphics is dependent on fingertip tactual feedback. Verbal descriptions are dependent on listening. However, the integration of the multiple modalities has been challenging at the least. Integrating multiple modalities has been studied in HCI (human computer interaction) research to enhance the processing and understanding of complex information. A few studies have focused on the integration of hearing and tactual feedback to assist the visually impaired. For example, a method to provide simple visual information, such as navigating bar

charts (W. Yu et al., 2002) has been suggested. In our study, histology images that have both educational and clinical relevance are evaluated.

By building on the assumption that images can be encoded through several features, and each feature can only be represented by one modality; the objective of this work is to find the optimal mapping between feedback modality and image feature in terms of human performance on image navigation and recognition. This problem can be represented as:

$$\max f(i, j)$$

where f is a metric that evaluates the outcome and performance of a multimodal system. Different image features are denoted by i , and j represents a feedback modality. The indices of i and j indicates the selection of mapping from one image feature to one feedback modality.

1.2.1 Research Questions

RQ1: What is the optimal mapping between feedback modality and image feature that leads to better task performance?

RQ2: Does this integrated method lead to better task performance than single-modality methods currently provided to blind students?

1.3 Contribution

In this thesis, the relationship between image features and sensory modalities is studied which could be applied to different image perception based areas, such as virtual reality environments and vision sensorial substitution systems. The construction of Bayesian network reveals how various image features can affect each other. These causal relations

between different image features facilitate the progress of generating images accessible to BVI persons. With the conditional dependency probabilities computed from the image feature Bayesian network, the key information in an image can be identified by finding the feature of largest possibility, or by finding the feature which is the cause for most features. Empirical studies incorporating a real-time multimodal image perception approach described in this thesis have shown to effectively help BVI people to independently navigate and explore histology images. One of the advantages of this approach is it not only decreases the time and man power required in traditional print-out methods, but makes real-time image-based data interpretable by blind individuals. This HCI system can also be connected to a light microscope with a computer monitor output in order to render digitized image information to BVI users in real-time. A 6° of freedom haptic controller and peripheral vibrotactile device connected to the computer as well as the computer speakers are used for the user interface.

1.4 Thesis Structure

The rest of the thesis is divided as follows. Chapter 2 summarizes previous related work. Chapter 3 describes the methodology of the scientific approach applied to address the research questions in this thesis (1.2.1). An overview of the entire system with a succinct description of each module is first described. Then, the strategy applied to characterize images by key features and to construct feature relationships using a Bayesian network are illustrated. The last part of Chapter 3 describes the investigation of proper sensorial substitution by solving a linear assignment problem. Experiments and results are explained in Chapter 4. Finally, conclusions and future work are discussed in Chapter 5.

CHAPTER 2. LITERATURE REVIEW

This chapter is an overview of current research that pertains to this thesis. First, real-time sensory substitution that expresses the information conveyed by one sensory modality through another sensory modality is described. Different applications of real-time sensory substitution are illustrated. Then, an introduction of basic image processing techniques that are utilized to extract key features of images is mentioned. The Bayesian network and linear assignment problem used in this thesis are also illustrated here. At last, systems that utilized multiple sensory modalities are discussed.

2.1 Sensory Substitution

Blind or deaf people fail to see or hear because they lose the ability to transmit the sensory signals from the sensory modality to their brain (Paul Bach-y-Rita & W. Kercel, 2003). Therefore, to replace the functionality of an impaired sensory modality, other functioning sensory systems must be utilized to alternatively convey the missing sensory information. This is called sensory substitution. This concept was first introduced in 1969 to describe blind persons perceiving images using tactile images (Bach-Y-Rita, Collins, Saunders, White, & Scadden, 1969). Through the vibrations of four hundred solenoid stimulators arranged in a twenty by twenty array that presses against the skin of the back (figure 2.1), participants were able to distinguish and identifying different objects. This required twenty to forty hours of extensive training. Tactile and audio sensory

substitutions are the two most popular sensory substitution approaches currently most studied. In sections below, tactile and audio sensory substitutions for vision sensory systems are discussed.

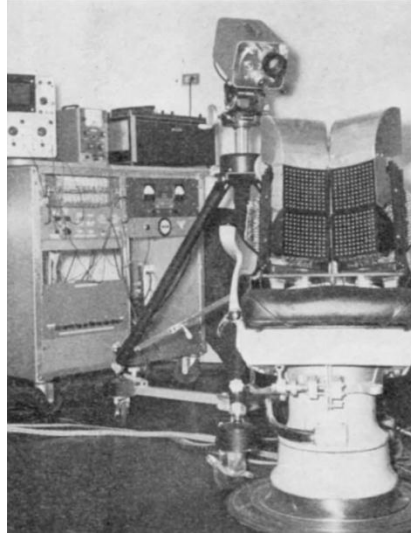


Figure 2.1 Vision substitution system with vibration stimulators.

2.1.1 Tactile Sensory Substitution

Tactile sensory substitution can first be categorized as two different groups based on the two different types of stimulators it utilizes: electrotactile or vibrotactile. Both of them have strengths and drawbacks.

Obvious from the literature, electrotactile stimulators generate electronic voltage to stimulate the touch nerve endings in the skin. Different sensations will be felt according to various voltages, currents and waveform. It can also be affected by the material, size of the contact devices and the skin condition of the contact location (Kaczmarek, Webster, Bach-y-Rita, & Tompkins, 1991). A variety number of body areas can be utilized to receive electrotactile stimulation, such as back, abdomen, fingers, forehead, tongue and the roof of the mouth (Paul Bach-y-Rita, 2004). Due to different impedance of different

skin areas, high voltage stimulation may be applied if the contact is located at high impedance areas. However, this is not considered as a safe approach. The tongue and roof of the mouth are then proposed to be a better place to receive electrotactile stimulation since it is proved that low currents and voltages can be felt at those areas (P. Bach-y-Rita et al., 1998; Tang & Beebe, 2003). Using the tongue as the receptor has been approved as assistance to the blind in the United Kingdom and utilized in clinical experiments for a number of applications (“HowStuffWorks ‘BrainPort,’” n.d.). The major disadvantages of electrotactile stimulator would be the distress caused to participants while the voltage is high or the duration of stimulation is too long. Thresholds of sensation and pain or P/S , can be considered as a key indicator to determine the properness of parameters setting. In Table 2.1 (Kaczmarek et al., 1991) summarizes the results of some experiments which indicate a best range of parameters setting to satisfy the sensation and pain threshold.

Vibrotactile stimulations take advantage of mechanical vibratory somatosensation through the skin. Vibration sensations are perceived according to different vibration frequencies, normally ranges from 10 to 500Hz (Kaczmarek et al., 1991). The Optacon™ was manufactured in 1971. It is a vision sensory substitution device that realized real-time paper material reading by providing feedbacks through vibration on index finger. Figure 2.2 shows an example of using Optacon reading a book. The Optacon consists of a small camera and a 24 by 6 array that can vibrate according to the image that is being viewed through the camera. Blind people place their index fingers onto the dynamic tactile array, and use their other hand to move the camera across a line of print. Although vibrotactile stimulation is more safe than electrotactile stimulation, the main problem of

vibrotactile stimulation is that participants may physiologically adapt to the tactile sense rapidly (Way & Barner, 1997).

Table 2.1 Summary of electrotactile sensation thresholds and pain/sensation current ratios

Electrode type	Body location	Electr. Area (mm ²)	Wave-form	Freq. (Hz)	Pulse width limits (ms)	Sensation Current (mA)	Sensation Charge (nC)	P/S
Silver coaxial	Abdomen	15.9	M-	60/200	0.002	20	40	8
					0.7	0.1	70	
SS coaxial gelled	Trunk	8.42	M	(a)	0.1	1.5	150	1.6
	Fingertip	8.42	M	(a)	0.1	6	600	
SS/aluminum coaxial	Abdomen	0.785	M	50	0.25	0.4	100	6.25
Steel electrode pair	Fingertip	0.0078	M	(b)	0.5	(c) 0.2	100	1.5
						(d) 1.0	500	
Coaxial	Forearm Back abdomen	7.07	PT	25	1	17	17	8.4
					100	2.5	250	

Waveforms: M is monophasic, + or – indicated if known; PT is the pulse train

Comments: (a) Best frequency 1-100 Hz; (b) Best frequency 1-200 Hz;

(c), (d) 0.79 and 6.35 mm electrode spacing.

Tactile substitution can also be divided as several categories in terms of the sensory channel it replaces, such as tactile-visual substitution and tactile-auditory substitution. Since people that are blind or visually impaired is the target group in this thesis, only tactile-visual substitution is introduced in the following section.



Figure 2.2 Example of reading paper material using Optacon.

2.1.1.1 Tactile-visual substitution

Since tactile-vision sensory substitution (TVSS) was first introduced by Bach-y-Rita in 1969, studies and applications of it have not paused. From image perception to video understanding, from obstacle detection to way finding, tactile-vision sensory substitution has been utilized in various ways helping BVI people succeed not only in performing activities of daily living (ADL), but in academic and occupational activities as well (Paul Bach-y-Rita, 2004; Fritz, Way, & Barner, 1996; Johnson & Higgins, 2006; Nguyen et al., 2013; Owen, Petro, D'Souza, Rastogi, & Pawluk, 2009; Rastogi & Pawluk, 2013).

Tongue has been shown to be sensitive to electrotactile stimulations at low electronic voltage and currents. Various applications have been studied utilizing the tongue due to its high concentration of sensory receptors on its surface (Paul Bach-y-Rita & Kaczmarek, 2002; Paul Bach-y-Rita & W. Kercel, 2003). Figure 2.3 shows an example of TVSS using the tongue as the visual substitute modality. Cross-modal brain plasticity is also examined using electrotactile stimulations on the tongue (Pitro, Moesgaard, Gjedde, &

Kupers, 2005; Sampaio, Maris, & Bach-y-Rita, 2001). Wireless electro tactile devices have been studied to give BVI persons more freedom to navigate the environment (Nguyen et al., 2013). A recent research that takes advantages of smartphones (Kajimoto, Suzuki, & Kanno, 2014) reveals more possibilities and directions that TVSS can be applied. The system consists of an electro tactile display with 512 electrodes, a smartphone and an LCD (shown in Figure 2.4). Participants were able to get a view of the surroundings by taking photos using the camera on the smartphone. Images are then converted through the optical sensors beneath each electrode. This low cost but powerful system gives us a hint to connect current assistive technologies with mobile devices which become increasingly popular in recent years. Although TVSS succeed in helping BVI people navigate the environment and perceiving images, the low resolution of tactile somatosensation compared to the visual system has always been a main drawback of this method. The low resolution of tactual sensory compared to visual limits blind or visually impaired people to access complex visual information. Studies have shown the ratio of tactual to visual bandwidth is around 1 to 1000, which means the capacity of tactual sense to receive and perceive is much less than vision (Way & Barner, 1997). Usually, a TVSS user must move the camera connected with the TVSS all around to identify an object. Therefore, to improve the capabilities of conventional tactile-vision sensory substitution and decrease the drawbacks of low resolution of tactile displays, image processing and trajectory tracking algorithms have also been studied to help BVI explore the environment (Hsu et al., 2013).



Figure 2.3 Tactile-vision substitution through the tongue.



Figure 2.4 Mobile TVSS system with the use of smartphones.

2.1.2 Auditory Sensory Substitution

Instead of tactual feedback, auditory sensory substitution systems take advantages of auditory feedback to compensate for the lack of other sensory modalities. Visual or tactile

information are detected and transformed into auditory signals. Auditory-visual substitution can be considered as the most popular auditory sensory substitution system.

2.1.2.1 Auditory-visual substitution

To take advantages of auditory sensory, the frequency of auditory pitch, binaural intensity and phase differences, sound loudness, specific sets of tones are mapped to different image properties (Capelle et al., 1998). Spontaneous mappings were found not only between auditory pitch and object location, but between auditory pitch and object size as well (Evans & Treisman, 2010). In addition to auditory pitch, sound loudness can help convey visual information as well. It was found that loud sounds facilitate the perception of large objects, while soft sounds can improve the perception of small ones (Marks, 1987; Smith & Sera, 1992). Researches have also been conducted to convey live video through auditory pitch and loudness (Meijer, 1992). In most auditory-visual substitution systems, only grayscale images are utilized and color information is not conveyed. Recently in 2012, a new sensory substitution system “EyeMusic” was released. It can not only represent real-time visual information through small computer or smartphone with stereo headphones, but can represent color information through different musical instruments as well. Due to the limitation of differentiating among different musical instruments, only five colors are conveyed: white, blue, red, green and yellow (Sami Abboud, 2014).

Trained early blind participants showed increased performance in localization and object recognition (Arno et al., 1999) through auditory-visual substitution. However, auditory-vision substitution requires the memorization of different audio forms and extensive

training is required to map from different audio stimulus to visual cues (Arno et al., 1999). In addition, the focus on auditory feedback can decrease subjects' ability to get information from the environment (Meers & Ward, 2005).

2.2 Image Processing

Current image processing methods have already provided various ways to make images easy to perceive by visually impaired people due to the limitation of auditory and tactual sensory modality compared with visual system (Way & Barner, 1997). According to different goals and types of images, the choice of images processing techniques differs for different applications and research. For image enhancement and simplification for visually impaired people, color image to grayscale, edge detection and texture analysis are several commonly applied image processing techniques (Rastogi & Pawluk, 2013; Way & Barner, 1997). These image processing techniques are utilized in this thesis research and discussed in following sections.

2.2.1 Color to Grayscale

Since most of current assistive technologies convey intensity information to blind or visually impaired people, intensity can be considered as the most important element for image processing techniques. Converting a color image to a grayscale one is a basic step to convey visual information to BVI persons (Ikei, Wakamatsu, & Fukuda, 1997; Way & Barner, 1997). Edge detection algorithms calculate significant changes in intensity between nearby pixels. Textures are recognized by different placement and repeat of intensity values.

A color image I of width w and height h can be represented as a three-dimensional array of size $w \times h \times 3$, where each of the three dimensions represents a different color channel: red, green and blue. Figure 2.5 shows an example of the three channels and the converted grayscale one of a color image. More formally, the values in each channel are represented as R_{mn} , G_{mn} and B_{mn} , where $1 \leq m \leq w$ and $1 \leq n \leq h$. To convert a color image to grayscale, a common strategy is to calculate the weighted sums of each pixel's RGB values and have that value represent the grayscale equivalent quantity. This conversion is described as

$$I_g = \mu \begin{bmatrix} R_{mn} \\ G_{mn} \\ B_{mn} \end{bmatrix} \quad (2.1)$$

where $1 \leq m \leq w$ and $1 \leq n \leq h$.

I_g is the grayscale image and μ is the weighting coefficient, which is a 1×3 array. Different algorithms lead to different weighting coefficients. There are two most popular weighting coefficients currently applied. One is to take the average of all R, G and B channels. This average-method does not show accurate results as human visual system. Since the cone density in human eyes is not uniform across colors, human eyes are more sensitive to green light, followed by red and blue. Therefore, to correct for human visual system, a method normally named as "luminance" is introduced. This luminance-method takes green channel as the most important factor, so that the G channel has the largest weighting. MATLAB utilize this luminance method and the value of μ (and commonly adopted) is [0.2989, 0.5870, 0.1140] as an example.



Figure 2.5 Examples of color image to grayscale conversion. (a) Original image; (b) Converted grayscale image; (c) Red channel; (d) Green channel; (e) Blue channel.

2.2.2 Edge Detection

Edge detection can give a clear impression of the shape and size of objects in an image. The representation of edges in an image plays an important role in image navigation for the visually impaired or blind. Specially, it alleviates the load in what respects to distinguishing objects and background, and tracing shapes and sizes.

Edge detection is a process that can locate and highlight sharp discontinuities in pixel intensity, which represent boundaries of objects in an image. Classical strategies of edge detection algorithms apply a 2D mask throughout the image (a process called convolution) (Heath, Sarkar, Sanocki, & Bowyer, 1997). This 2D mask is also called an operator, which is sensitive to large gradients in an image while ignoring area of similar pixel intensities (G & S, 2011). Current researches have provided us a variety of operators that performs well for different types of edges and images. Comparisons and surveys are also published as guidance on how to choose the algorithm that fits best different applications

(Davis, 1975; Peli & Malah, 1982). Sobel, Prewitt, Roberts, Laplacian of Gaussian, Zero-Cross and Canny are several popular edge detection operators that have been implemented by various programming languages. Classical operators, such as Sobel and Prewitt, and Zero-Cross which is based on second directional derivative of an image are simple and fast; however, these operators are sensitive to noise and inaccurate when images become complicated. Some Gaussian operators, like Canny, perform better when facing noise in images and provide more accurate and localized detection results. However, it is time consuming and of relatively high computational complexity (G & S, 2011). Figure 2.1 below shows a visual comparison between these edge detection algorithms.

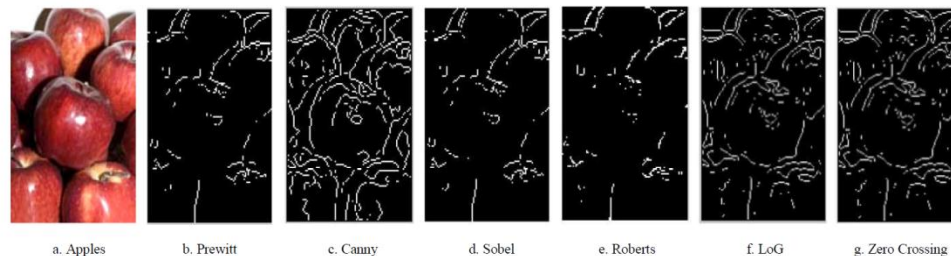


Figure 2.6 Results of different edge detection algorithms

From Figure 2.6 we can observe that Canny algorithm shows the best result. This is because the research on this algorithm followed three criteria to improve the performance of edge detection algorithm in his times. The first criterion is “Good Detection”, which means high hit rate and low error rate. The second criterion “Good localization” aims to mark the edge points as close as possible to the true edge. And the third criterion is “Only one response to a single edge” (Canny, 1986).

2.2.3 Texture Analysis

Image texture in this thesis refers to the texture of an object from the image processing stand point, and not necessary from the perceptual point of view. The texture discussed in this thesis represents spatial arrangement of color and intensity in images or particular regions in an image (Stockman & Shapiro, 2001). It can be considered as one of the main features that can characterize an image or an object in the image (Wechsler, 1980). Therefore, image texture is always utilized to help in segmentation and classification of images. For instance, texture analysis has been used to segment different information on a page layout from text regions to non-text regions (Jain & Zhong, 1996). Text regions have specific textures since they follow a unique spatial arrangement rule that each text lines are of the same orientation and the same spacing between them. Texture analysis has also been adopted to facilitate mountain vegetation mapping (Dobrowski, Safford, Cheng, & Ustin, 2008) and map construction. From satellite photos (see Figure 2.7), terrain of different vegetation communities represents different textures. Since texture analysis has been proved succeeded in classify different objects and regions in an image, various applications are studied utilizing texture analysis to help blind or visually impaired people distinguish different objects. A sonar aid was developed to help blind people navigate the environment providing feedbacks of objects surface textures (Kay, 1974). Text detection from natural scene images using texture analysis approaches were also studied to help blind or visually impaired people read print materials (Ezaki, Bulacu, & Schomaker, 2004; Hanif & Prevost, n.d.). A wearable real-time vision substitution system that utilized texture analysis to filter important environment elements was also

studied to help blind people travel (Balakrishnan, Sainarayanan, Nagarajan, & Yaacob, n.d.).

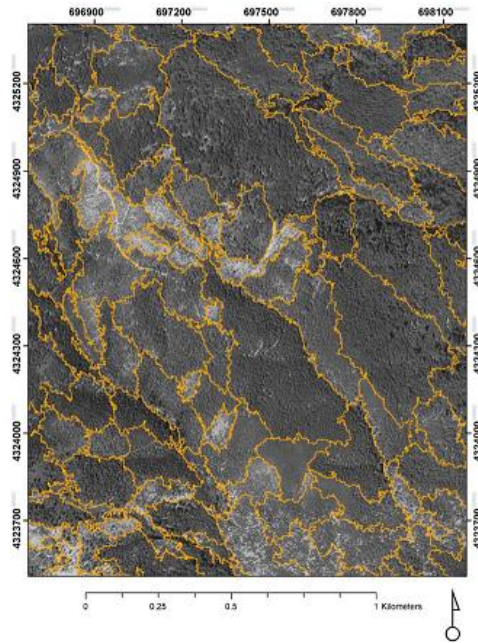


Figure 2.7 Vegetation communities derived from automated segmentation based on image texture analysis

There are two different approaches, statistical and structural approaches, developed to accomplish texture analysis tasks since texture can be defined by two ways. One definition describes texture as images shown stochastic structure. The other definition characterizes texture as patterns that show repeated manner over a region of the image. Besides statistical and structural, impressionistic and deliberate are two other names that used to describe these two approaches as well (Lipkin, 1970).

2.2.3.1 Statistical Texture Analysis Approach

Statistical texture analysis approaches take image textures as quantitative measurements that can be computed through analysis of intensity and color relationships between pixels. One statistical method is called first-order statistics. It calculates the gray level differences between image pixels and estimates the probability density for these differences (Haralick, Shanmugam, & Dinstein, 1973). The other statistical method is called second-order statistics, which is the most used statistical method for texture classification (Sutton & Hall, 1972). It is known as co-occurrence matrices as well. Two parameters are used in this co-occurrence matrices method, a distance and an angle. It discovers the spatial relations between similar gray levels. After the first and second order statistics are computed, several features can be extracted to characterize one texture. Mean, variance, coarseness, skewness and kurtosis are common measurements applied. Besides these two approaches, Fourier analysis has also been utilized to investigate textures since Fourier transformation deals with frequency domain (Lendaris & Stanley, 1970). Experiments have been conducted to test the performance of Fourier analysis (Bajcsy, 1973). The experimental results show that it can provide global information but shows weakness in analyzing local information (Wechsler, 1980). However, Fourier analysis is computational expensive and problem arises when it deals with non-square region. In 1975, Mary M. Galloway introduced a new way to analysis and to classify image textures, which is called gray level run lengths approach. This method first finds connected pixels of the same gray level and then use the lengths of those connected pixels and the distribution of the lengths as measurements to characterize an image texture (Galloway, 1975). More features were introduced in later studies for the run

lengths method. For instance, the gray value distribution of the runs and the percentage for runs of same length are two popular studied features (Chu, Sehgal, & Greenleaf, 1990). In a review paper that compares the four texture analysis methods mentioned above, it states that the co-occurrence matrices method performs the best, followed by first-order statistics, Fourier analysis and gray level run length method (Connors & Harlow, 1980).

2.2.3.2 Structural Texture Analysis Approach

Different from statistical approach, structural texture analysis approach normally deal with artificial image textures (see Figure 2.8). Structural approach assumes that textures are consist of a set of primitive units that can be easily identified (Wechsler, 1980). According to Fumiaki Tomita and Saburo Tsuji, structural texture analysis consists of extraction of texture elements (or primitives), shape analysis of texture elements and estimation of placement rule of texture elements (Tomita & Tsuji, 1990). The texture elements are defined by a simple shape region of uniform grey level. After these elements are extracted, brightness, area, size, directionality and curvature are computed as properties of an element. Classification of these primitives can then be performed according to the properties computed.

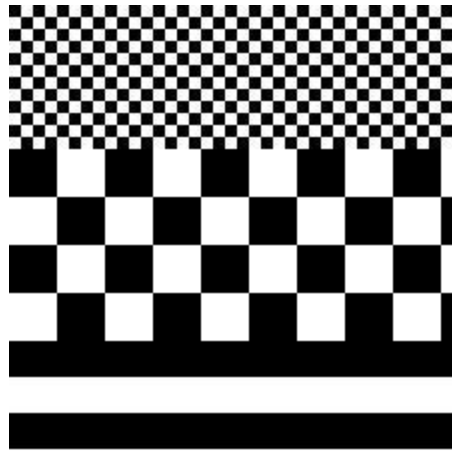


Figure 2.8 Artificial image texture examples.

Structural texture analysis approach is not widely utilized as statistical method. Identifying the primitives in a right manner is not an easy case if it does not deal with artificial image textures or simple textures. Also, the investigation of possible placement rule of texture primitives is still a challenge (Wechsler, 1980).

2.3 Multimodal Sensory Interpretation

Multimodal sensory interpretation involves the integration of multiple sensory signals to convey or retrieve information. Although multimodal sensory integration did not gain much attention until the twentieth century as an area of academic study, it can be discovered in all aspects of human's life (Kress, 2009). In one example, body language (e.g. prosodic gestures) have meaning when are accompanied by speech and facial expressions. Gestures also are used to emphasize verbal content. In the context of the blind or visually impaired persons, multimodality has been adopted in a number of systems (Lécuyer et al., 2003; W. Yu et al., 2002; Wai Yu, McAllister, Murphy, Kuber, & Strain, n.d.) to convey visual information through alternative channels/modalities, such

as integrating both tactual and auditory sensory that are functional to the subject, and therefore it provides an effective form of interaction.

Multimodal sensory substitution approaches have been recognized as the most effective way to surmount the obstacles that visually impaired people may meet when accessing image information. Multimodal sensory substitution methods are reported to have the ability to maximize the benefits of each singular modality through the interaction of each modality and enhance the accessibility (Jacko et al., 2003; Wai Yu et al., n.d.). The classic drag-and-drop tasks were tested using multimodal feedbacks from auditory, tactual and haptics. The experiment results indicated significant performance over single visual feedback for both visually impaired and sighted users.

A *Multimodal System* can be defined as a system that integrates multiple human sensory modalities, such as visual, auditory and haptic/kinesthetic signals (Blattner & Glinert, 1996). These modalities can be taken as both input (control forms) and output (feedback). Control a system using speech, hand gestures and eye gazes can be examples of input usage. An example of an input multimodal system is given by Koons et al. In that system, speech, gesture and gaze, are complemented to evoke an action. The spoken command “Move the blue circle there” is recognized through a discrete word recognition system and the direction “there” is defined by a pointing gesture using a hand data glove, and gaze direction using eye tracker (Koons, Sparrell, & Thorisson, 1993).

For BVI persons, multiple modalities have been used as system output that convey information to users. For visually impaired or blind people browsing the web or graphs’ exploration, a multimodal system that integrated both audio and haptic sensory feedback was developed to express visual information in real-time (W. Yu et al., 2002; Wai Yu et

al., n.d.). In this system, most participants used haptic feedback as navigation guide, while auditory feedback were used to provide information of a certain object, such as bars height in a bar chart. Another multimodal system was developed to help BVI people explore and navigate in virtual environment, named as HOMERE (Lécuyer et al., 2003). It integrated force feedback from a virtual blind cane, a thermal feedback simulating a virtual sun and an auditory feedback according to specific events in the environment. Figure 2.9 shows the experimental environment of this system. Besides the integration of auditory and haptics feedback, tactual feedback is also integrated with auditory to help BVI users browse graphical information, such as diagrams and pie charts (Wall & Brewster, 2006). This system is called Tac-tiles. It provided tactual feedback on fingertips by a dynamic tactile pin-array, with auditory feedback through speech or non-speech audio cues.



Figure 2.9 HOMERE multimodal system for virtual environment navigation for persons blind or visually impaired.

2.4 Bayesian Network

To construct the inference relation between different image features in this thesis, a Bayesian network is employed. This section introduces the basic concepts of Bayesian network, the methods to construct Bayesian networks, and how it has been applied in related research areas.

A Bayesian network, also known as a belief network, is a type of statistical model that describes the probabilistic dependencies between a set of variables (Heckerman, Geiger, & Chickering, 1995). Figure 2.10 shows a simple Bayesian network.

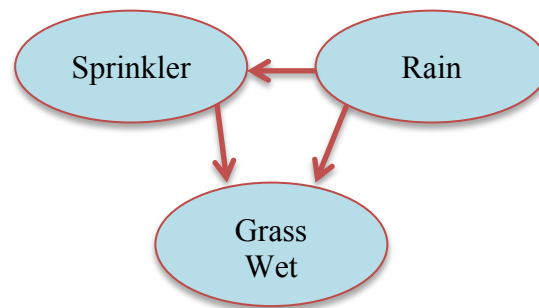


Figure 2.10 A simple Bayesian network.

In Figure 2.10, there are three events: sprinkler, rain and grass wet. Assume both the sprinkler and rain can cause grass to get wet, and the rain can cause the sprinkler. In a Bayesian network, these three events can be regarded as three variables. These three variables are of Boolean type, either true or false with a probability distribution. Therefore, the joint probability of this model can be formulated as:

$$P(G, S, R) = P(G | S, R) P(S | R) P(R) \quad (2.2)$$

where G represents grass wet, S represents sprinkler and R represents rain. The probability that one variable causes the other variable can then be easily calculated.

Bayesian network can also be denoted as a directed acyclic graph (DAG) that has a conditional probability distribution (CPD) with it, P . Variables are named as nodes in a graph and the dependency relationships are denoted as directed edges. Therefore, a Bayesian network can be denoted as $B = (G, P)$ where G represents the graph and P represents the conditional probability distribution.

To construct a Bayesian network or Bayesian model, a DAG is first developed. To construct the DAG of Bayesian structure, there are two methods that are most adopted. The Expert-based modeling method generates the Bayesian structure by human experts. Current literature and experts' experience and opinions are taken into account to construct the model (Yu-Ting Li & Juan P. Wachs, 2014a). The other approach to construct a Bayesian structure utilizes Genetic Algorithms (GA). Several structures are first generated randomly. The best structure is then selected with Genetic Algorithms and observation data. Once the Bayesian structure is constructed, the probability of each variable can be calculated with observation data. The states of certain variables can then be inferred when evidence variables are observed. This process of calculating probability distribution of variables based on observed evidence is called probabilistic inference (Heckerman, 2008).

Bayesian network has been utilized to finish various tasks. It was used as a prediction model to estimate the maintainability for object-oriented systems (van Koten & Gray, 2006). Bayesian network shows higher accuracy when compared with commonly used regression-based models. Bayesian network has also been applied to speech recognition. By using Bayesian network, long-term articulatory and acoustic context can be explicitly represented with the cooperation of hidden Markov models (HMMs) (Zweig & Russell,

1998). It can analyze different factors together based on the conditional probabilistic dependencies and reveal uncertainty related to any strategy proposed. Bayesian network has also contributed to biology related field, such as modeling gene regulatory networks, protein structure and analyzing gene expression (N. Friedman, Linial, Nachman, & Pe'er, 2000).

Bayesian network has also been applied in related image feature extraction area. A dynamic Bayesian network was generated to perform autonomous 3D model reconstruction task from single 2D image (Delage, Lee, & Ng, 2006). Object detection was achieved by constructing dependency relations among different image features through a Bayesian network (Schneiderman, 2004). Bayesian networks have also been applied for semantic image understanding and image interpretation by constructing a probability distribution function among various image and object features (Kumar & Desai, 1996; Luo, Savakis, & Singhal, 2005). Pertaining to this thesis work, Bayesian network has been studied to model the relationships between basic visual features in an image. Bayesian network is utilized to infer the causal relations between identity and the position of features in visual scenes (Chikkerur, Serre, Tan, & Poggio, 2010). Chikkerur's study indicates that spatial information can reduce the uncertainty in shape information.

2.5 Linear Assignment Problem

Linear Assignment Problem (LAP) is utilized in this thesis to model the relationship between image features and sensory modalities. This section introduces the concepts

involved in LAP and various algorithms that solved this problem. Applications of this problem are also discussed.

The Linear Assignment Problem (LAP) (Munkres, 1957) is considered as one of the most basic optimization problems in operation research and combinatorial optimization (Burkard, Dell'Amico, & Martello, 2009). The goal is to find out the maximum or minimum weight matching in a weighted bi-graph. Bi-graph is a graph that the vertices in it can be divided into two independent sets, U and V , and edges only connect a vertex in U to one vertex in V . There are no edges inside one set (see Figure 2.11). A bi-graph can be denoted as $G=(U,V;E)$, where G represents the graph, U and V represents two sets of vertices and E denotes the edges in this graph.

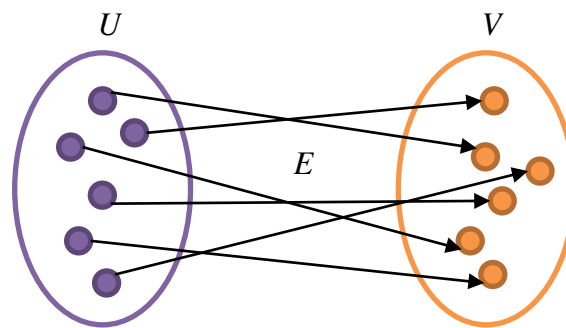


Figure 2.11 Linear assignment problem bi-graph.

A classical scenario of linear assignment problem is that there are n men and n jobs and each man's completion time on each job are given. The objective of this problem is to find out the optimal assignment of men to jobs which makes the total completion time for all jobs a minimum.

To construct a mathematical model for the assignment problem, a cost matrix, $C=(c_{ij})$, should first be defined as a $n \times n$ matrix that the cost of assign row i to column j is c_{ij} .

Then the linear assignment problem is described as an assignment whereas each row i is assigned to one column j in matrix $C=(c_{ij})$, in a way that the total cost can be minimized.

Let a binary matrix $X=(x_{ij})$ such that

$$x_{ij} = \begin{cases} 1, & \text{if there is assignment from } i \text{ to } j \\ 0, & \text{otherwise} \end{cases} \quad (2.3)$$

Then, the linear assignment problem is defined as

$$\min \sum_{i=1}^n \sum_{j=1}^n c_{ij}x_{ij} \quad (2.4)$$

And

$$\sum_{i=1}^n x_{ij} = 1 \quad (2.5)$$

$$\sum_{j=1}^n x_{ij} = 1 \quad (2.6)$$

To solve this problem, many algorithms have been developed, starting with Easterfield in 1946. In 1952, Votaw and Orden first named this problem as the “assignment problem” (Burkard et al., 2009). The “Hungarian algorithm” is one of the most well-known combinatorial optimization algorithms that can solve the LAP in time complexity of $O(n^4)$. The first computer code for solving the linear assignment problem was published based on the Munkres algorithm (Silver, 1960) in 1960. Later in the 1960s, the first $O(n^3)$ algorithm, which is also the best time complexity one, was developed by (Dinic & Kronrod, 1969). Other $O(n^3)$ algorithms were studied in the following years as well, such as shortest path computations on reduced costs (Edmonds & Karp, 1972; Tomizawa, 1971) and primal simplex algorithm (Akgül, 1993).

Applications of linear assignment problem are studied across various areas in real world. Shortest routes were computed to make phones efficiently communicating with multiple satellites and ground stations by solving a linear assignment problem (Burkard, 1986). LAP was also applied to determine entry and exit terminals setting in transportation center, like train station and airport, to minimize the density of routing (Burkard, 1986).

CHAPTER 3. METHODOLOGY

This section discusses the main methods used to determine the best mapping between image features and feedback modalities in order to explore histology images using multimodal sensory substitution. The main components of the system architecture of the multimodal navigation system are presented in Figure 3.1. Histology images of blood smears are considered as the input of the entire system. Seven features are extracted to describe this input histology image. These features are classified into two groups, primary features (see Module (a)) and peripheral features (see Module (b)). The purple arrows between these features indicate “cause-effect” relationship, or what the evidence is and what is inferred from this evidence through a Bayesian network. After extracting the primary features from the input image (see Module (a)), the output of the system is the tangible expression of the extracted features through different modalities (see Module (c)). These modalities, in turn, are assigned to specific devices (e.g. haptic device) used to manipulate and explore the image (using the Linear Assignment Problem (LAP)). The orange arrows in the picture express one possible assignment, which is not the final assignment. The two key components, image feature Bayesian network and modality assignment problem, in the system architecture are then well-illustrated in the following sections.

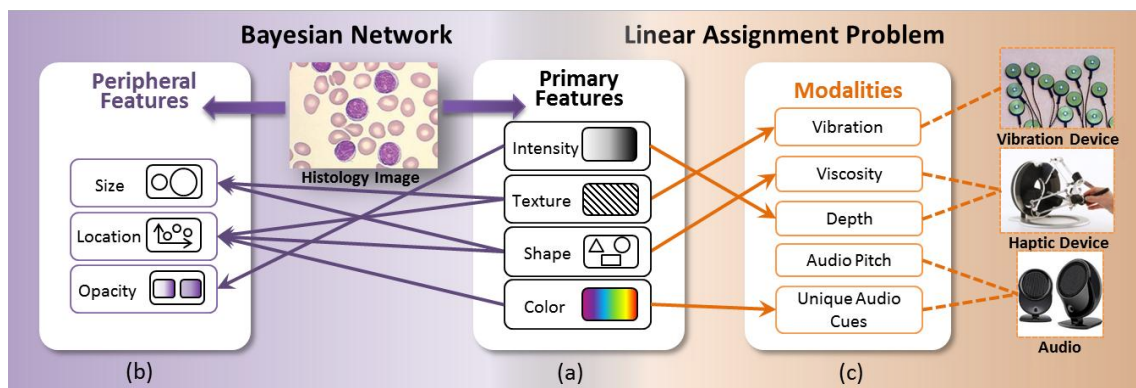


Figure 3.1 System Architecture: Module (a) Primary Image Features; Module (b) Peripheral Image Features; Module (c) Sensory Modalities.

3.1 Image Feature Bayesian Network

The feature of interest for an image varies depending not only on the theme of an image, but the function of image as well. In this research, the main focus is on histology images as educational tools, and those are utilized as test images in the various experiments conducted. For histology images, features of interest may be focused on those features that can characterize a cell and differentiate it from other kinds of cells. More specifically, seven features were used to encapsulate the content of histology images in a compact manner. The objects' *location*, *intensity*, *texture*, *shape*, *color*, *size* and *opacity* are the key perceptual information that was found necessary for blind or visually impaired people to understand histology images. This is supported by previous research conducted in the area of perception of the blind (Chaudhuri, Rodenacker, & Burger, 1988; Cruz-Roa, Caicedo, & González, 2011). These seven features were classified into two groups: primary and peripheral. Intensity, texture, shape and color are categorized as primary features that can be directly mapped to specific modalities, while location, size and opacity are classified as peripheral features since they can only be acquired through

experience, or inferred through the frequency of occurrence of primary features (Howard, 1958).

3.1.1 Primary Feature Extraction

Primary features are extracted from images using image processing algorithms, such as color image to grayscale conversion, texture analysis and edge detection. These features are discussed in the following subsections.

3.1.1.1 Intensity

Intensity information represents the brightness in an image. It can be considered as the most important feature of an image to persons blind or visually impaired. Most of current assistive technologies help the BVIs to see images through the delivery of image intensities (Lescal, Rouat, & Voix, 2013; Marks, 1987; Meijer, 1992).

Computing the combined intensity of each pixel is a basic operation that requires converting a color image into a grayscale one. Intensity of each pixel is computed through the summation of weighted RGB values, as in

$$Intensity = 0.2989 * R + 0.5870 * G + 0.1140 * B \quad (3.1)$$

where R, G and B represents the value in red, green and blue channels of a pixel, respectively, in RGB color space. Green channel takes more weight since human eyes are more sensitive to green lights. One example is shown in figure 3.2.

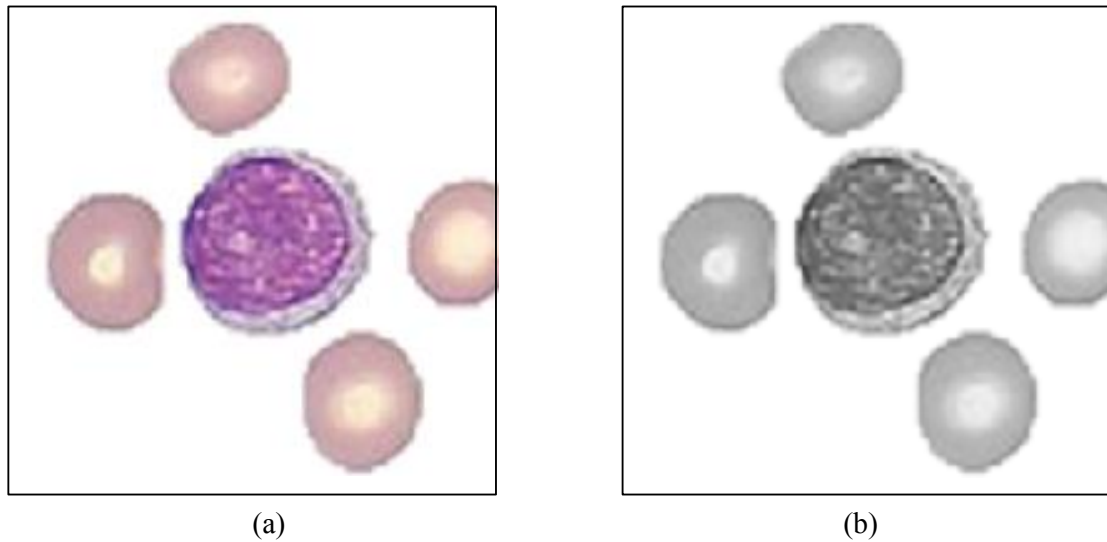


Figure 3.2 Intensity computed from color information. (a) Color image (b) Grayscale image of (a).

3.1.1.2 Texture

Texture in this thesis refers to the texture of an object from the image processing stand point, and not necessary from the perceptual point of view. Namely, texture here refers to the spatial arrangement of intensities displayed by an object in an image. Discrimination among object's textures enabled BVI persons to perceive difference between classes of objects. Figure 3.3 shows an example for objects with different textures. Figure 3.3 (a) is a blood smear image showing both white and red blood cells. From the image, it can be observed that all red blood cells (see Figure 3.3 (b)) show a uniform texture which is light in the center and dark in the periphery. However, white blood cell (see Figure 3.3 (c)) shows distinct textures from red blood cells due to striking intracellular structures, most notably the dark purple cell nucleus at the center.

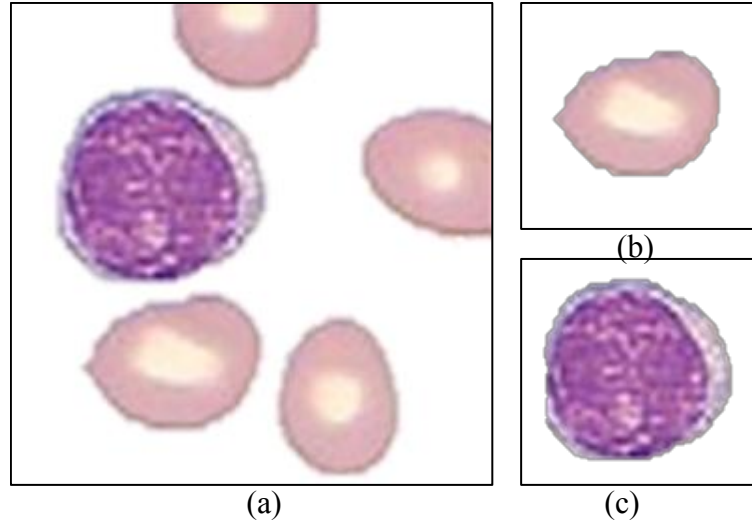


Figure 3.3 Different textures compared between red blood cells (b) and white blood cells (c) in a blood smear image (a).

Discrimination between objects' textures is done using gray-level co-occurrence matrices (Davis, Johns, & Aggarwal, 1979). To define a gray-level co-occurrence matrix $P[i,j]$, a displacement vector $d=(dx,dy)$ should first be defined where the entries dx , dy correspond to displacement unit in x and y direction. The values $P(i,j)$ are obtained by counting the number of pairs of pixels that displaced by d having gray values i and j . After the matrix is generated, there are several statistics that are used to characterize texture (Clausi, 2002). For example, *entropy* indicates the randomness of gray-level distribution, which is defined as:

$$Entropy = -\sum_i \sum_j P[i,j] \log P[i,j]. \quad (3.2)$$

Other texture related features are *energy*, *contrast* and *homogeneity*, and are defined as follows:

$$Energy = \sum_i \sum_j P^2[i,j] \quad (3.3)$$

$$Contrast = \sum_i \sum_j (i - j)^2 P[i, j] \quad (3.4)$$

$$Homogeneity = \sum_i \sum_j \frac{P[i, j]}{1 + |i - j|} \quad (3.5)$$

All objects in the image are identified using these four statistics. The distance of these statistics between different objects is then computed to find out similarity between different textures. The distance metric is defined as:

$$d_{ij} = (\eta_i - \eta_j)^2 + (E_i - E_j)^2 + (C_i - C_j)^2 + (H_i - H_j)^2 \quad (3.6)$$

where d_{ij} is the statistics distance between object i and j . η denotes Entropy, E represents Energy, C is Contrast and H stands for Homogeneity.

A threshold is set to distinguish between similar and different textures. In our work, this threshold was determined empirically.

3.1.1.3 Shape

Shape is another crucial image feature for BVI persons to enable greater understanding of the salient characteristics of an object. For histology, shape of cells can help distinguish different cell types. Figure 3.4 shows an example of normal red blood cells and sickle cells. Normal red blood cells (Figure 3.4 (a)) are round shaped, while sickle cells (Figure 3.4 (b)) are flat arcs and show distinct differences in both directions.

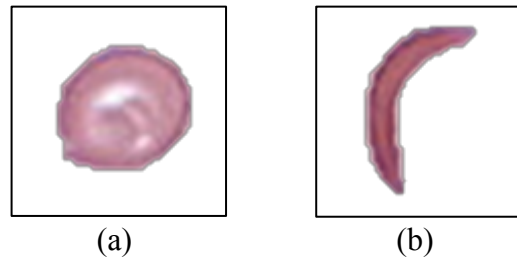


Figure 3.4 Shape difference between normal red blood cells and sickle cells.

To characterize the shape or boundary of an object in a color image, a conversion from color to grayscale is first performed, and then the Canny edge detection algorithm (Canny, 1986) was utilized. Finally chain code is used to represent the shape in a compact fashion.

Figure 3.5 shows the edge detection results of tested blood smear images.

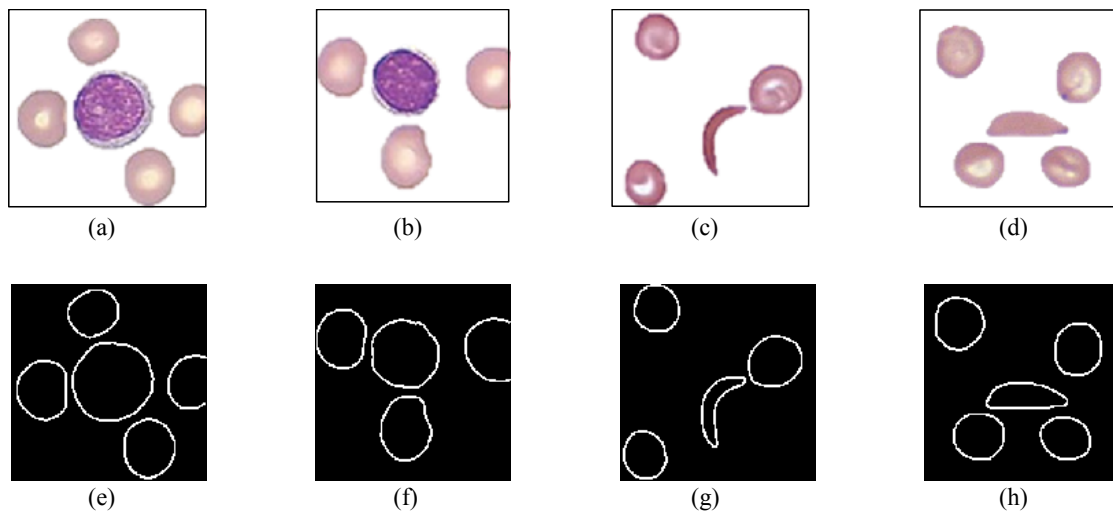


Figure 3.5 Edge detection results for 4 tested blood smear images. (a) ~ (d): Original images; (e) ~ (h) corresponding detected edges.

3.1.1.4 Color

Color information is obtained by brightness normalized RGB values of pixels. Given a color of (R, G, B) , where R , G and B represents the intensity of Red, Green and Blue channels. Normalized RGB values are then computed by following equations:

$$r = \frac{R}{R+G+B} \quad (3.7)$$

$$g = \frac{G}{R+G+B} \quad (3.8)$$

$$b = \frac{B}{R+G+B} \quad (3.9)$$

where r , g and b denotes normalized R , G and B values, which can be also interpreted as the proportion of red, green and blue in the color.

The use of normalized RGB values removes the effect of any intensity variance, which also facilitates the mapping to other sensory modalities.

3.1.2 Peripheral Feature Extraction

Since some image features cannot be properly conveyed by certain or only one sensory modality, extracted image features from histology images are classified into two groups, primary features and peripheral features. Primary features are mapped to other sensory modalities, while peripheral features are inferred from primary features through a Bayesian network. The Bayesian network is generated to infer the probability of the peripheral features based on evidences exhibited by the occurrence and the amount of the primary features. The construction of Bayesian network is a three-step approach. In the first step, expert-based modeling is used to generate fundamental structures of this network. Several candidate structures were generated by human experts during this step.

Since each link in the network associates with a conditional probability of inferring a child node from a parent, then in the second step, a probability function is applied to calculate the probability of each link from observations obtained through experiments. At last, a Bayesian scoring function was applied to find out the optimal structure between candidate structures generated by human experts (Yu-Ting Li & Juan P. Wachs, 2014a, 2014b).

3.1.2.1 Expert-based Modeling

In the first step, Expert-based modeling is used to generate fundamental structures of this network. Our expert is a postdoctoral researcher who is blind with no functional sight and a doctorate degree in chemistry. He is also Braille literate with extensive experience using assistive technologies for the BVI community. In this Bayesian network, there are seven nodes where each node represents the perception of an image feature. All of these nodes are of type Boolean, which in this case means whether a certain feature is perceived or not. The definition and states of each node are summarized in Table 3.1.

Table 3.1 Definition of Discrete States for Each Node

Node	Description (Perception of feature)	States
n_1	Intensity	{ <i>True, False</i> }
n_2	Texture	{ <i>True, False</i> }
n_3	Shape	{ <i>True, False</i> }
n_4	Color	{ <i>True, False</i> }
n_5	Size	{ <i>True, False</i> }
n_6	Location	{ <i>True, False</i> }
n_7	Opacity	{ <i>True, False</i> }

Three candidate structures (shown in Figure 3.2) are generated by the BVI subject during this step. A blind scientist was recruited to link this two groups of features based on his

own experience and current literatures. The blind scientist was first presented with these seven features classified in two different groups. Then questions were asked that given one feature from the primary features group, what features from peripheral group may be perceived. By asking these questions, the blind scientist was required to generate any possible combinations of links between these two groups of features. In this thesis, the blind scientist came up with three structures shown below.

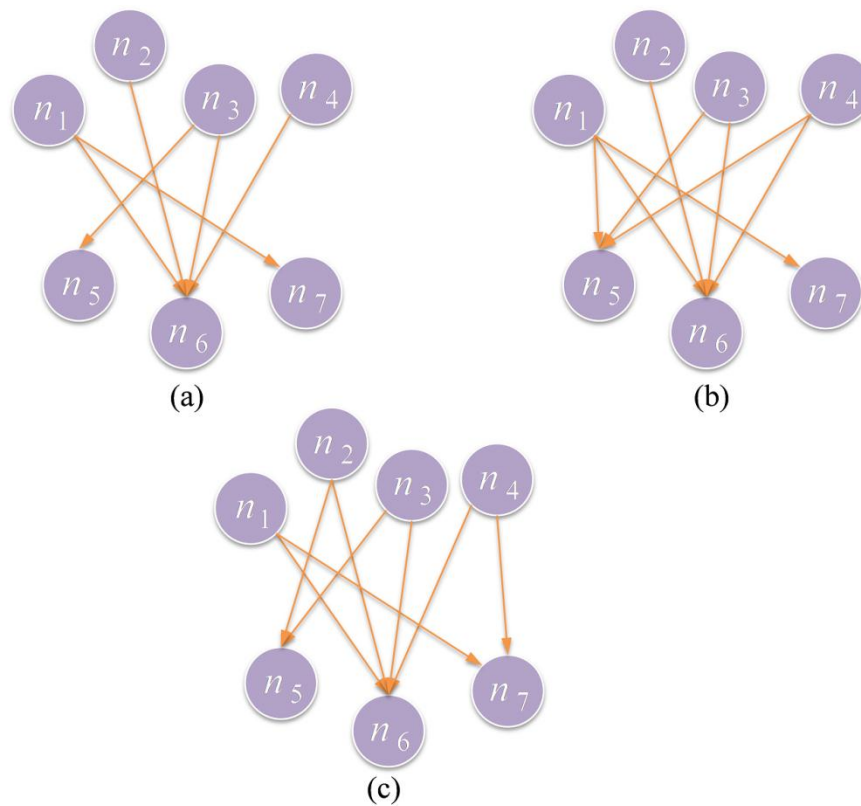


Figure 3.6 Candidate Bayesian structures generated by typical user.

3.1.2.2 Probability Calculation

Each link in the network is associated with a conditional probability of inferring a child node from a parent. The probability function to calculate the probability of each link is

computed from observations, which in turn are obtained through experiments. The probability function is defined as:

$$P(n_i | n_j) = \frac{P(n_i, n_j)}{P(n_j)} = \frac{N(n_i = 1, n_j = 1)}{N(n_j = 1)} \quad (3.10)$$

where $N(x)$ counts the number of observations that satisfied condition x .

3.1.2.3 Bayesian Network Optimization

At last, to find out the optimal structure between candidate structures generated by human experts, a Bayesian scoring function (Nir Friedman, 1997) is defined:

$$score(D, G) = P(D | G) = \prod_{i=1}^N \prod_{j=1}^{q_i} \frac{\Gamma(N_{ij})}{\Gamma(N_{ij} + M_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(a_{ijk} + s_{ijk})}{\Gamma(a_{ijk})} \quad (3.11)$$

where

$$N_{ij} = \sum_{k=1}^{r_i} a_{ijk} \quad (3.12)$$

$$M_{ij} = \sum_{k=1}^{r_i} s_{ijk} \quad (3.13)$$

D represents the observation dataset obtained through experiments, G represents the candidate Bayesian structure and N is the number of nodes in the network. q_i is the number of possible values of node i 's predecessors; r_i is the number of different values of node i ; a_{ijk} is the parameter of a Bayesian network with Dirichlet distribution; s_{ijk} is the number of tuples in the dataset where node i is equal to k and its predecessors are in j th instantiation. Γ denotes the gamma function. The probability density function of Dirichlet distribution can be represented as:

$$\frac{1}{B(\alpha)} \prod_{i=1}^k x_i^{\alpha_i - 1} \quad (3.14)$$

where $\alpha = (\alpha_1, \dots, \alpha_k)$ and

$$B(\alpha) = \frac{\prod_{i=1}^k \Gamma(\alpha_i)}{\Gamma\left(\sum_{i=1}^k \alpha_i\right)} \quad (3.15)$$

The optimal candidate Bayesian structure is the one that maximizes equation 3.11. Figure 3.7 shows the optimal structure with conditional probabilities defined in equation 3.10.

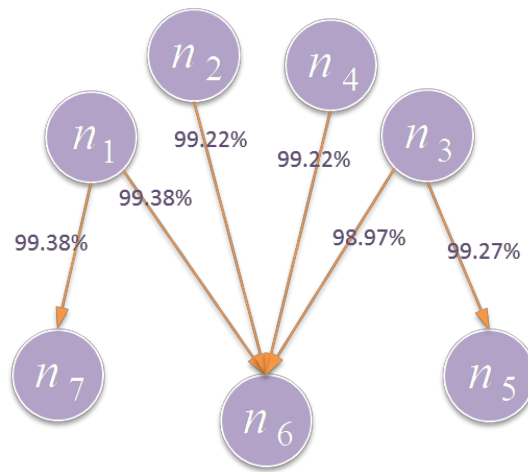


Figure 3.7 Optimal Bayesian structure.

3.2 Modality Assignment Problem

To express the four primary features discussed in Chapter 3.1.1 to BVI persons, five sensory modalities (vibration, viscosity, depth, audio pitch and unique audio cues) were selected in this study taking different manifestations for each sensory type. Since only one modality can be used to represent one feature, the mapping problem between primary features and modalities can be considered as a linear assignment problem (see Figure 3.1 Module (b) and Module (c)). There will be a particular cost for mapping one feature to one modality; therefore, the optimum mapping combination can be generated by finding

the match that has the minimum total cost. Also, not all modalities need to be candidates for each feature because some modalities may not be applicable to a certain feature in terms of its property and the modality's property. Table 3.2 shows the candidate modalities for each feature.

Table 3.2 Candidate Modalities for Each Feature

Feature/ Modality	M1: Vibration	M2: Viscosity	M3: Depth	M4: Audio Pitch	M5: Unique Audio Cues
F1: Intensity	✓	✓	✓	✓	
F2: Texture	✓	✓	✓	✓	
F3: Shape	✓	✓	✓	✓	
F4: Color	✓				✓

3.2.1 Problem Definition

The formal definition of this assignment problem is: Given two sets, F , represents primary features of size 4 and M , denotes modalities of size 5, together with a cost function $C: F \times M \rightarrow \mathfrak{R}$. Find a bijection function $g: F \rightarrow M$ such that the cost function is minimized:

$$\min \sum_{i \in F} \sum_{j \in M} C(i, j) x_{ij} \quad (3.16)$$

subject to the constraints:

$$\sum_{j \in M} x_{ij} = 1 \text{ for } i \in F, \quad (3.17)$$

$$\sum_{i \in F} x_{ij} = 1 \text{ for } j \in M. \quad (3.18)$$

Variable x_{ij} denotes the assignment of feature i to modality j , taking value 1 if there is an assignment and 0 otherwise. According to Table 3.2, the cost of no assignment between i and j represented in the cost matrix $C(i, j)$ are set to be infinity.

3.2.2 Cost Weighing

After the objective function is defined, the individual cost c_{ij} in the cost matrix $C(i,j)$ for conveying image feature i through sensory modality j , need to be computed. In this thesis, the individual cost is calculated through data obtained from empirical experiments. Human subjects were recruited to perform the same tasks using feedbacks from all candidate modalities. The human performance was evaluated through response time and error rate. And then for each feature, all the candidate sensory modalities were ranked by their performance. The modality that showed higher performance has a higher ranking. The ranking of the modality was then used as individual cost in the cost matrix. For example, if the ranking of candidate modalities for feature intensity is: vibration > audio pitch > viscosity > depth, then the costs of mapping intensity to these modalities would be: vibration of cost 1, audio pitch of cost 2, viscosity of cost 3 and depth of cost 4. The following expression shows what the matrix looks like:

$$C(1, j) = [1 \ 3 \ 4 \ 2 \ \infty] \quad (3.19)$$

Since the objective is to minimize the total cost, the smaller digit indicates higher ranking and better option.

3.2.3 Linear Assignment Algorithm

An extension of Munkres or Hungarian Algorithm (Bourgeois & Lassalle, 1971) is applied to solve this problem with a rectangle cost matrix since the number of features is different from the number of available modalities. Besides adding lines of zero elements to the rectangle cost matrix to make it a square one and then apply the Munkres algorithm,

this study takes advantage of Bourgeois and Lassalle's work and can be described as follows.

First of all, there are several preliminary steps before real algorithm begins. k is defined to represent the minimum number of rows and columns.

$$k = \min(n, m) \quad (3.20)$$

where n is the number of rows and m is the number of columns.

If the number of rows is greater than the number of columns, in each column, subtract the smallest item in it from each item in this column. Otherwise if the number of columns is greater than the number of rows, subtract the smallest item in each row from every item in that row.

Step 1: Then after these preliminary steps, the first step is the same as the first step of Munkres algorithm. From left to right and up to down, find zeroes in the resulting matrix from preliminary step. Mark the zero as star, 0^* , if there is no zero in its row or column been marked as star. Repeat this procedure for all zeroes.

Step 2: The second step is to cover every column that contains a zero star 0^* . If k columns are covered, then the starred zeroes are the expected independent set. Otherwise, further steps should be performed.

Step 3: If the assignment is not completed by the second step, following steps should be followed which are the same as those steps in Munkres algorithm. The third step is to choose a uncovered zero and prime it to be $0'$. If no zero is marked as star in its row, step 4 which is a sequence of changing between starred and primed zeroes should be followed. Repeat this step until all zeroes are covered. Then step 5 should be followed.

Step 4: Z_0 is defined as an uncovered prime zero $0'$. Z_1 is defined as the star zero 0^* in Z_0 's column. And Z_2 denotes the prime zeroes $0'$ in Z_1 's row. Repeat finding uncovered prime zeroes $0'$ until we find one with no star zero 0^* in its column. Unstar every starred zero 0^* in this process and star those primed zeroes. Erase all primes and not-covered line. Then follow step 2.

Step 5: h is defined as the smallest uncovered individual in the matrix. Add h to every covered row and then subtract it from each not-covered column. Then go back to step 3.

CHAPTER 4. EXPERIMENTS AND RESULTS

4.1 Experiments

To validate the approach presented, two experiments have been conducted. The goal of Experiment 1 is to determine the proper costs in the cost matrix $C(i,j)$ illustrated previously in Chapter 3 from subjects' task performance. Different modalities were compared to rank and each modality was matched to every feature through human performance testing. The ranking was then applied as individual costs in the cost matrix. The best matching of modality and the associated feature to it was computed using the cost matrix solving equation 3.14. Experiment 2 compared this multimodal sensory substitution system with a traditional tactile paper approach using specialized thermal capsule paper.

A computer-based image perception system with multimodal input/output channel was developed to support the main tasks in this research. Subjects were able to interact with the images shown on a screen through a haptic device (Force Dimension® Omega 6) with stylus end-effector. This device is utilized as a mouse pointer when force feedback is not activated. When force feedback is deployed, modality “Depth” and “Viscosity” were provided through this haptic device. The “Vibration” feature was experienced through Tactors (Engineering Acoustics, Inc.) felt by the fingertips of the opposite hand. “Audio Pitch” and “Unique Audio Cues” were generated from the computer speaker (see

Figure 3.1). Approval for human subjects testing from Institutional Review Board (IRB) is attached in Appendix A.

4.1.1 Experiment 1: Finding the Rank of Modalities

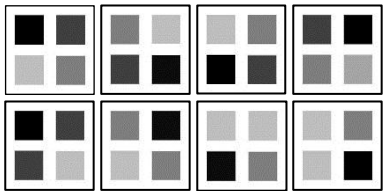
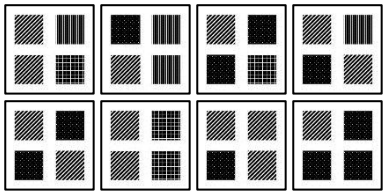
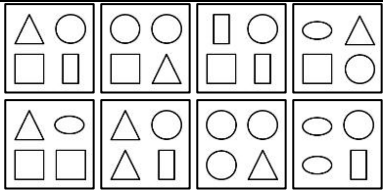
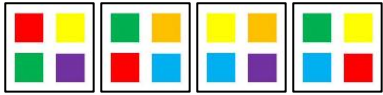
The objective of experiment 1 is to find out the rank of modalities for each image feature. Six blindfolded subjects and one blind subject were recruited for this experiment. A Within-participants experiment was adopted so each subject was presented with all test conditions. Also, each subject was required to test all four primary features since the test on each feature is independent on other features. For each mapping from one feature to one modality, two images (highlighting a specific feature) were deployed. To alleviate the learning effect (Purchase, 2012), different test images were presented for each modality in this experiment. Two tasks were performed for each image and a post-experiment survey was distributed after each group. Two different tasks were designed to permit generalizability. If only one task was used, the conclusions made would hold only for that particular task. One of the tasks performed in experiment 1 required participants to explore the whole image, which tested the performance of image navigation. The other task performed required participants to compare two specific objects, which tested the performance of differentiate certain features.

For feature intensity, since it was mapped to four modalities and for each mapping two test images were used, eight test images were deployed in total. Also, two tasks were performed for each test image, so there were sixteen trials for feature intensity. Since feature texture and shape were mapped to four modalities and feature color was mapped

to two modalities (see Table 3.2), there were 56 trials total ($16+16+16+8=56$) in experiment 1 for each subjects.

Response time and errors were recorded to evaluate human performance, which were used to compute the ranking of modalities for each feature. The test images and tasks for each feature are summarized in Table 4.1.

Table 4.1 Summary of Test Images and Tasks for Each Primary Feature

Features	Test Images	Tasks
Intensity		<ol style="list-style-type: none"> 1. Which is the darkest object in this image? 2. Compare the darkness difference between the left two objects and the right two objects. Which difference is larger?
Texture		<ol style="list-style-type: none"> 1. Are the left two objects of the same texture? 2. How many different textures are in this image?
Shape		<ol style="list-style-type: none"> 1. What is the shape of the top left object? 2. How many different shapes are in this image?
Color		<ol style="list-style-type: none"> 1. Are the left two objects of the same color? 2. How many different colors are in this image?

(A single task for each feature is comprised of a random selection of half of these test images.)

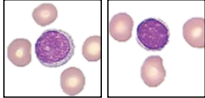
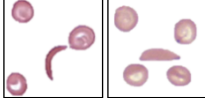
Response time and number of error answers of the tasks were recorded to evaluate human performance. The human performance of modalities for each feature was then used to decide the ranking of candidate modalities. The modality that showed less performance time and higher accuracy was ranked higher than the one of longer response time and lower accuracy. The rankings were then considered as the individual costs in the cost matrix to generate the optimal mapping between image features and sensory modalities.

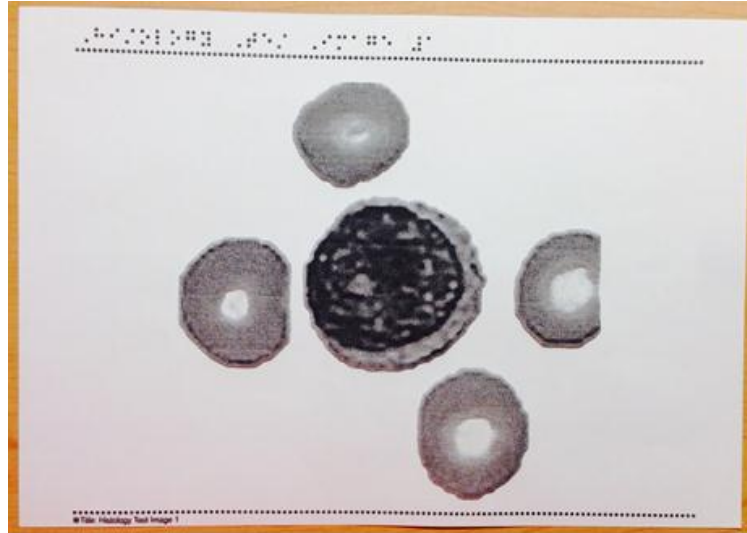
For detailed experiment procedures, please see Appendix B.

4.1.2 Experiment 2: Comparing with print-out tactile paper

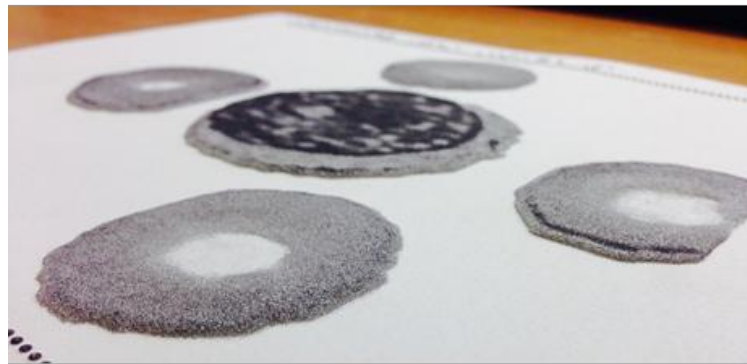
This experiment validated the effectiveness and efficiency of the studied multimodal image perception method with respect to a standard method to convey perceptual information to BVI persons – a print-out tactile paper (see examples in Figure 4.1). Only intensity information is used to generate a tactile paper. In Figure 4.1 (b), it can be observed that dark regions are raised and higher than light regions. Within-participants experiments were applied with ten blindfolded subjects and four blind subjects in which each subject was presented with both methods. The order of method tested was randomized to decrease learning effect. Blood smear images of two different interests were tested. One is to differentiate between red blood cells and white blood cells. The other is to distinguish between normal red blood cells and sickle cells. Test images and tasks are shown in Table 4.2. Response time and errors are recorded to evaluate human performance. For detailed experiment procedures, please see Appendix B.

Table 4.2 Test Images and Tasks for Experiment 2

Test Images	Tasks	
<p>Red blood cell</p> <p>vs.</p> <p>White blood cell</p>		<ol style="list-style-type: none"> 1. Which is the white blood cell in this image? 2. How many red blood cells in this image?
<p>Sickle cell</p>		<ol style="list-style-type: none"> 1. Which is the sickle cell in this image? 2. How many normal red blood cells in this image?



(a)



(b)

Figure 4.1 Tactile paper using specialized thermal capsule paper.

4.2 Results

4.2.1 Experiment 1: Finding the Rank of Modalities

The mean response times and error rates of each matching between modality and feature are shown from Figure 4.2 to Figure 4.5 in groups of features. ANOVA tests and *t*-tests were first performed to examine the significant differences in performance between different modalities. Data indicating no difference was not considered to determine the ranking of modalities. Since two tasks were performed for each feature (see Table 4.1),

response time was then evaluated to rank the modalities for both tasks. The modality shows less response time ranked higher than the one has longer response time. If any of the ranking based on response time does not result in a preference for the two tasks, error rate was then considered to determine the ranking of modalities. The modality of lower error rate ranked higher than the one of higher error rate.

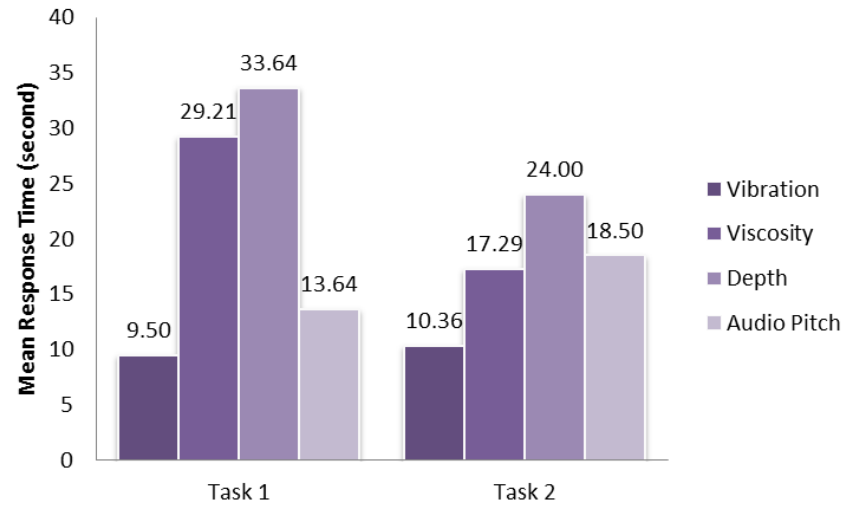
4.2.1.1 Intensity

To investigate how significant each modality is different from others, one-way ANOVA tests are performed. The results show significant differences in performance between the four modalities. The F value of response time for task 1 and 2 are $F_1(3,52)=12.15$, $p=3.87e^{-6}$ and $F_2(3,52)=3.05$, $p=0.04$, and the F value of error rate for task 1 and 2 are $F_1(3,52)=3.03$, $p=0.04$ and $F_2(3,52)=2.98$, $p=0.04$. Since all p -values are less than 0.05, there is significant difference between the four modalities.

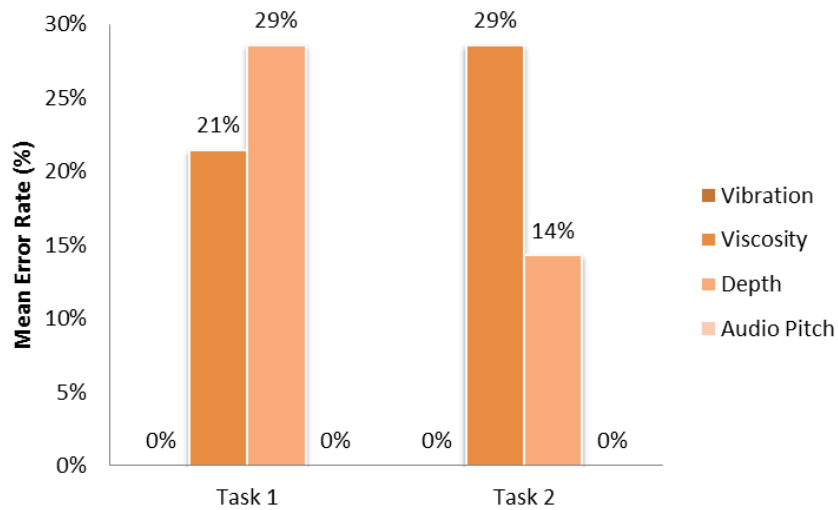
In Figure 4.2 (a), it was observed that for task 1, best performance was achieved through modality vibration, followed by audio pitch, viscosity and depth. However, for task 2, viscosity showed better performance than audio pitch while the other two remains the same. To solve the conflict in task 1 and 2, error rate was then considered. From Figure 4.2 (b), it was observed that for both tasks, audio pitch showed higher accuracy than viscosity. In this case, the ranking of modalities for feature intensity is (from best to worst): vibration, audio pitch, viscosity and depth. The cost matrix for feature intensity would be:

$$C(1, j) = [1 \quad 3 \quad 4 \quad 2 \quad \infty].$$

Further analysis was performed to study the relation between response time and error rate for task 2 since results of modality viscosity and depth shown in response time and error rate display conflicting performance (the improvement of one worsen the other). For Task 2, the response times indicate viscosity performs better than depth, however in error rate, depth showed better performance than viscosity. Therefore, the correlation coefficient was calculated to determine the nature of their relationship. The correlation coefficient between time/error values of task 2 is 0.28, which means the errors are not a direct consequence of fast responses.



(a)



(b)

Figure 4.2 Response time and error rate for feature Intensity.

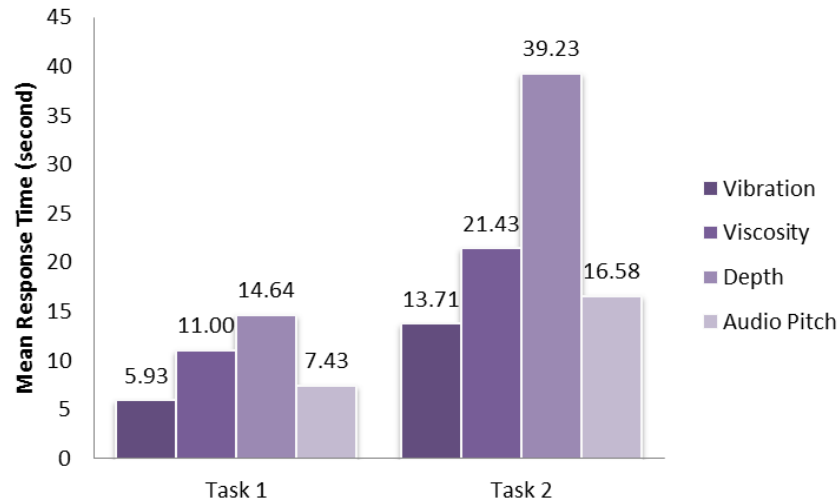
4.2.1.2 Texture

In spite of the difference in error rate shown in Figure 4.3, it can be observed from the F value of response time for task 1 and 2 ($F_1(3,52)=4.10, p=0.01$ and $F_2(3,49)=6.75, p=0.0006$) and the F value of error rate for task 1 and 2 ($F_1(3,44)=3.05, p=0.04$ and

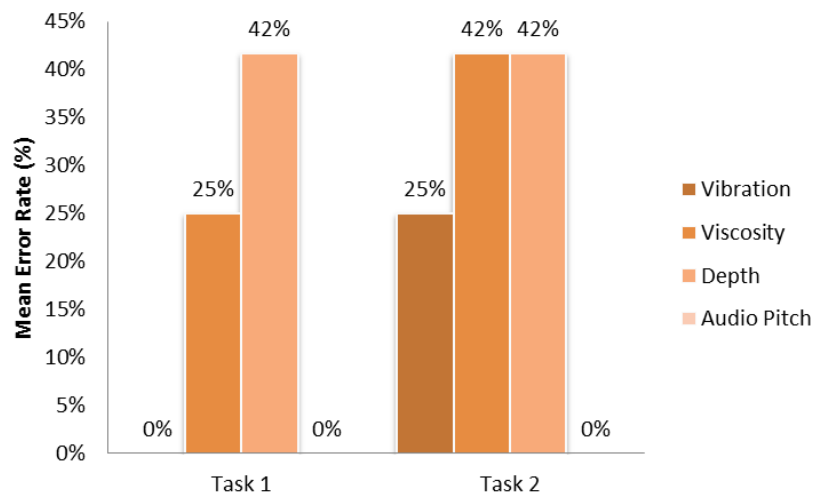
$F_2(3,44)=2.53, p=0.07$), that the difference in error rate for task 2 is not significant when comparing the four modalities. Therefore, error rate for task 2 was not considered when deciding the ranking of four modalities.

From response time for both tasks (see Figure 4.3 (a)), it is apparent that vibration performs best, followed by audio pitch, viscosity and depth. Also, observed from Figure 4.3 (b), the error rate of task 1 showed similar ranking as response time. Vibration and audio pitch showed 100% accuracy, followed by viscosity and depth. Therefore, considered both response time and error rate, the ranking of modalities for feature texture is: vibration, audio pitch, viscosity and depth. The cost matrix of feature texture is represented as:

$$C(2, j) = [1 \ 3 \ 4 \ 2 \ \infty].$$



(a)



(b)

Figure 4.3 Response time and error rate for feature Texture.

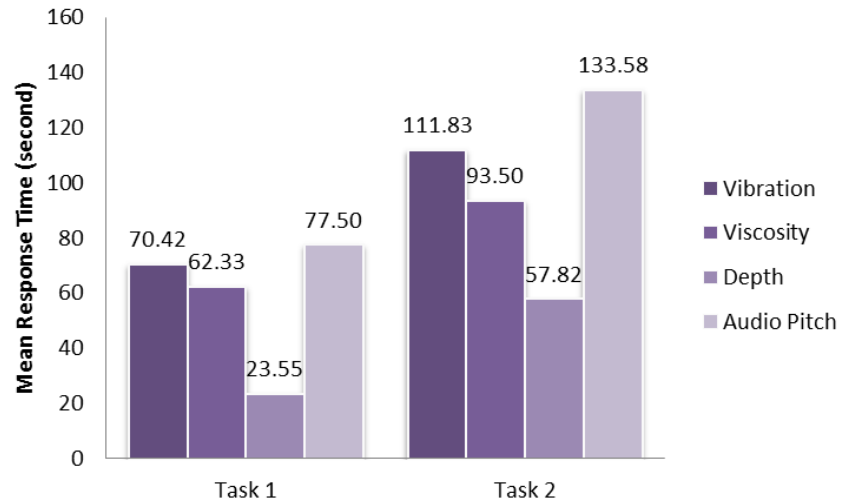
4.2.1.3 Shape

The F value of response time for task 1 and 2 are $F_1(3,43)=3.51$, $p=0.02$ and $F_2(3,43)=3.61$, $p=0.02$, and the F value of error rate for task 1 and 2 are $F_1(3,44)=3.27$, $p=0.03$ and $F_2(3,44)=2.52$, $p=0.07$. Since in task 2, the error rate did not show significant

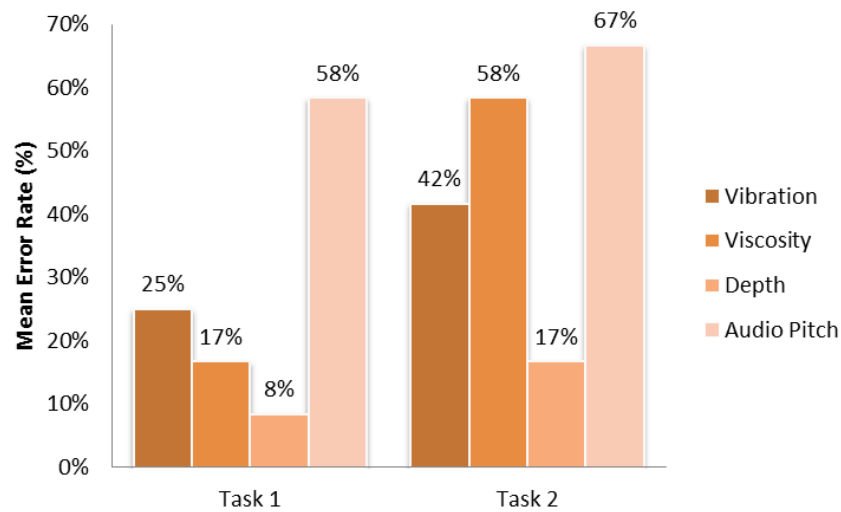
difference between the four modalities, it was not considered in deciding the rank of modalities for feature shape.

From Figure 4.4 (a), it can be observed that in both tasks, depth showed shortest response time, followed by viscosity, vibration and audio pitch. From task 1 error rate (see Figure 4.4 (b)), it indicated same ranking of modalities as response time. Depth showed lowest error rate, followed by viscosity, vibration and audio pitch. Therefore, the ranking of modalities for feature is: depth, viscosity, vibration and audio pitch. The cost matrix for feature shape is represented as:

$$C(3, j) = [3 \quad 2 \quad 1 \quad 4 \quad \infty].$$



(a)



(b)

Figure 4.4 Response time and error rate for feature Shape

4.2.1.4 Color

A *t*-test was performed for feature color on response time since there are only two candidate modalities and there is no difference between these two candidates on error rate (see Figure 4.5 (b)). The *p*-value for task 2 is smaller than 0.05, while it is not the case for

task 1 ($p_1=0.14$, $p_2=0.04$), which means there is significant difference between these two modalities for task 2. However, no conclusion can be drawn for task 1. Therefore, the response time for task 1 was not considered to decide the ranking of modalities for feature color. Figure 4.5 (a) shows that unique audio cues are associated to higher performance than vibration when representing feature color, since it has less response time for task 2. In this case, unique audio cues have been shown to be the optimal candidate for feature color. The cost matrix for feature color is shown as:

$$C(4, j) = [2 \quad \infty \quad \infty \quad \infty \quad 1].$$

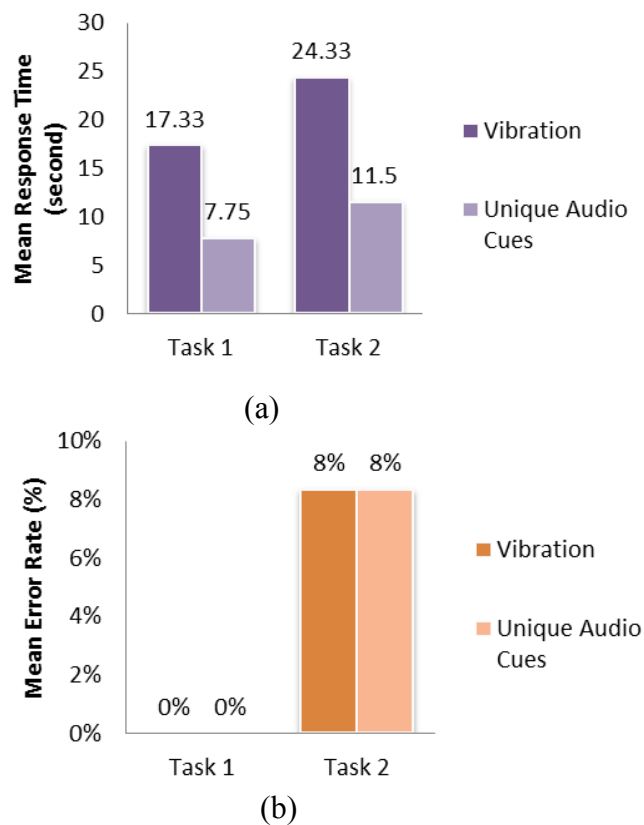


Figure 4.5 Response time and error rate for feature Color

4.2.1.5 Cost Matrix

The ranking of modalities for each feature is summarized below. Smaller digit indicates higher ranking and better performance.

- Intensity: (1) vibration, (2) audio pitch, (3) viscosity, (4) depth.
- Texture: (1) vibration, (2) audio pitch, (3) viscosity, (4) depth.
- Shape: (1) depth, (2) viscosity, (3) vibration, (4) audio pitch.
- Color: (1) unique audio cues, (2) vibration.

Since the individual cost used in the cost matrix is the ranking of modalities for each feature, the cost matrix is defined as

$$C(i, j) = \begin{bmatrix} 1 & 3 & 4 & 2 & \infty \\ 1 & 3 & 4 & 2 & \infty \\ 3 & 2 & 1 & 4 & \infty \\ 2 & \infty & \infty & \infty & 1 \end{bmatrix}$$

The optimal matching between modality and feature computed by the extended Munkres Algorithm (Bourgeois & Lassalle, 1971) is shown in Fig. 4.6.

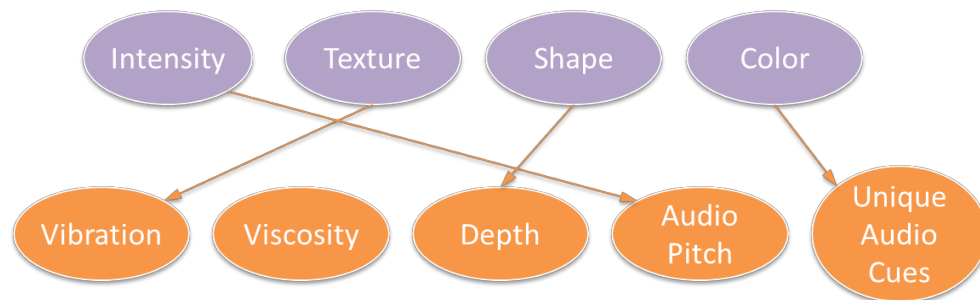
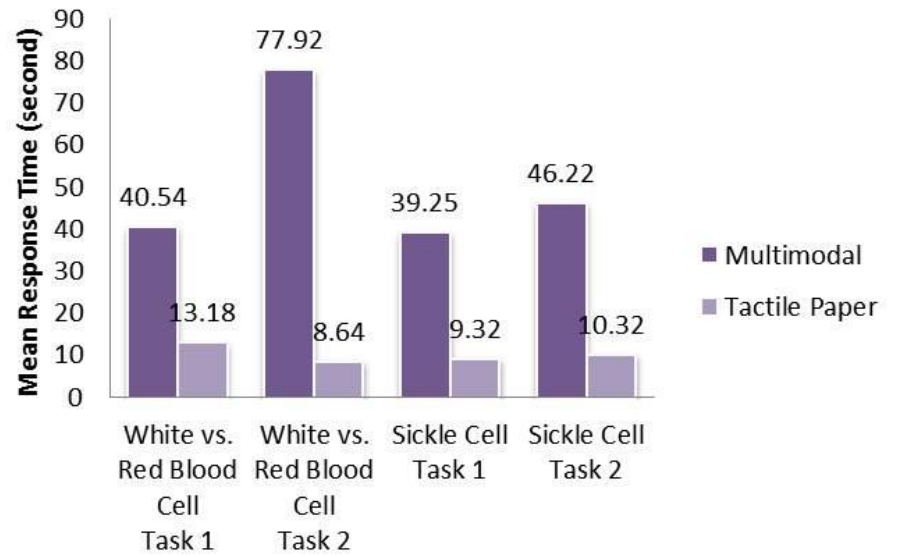


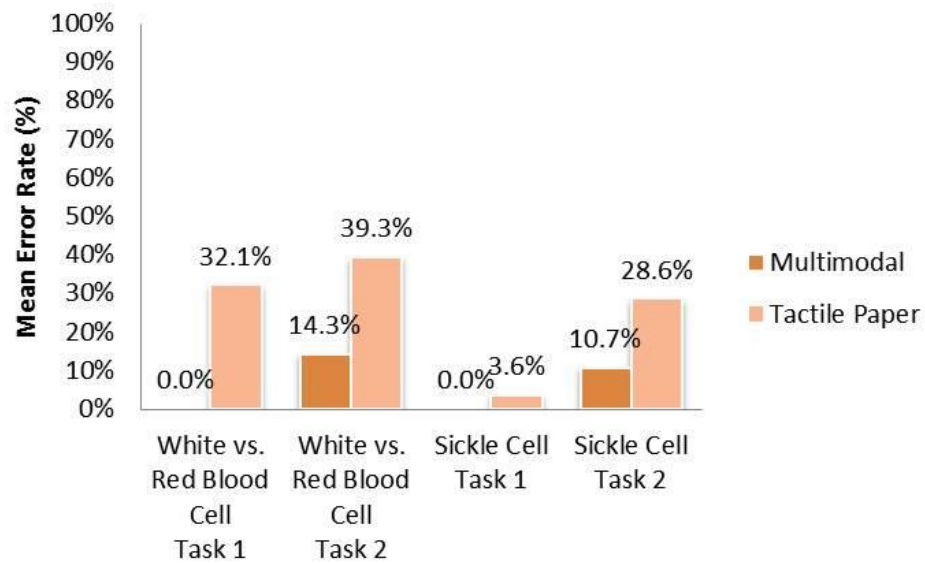
Figure 4.6 Optimal matching of modalities and primary features

4.2.2 Experiment 2: Comparing with print-out tactile paper

A *t*-test is performed to validate the difference between the response time and error rates of two methods. The *p*-values of response time for the four tasks are: $p_1=1.7e^{-5}$, $p_2=8.4e^{-9}$, $p_3=1.7e^{-6}$ and $p_4=2.5e^{-9}$. All these *p*-values indicated significant difference between multimodal method and tactile paper. The *p*-values of error rate for the four tasks are: $p_1=7.4e^{-4}$, $p_2=0.03$, $p_3=0.32$ and $p_4=0.09$. Since the error rate for the last two tasks were not significant different, it was not considered. Figure 4.7 shows the mean response time and error rate for all the tasks in experiment 2. Although the multimodal method requires more time to accomplish one task, it shows higher accuracy in all tasks. And the error rate for the first two tasks indicated significant better performance of multimodal method with accuracy of 32.1% higher of task 1 and 29.17% higher of task 2 in differentiating white blood cells and red blood cells.



(a)



(b)

Figure 4.7 Response time and error rate for all tasks in experiment 2.

To determine whether the higher accuracy came from longer response time, the correlation coefficient is calculated. Since the correlation value between time/error is positive (0.0064), it can be concluded that lower error rate is not a consequence of longer

response time. We conclude that the multimodal method has higher accuracy because it provides a perceptually rich way to interpret images compared to using tactile paper. It was also observed during the experiments, that when using the multimodal method, most of the response time was taken to explore the image. Once a cell was located, it required little effort for the participants to recognize the type of cell. Briefly, a key factor that makes navigation of an image effortful, is the single-hand operation and loss of point of reference. In the traditional tactile paper method, both hands are utilized to locate an object and interpret its features. Also, the point of reference can be easily retrieved by the relative position of both hands.

In the second experiment, error rate is considered a more important factor than response time, because the intent of our system is to be used as an educational tool to help BVI students and scientists learn and recognize the features of different histological images. Therefore, we believe it is more important for users to correctly identify intracellular structures, the shape of various types of cells, and make comparisons between cell types rather than improving the speed BVI students or scientists can recognize cell features.

CHAPTER 5. CONCLUSIONS AND FUTURE WORK

In this thesis, a real-time multimodal image perception approach is developed to convey visual information to blind or visually impaired people. Images characterized by seven key features; intensity, color, shape, texture, etc., are represented through three different sensory channels of hearing, haptics and vibrotactility. A Bayesian network is constructed to infer the relationship between primary and peripheral image features. Linear assignment algorithm is utilized to optimize the matching between image features and sensorial modalities. This novel approach not only decreases the time and man power required to create traditional tactile print-outs, but allows computer-based image data to be interpretable by blind individuals in real-time. This HCI system can be connected to a light microscope or other scientific instruments that employ a computer monitor output in order to represent digitized image information to BVI users in real-time. This substitution of visual scientific data with other sensory outputs can allow students and scientists who are BLV to participate in many kinds of scientific investigation and experimentation that are currently unavailable to them. This ability allows them to understand visual scientific data that they had generated themselves, which is critical when conducting independent research. In addition, alternative sensorial perception of data that is typically rendered visually may provide unique or more in depth understanding of the results from the scientific instrument.

5.1 Possible Changes to Bayesian Network

In this thesis, the Bayesian network constructed to infer peripheral image features from primary image features is based on a selection of candidate structures generated by human experts. Three candidate structures were generated by one science expert who is blind based on his experience and current literature. However, it is important to question whether instead of obtaining multiple responses from one expert, the approach should be getting one response from multiple experts that are BVI. Therefore, more human experts can be recruited in the future to generate candidate structures between peripheral and primary image features. Besides expert-based modeling, Genetic algorithm (GA) can also be utilized to generate candidate structures where the dependencies between nodes are generated following Genetic algorithm's operations. The process generating Bayesian networks using Genetic algorithm is called evolution-based modeling (Yu-Ting Li & Juan P. Wachs, 2014a). The initial population of candidate structures is generated randomly. Then genetic operators, such as crossover and mutation, are used to generate new generations of candidate structures based on a selected portion of last generation. To select a portion of candidate structures from last generation, the score function (equation 3.11) shown in Chapter 3.1.2.3 is used. The structures that show higher score than their antecedents are selected. The number of iterations is set according to empirical study requirements.

5.2 Expanding the Modality Assignment Problem

In addition to relating primary and peripheral image features through a Bayesian network, a linear assignment problem (LAP) was used to assign image features to different sensory

modalities based on using a cost weighing approach. The Cost matrix was constructed by evaluating subjects' performance using a series of test images for each primary image feature. Besides utilizing linear assignment problem, our assignment problem between image features and sensory modalities can be extended to a quadratic assignment problem (QAP) as well, taking consideration of more factors, such as the inference relations among primary image features and the influence of one sensory modality to another. A quadratic assignment problem is also one of the fundamental optimization problems in deciding combinations that solves assignment problem with linear costs between two parties. With solutions of linear assignment problem, only the costs between each pair of the two parties are considered to make the assignments. To get a more accurate assignment solution, more information should be taken into consideration. By extended our problem to a quadratic assignment problem, two more matrices are required. According to this thesis's problem, one matrix is required to represent the relationships of the four primary image features and the other matrix is required to indicate the relationships between the five output sensory modalities. This quadratic assignment problem considers not only the linear cost between image features and sensory modalities, but also the inherent relations among features and modalities themselves. Since QAP also takes in consideration of the inner relation among both primary image features and sensory modalities, besides the linear cost between primary image features and sensory modalities, it should provide a more accurate assignment than the LAP.

5.3 Considerations for Future Experiments

Besides the possible improvements of the methodologies applied in this thesis, the experiments deployed can be extended as well. Our real-time multimodal image perception system was compared to traditional tactile images with ten blindfolded and four blind subjects. Some researchers have shown no performance difference between blind and blindfolded subjects, while others find blind have a better developed sense (*Enhancing performance for action and perception*, 2011). Therefore, it is still necessary to recruit more blind participants to test our system. Besides print-out tactile images, other assistive technologies can be compared with our system as well. 3D printing tactile plates can be an option. 3D printed plates convey more information than tactile images, such as intensity, object patterns and relative relationships between objects (Greg J. Williams et al., 2014). Although empirical experiments showed that our multimodal system provides more visual information to BVI people than tactile images, it is worthwhile to test whether this is true compared to other currently available assistive technologies and how much more visual information can our system convey.

5.4 Possible Improvements for Human-computer Interaction

From experimental results, it is observed that this multimodal approach takes more time than tactile paper to navigate and explore an image; however empirical experiments showed it higher accuracy in recognizing and analyzing objects within blood smear images. Observed from conducted experiments, most participants searched test images in a sequential manner that from left to right and top to bottom. Since they lost their point of reference when using only one hand, they might go repeatedly to the same place they

already searched or miss some area that they have not yet reached. So they need a lot of time to go back and forth checking if they missed anything, which makes navigation and exploration very time consuming. Therefore, in the future, we will research interaction methods to improve performance in image navigation. As discussed in Chapter 4.2.2, the utilization of single hand examination is considered as the main reason that makes the navigation and exploration of images so time consuming, compared to using both hands during the examination of traditional tactile paper images. More specific, in our multimodal image perception system, participants interact with the images through a haptic device with stylus end-effector. This stylus end-effector makes the interaction based on only one pixel in the image. While during the traditional tactile method, BVI persons interact with the tactile image using both their hands with a relatively large receptor area at their fingertips flexibly searching nearby areas. In our conducted experiments, participants were able to navigate the tactile image using both their hands. To validate the hypothesis that single-hand interaction makes image navigation more time consuming, further experiments can be conducted to test the response time when participants can interact with tactile images using only one finger compared to the stylus end-effector used in our multimodal system. Besides this one-finger stimulation experiment, a 7 degree of freedom (DOF) haptic device with a gripper end-effector (see Figure 5.1) can be utilized to improve the performance for image navigation and exploration. The 7 DOF haptic device with a gripper end-effector provides force feedback on the gripper as well. Both fingers on the gripper can feel the force feedback. A point of reference may be retrieved since two fingers are utilized and the relative positions of objects in the image can be inferred by the displacement of both fingers.

Besides single hand operation, a bimanual system with two haptic devices used concurrently by both hands is also a possible solution to improve the performance of image navigation.

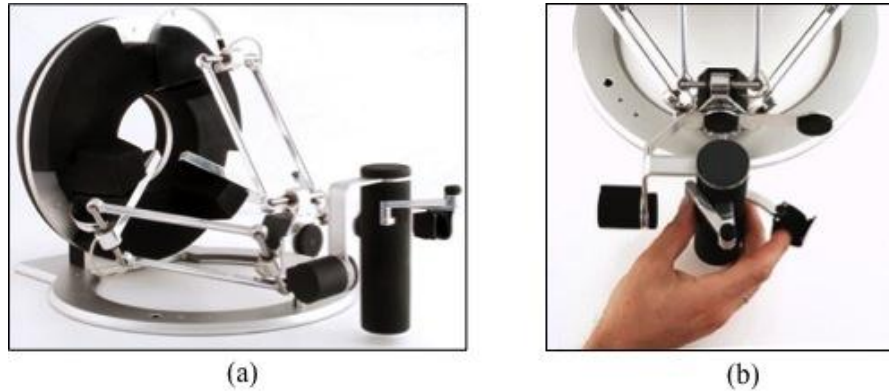


Figure 5.1 Force Dimension 7 DOF haptic device with a gripper end-effector.

Another logistical difference when interpreting tactile paper using one's hands is that the raised images on the paper are typically laid on a horizontal surface. However, when using our method, images are presented vertically on the computer screen. Whether the direction of perceiving an image will affect participants' performance can be considered as a key point in future research.

BIBLIOGRAPHY

BIBLIOGRAPHY

- Akgül, M. (1993). A genuinely polynomial primal simplex algorithm for the assignment problem. *Discrete Applied Mathematics*, 45(2), 93–115. doi:10.1016/0166-218X(93)90054-R
- American Foundation for the Blind. (n.d.). *Statistical Snapshots from the American Foundation for the Blind*. Last modified Jan, 2013. Accessed April 12, 2013.
- Arno, P., Capelle, C., Wanet-Defalque, M.-C., Catalan-Ahumada, M., & Veraart, C. (1999). Auditory coding of visual patterns for the blind. *Perception*, 28(8), 1013 – 1029. doi:10.1068/p2607
- Bach-y-Rita, P. (2004). Tactile sensory substitution studies. *Annals of the New York Academy of Sciences*, 1013, 83–91.
- Bach-Y-Rita, P., Collins, C. C., Saunders, F. A., White, B., & Scadden, L. (1969). Vision Substitution by Tactile Image Projection. *Nature*, 221(5184), 963–964. doi:10.1038/221963a0
- Bach-y-Rita, P., & Kaczmarek, K. A. (2002, August 6). Tongue placed tactile output device.
- Bach-y-Rita, P., Kaczmarek, K. A., Tyler, M. E., & Garcia-Lara, J. (1998). Form perception with a 49-point electrotactile stimulus array on the tongue: A technical note. *Journal of Rehabilitation Research and Development*, 35(4), 427–430.

- Bach-y-Rita, P., & W. Kercel, S. (2003). Sensory substitution and the human-machine interface. *Trends in Cognitive Sciences*, 7(12), 541–546. doi:10.1016/j.tics.2003.10.013
- Bajcsy, R. (1973). Computer Description of Textured Surfaces. In *Proceedings of the 3rd International Joint Conference on Artificial Intelligence* (pp. 572–579). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. Retrieved from <http://dl.acm.org/citation.cfm?id=1624775.1624844>
- Balakrishnan, G., Sainarayanan, G., Nagarajan, R., & Yaacob, S. (n.d.). *Wearable Real-Time Stereo Vision for the Visually Impaired*.
- Blattner, M. M., & Glinert, E. P. (1996). Multimodal integration. *IEEE MultiMedia*, 3(4), 14–24. doi:10.1109/93.556457
- Bourgeois, F., & Lassalle, J.-C. (1971). An Extension of the Munkres Algorithm for the Assignment Problem to Rectangular Matrices. *Commun. ACM*, 14(12), 802–804. doi:10.1145/362919.362945
- Bradley S. Duerstock, Lisa Hillard, & Deana McDonagh. (2014). Technologies to Facilitate the Active Participation and Independence of Persons with Disabilities in STEM from College to Careers. In *From College to Careers: Fostering Inclusion of Persons with Disabilities in STEM* (pp. 5–30). Washington, DC.
- Burch, D., & Pawluk, D. (2011). Using multiple contacts with texture-enhanced graphics. In *2011 IEEE World Haptics Conference (WHC)* (pp. 287–292). doi:10.1109/WHC.2011.5945500

- Burkard, R. E. (1986). Assignment problems: Recent solution methods and applications. In A. Prékopa, J. Szelezsáan, & B. Strazicky (Eds.), *System Modelling and Optimization* (pp. 153–169). Springer Berlin Heidelberg. Retrieved from <http://link.springer.com/chapter/10.1007/BFb0043834>
- Burkard, R. E., Dell’Amico, M., & Martello, S. (2009). *Assignment Problems*. SIAM.
- Canny, J. (1986). A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-8*(6), 679–698. doi:10.1109/TPAMI.1986.4767851
- Capelle, C., Trullemans, C., Arno, P., & Veraart, C. (1998). A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution. *IEEE Transactions on Bio-Medical Engineering*, *45*(10), 1279–1293. doi:10.1109/10.720206
- Chaudhuri, B. B., Rodenacker, K., & Burger, G. (1988). Characterization and featuring of histological section images. *Pattern Recognition Letters*, *7*(4), 245–252. doi:10.1016/0167-8655(88)90109-2
- Chikkerur, S., Serre, T., Tan, C., & Poggio, T. (2010). What and where: A Bayesian inference theory of attention. *Vision Research*, *50*(22), 2233–2247. doi:10.1016/j.visres.2010.05.013
- Chu, A., Sehgal, C. M., & Greenleaf, J. F. (1990). Use of gray value distribution of run lengths for texture analysis. *Pattern Recognition Letters*, *11*(6), 415–419. doi:10.1016/0167-8655(90)90112-F

- Clausi, D. A. (2002). An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of Remote Sensing*, 28(1), 45–62. doi:10.5589/m02-004
- Connors, R. W., & Harlow, C. A. (1980). A Theoretical Comparison of Texture Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-2(3), 204–222. doi:10.1109/TPAMI.1980.4767008
- Cruz-Roa, A., Caicedo, J. C., & González, F. A. (2011). Visual pattern mining in histology image collections using bag of features. *Artificial Intelligence in Medicine*, 52(2), 91–106. doi:10.1016/j.artmed.2011.04.010
- Davis, L. S. (1975). A survey of edge detection techniques. *Computer Graphics and Image Processing*, 4(3), 248–270. doi:10.1016/0146-664X(75)90012-X
- Davis, L. S., Johns, S. A., & Aggarwal, J. K. (1979). Texture Analysis Using Generalized Co-Occurrence Matrices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(3), 251–259. doi:10.1109/TPAMI.1979.4766921
- De Volder, A. G., Toyama, H., Kimura, Y., Kiyosawa, M., Nakano, H., Vanlierde, A., ... Senda, M. (2001). Auditory triggered mental imagery of shape involves visual association areas in early blind humans. *NeuroImage*, 14(1 D), 129–139. doi:10.1006/nimg.2001.0782
- Delage, E., Lee, H., & Ng, A. (2006). A Dynamic Bayesian Network Model for Autonomous 3D Reconstruction from a Single Indoor Image. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Vol. 2, pp. 2418–2428). doi:10.1109/CVPR.2006.23

- Dinic, E. A., & Kronrod, M. A. (1969). An algorithm for the solution of the assignment problem. In *Soviet Math. Dokl* (Vol. 10, pp. 1324–1326).
- Dobrowski, S. Z., Safford, H. D., Cheng, Y. B., & Ustin, S. L. (2008). Mapping mountain vegetation using species distribution modeling, image-based texture analysis, and object-based classification. *Applied Vegetation Science*, *11*(4), 499–508. doi:10.3170/2008-7-18560
- Easterfield, T. E. (1946). A combinatorial algorithm. *Journal of the London Mathematical Society*, *1*(3), 219–226.
- Edmonds, J., & Karp, R. M. (1972). Theoretical Improvements in Algorithmic Efficiency for Network Flow Problems. *J. ACM*, *19*(2), 248–264. doi:10.1145/321694.321699
- Enhancing performance for action and perception: multisensory integration, neuroplasticity & neuroprosthetics.* (2011). Elsevier.
- Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, *10*(1). doi:10.1167/10.1.6
- Ezaki, N., Bulacu, M., & Schomaker, L. (2004). Text detection from natural scene images: towards a system for visually impaired persons. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004* (Vol. 2, pp. 683–686 Vol.2). doi:10.1109/ICPR.2004.1334351
- Friedman, N. (1997). Learning Belief Networks in the Presence of Missing Values and Hidden Variables. In *Proceedings of the Fourteenth International Conference on Machine Learning* (pp. 125–133). Morgan Kaufmann.

- Friedman, N., Linial, M., Nachman, I., & Pe'er, D. (2000). Using Bayesian networks to analyze expression data. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, 7(3-4), 601–620. doi:10.1089/106652700750050961
- Fritz, J. P., Way, T. P., & Barner, K. E. (1996). Haptic Representation of Scientific Data for Visually Impaired or Blind Persons. In *In Technology and Persons With Disabilities Conference*.
- G, N. V., & S, H. K. (2011). Study and comparison of various image edge detection techniques used in quality inspection and evaluation of agricultural and food products by computer vision. *International Journal of Agricultural and Biological Engineering*, 4(2), 83–90. doi:10.3965/ijabe.v4i2.277
- Galloway, M. M. (1975). Texture analysis using gray level run lengths. *Computer Graphics and Image Processing*, 4(2), 172–179. doi:10.1016/S0146-664X(75)80008-6
- Greg J. Williams, Ting Zhang, Alex Lo, Gonzales, A., Baluch, D.P., & Bradley S. Duerstock. (2014). 3D Printing Tactile Graphics for the Blind: Application to Histology. In *Annual Rehabilitation Engineering Society of North America Conference 2014*. Indianapolis, IN.
- Hanif, S. M., & Prevost, L. (n.d.). *Texture Based Text Detection in Natural Scene Images: A Help to Blind and Visually Impaired Persons*.
- Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-3(6), 610–621. doi:10.1109/TSMC.1973.4309314

- Heath, M. D., Sarkar, S., Sanocki, T., & Bowyer, K. W. (1997). A robust visual method for assessing the relative performance of edge-detection algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*(12), 1338–1359. doi:10.1109/34.643893
- Heckerman, D. (2008). A Tutorial on Learning with Bayesian Networks. In P. D. E. Holmes & P. L. C. Jain (Eds.), *Innovations in Bayesian Networks* (pp. 33–82). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-85066-3_3
- Heckerman, D., Geiger, D., & Chickering, D. M. (1995). Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, *20*(3), 197–243. doi:10.1007/BF00994016
- Heller, M. A. (2002). Tactile picture perception in sighted and blind people. *Behavioural Brain Research*, *135*(1–2), 65–68. doi:10.1016/S0166-4328(02)00156-0
- Howard, S. (1958). *Principles of perception* (Vol. xii). Oxford, England: Harper.
- HowStuffWorks “BrainPort.” (n.d.). *HowStuffWorks*. Retrieved June 10, 2014, from <http://science.howstuffworks.com/brainport.htm>
- Hsu, B., Hsieh, C.-H., Yu, S.-N., Ahissar, E., Arieli, A., & Zilbershtain-Kra, Y. (2013). A tactile vision substitution system for the study of active sensing. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 3206–3209). doi:10.1109/EMBC.2013.6610223
- Ikei, Y., Wakamatsu, K., & Fukuda, S. (1997). Vibratory tactile display of image-based textures. *IEEE Computer Graphics and Applications*, *17*(6), 53–61. doi:10.1109/38.626970

- Jacko, J. A., Scott, I. U., Sainfort, F., Barnard, L., Edwards, P. J., Emery, V. K., ... Zorich, B. S. (2003). Older Adults and Visual Impairment: What Do Exposure Times and Accuracy Tell Us About Performance Gains Associated with Multimodal Feedback? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 33–40). New York, NY, USA: ACM. doi:10.1145/642611.642619
- Jain, A. K., & Zhong, Y. (1996). Page segmentation using texture analysis. *Pattern Recognition*, 29(5), 743–770. doi:10.1016/0031-3203(95)00131-X
- Johnson, L. A., & Higgins, C. M. (2006). A Navigation Aid for the Blind Using Tactile-Visual Sensory Substitution. In *28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2006. EMBS '06* (pp. 6289–6292). doi:10.1109/IEMBS.2006.259473
- Kaczmarek, K. A., Webster, J. G., Bach-y-Rita, P., & Tompkins, W. J. (1991). Electrotactile and vibrotactile displays for sensory substitution systems. *IEEE Transactions on Biomedical Engineering*, 38(1), 1–16. doi:10.1109/10.68204
- Kajimoto, H., Suzuki, M., & Kanno, Y. (2014). HamsaTouch: Tactile Vision Substitution with Smartphone and Electro-tactile Display. In *Proceedings of the Extended Abstracts of the 32Nd Annual ACM Conference on Human Factors in Computing Systems* (pp. 1273–1278). New York, NY, USA: ACM. doi:10.1145/2559206.2581164
- Kay, L. (1974). A sonar aid to enhance spatial perception of the blind: engineering design and evaluation. *Radio and Electronic Engineer*, 44(11), 605. doi:10.1049/ree.1974.0148

- Koons, D. B., Sparrell, C. J., & Thorisson, K. R. (1993). Intelligent Multimedia Interfaces. In M. T. Maybury (Ed.), (pp. 257–276). Menlo Park, CA, USA: American Association for Artificial Intelligence. Retrieved from <http://dl.acm.org/citation.cfm?id=162477.162508>
- Kress, G. (2009). *Multimodality: A Social Semiotic Approach to Contemporary Communication*. Routledge.
- Kumar, V. P., & Desai, U. B. (1996). Image interpretation using Bayesian networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(1), 74–77. doi:10.1109/34.476423
- Lécuyer, A., Mobuchon, P., Mégard, C., Perret, J., Andriot, C., & Colinot, J.-P. (2003). HOMERE: A Multimodal System for Visually Impaired People to Explore Virtual Environments. In *Proceedings of the IEEE Virtual Reality 2003* (p. 251–). Washington, DC, USA: IEEE Computer Society. Retrieved from <http://dl.acm.org/citation.cfm?id=832289.835979>
- Lendaris, G. G., & Stanley, G. L. (1970). Diffraction-pattern sampling for automatic pattern recognition. *Proceedings of the IEEE*, 58(2), 198–216. doi:10.1109/PROC.1970.7593
- Lescal, D., Rouat, J., & Voix, J. (2013). Sensorial substitution system from vision to audition using transparent digital earplugs. *Proceedings of Meetings on Acoustics*, 19(1), 040014. doi:10.1121/1.4799670
- Lipkin, B. S. (1970). *Picture Processing and Psychopictorics*. Elsevier.

- Luo, J., Savakis, A. E., & Singhal, A. (2005). A Bayesian network-based framework for semantic image understanding. *Pattern Recognition*, 38(6), 919–934.
doi:10.1016/j.patcog.2004.11.001
- Marks, L. E. (1987). On cross-modal similarity: auditory-visual interactions in speeded discrimination. *Journal of Experimental Psychology. Human Perception and Performance*, 13(3), 384–394.
- Meers, S., & Ward, K. (2005). A vision system for providing the blind with 3D colour perception of the environment. *Faculty of Informatics - Papers (Archive)*. Retrieved from <http://ro.uow.edu.au/infopapers/436>
- Meijer, P. B. L. (1992). An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering*, 39(2), 112–121.
doi:10.1109/10.121642
- Munkres, J. (1957). Algorithms for the Assignment and Transportation Problems. *Journal of the Society for Industrial and Applied Mathematics*, 5(1), 32–38.
- National Science Board. (2014). *Science and Engineering Indicators 2014*.
- Nguyen, T. H., Nguyen, T. H., Le, T. L., Tran, T. T. H., Vuillerme, N., & Vuong, T. P. (2013). A wireless assistive device for visually-impaired persons using tongue electrotactile system. In *2013 International Conference on Advanced Technologies for Communications (ATC)* (pp. 586–591).
doi:10.1109/ATC.2013.6698183

- Owen, J. M., Petro, J. A., D'Souza, S. M., Rastogi, R., & Pawluk, D. T. V. (2009). An improved, low-cost tactile “mouse” for use by individuals who are blind and visually impaired. In *Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility* (pp. 223–224). New York, NY, USA: ACM. doi:10.1145/1639642.1639686
- Peli, T., & Malah, D. (1982). A study of edge detection algorithms. *Computer Graphics and Image Processing*, 20(1), 1–21. doi:10.1016/0146-664X(82)90070-3
- Ptito, M., Moesgaard, S. M., Gjedde, A., & Kupers, R. (2005). Cross-modal plasticity revealed by electrotactile stimulation of the tongue in the congenitally blind. *Brain*, 128(3), 606–614. doi:10.1093/brain/awh380
- Purchase, H. C. (2012). *Experimental Human-Computer Interaction: A Practical Guide with Visual Examples*. Cambridge University Press.
- Rastogi, R., & Pawluk, D. T. V. (2013). Dynamic tactile diagram simplification on refreshable displays. *Assistive Technology: The Official Journal of RESNA*, 25(1), 31–38.
- Sami Abboud, S. H. (2014). EyeMusic: Introducing a “visual” colorful experience for the blind using auditory sensory substitution. *Restorative Neurology and Neuroscience*. doi:10.3233/RNN-130338
- Sampaio, E., Maris, S., & Bach-y-Rita, P. (2001). Brain plasticity: “visual” acuity of blind persons via the tongue. *Brain Research*, 908(2), 204–207. doi:10.1016/S0006-8993(01)02667-1

- Schneiderman, H. (2004). Learning a restricted Bayesian network for object detection. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004* (Vol. 2, pp. II-639-II-646 Vol.2). doi:10.1109/CVPR.2004.1315224
- Silver, R. (1960). An Algorithm for the Assignment Problem. *Commun. ACM*, 3(11), 605–606. doi:10.1145/367436.367476
- Smith, L. B., & Sera, M. D. (1992). A developmental analysis of the polar structure of dimensions. *Cognitive Psychology*, 24(1), 99–142. doi:10.1016/0010-0285(92)90004-L
- Stockman, G., & Shapiro, L. G. (2001). *Computer Vision* (1st ed.). Upper Saddle River, NJ, USA: Prentice Hall PTR.
- Sutton, R. N., & Hall, E. L. (1972). Texture Measures for Automatic Classification of Pulmonary Disease. *IEEE Transactions on Computers*, C-21(7), 667–676. doi:10.1109/T-C.1972.223572
- Takagi, N. (2009). Mathematical figure recognition for automating production of tactile graphics. In *IEEE International Conference on Systems, Man and Cybernetics, 2009. SMC 2009* (pp. 4651–4656). doi:10.1109/ICSMC.2009.5346749
- Tang, H., & Beebe, D. J. (2003). Design and microfabrication of a flexible oral electrotactile display. *Journal of Microelectromechanical Systems*, 12(1), 29–36. doi:10.1109/JMEMS.2002.807478
- Tomita, F., & Tsuji, S. (1990). Structural Texture Analysis. In *Computer Analysis of Visual Textures* (pp. 71–82). Springer US. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4613-1553-7_5

- Tomizawa, N. (1971). On some techniques useful for solution of transportation network problems. *Networks*, *1*(2), 173–194. doi:10.1002/net.3230010206
- Van Koten, C., & Gray, A. R. (2006). An application of Bayesian network for predicting object-oriented software maintainability. *Information and Software Technology*, *48*(1), 59–67. doi:10.1016/j.infsof.2005.03.002
- Votaw, D. F. (1952). *The personnel assignment problem*.
- Wall, S. A., & Brewster, S. A. (2006). Tac-tiles: Multimodal Pie Charts for Visually Impaired Users. In *Proceedings of the 4th Nordic Conference on Human-computer Interaction: Changing Roles* (pp. 9–18). New York, NY, USA: ACM. doi:10.1145/1182475.1182477
- Ward, J., & Meijer, P. (2010). Visual experiences in the blind induced by an auditory sensory substitution device. *Consciousness and Cognition*, *19*(1), 492–500. doi:10.1016/j.concog.2009.10.006
- Way, T. P., & Barner, K. E. (1997). Automatic visual to tactile translation. I. Human factors, access methods and image manipulation. *IEEE Transactions on Rehabilitation Engineering*, *5*(1), 81–94. doi:10.1109/86.559353
- Wechsler, H. (1980). Texture analysis — a survey. *Signal Processing*, *2*(3), 271–282. doi:10.1016/0165-1684(80)90024-9
- Wong, P. C., & Thomas, J. (2004). Visual Analytics. *IEEE Computer Graphics and Applications*, *24*(5), 20–21. doi:10.1109/MCG.2004.39
- Yu, W., McAllister, G., Murphy, E., Kuber, R., & Strain, P. (n.d.). Developing Multimodal Interfaces for Visually Impaired People to Access Internet. In *PROC. 8 TH ERCIM WORKSHOP – USER INTERFACES FOR ALL*.

- Yu, W., Reid, D., & Brewster, S. A. (2002). Web-based multimodal graphs for visually impaired people. In S. Keates (Ed.), . Presented at the 1st Cambridge Workshop on Universal Access and Assistive Technology (CWUAAT), Cambridge, England: Springer Verlag. Retrieved from <http://eprints.gla.ac.uk/3228/>
- Yu-Ting Li, & Juan P. Wachs. (2014a). A Bayesian Approach to Determine Focus of Attention in Spatial and Time-Sensitive Decision Making Scenarios. In *AAAI-14 Workshop on Cognitive Computing for Augmented Human Intelligence*. Québec City, Québec, Canada.
- Yu-Ting Li, & Juan P. Wachs. (2014b). Linking Attention to Physical Action in Complex Decision Making Problems. In *the 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. San Diego, CA, US.
- Zweig, G., & Russell, S. (1998). Speech recognition with dynamic Bayesian networks. Retrieved from <http://www.aaai.org/Papers/AAAI/1998/AAAI98-024.pdf>

APPENDICES

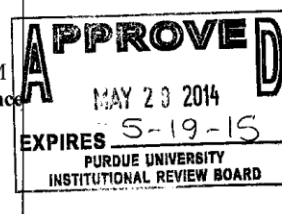
Appendix A Consent Form

Participants were required to read and sign a consent form before they start any experiments. The experimenter will read the consent form to participants blind or visually impaired. The approved consent form by Purdue University Institutional Review Board (IRB) is attached below.

Research Project Number _____

RESEARCH PARTICIPANT CONSENT FORM
Assistive Technology: Institute for Accessible Science
Bradley Duerstock
 Purdue University
 Biomedical Engineering

For IRB Office Use Only



Purpose of Research

The purposes of this research are 1) to develop and test assistive technologies that support persons with disabilities in pursuing education and careers in science and engineering and 2) to investigate the impact of such technologies on the education and career plans of persons with disabilities.

Specific Procedures

You will be asked to use assistive technologies to perform simulated laboratory activities (using everyday substances such as water or saline, exploring and navigating images through a haptic device with feedback from computer speakers and vibration devices attached to clothes) and answer questions about your comfort with and perceptions of the technology. Additionally, you will be observed and videotaped as you complete the activities.

Duration of Participation

Each session will last approximately one hour. You may participate in up to five sessions as your schedule permits.

Risks

The risks of participation in this research are no greater than every day activities. You will be performing normal laboratory tasks with inert (water) supplies. You will be able to perform the activities at your own pace and can stop at any time to rest during or between activities. You will not be asked to perform any activities that cause them discomfort. If you experience discomfort at any time for any reason, you can stop the activity.

Benefits

You will be testing technology that could possibly benefit your ability to perform laboratory activities. Additionally, your participation in the research may improve society's understanding of how technology can benefit persons with disabilities who are interested in pursuing science careers.

Confidentiality

The project's research records may be reviewed by the National Institutes of Health and by departments at Purdue University responsible for regulatory and research oversight. Photographs or digital images and/or video you performing the tests will be used for scientific publications and conference presentations with explicit permission from you. No records will ever be kept associating your name or any other identifiable personal information with these images. All identifiable data and research records will be stored in a locked cabinet. Electronic video recordings will be stored on a secure, password protected server which only the research team will have access to. Each participant will be assigned an arbitrary code that is linked to their identity. The key for these codes will be kept in a location separate from the data. Only the research team will have access to the identified data.

Voluntary Nature of Participation

You do not have to participate in this research project. If you agree to participate you can withdraw your participation at any time without penalty.

Research Project Number _____

Contact Information:

If you have any questions about this research project, you can contact Bradley Duerstock (bsd@purdue.edu; 765-496-2364). If you have concerns about the treatment of research participants, you can contact the Institutional Review Board at Purdue University, Ernest C. Young Hall, Room 1032, 155 S. Grant St., West Lafayette, IN 47907-2114. The phone number for the Board is (765) 494-5942. The email address is irb@purdue.edu.

_____ I give permission to use my photographs or video images for presentations and publications.

_____ I do not give permission to use my photographs or video images for presentations and publications.

Documentation of Informed Consent

I have had the opportunity to read this consent form and have the research study explained. I have had the opportunity to ask questions about the research project and my questions have been answered. I am prepared to participate in the research project described above. I will receive a copy of this consent form after I sign it.

Participant's Signature

Date

Participant's Name

Researcher's Signature

Date

Appendix B Experiment Procedures

Experiment 1:

Each participant was presented with all four image features. The experiment order to test modalities for each feature was randomized. Also, the experiment order to test each candidate modality for a feature was also randomized. Blindfolded subjects were required to put on an eye mask before they start the experiment. Before each new mapping from image features to sensory modalities was tested, a training trial was first performed. The two tasks used in real experiments were also tried in the training trial. Therefore, participants knew what they were going to find in the experiments. Participants were asked to press a button on the haptic device they use right before they began a task and press the button again once they got the answer. The button was used to record response time for each task. After each task was performed and the end-of-task button was pressed, participants were required to speak out the answer they got. Then they could start the next task once they pressed the start-of-task button. The right answers were released to participants in the training trial, while kept sealed in real experiments. After the experiments for each feature, participants could choose to have a short break or not. The whole experiment normally lasted from one hour to one and a half hours.

Experiment 2:

The second experiment followed the similar process of the first experiment. Each participant was presented with both multimodal method and print-out tactile images. Also, the test order was randomized to decrease learning effect. Training trials were performed

before any real experiments. While they were testing with tactile images, the response time was recorded by experimenters. After the whole experiments, feedbacks from participants were collected to analyze how people like the multimodal method studied in this thesis and what do they think about this method. This experiment normally took from half an hour to one hour.

VITA

VITA

Ting Zhang
School of Industrial Engineering, Purdue University

Education

B.S., Software Engineering, 2012, Zhejiang University, Hangzhou, P.R. China
M.S.I.E, Industrial Engineering, 2014, Purdue University, West Lafayette, Indiana

Work Experience

01/2014 – 05/2014	Teaching Assistant, School of Industrial Engineering, Purdue University, West Lafayette, IN.
01/2013 – Present	Research Assistant, School of Industrial Engineering, Purdue University, West Lafayette, IN.

Publications

1. **Ting Zhang**, Greg J. Williams, Bradley S. Duerstock, Juan P. Wachs, “Multimodal Approach to Image Perception of Histology for the Blind or Visually Impaired”, *In Proceedings of the 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2014* (to be appeared).
2. Williams, G.J., **Zhang, T.**, Lo., A., Gonzales, A., Baluch, D.P., Duerstock, B.S. “3D Printing Tactile Graphics for the Blind: Application to Histology”, *In Proc. of Annual Rehabilitation Engineering Society of North America Conference 2014*, June 11-15, 2014.
3. Yu-Xin Ma, Jia-Yi Xu, Di-Chao Peng, **Ting Zhang**, Cheng-Zhe Jin, Hua-Min Qu, Wei Chen, and Qun-Sheng Peng. *A Visual Analysis Approach for Community Detection of Multi-Context Mobile Social Networks*. [J] *Journal of Computer Science and Technology*, 2013, V28(5): 797-809

2nd Page of VITA

Academic Experience

- 06/2013- Present** **A Multimodal Image Perception System for Visually Impaired People**
 Research Assistant, School of Industrial Engineering, Purdue University
- Developing a system conveys 2D image features, like shape, intensity and color, to visually impaired people through multiple modalities other than visual, such as haptics and auditory.
- 11/2012-05/2013** **Independent Light Microscope Operation for Students with Mobility and Visual Impairments**
 Research Assistant, School of Industrial Engineering, Purdue University
- Implemented function of auto loading slides onto the microscope.
 - Implemented function of autofocus and auto-exposure.
 - Developed the webpage interface.
- 09/2012-12/2012** **HMM Based Approach One Shot Gesture Recognition**
 School of Industrial Engineering, Purdue University
- Participated in implementing a gesture recognition system using only one video as training sample.
 - Developed an algorithm based on a Hidden Markov Model (HMM) and took HOG and HOF as features of each video frame.
- 08/2011-06/2012** **A Visual Analysis Approach for Context-Aware Community Detection of Mobile Social Networks**
 Visualization Analytics Group, CAD&CG Lab, Zhejiang University
- Participated in developing the community detection system, including visual representation of communities.
 - Modified current community detection algorithm based on context information.
 - Participated in designing and implementing two user studies.
- 07/2011-09/2011** **Methods for Inferring Semantic Information from GPS Logs of Human Mobility**
 Visualization and Interface Design Innovation Lab, the University of California-Davis
- Processed taxi GPS logs and developed a platform for conducting visual analytics on human mobility data based on GPS logs.
 - Compared GPS logs with Google Map routes to help users make decisions.

Volunteer Experience

- 01/2013-12/2013** **Boiler Out Volunteer Program**
- 03/2011** **The 2011 ACM Conference on Computer Supported Cooperative Work** Local Student Volunteer
- 10/2011** **The 8th National Para Games, P.R. China** Student Volunteer

PUBLICATIONS

PUBLICATIONS

1. **Ting Zhang**, Greg J. Williams, Bradley S. Duerstock, Juan P. Wachs, “Multimodal Approach to Image Perception of Histology for the Blind or Visually Impaired”, *In Proceedings of the 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2014* (to be appeared).
2. Williams, G.J., **Zhang, T.**, Lo., A., Gonzales, A., Baluch, D.P., Duerstock, B.S. “3D Printing Tactile Graphics for the Blind: Application to Histology”, *In Proc. of Annual Rehabilitation Engineering Society of North America Conference 2014*, June 11-15, 2014.