

**MULTIMODAL DIGITAL IMAGE EXPLORATION WITH  
SYNCHRONOUS INTELLIGENT ASSISTANCE FOR THE BLIND**

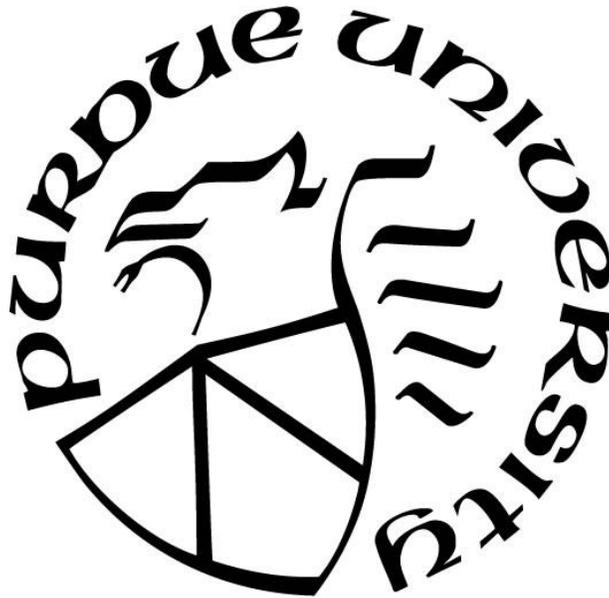
by  
**Ting Zhang**

**A Dissertation**

*Submitted to the Faculty of Purdue University*

*In Partial Fulfillment of the Requirements for the degree of*

**Doctor of Philosophy**



School of Industrial Engineering

West Lafayette, Indiana

May 2020

**THE PURDUE UNIVERSITY GRADUATE SCHOOL  
STATEMENT OF COMMITTEE APPROVAL**

**Dr. Bradley S. Duerstock, co-Chair**

School of Industrial Engineering, School of Biomedical Engineering

**Dr. Juan P. Wachs, co-Chair**

School of Industrial Engineering

**Dr. Dan Goldwasser**

School of Computer Science

**Dr. Tahira Reid**

School of Mechanical Engineering

**Approved by:**

Dr. Abhijit Deshmukh

*This dissertation is dedicated to my parents and family.*

## ACKNOWLEDGMENTS

Foremost, I would like to express my heartfelt 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 dissertation. 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. Dan Goldwasser and Prof. Tahira Reid for providing many valuable insights and comments that improved the content of this dissertation.

I also really appreciate the support and help from my dearest friends and lab mates, Jackie Cha, Glebys Gonzalez, Shruthi Suresh, Tong Zhu, Yuling Gao, Siming Xu, Hairong Jiang, Yu-Ting Li, Daniela Chanci Arrubla, Juan Antonio Barragan, Oscar Rincon Guevara, Mandira Marambe, Naveen Madapana, Edgar Rojas-Munoz, Shann Palaniappan, Natalia Sánchez, Chenxi Xiao, Xingguang Zhang.

This dissertation is supported by the National Institute for Health Director's Pathfinder Award to Promote Diversity in the Scientific Workforce (1DP4GM096842-01) and the Partnerships for Innovation program through the National Science Foundation. I would also thank the State of Indiana through support of the Center for Paralysis Research at Purdue University.

# TABLE OF CONTENTS

LIST OF TABLES.....	9
LIST OF FIGURES.....	10
ABSTRACT.....	13
1. INTRODUCTION.....	14
1.1 Significance.....	15
1.2 Definitions.....	15
1.3 Research Problem.....	16
1.3.1 Research Question 1 (RQ1).....	16
1.3.2 Research Question 2 (RQ2).....	16
1.3.3 Research Question 3 (RQ3).....	17
1.4 Overview of the Document’s Structure.....	17
2. LITERATURE REVIEW.....	18
2.1 Assistive Technologies for Image Exploration.....	18
2.1.1 Image Exploration Using Physical Media.....	19
2.1.2 Digital Image Exploration.....	20
2.2 Tactile Exploration Procedures.....	22
2.2.1 EPs for Internal Object Properties.....	22
2.2.2 EPs to Measure Spatial Relations.....	23
2.3 Assisting Strategies for Exploration-based Tasks.....	26
2.3.1 Computer-aided Image Exploration.....	26
2.3.2 Assistive Strategies for 3D-space Exploration.....	27
2.4 Sensory Substitution.....	27
2.4.1 Tactile-visual Substitution.....	28
2.4.2 Auditory-visual Substitution.....	28
2.5 Summary.....	29
3. METHODOLOGY.....	30
3.1 Modeling Exploration Procedures.....	30
3.1.1 Temporal Representation of Data.....	33
Angles.....	34

Contextual Information .....	34
Reference switch.....	35
3.1.2 Spike-timing Neural Network .....	35
General Network Configuration. ....	36
Number of Neurons .....	37
Network Training.....	38
3.1.3 Classification .....	39
Representation of SNN-Encoded Features.....	39
Distance Metrics .....	40
3.1.3.1 Decision Fusion using Dempster-Shafer Theory (DST) .....	41
3.2 Designing an Assistance Model.....	42
3.2.1 The goals of exploration procedures .....	43
3.2.2 Exploratory Aids .....	45
Contour Compass.....	45
Path Shortcut .....	46
Border Projection.....	47
Contour Neighbors.....	48
3.3 Sensory-substituted User Interface .....	49
3.4 Performance measurement .....	51
3.4.1 Task performance .....	51
Task completion time .....	51
Accuracy of image understanding .....	51
3.4.2 Workload Assessment.....	52
Rating scales .....	52
Physiological measurements .....	53
3.5 Summary .....	53
4. EXPERIMENTS AND RESULTS .....	54
4.1 Learning Exploration Procedures .....	54
4.1.1 Data collection.....	54
4.1.2 Evaluation .....	55
Table 4.1 The precision rate, recall rate and f1 score of the results. ....	56

Time complexity analysis.....	57
Comparisons with DTW and HMM .....	58
Recognition of partial EPs .....	60
Training SNNs with different properties of neurons .....	61
4.2    Developing an Assistance Model .....	63
4.2.1    Data collection.....	64
4.2.2    Results .....	64
4.3    Sensory Substituted User Interface.....	64
4.3.1    Participants .....	65
4.3.2    Tasks and Measurements .....	65
Task 1: Evaluation of BP and CN.....	65
Task 2: Evaluation of CC and PS .....	67
Subjective Measurements.....	67
4.3.3    Procedure .....	68
4.3.4    Experimental Results .....	68
Border Projection.....	68
Contour Neighbors.....	70
Contour Compass.....	72
Path Shortcut .....	74
4.3.5    Discussion.....	76
Comparisons of feedback .....	77
Comparisons of exploratory aids .....	77
Exploratory aids integration .....	78
4.4    Evaluation of the computer-aided image exploration system.....	78
4.4.1    Participants .....	78
4.4.2    Tasks and Measurements .....	79
4.4.3    Experiment Apparatus .....	79
4.4.4    Procedure .....	80
4.4.5    Experimental Results .....	81
Task performance .....	81
Workload Assessment .....	83

4.4.6	Discussion.....	86
	Comparisons with the baseline .....	86
	Comparisons with human assistant.....	86
4.5	Summary .....	87
5.	CONCLUSIONS AND FUTURE WORK.....	88
5.1	Adaptive Human Computer Interface .....	89
5.2	Virtual Reality for Individuals with Visual Impairments.....	90
5.3	Exploratory Aids for Outdoor Exploration .....	90
	REFERENCES .....	91
	VITA.....	100
	PUBLICATIONS .....	101

## LIST OF TABLES

Table 2.1 Summary of observed exploration procedures from literature. ....	25
Table 3.1 Classified exploration procedures. ....	31
Table 3.2 Rules for contextual information assignment. ....	34
Table 3.3 Goals of exploration procedures. ....	44
Table 3.4 Summary of exploratory aids. ....	49
Table 4.1 The precision rate, recall rate and f1 score of the results. ....	56
Table 4.2 Classification accuracies with different type of neurons. ....	62
Table 4.3 Testing Images for Task 1. ....	66
Table 4.4 Testing Images for Task 2. ....	67
Table 4.5 Test images to evaluate the system with exploratory aids. ....	81

## LIST OF FIGURES

Figure 2.1 Example tactile graphics on a microcapsule paper. (a) Original image; (b) Swelled tactile graphics of (a). .....	19
Figure 2.2 Example of embossed tactile graphics. (a) Original line graph; (b) Embossed tactile graphics of (a). .....	20
Figure 2.3 Examples of off-the-shelf haptic devices. (a) Logitech WingMan Force Feedback Mouse; (b) Force Dimension Omega 6. ....	21
Figure 3.1 System Architecture .....	30
Figure 3.2 Proposed framework to learn exploration procedures.....	33
Figure 3.3 Example image and its index map. (a) Original image; (b) Index map. ....	35
Figure 3.4 Example Spiking Neural Network. The nodes are neurons and the links are synapses that connect the neurons. The width of the link represents the weight, while the color represents the conduction delay.....	37
Figure 3.5 Firing pattern of neurons. The top raster plot shows the fired neurons at each time stamp and the bottom plot shows the value of input data at each time stamp. ....	38
Figure 3.6 Examples of synapse weights fluctuation. ....	39
Figure 3.7 Histogram of the goals for exploration procedures. ....	45
Figure 3.8 Examples of path shortcut. (a) Shortcut between two neighbors; (b) No shortcut for object 1 and 3; (c) Shortcut from object 2 to the top, right and bottom image border and no shortcut from object 2 to the left image border. ....	47
Figure 3.9 Examples of border projection. (a) Aids on the nearest image border; (b) No aids for blocked objects. ....	48
Figure 3.10 Sensory substituted interface. ....	50
Figure 3.11 Pipeline for image processing (a) replicated image, and (b) original image. ....	52
Figure 4.1 A haptic-based image exploration system. ....	55
Figure 4.2 Examples of captured EP AP performed by participants. ....	55
Figure 4.3 Confusion matrix of classification accuracy for the proposed framework with 95.89% accuracy.....	56
Figure 4.4 Classification accuracy over different number of training samples. ....	57
Figure 4.5 Confusion matrix of recognition accuracy for SNN with NHNF descriptor, with accuracy 88.68%. ....	58
Figure 4.6 Confusion matrix of recognition accuracy for DTW, with accuracy 61.30%. ....	59

Figure 4.7 Confusion matrix of recognition accuracy for HMM, with accuracy 28.70%. .....	60
Figure 4.8 Examples of complete and incomplete EPs. ....	61
Figure 4.9 The responses to stimulus of different types of neurons. ....	63
Figure 4.10 Normalized histograms of goals for image exploration. ....	64
Figure 4.11 Average task completion time for Task 1 with Border Projection. ....	68
Figure 4.12 Level of helpfulness of Border Projection. ....	69
Figure 4.13 Level of interruption of Border Projection. ....	70
Figure 4.14 Average task completion time for Task 1 with Contour Neighbors. ....	70
Figure 4.15 Level of helpfulness of Contour Neighbors. ....	71
Figure 4.16 Level of interruption of Contour Neighbors. ....	72
Figure 4.17 Performance comparisons among the baseline and Contour Compass via three types of sensory feedback: (a) Average task completion time; (b) Replication accuracy. ....	73
Figure 4.18 Level of helpfulness of Contour Compass. ....	73
Figure 4.19 Level of interruption of Contour Compass. ....	74
Figure 4.20 Performance comparisons among the baseline and Path Shortcut via three types of sensory feedback: (a) Average task completion time; (b) Replication accuracy. ....	75
Figure 4.21 Level of helpfulness of Path Shortcut delivered through magnet attraction, sound and vibration. ....	76
Figure 4.22 Level of interruption with Path Shortcut delivered through magnet attraction, sound and vibration. ....	76
Figure 4.23 Multimodal image exploration system with four exploratory aids. ....	79
Figure 4.24 Setup for physiological data collection. (a) The participant is wearing a ear clip to collect heart rate data and sensors under the index and middle finger to measure skin conductance; (b) a close-up view of the sensors. ....	80
Figure 4.25 Comparisons of average task completion time among the three interfaces: baseline, human assistant and proposed computer assistant. ....	82
Figure 4.26 Comparisons of average replication accuracy among the three interfaces: baseline, human assistant and proposed computer assistant. ....	83
Figure 4.27 Comparisons of average mental workload measured by NASA TLX questionnaire among the three interfaces: baseline, human assistant and proposed computer assistant. ....	84
Figure 4.28 Comparisons among the three interfaces: baseline, human assistant and proposed computer assistant in terms of (a) normalized average heart rate and (b) normalized average R-R intervals. ....	85

Figure 4.29 Comparisons among the three interfaces: baseline, human assistant and proposed computer assistant in terms of normalized average SCR frequency. ....85

## ABSTRACT

Emerging haptic devices have granted individuals who are blind the capabilities to explore images in real-time, which has always been a challenge for them. However, when only haptic-based interaction is available, and no visual feedback is given, image comprehension demands time and major cognitive resources. This research developed an approach to improve blind people's exploration performance by providing assisting strategies in various sensory modalities, when certain exploratory behavior is performed. There are three fundamental components developed in this approach: the user model, the assistance model, and the user interface. The user model recognizes users' image exploration procedures. A learning framework utilizing spike-timing neural network is developed to classify the frequently applied exploration procedures. The assistance model provides different assisting strategies when certain exploration procedure is performed. User studies were conducted to understand the goals of each exploration procedure and assisting strategies were designed based on the discovered goals. These strategies give users hints of objects' locations and relationships. The user interface then determines the optimal sensory modality to deliver each assisting strategy. Within-participants experiments were performed to compare three sensory modalities for each assisting strategy, including vibration, sound and virtual magnetic force. A complete computer-aided system was developed by integrating all the validated assisting strategies. Experiments were conducted to evaluate the complete system with each assisting strategy expressed through the optimal modality. Performance metrics including task performance and workload assessment were applied for the evaluation.

# 1. INTRODUCTION

Traditional image presentations for the blind or visually impaired (BVI) individuals, such as braille and tactile graphics, are printed on physical media (similar to tactile paper and 3-D printouts) that cannot convey the complexity of visual information associated with scientific imaging and to do so in real-time (Csapó, Wersényi, Nagy, & Stockman, 2015). Tactile graphics require images to be simplified as grayscale line drawings first, which omits detailed information and can be biased depending on the person generating them. Moreover, tactile image production is time consuming and expensive. The whole process to produce one tactile image can take up to several hours (Sheppard & Aldrich, 2001). Emerging technologies, such as low cost, compact and precise haptics devices, have granted individuals who are blind the capability to explore digital images in real-time (Crossan & Brewster, 2006; Iglesias et al., 2004; Zhang, Duerstock, & Wachs, 2017). Various interfaces utilizing haptics feedback have been developed and studied substituting visual information, with alternative channels of information (Dewhurst, 2009).

However, when only haptic-based interaction is available and no visual feedback is given, image comprehension demands time and major cognitive resources (Zhang et al., 2017). Different from viewing images with vision, the understanding of an image through tactile sensation is limited to the area of contact and completed by mentally stitching together the fragments. Users need to come up with compensation mechanisms and strategies to overcome this limitation and become proficient in image exploration. For example, to distinguish the shapes of objects on a tactile paper, people who are blind will repetitively move their finger along the raised lines that indicate the contour of objects.

To facilitate such haptic-based image analysis, there is a need for new computational models and techniques that enable the understanding of the underlying user strategies, and provide assistance through the process of image exploration. There are three fundamental components to design and develop such an intelligent human-machine interaction system: the user model, the assistance model, and the user interface. In the context of image exploration for the blind community, the user model recognizes the exploratory behavior of a user. User behavior in this scenario refers to the exploration procedures (EPs) that users adopt during image exploration (Ungar, Blades, & Spencer, 1995; Vinter, Fernandes, Orlandi, & Morgan, 2012). The assistance model provides synchronized guidance based on the performed user behavior, and then deliver

guidance through the user interface. The user interface determines the way that users receive such guidance, such as acoustic cues or haptics feedback.

## 1.1 Significance

Scientific image comprehension has been a roadblock for BVI individuals, in particular academically. In instructional settings, images are the major and popular way to illustrate concepts for STEM subjects. BVI students are often left behind in class and discouraged to pursue STEM related careers, which can potentially provide a higher-paid job that leads to higher quality of life (Rothwell, 2013). Real-time systems that allow people who are BVI to explore images are currently based on either haptic devices that have one contact points, or touch screens with two or three contact points. As for tactile interaction, the understanding of an image is achieved through the information covered by the exploration, the efficiency of image exploration is decreased due to the limitation of contact points. The developed system with intelligent assistance aims to enhance and facilitate the image exploration experience of BVI users.

Developing such an intelligent assisting mechanism in the scope of image exploration for BVI users, also paves the way for other research that aims to enhance the life of BVI people. The dissociation occurred in instructional setting, also occurs in virtual interaction. While social networks grow, the current technologies have not made it possible for BVI people to participate due to the limitation handling visual information, which is a major component of social media (e.g. pictures in Facebook). The proposed system could also potentially improve blind people's social status by providing them the equal opportunities to access visual information. This proposed intelligent assisting mechanism can also be adapted for real world exploration, such as outdoor and indoor navigation systems.

Whereas is for social, scientific or navigation purposes, the proposed approach has the potential to help the BVI users to build a mental map of the visual information more efficiently than what has been currently offered to this market.

## 1.2 Definitions

In this section, some important terms used throughout this document are defined.

1. **Image Exploration:** People who are BVI explore an image to understand its content.

2. **Exploration procedures:** Repetitive motions of users moving a pointer on a 2D image, are classified into several categories and defined as exploration procedures.
3. **Exploratory Aids:** Strategies provided by the system to enhance the image exploration for the user.
4. **Sensory Substitution:** The process of delivering information from a different sensory modality than the one that is often used due to preference or lack of availability.
5. **Haptic-based:** interfaces using force based control.

### 1.3 Research Problem

The design and implementation of an interaction scheme that provides adaptive assistance for haptic-based image exploration for individuals who are BVI includes: (a) developing a user model that can recognize user behavior during exploration; (b) developing an assistance model that provide exploratory aids based on the user model; and (c) compare and validate different user interfaces.

#### 1.3.1 Research Question 1 (RQ1)

*How to learn effective exploration procedures for individuals who are BVI?*

User behavior in the context of haptic-based image exploration, refers to the exploration procedures commonly adopted by users who are BVI. By answering this question, it sets the footstone to develop an intelligent assisting system.

#### 1.3.2 Research Question 2 (RQ2)

*What's the effective mechanism for assistance?*

Having identified effective exploration procedures (RQ1), proper exploratory aids need to be designed. Furthermore, various sensory modalities could be used to deliver such strategies. However, some modalities may be easier accessed and perceived than others which can lead to an overall better experience and performance. This question aims to understand what are the proper exploratory aids and what is the optimal sensory modality to express the strategies. Experiments and surveys will be conducted to compare different user interfaces.

### 1.3.3 Research Question 3 (RQ3)

*What are the metrics to measure the performance of the proposed system?*

Once we are able to recognize user behavior (RQ1) and deliver assisting strategies through optimal sensory modalities (RQ2), the proposed system needs to be evaluated and compared with other approaches to assess its potential benefit.

## 1.4 Overview of the Document's Structure

This chapter introduced the researched problem and the motivation behind it. Chapter 2 provides a review of the literature on the topics related to this research. Chapter 3 explains the proposed idiosyncrasy used to develop the proposed system. The following chapter discusses the experiments and results. Finally, in Chapter 5, conclusion and future work is discussed.

## **2. LITERATURE REVIEW**

This chapter gives an overview of the state of art research related to this proposal. First, we reviewed current assistive technologies that help individuals who are BVI to understand images, including their advantages and limitations in terms of the modality expressiveness and easiness of interaction. Another important aspect in our work involves learning and recognizing BVI's users behaviors while they are exploring images. User behaviors are referred as exploration procedures (EPs) throughout this proposal. In the second section, various EPs are summarized from the literature. Different aspects of the EPs are also discussed, including what the user's goals are when using the EP, their mode of exploration (using fingers or palm), and their applicability in different exploration tasks. The third part of this literature review includes different assisting strategies to enhance the exploration experience. While enhancing user performance during image exploration by providing guidance has not drawn a lot of attention in past studies, studies has been done to facilitate BVI individuals exploring real world spaces. Learning from the state of the knowledge about strategies applied to 3D exploration task can provide a good starting point to the design of assisting strategies for 2D image exploration. Last, sensory substitution is discussed, including substituting vision with either audio, touch or a mix of them.

### **2.1 Assistive Technologies for Image Exploration**

Ranging from low-tech tactile papers to high-tech real-time sensory substitution systems, there has been a substantial amount of work developing assistive technologies to allow images be accessible to BVI individuals. The state of the art in image exploration techniques can be summarized into two categories: exploring images physically or digitally. Physical image representation includes traditional tactile graphics and modern 3D-printed models, while digital image representation utilizes computer generated virtual graphics with haptics and audio feedback. In the following sections, techniques supporting physical image exploration are introduced, followed by advanced digital image exploration systems.

### 2.1.1 Image Exploration Using Physical Media

Tactile graphics is the most common approach to deliver two-dimensional visual information to individuals who are BVI, such as diagrams, pictures and charts. They are widely used in instructional settings to teach students understand visual based concepts, such as geometry, physics, chemistry and other STEM related fields (Sheppard & Aldrich, 2001). Tactile graphics can be generated in many different ways, including manually created with crafting materials, printed on a microcapsule paper, thermoformed with drawing molds, or embossed with braille embossers (Gupta, Balakrishnan, & Rao, 2017). Printed on a microcapsule paper is a widely used approach as it's relatively easy to make and cost less than other forms (Yu & Brewster, 2002). The microcapsule paper swells when heated, to different height based on the color intensity on the image. Darker areas appear higher than lighter areas. Figure 2.1 shows an example of the original image and its tactile graphics on a microcapsule paper.

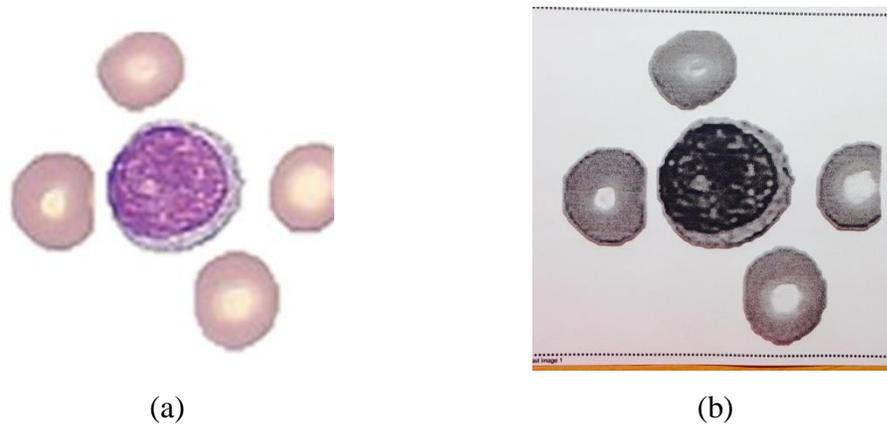
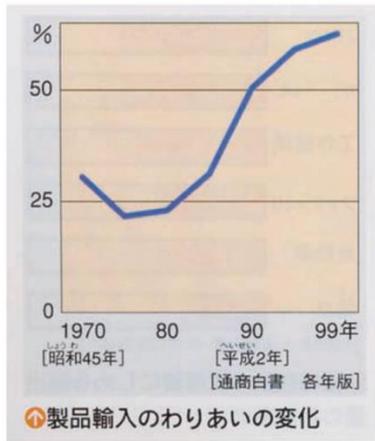
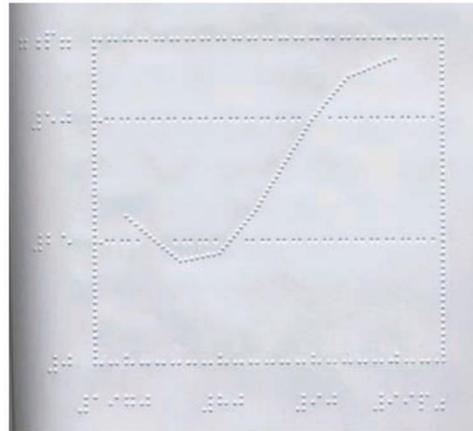


Figure 2.1 Example tactile graphics on a microcapsule paper. (a) Original image; (b) Swelled tactile graphics of (a).

As technologies advance, embossers used to print braille were upgraded to print tactile graphics with raised dots. Advanced embossers can create dots at different heights, enabling a richer content on the image (Kouroupetroglou et al., 2016). Figure 2.2 shows an example of the image, and its tactile graphics (Kaneko & Ooduchi, 2010).



(a)



(b)

Figure 2.2 Example of embossed tactile graphics. (a) Original line graph; (b) Embossed tactile graphics of (a).

Although tactile graphics is the most widely applied image representation for BVI people, it can not be used to deliver complex visual information and often requires additional description from human assistants. In addition to tactile paper, 3D printing technology becomes an alternative as 3D printers becomes more affordable (Stangl, Kim, & Yeh, 2014). Studies investigated different factors that could affect the interpretation of 3D printed images, including the height and size of 3D printed objects compared to the original images, and the printed roughness of material (Williams et al., 2014). It was found although 3D printed tactile images can represent more complicated visual information than tactile paper, it requires fine adjustment of the 3D-printing parameters and additional 3D modeling techniques to convert 2D images into 3D objects.

### 2.1.2 Digital Image Exploration

With the advance of touchscreen devices, tactile displays became handy, since they can provide a tactile sensation to fingers while moving on a touch screen. TeslaTouch (Xu, Israr, Poupyrev, Bau, & Harrison, 2011) is an example of tactile displays that generates voltage differences between the finger and the touch screen. This inexpensive and easily maintained device supports various tactile sensations, which facilitates the exploration and interpretation of complex images. However, as opposed to tactile graphics, fingers need to actively sliding on the touch screen to feel the sensation.

Besides touch screens, force feedback devices have draw attention in the field for image exploration by BVI individuals (Colwell, Petrie, Kornbrot, Hardwick, & Furner, 1998; Wies, O’Modhrain, Hasser, Gardner, & Bulatov, 2001; Zhang et al., 2017). Figure 2.3 shows two examples off-the-shelf devices that can provide force feedback, the Logitech WingMan Force Feedback Mouse (a) and Force Dimension Omega 6 (b). In this proposal, we utilized a multimodal image exploration system that uses Omega 6 as its haptic device (Zhang et al., 2017). The Force Dimension Omega 6 has six degrees of freedom and can provide force feedback in a 3D space.



Figure 2.3 Examples of off-the-shelf haptic devices. (a) Logitech WingMan Force Feedback Mouse; (b) Force Dimension Omega 6.

Instead of using fingers, BVI individuals can feel the virtual graphic rendered by the computer, by moving the haptic device across the image. Tactile feedback is provided according to the position of the haptic device. Yu and Brewster developed a multimodal system that people can use to explore bar graphs using a haptic device, together with auditory feedback (Yu & Brewster, 2002). Experimental results indicated significant higher accuracy in understanding the bar charts through the haptic interface compared to traditional tactile diagram. Interactive systems have been developed using haptic devices and sound to facilitate the learning of visual concept, such as astronomy (Tuominen, Kangassalo, Hietala, Raisamo, & Peltola, 2008) and maps (Kaklanis, Votis, & Tzovaras, 2013). Students showed a more accurate understanding of the concepts and more interests in learning the complex visual concept. Another haptic-based interface was developed to test the applicability of haptic devices to represent geometries, such as line drawing and curves, and maps and floor plans (Sjöström, Danielsson, Magnusson, & Rasmus-Gröhn, 2003). In the test for geometric graphs, it was reported that 96% of the participants

recognized a particular part of a curve, while the rate of success decreased as the complexity of line drawings increased.

## **2.2 Tactile Exploration Procedures**

As opposed as to using vision to perceive images, exploring visual information through touch sensation is more difficult and complicated. It requires a lot more memory as well as well-developed procedures (Hatwell, Streri, & Gentaz, 2003). As tactile feedback is received through the area of skin that is in contact with the stimulus, the amount of information obtained is limited based on the exploration trajectories that they follow. Therefore, blind individuals have developed different exploration procedures (EPs) to make more effective their interaction with tactile information. It has been observed that these exploration procedures (EPs) are developed with the growth of age and experience (Vinter et al., 2012).

Studies focusing on what type of EPs are adopted by the blind community have been conducted in two major groups of visual properties: local and global visual information. Studies relying on local visual information consider the internal properties of a visual object, such as the shape, size, texture and color to characterize an object. Whereas, global visual information studies (also referred as spatial information) relies on representing the relations between objects, and the object's absolute location to characterize it. In the following sections, different EPs are discussed in terms of the type of visual information that is been used.

### **2.2.1 EPs for Internal Object Properties**

Visual information by touching real-world objects or elevated 2D tactile images are the most common way for individuals who are blind to understand that information. A lot of studies have been conducted to understand the procedures that BVI individuals naturally adopt when exploring objects' internal properties using their hands. Lederman & Klatzky (Lederman & Klatzky, 1987, 1990, 1993) summarized six exploration procedures including dimensions, texture, hardness, weight, and shape. For example, "lateral motion" or "surface sweeping" referred in (Davidson, 1972) is a back-and-forth movement across a small area of surface. People use this procedure to understand the texture, interior structure of an object. "Contour following" is another

often-used EP to understand the exact shape of an object. Different from “surface sweeping”, it is a motion that focuses on the boundaries of objects.

Some of these exploration procedures used for 3D objects can be also extended to 2D tactile images. Vinter et al. summarized seven EPs from studies with children who are BVI using elevated 2D tactile images (Vinter et al., 2012). The study’s goal was to understand the relation between their applied exploration procedures and task performance. Children were asked to explore a two-dimensional tactile pattern using their hands and later draw the perceived pattern on a sheet of paper. Seven exploration procedures were summarized and studied. Procedure “contour following” is the most common procedure found in this task, and indicated a higher correlation with the recognition of object’s shape and size. “Surface sweeping” is the second most common procedure found in this study.

### **2.2.2 EPs to Measure Spatial Relations**

Besides understanding the internal properties of objects, measuring the spatial relations between objects is another challenging visual analysis task for BVI individuals. While peripheral information is accessible through vision, tactile sensation only provides feedback statically at the point of contact. In the experiments conducted by (Ungar et al., 1995) with children who are BVI, participants were asked to explore a circular board with different number of objects placed on it, and then replicate the setup on another board. Four modes of exploration were found in this experiment, including relative exploration, edges exploration, a mix of relative and edges exploration, and pointing & touching in turn. Relative exploration is a back-and-forth movement between two objects to get the position of one object relative to another. Edge exploration is a back-and-forth movement between one object and the boundary of the panel that holds the object. Observers use this type of procedure to get the location of an object relative to the boundary of its absolute reference frame. In the experiments, some participants used one of the EPs alone, while some of them performed both EPs during exploration. Pointing and touching in turn is a movement that children pointed and touched each object separately in turn, following an egocentric reference frame. Results revealed that by using both EPs, participants finished the task with higher accuracy, while the pointing & touching movement gave the least performance (Ungar et al., 1995).

An experiment conducted by (Gaunet, Martinez, & Thinus-Blanc, 1997) with adults who are BVI indicated similar results. Participants were asked to explore a square panel with five

objects of different shapes, and memorize the location of all objects. They were later asked if one of the objects has been moved. Two modes of exploration were identified in the experiment: a “cyclical” exploration and a “back and forth” exploration. Participants visited the objects one by one, and ended by returning to the first object visited by following the “cyclical” exploration. Conversely, with the “back and forth” exploration, participants repeatedly visited the same two objects with a back-and-forth trajectory, then moved to the next two objects. The “back-and-forth” exploration procedure resulted in a lower recall error compared with the “cyclical” exploration.

A more recent study has been conducted to understand the procedures applied by blind users when interacting with large touchscreens (Guerreiro et al., 2015). Counting the number of objects and locating tasks were performed by the participants to observe commonly used procedures. When the user was asked to find an object without prior knowledge, path scan was a commonly applied procedure. It consists of the user moving from left to right horizontally scanning the image. However, when prior knowledge was available to the user, the “to-the-point” procedure was used more often. The latter consists of the user going directly to the memorized location. The experimental results also indicated finding objects on a large surface without prior location knowledge is a time-consuming task, that would require additional support, such as giving instructions to locate nearest object or providing relation information of two neighbors (Kane et al., 2011).

Table 2.1 summarizes the EPs that were reported to be commonly used during tactile exploration. The table includes a verbal description of the EP, its related visual properties.

Table 2.1 Summary of observed exploration procedures from literature.

<b>Exploration Procedure</b>	<b>Description</b>	<b>Related Visual Property</b>	<b>Literature</b>
Contour Following	Dynamic edge following using finger movements	shape, size	(Lederman & Klatzky, 1987)
Enclosure of the global shape	Dynamic molding of the palm and/or fingers to the shape's contours	shape	(Lederman & Klatzky, 1987)
Enclosure of local shapes	Dynamic molding of the fingers to parts of the pattern.	partial shape	(Lederman & Klatzky, 1987)
Pinch procedure	Holding edges in a pincer grip between the thumb and one or more other fingers	size	(Davidson, 1972)
Surface sweeping	Dynamic and usually repetitive movement of one or more fingers or of the palm over the model's surface.	texture	(Davidson, 1972)
Static contact	Stationary contact with the surface without molding.	texture	(Lederman & Klatzky, 1987)
Relative exploration	Back-and-forth movement between two objects.	location	(Ungar et al., 1995)
Edge exploration	Back-and-forth movement between the object and the image boundary.	location	(Ungar et al., 1995)

## 2.3 Assisting Strategies for Exploration-based Tasks

The BVI community often encounters to perform exploration tasks which includes searching in the 2D space, over media as diverse as tactile images and webpages, or even as complex as explorations in the 3D space, such as indoor/outdoor environments. Strategies used by BVI individuals when exploring indoor and outdoor environments is a well-studied research topic, compared to exploration of images and webpages, where there is not much substantial work, in spite that real-world navigation is a more fundamental need for daily-living activities. What is known in terms of assisting strategies for 3D-space explorations provide valuable insights that can be extrapolated to image-based explorations, due to the similarities existed between the tasks. For example, exploring a room in order to understand its layout is somehow similar to exploring a tactile image with several items embedded on it. In this section, we summarize few computer-aided image exploration systems, and review studies that incorporated assisting strategies for BVI individuals. We discuss the requirements in order to adapt those strategies to image-based explorations.

### 2.3.1 Computer-aided Image Exploration

Computer-aided image exploration systems tracks user behavior, and provides guidance or exploration tips based on the current user status. For example, by tracking the position of user's finger, and recognizing the object that the user pointed at, the system can deliver information about the object to the user through speech (Blenkhorn & Evans, 1998; Fusco & Morash, 2015; Löttsch, 1994; Parkes, 1988; Suzuki, Stangl, Gross, & Yeh, 2017). Kurze developed a guidance system for tactile images that used eight vibrating elements indicating directional information (Kurze, 1999). The user was given a list of objects on the image, and two different guidance strategies were developed. The "tourist guide mode" lead users to objects following a sequence based on the object's importance. That system encouraged users to visit important object first. Users could follow the guidance provided by the system in order to explore the next object. The "scout mode" delivered guidance towards the object that the user specified beforehand. In that experiment, Kurze also compared guidance with vibration and sound, and found that vibrational guidance was more effective than auditory feedback. Later studies have proposed additional strategies to enhance blind users' exploration through large touchscreens. For example, "Edge Projection" is a strategy that

projects the location of objects onto the edges of the touchscreen. Users could explore these objects by sliding their finders along the horizontal or vertical boundary of the touchscreen (Kane et al., 2011). Experimental results revealed that “Edge Projection” significantly reduced the task exploration time.

### **2.3.2 Assistive Strategies for 3D-space Exploration**

Various navigational strategies have been developed to assist BVI people explore 3D spaces. One strategy commonly applied is providing close-by point of interest (POI) information along the exploration process (Abowd et al., 1997; Blum, Bouchard, & Cooperstock, 2012; Golledge, Klatzky, Loomis, Speigle, & Tietz, 1998; Golledge, Loomis, Klatzky, Flury, & Yang, 1991). L. Picinali et al. developed a system using acoustic cues to represent the surrounding environment to a given location (Picinali, Afonso, Denis, & Katz, 2014). In that system, as users were moving along a corridor, sounds simulating events were generated and rendered according to the distance between the user and the POI. For example, when the user would be approaching a restroom, a toilet flushing sound would be played and get louder as the user is getting closer to the restroom (Picinali et al., 2014). Another type of exploration systems allowed users to annotate their own POI during exploration and received notifications every time that they were approaching those POIs, such as the Talking Points (Stewart et al., 2008) system.

A different line of research focused on virtually generated spaces in order to facilitate the design and development of various assisting strategies. For example, users could trigger a “look-around” mode in a virtual environment navigation task to find out what objects are around them (Lahav et al., 2018). Different sound tones and beeps were assigned to represent objects, while force feedback was utilized to encode the distance between the user and its surrounding objects.

## **2.4 Sensory Substitution**

Blind people cannot see because they lose the ability to transmit the sensory signals to the brain (Bach-y-Rita & W. Kercel, 2003). However, it was proposed to replace the ability of a defective sensory modality, by other functioning sensory channel that can alternatively convey the missing sensory information to the brain. 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 times twenty array on the skin of their back, participants were able to distinguish different objects after they went through extensive training, (from 20 to 40 hours). Tactile and audio sensory substitutions are the two most popular sensory substitution approaches currently most studied, which are illustrated in the next section.

#### **2.4.1 Tactile-visual Substitution**

Since the introduction of tactile-vision sensory substitution (TVSS) idea by Bach-y-Rita in 1969, there has been ongoing research in that field as part of assistive technologies. From image perception to video understanding, from obstacles 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 (Hsu et al., 2013; Johnson & Higgins, 2006; Kaneko & Ooduchi, 2010; Way & Barner, 1997).

Wireless electrotactile devices have been also proposed as an aid to navigate a variety of environments (Nguyen et al., 2013). For example, (Kajimoto et al., 2014) takes advantages of smartphones to deliver directions through vibrational feedback. The system consists of an electrotactile display with 512 electrodes, a smartphone and an LCD. Participants were able to get a view of the surroundings by taking photos using the smartphone embedded camera. Images were processed by the smartphone and then converted to vibration through each electrode. While TVSS enabled BVI people navigate environments through image perception, the low resolution of tactile sensors compared to the visual system is the main drawback of this method.

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).

#### **2.4.2 Auditory-visual Substitution**

The auditory channel was also used to convey image properties. The frequency of auditory pitch, binaural intensity and phase differences, sound loudness, and specific sets of tones have been mapped to different image properties (Capelle, Trullemans, Arno, & Veraart, 1998). Mappings were found to be useful not only between auditory pitch and object location, but between

auditory pitch and object size (Evans & Treisman, 2011). In addition to auditory pitch, sound level 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). Research has also been conducted to convey live video to an individual through auditory pitch and loudness (Meijer, 1992). In most auditory-visual substitution systems, only grayscale images were utilized and color information was discarded all together. More recently, a sensory substitution system, called “EyeMusic”, was able to convey real-time visual information through small computer or smartphone using sound delivered through stereo headphones. Furthermore, it would represent color information through different musical instruments as well. Due to the difficulty associated with the mapping between different musical instruments and color, only five colors were conveyed in a distinctive fashion: white, blue, red, green and yellow (Abboud, Hanassy, Levy-Tzedek, Maidenbaum, & Amedi, 2014).

Early blind participants showed increased performance in localization and object recognition (Arno, Capelle, Wanet-Defalque, Catalan-Ahumada, & Veraart, 1999) through auditory-visual substitution when they were trained as opposed to those that were not trained. Training is key to the success of such systems since auditory-vision substitution involves the memorization of audio patterns and their associated visual cues (Arno et al., 1999). In addition, when the users focus on auditory feedback, it can decrease their awareness about the environment around them, thus leading to potential safety concerns (Meers & Ward, 2004).

## **2.5 Summary**

This literature review presented the state-of-art in assistive technologies designed to support BVI individuals in image comprehension. Then, the concept of exploration procedures is presented together with a discussion of commonly applied exploration procedures. Then, I presented an overview of assisting strategies aimed to enhance image exploration performance. In this context, I discussed the challenges associated with the design of accessible real-time image exploration devices. This review also indicates that the hardware at this point is mature to facilitate image exploration, including commodity components and haptics, but there is a need for algorithms that can adapt to the cognitive processes and learning stage of the operator without significant customization.

### 3. METHODOLOGY

Portions of this Chapter have been published.

T. Zhang, B. S. Duerstock and J. P. Wachs, "Classification of Blind Users' Image Exploratory Behaviors Using Spiking Neural Networks," in IEEE Transactions on Neural Systems and Rehabilitation Engineering. DOI: 10.1109/TNSRE.2019.2959555

In this chapter, several approaches are proposed to address the previously posed research questions, and such methods are described in detail. The system architecture is illustrated in Figure 3.1. To address RQ1, which relates to modeling exploration procedures used by users who are BVI, a framework utilizing Spike-timing Neural Network (SNN) is developed. Once we are able to understand the user's behavior by recognizing exploration procedures, RQ2 is addressed. RQ2 focuses on the discovery of effective assisting mechanism, which is tackled through designing and comparing different exploratory aids through user studies. In RQ3, quantitative and qualitative analysis will then be conducted to evaluate the proposed assistance system. Our proposed system were validated and compared with similar interfaces, including with and without a human assistant.

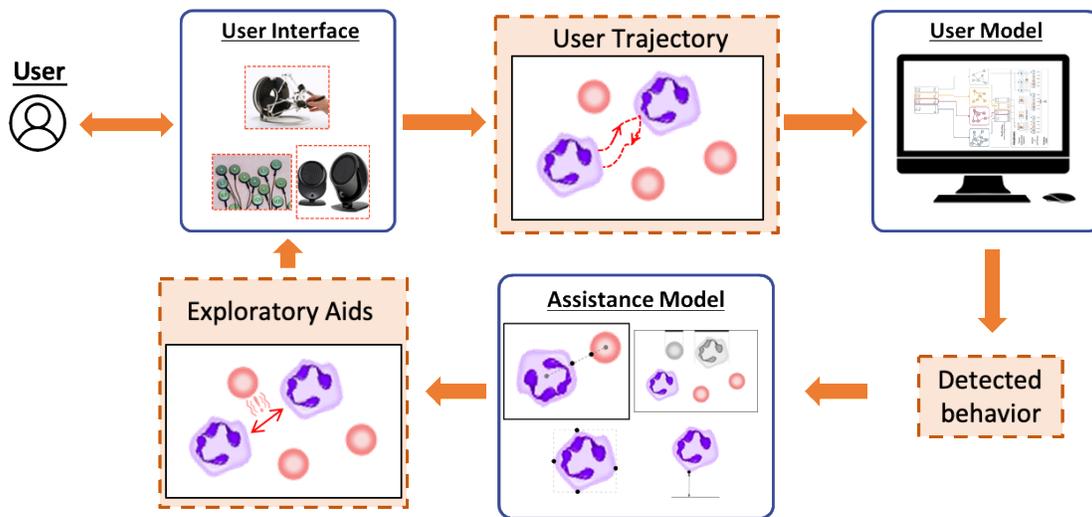


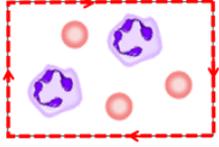
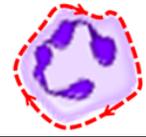
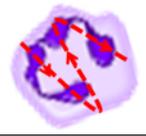
Figure 3.1 System Architecture

#### 3.1 Modeling Exploration Procedures

The first step towards the development of a user-adaptive system is the modeling of commonly applied exploration procedures by users who are BVI. Table 3.1 summarizes five

exploration procedures (EPs) that are used by individuals who are BVI to explore images with tactile sensation using a haptic device (Zhang, Zhou, Duerstock, & Wachs, 2018).

Table 3.1 Classified exploration procedures.

<b>Exploration Procedure</b>	<b>Description</b>	<b>Visualization<sup>a</sup></b>
Frame Following (FF)	Trace the boundary of the image to obtain the image size.	
Contour Following (CF)	Trace the boundary of objects on the image to learn the size and shape of objects.	
Surface Sweeping (SS)	Back-and-forth movement inside objects to learn the feature of objects.	
Relative Positioning (RP)	Back-and-forth movements between objects to obtain their relative locations.	
Absolute Positioning (AP)	Back-and-forth movements between objects and the image boundary to obtain their absolute locations on the image.	

Through this thesis, blood smear images are used as an example. Dotted red lines with arrows represent the trajectory of a user movement. Red circles represent red blood cells, and purple contours represent white blood cells.

Table 3.1 shows EPs and their associated spatio-temporal patterns. The Spike-timing Neural Network with plasticity (SNN) algorithm has proven to be an effective approach for modeling spatio-temporal patterns (Rekabdar, Nicolescu, Nicolescu, & Louis, 2017; Rekabdar, Nicolescu, Nicolescu, Saffar, & Kelley, 2016). SNN were applied to classify scale- and translation-invariant spatio-temporal patterns successfully (Rekabdar et al., 2016). It has also been shown that

SNN is suitable for early prediction, which is a desired property in intelligent systems (Rekabdar et al., 2017).

Previous work on spatio-temporal modeling using SNN focused on one dimensional feature based on trajectory, such as angles of movements or traversed pixels (Rekabdar et al., 2016). This does not, however, apply to our scenario since EPs are both characterized by trajectory and context. For example, the context indicates whether the trajectory is inside an object (SS) or between two objects (RP). Therefore, in our case a multimodal SNN approach is required. In this thesis, we developed a framework using multiple SNNs and a classification scheme to recognize five EPs shown in Table 3.1. In our framework, a sequence of multidimensional feature vectors is extracted from each sample of EP. The features used represent the angles of movements, trajectory context and users' reference during the exploration. The features representing the trajectories' angles and context information are further encoded through the training of SNNs. These features are entered to the SNNs for training acting as the stimulus of the network. Once the SNNs are trained, the characteristic responses of each EP to the SNNs are encoded as model strings, serving as the templates for each EP. Those templates are further used for classification, together with the features representing the reference point of interaction. For classification, we adapted a modified version of Dynamic Time Warpping (DTW) with Longest Common Subsequence (LCS) as the similarity measure between model strings. Dempster-Shafer Theory (DST) was then applied at last to fuse the beliefs from multiple features into a final decision (i.e., the predicted type of EP). The developed framework is also illustrated in Figure 3.2.

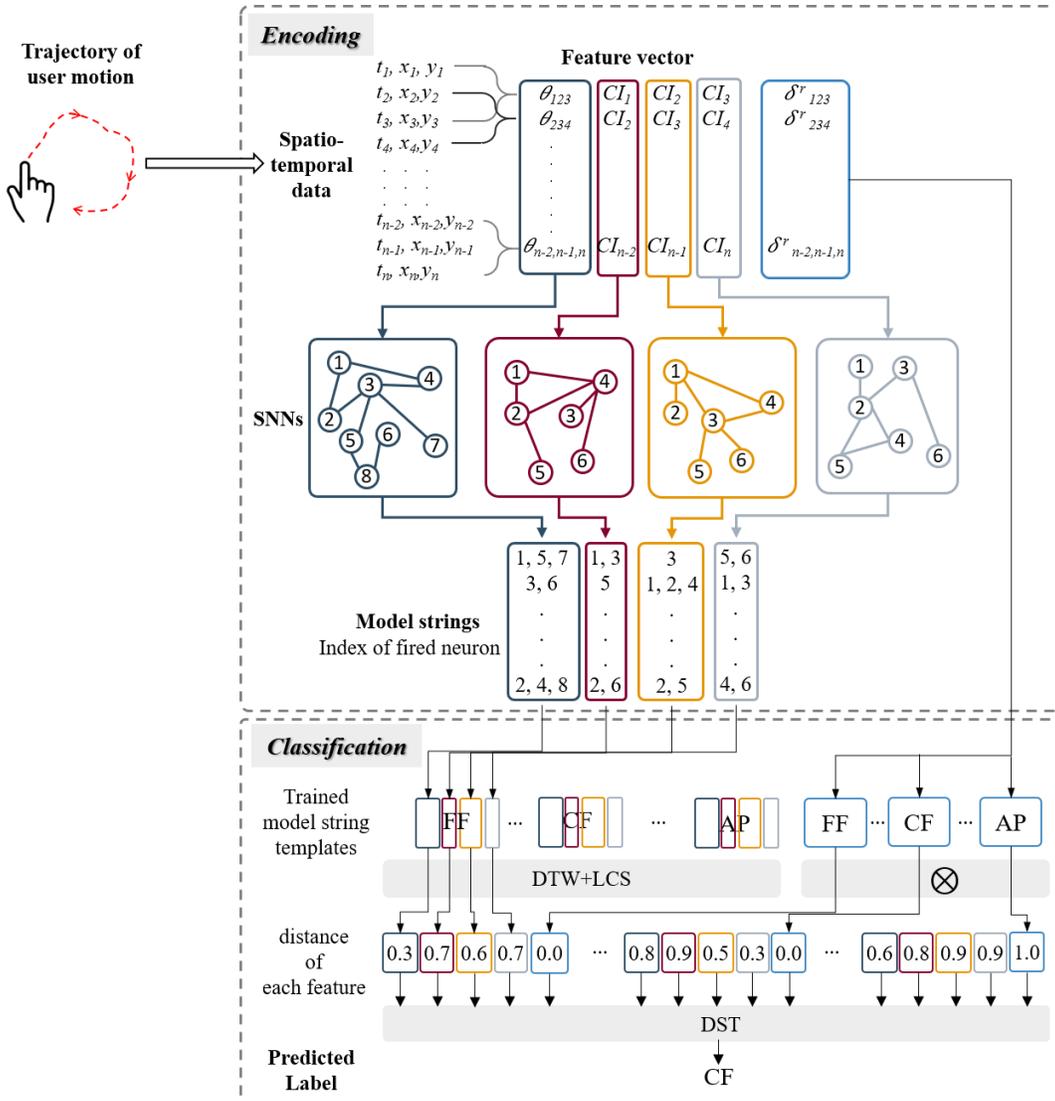


Figure 3.2 Proposed framework to learn exploration procedures.

### 3.1.1 Temporal Representation of Data

The five EPs for image exploration are characterized by spatio-temporal patterns. These patterns are obtained from trajectory data collected from a multimodal interface (Zhang et al., 2017) during the exploration of histological images by users who are blind. Each sequence of EP has an associated sequence of time and the cursor's x-and y-position on an image  $(t_i, x_i, y_i)$ , where  $i$  ranges from 1 to the length of the EP. Translation, rotation, and scale-invariant features are then extracted from the spatial-temporal patterns, in the form of three types: angles, contextual information and the reference switch. Below is explained how each feature vector is computed.

## Angles

For every three consecutive time frames  $t_{i-1}$ ,  $t_i$  and  $t_{i+1}$ , two vectors  $v_i$  and  $v_{i+1}$  are obtained following (3-1), indicating the moving directions from  $t_{i-1}$  to  $t_i$ , and from  $t_i$  to  $t_{i+1}$ .

$$\begin{aligned} v_i &= (x_{i-1} - x_i, y_{i-1} - y_i) \\ v_{i+1} &= (x_i - x_{i+1}, y_i - y_{i+1}) \end{aligned} \quad (3-1)$$

The angle between these two vectors is then computed following (3-2).

$$\theta_{i-1,i,i+1} = \cos^{-1} \frac{v_i \cdot v_{i+1}}{\|v_i\| \|v_{i+1}\|} \quad (3-2)$$

## Contextual Information

The user will choose EPs according to the content of image, including the object size, location and shape. For example, to recognize EP CF (contour following), we need to find whether the traversed pixels belong to an object boundary, regardless of the specific shape of the object. To achieve shape- and location- invariance feature representation, we assign an index for each unique contextual information ( $CI_k$ ) for a given time frame  $k$ , following the rules explained in Table 3.2. Figure 3.3 shows an example of the image and its index map.

Table 3.2 Rules for contextual information assignment.

Index	Contextual Information
1	Background
2	Object contour
3	Object Inside Area
4~8	Pixels starting from 1 to 5 away outwards from the object contour
9	Image boundary
10~13	Pixels starting from 1 to 4 pixels away from the boundaries of an image

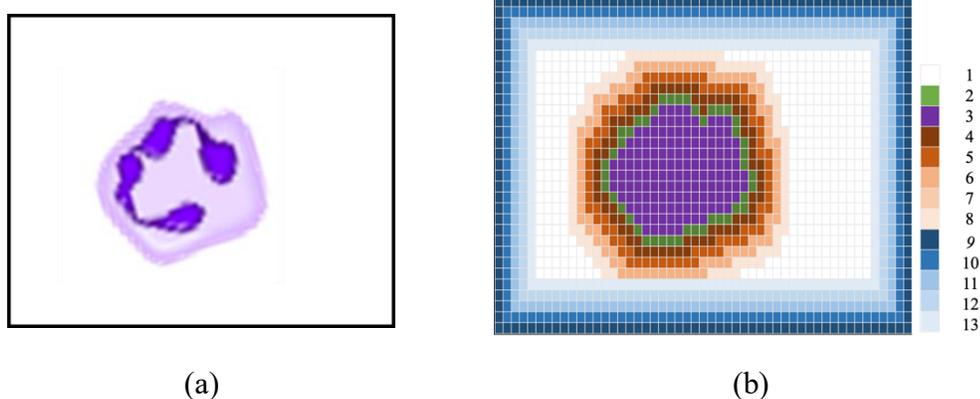


Figure 3.3 Example image and its index map. (a) Original image; (b) Index map.

### Reference switch

When people explore a visual landscape, such as an image, using only tactile information, features or objects are required as reference points. Algorithmically, this can be resembled by applying image processing techniques to segment the different objects, whereas each object gets unique IDs assigned (ranging from object 1 to  $n$ , where  $n$  is the number of objects in the image) -- for more details please refer to (Zhang et al., 2017). Once the objects are recognized, a reference point is defined as the object that the user is in contact with, including the image boundary and object  $i$  ( $i=1 \dots n$ ). The last reference point  $r_i$  is annotated for each time frame  $t_i$ . The altering of reference point ( $\delta^r$ ) is computed based on three consecutive time frames, following (3-3).

$$\delta^r_{i-1,i,i+1} = \begin{cases} 0, & r_{i-1} = r_i = r_{i+1} \\ 1, & \text{otherwise} \end{cases} \quad (3-3)$$

Given three consecutive time frames  $t_{i-1}$ ,  $t_i$  and  $t_{i+1}$ , the spatio-temporal information is then represented using the feature vector (3-4):

$$f = (\theta_{i-1,i,i+1}, CI_{i-1}, CI_i, CI_{i+1}, \delta^r_{i-1,i,i+1}) \quad (3-4)$$

where, the angle  $\theta_{i-1,i,i+1}$  is computed using (3-2), the context index ( $CI_i$ ) is calculated for each time frame  $t_i$ , reference switch  $\delta^r_{i-1,i,i+1}$  is obtained following (3-3).

### 3.1.2 Spike-timing Neural Network

SNNs were trained to encode the extracted features in (3-4), except the last feature “reference switch”. As defined in (3-3), it may be unnecessary computational expensive to encode  $\delta^r_{i-1,i,i+1}$  using SNN as it is a Boolean variable; alternatively this feature can be used for

classification directly. Therefore, four SNNs were trained in this work, with one network for each feature. All the networks have the same configuration except the number of neurons, which is dependent on the range of the specific feature.

### ***General Network Configuration.***

In an SNN, neurons are mutually connected through the synapses, and can be configured with the number of neurons, the connectivity among neurons, the conduction delay of each synapse, and the weights of synapses, as defined in Izhikevich (Izhikevich, 2006). Izhikevich’s model was used in this work considering its ability to simulate different neuron dynamics while requiring relatively small computational power, compared to other widely used models, such as “Integrate-and-Fire” and “HodgkinHuxley” (Izhikevich, 2004b). There are two types of neurons in an SNN, the excitatory and inhibitory neurons, of which the amount has a ratio of 4:1. The number of neurons for each of the four SNNs is then determined based on the range of the feature it is encoding and that is discussed in the next section. These neurons are mutually connected and the four SNNs developed in this work have a connectivity of 10% among all neurons. The synapses can have different conduction delay that partially determines the firing pattern of neurons. The differences between firing patterns of input data are then used for representation and further classification. The conduction delay is randomized from 1 *ms* to 20 *ms* for each synapse. The synaptic weights are configured as +6 for excitatory neurons, and -5 for inhibitory neurons initially. The weights are then updated using the rule of the spiking dependent plasticity (STDP) (Izhikevich, 2006). Based on the time-locked firing patterns of neurons, the STDP rule boosts or degrades inter-neuron connections by increasing or decreasing the synaptic weights.

Figure 3.4 illustrates an example of the trained SNN. The white nodes are excitatory neurons and the gray nodes are inhibitory ones. The width of the link indicates the synaptic weight where thicker link has larger synaptic weight. The dotted link indicates the synapses with a negative weight. The color of the link indicates the conduction delay which is similar to the color scheme of a traffic map that green indicates short delay, while red indicates longer delay. Edge bundling technique was applied to reduce the complexity of the graphic representation.

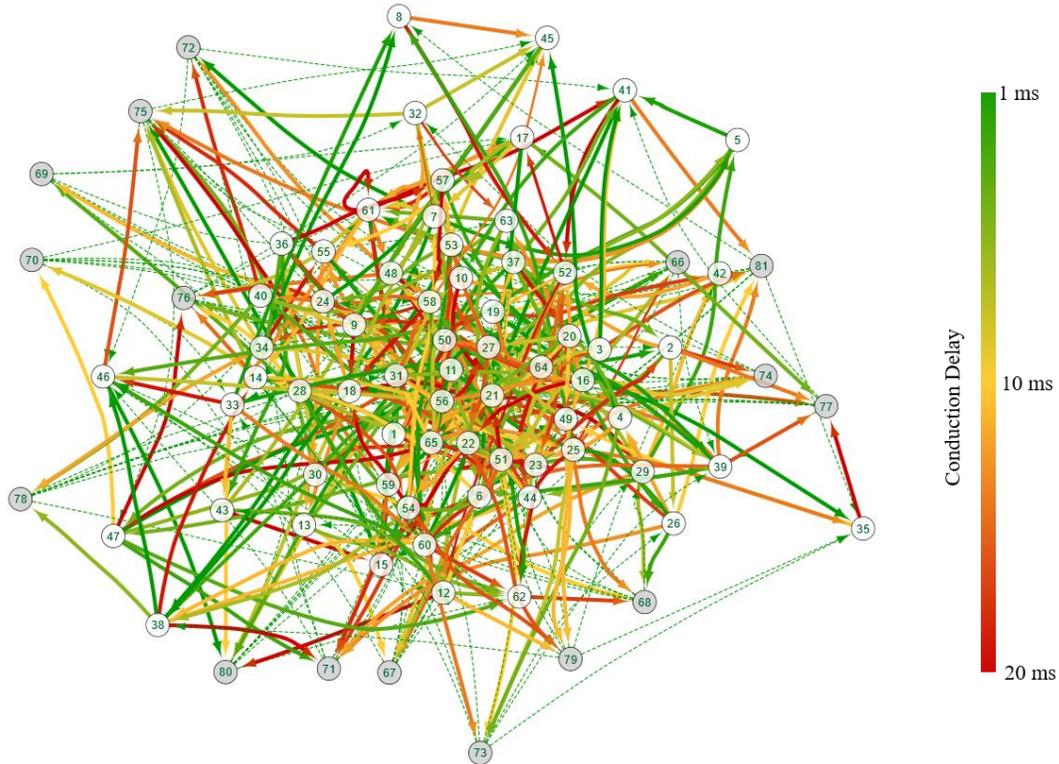


Figure 3.4 Example Spiking Neural Network. The nodes are neurons and the links are synapses that connect the neurons. The width of the link represents the weight, while the color represents the conduction delay.

### ***Number of Neurons***

The number of neurons for an SNN is dependent on the range of the feature value. The degree of feature angle  $\theta_{i-1,i,i+1}$  ranging from 0 to 180 inclusive, is divided into 19 sets. Five excitatory neurons are allocated for each set, resulting in 95 excitatory neurons (Rekabdar, Nicolescu, Kelley, & Nicolescu, 2014). Twenty-three inhibitory neurons are allocated as the ratio between excitatory and inhibitory neurons is 4 to 1 as discussed above. The SNN encoding feature angle has a total of 118 neurons, with 95 excitatory and 23 inhibitory neurons.

For feature contextual information (CI<sub>i</sub>), the index ranges from 1 to 13 as defined in Table 3.2. Similarly to the configuration for feature angle, each index is allocated with 5 excitatory neurons, thus leading to a network of 81 neurons with 65 excitatory and 16 inhibitory neurons.

## Network Training

To train an SNN using STDP rule, we stimulate the neurons based on the sequence of feature vectors (as in (3-3)) generated by each EP sample, with an interval of  $1ms$ . For example, if the angle  $\theta_j$  in the first frame of EP FF is associated with neurons 1 to 5, neuron 1 is activated with a  $20mA$  current, followed by neuron 2 to 5, at  $1ms$  intervals. For an EP with  $n$  time frames, a pattern of  $5n ms$  is transmitted into the network. Figure 3.5 shows an example of how neurons were fired given the input sequence of feature angle. The raster plot on the top indicates the firing patterns of neurons, while the bottom plot shows the input values of the angle feature over time.

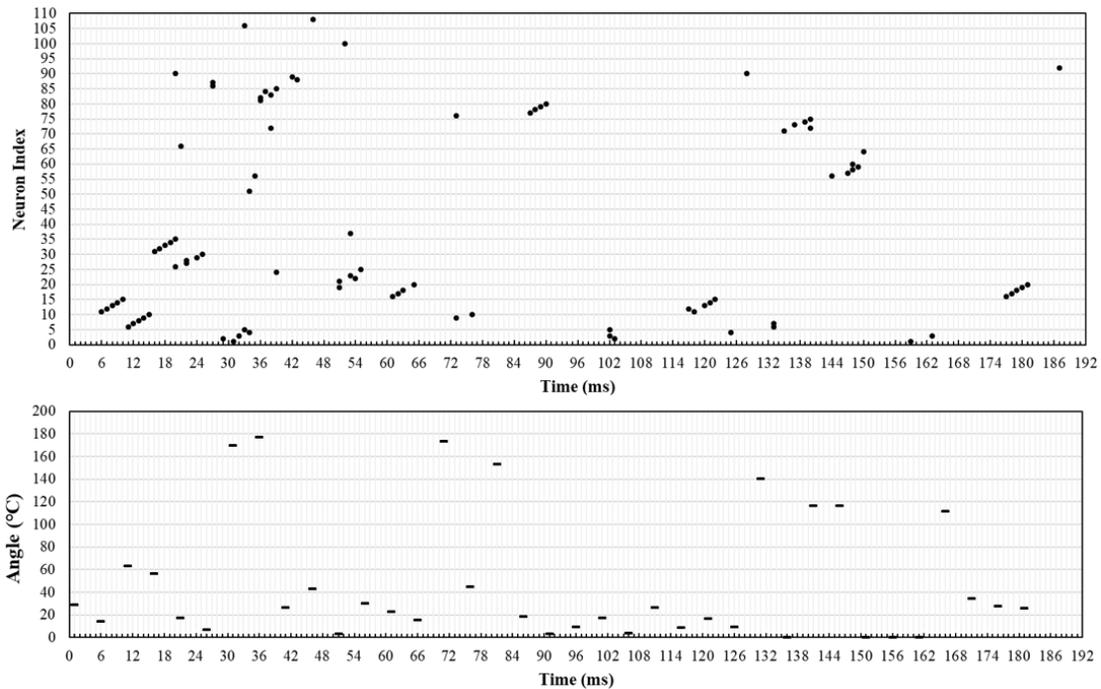


Figure 3.5 Firing pattern of neurons. The top raster plot shows the fired neurons at each time stamp and the bottom plot shows the value of input data at each time stamp.

To train the SNN for the 5 EPs, we provide  $m$  training samples for each EP. One round of training involves stimulating the network with  $5m$  training samples one by one. This follows the order of sample 1 of EP FF, sample 1 of EP CF, till sample 1 of EP AB, continued from sample 2 of EP FF. The STDP rule is used during the training process to update the synaptic weights for all neurons. During each round of training, the summation of the synaptic weights of all neurons is calculated. For two consecutive rounds of training, the differences between the synaptic weights,  $\Delta W$ , are computed following (3-5), where  $w_p^i$  is the weight of synapse  $p$  at training round  $i$ , and

the SNN has a total number of  $n$  synapses. The training is completed when synaptic weights remain stable. Figure 3.6 shows an example of the fluctuation of synaptic weights over different rounds of training. The training of the network is completed around 30 rounds.

$$\Delta W = \sum_{p=1}^n \|w_p^i - w_p^{i+1}\|^2 \quad (3-5)$$

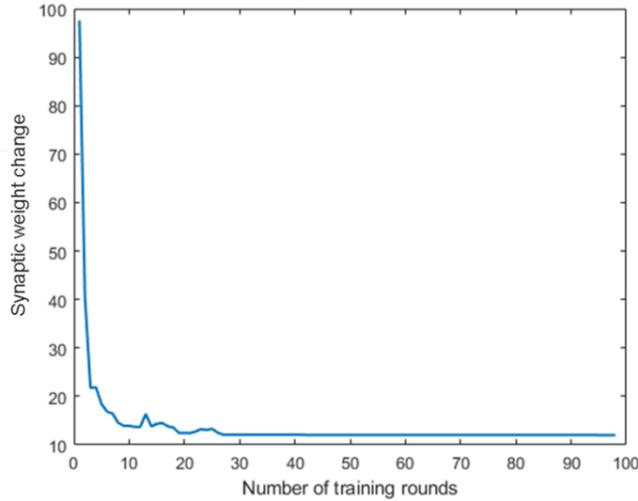


Figure 3.6 Examples of synapse weights fluctuation.

### 3.1.3 Classification

The features encoded by SNNs, including the angle and  $CI$  are used together with feature “reference switch”  $\delta^r$  to classify the various EPs. A distance-based metric is used to measure the similarities between samples. For SNN-encoded features, a modified dynamic time warping (DTW) is used to compute the distance, while for feature  $\delta^r$ , the distance is computed as the difference between  $\delta^r$  over time. Then, Dempster-Shafer Theory (DST) is applied to determine the type of EP by combining the distances obtained from all the features.

#### ***Representation of SNN-Encoded Features***

For SNN-encoded features, we record the indices of fired neurons (voltage  $\geq 30\text{mV}$ ) at each time stamp of the training stimulus. If no neurons are fired at a certain time, that time frame is skipped. A stimulus pattern is then represented as a string of fired neurons, which is defined as a ***model string*** (Rekabdar et al., 2017). The neurons that are fired at the same time is defined as a

**character.** The following is an example of a model string:  $\{(1,2,5), (4,6), (3,5), (1,7,8)\}$ . It indicates that neuron 1, 2 and 5 initially fired at the same time, followed together by neuron 4 and 6. Then, neuron 3 and 5 fired simultaneously. At last, neuron 1, 7 and 8 are fired at the same time.

To build the model string for each sample, we stimulate the associated neuron in the trained network for each time stamp, same as the training phase. Different from training, the synapse weights stay constant and the fired neurons for each time step are recorded.

For each feature  $k$  ( $k=1,2,3,4$ ), a modification of DTW with LCS as the distance function is applied to compute the distance  $d_k$  between a testing sample and a training sample. We compute the distance of one testing sample with all  $5m$  training samples, where  $m$  is the number of training samples for each EP.

### ***Distance Metrics***

To compute distance between model strings generated from SNNs, LCS is applied as the distance function for DTW to compute the similarity between two characters. We use DTW to align model strings of different lengths and compute the distance between two model strings. For DTW, smaller distance indicates higher similarity. However, for LCS, a larger value indicates higher similarity. Therefore, we invert the sign of similarity value between two characters and use it as the distance in DTW. To align strings with different length, a sliding window of size 15 is used in this work. If the string length is less than 15, the string length is used instead.

To calculate the distance between two samples  $a$  and  $b$  in the logic feature  $\delta^r$  dimension, we use the difference of summation over time, following (3-6), where sample  $a$  has  $p$  time frames and sample  $b$  has  $q$  time frames.

$$d^r_{a,b} = \begin{cases} 0.0, & \sum_{t=1}^p \delta^r_{a,t} = 0 \text{ and } \sum_{t=1}^q \delta^r_{b,t} = 0 \\ 0.0, & \sum_{t=1}^p \delta^r_{a,t} > 0 \text{ and } \sum_{t=1}^q \delta^r_{b,t} > 0 \\ 1.0, & \text{otherwise} \end{cases} \quad (3-6)$$

In (3-6), condition 1 indicates there is no reference switch in both sample  $a$  and  $b$ , while condition 2 means that user's reference was switched in both samples  $a$  and  $b$ . The distance between sample  $a$  and  $b$  is 0 in these two cases.

### 3.1.3.1 Decision Fusion using Dempster-Shafer Theory (DST)

Each sample of EP is characterized by multidimensional feature vectors. The DST fusion approach was used here to combine the degree of beliefs derived from multiple evidences (given by the individual features within the vector) into a final decision. DST is a form of decision-level fusion (e.g. early or late fusion) with low computational cost and robust to noise (Arif, Brouard, & Vincent, 2006). It is also easily extendable, that allows for incremental addition and/or subtraction of features. Due to such flexibility, it has been used for decision fusion in visual tracking (Li et al., 2013) and human activity recognition (Rottensteiner, Trinder, Clode, Kubik, & Lovell, 2004).

Each feature contributes to determine the EPs' class. The contribution for each feature is characterized by the Basic Belief Assignment (BBA) function  $m_k(y)$ , which is the belief for EP of type  $y$ , where  $y \in \{FF, CF, SS, RE, AB\}$ . Since there are five features in  $f$  defined in (3-4), we have four BBA functions,  $k=1,2,3,4,5$ . To compute the belief of each feature for each type of EP, the average DTW distance is used between the unknown sample and all the training templates of a certain EP class. Smaller DTW distance  $d$  indicates greater similarity, while greater value shows larger belief. Therefore, the negative value of the average distance represents the belief amount (greater similarity  $\rightarrow$  smaller  $d \rightarrow$  larger negative  $d \rightarrow$  larger BBA).

To combine belief from different sources, the Dempster's Rule of Combination (DRC) calculates a joint BBA  $m_{1,2}(\cdot)$  from two independent beliefs  $m_1(\cdot)$  and  $m_2(\cdot)$ , according to:

$$m_{1,2}(y) = \frac{1}{1 - K} \sum_{y_1 \cap y_2 = y} m_1(y_1) m_2(y_2) \quad (3-7)$$

where  $y_1, y_2, y \in \{FF, CF, SS, RP, AP\}$ .  $K = \sum_{y_1 \cap y_2 = \emptyset} m_1(y_1) m_2(y_2)$  is a measure of disagreement between the two beliefs  $m_1(\cdot)$  and  $m_2(\cdot)$ . A large  $K$  value implies strong disagreement between the two beliefs regarding the EP's class. The DRC process combines two BBA at a time, and the resultant joint BBA is combined with the remaining BBAs repeatedly, one at a time, to reach the final belief. In this study, five BBA functions through DTW comparisons of each feature (i.e., initializing  $m_1 \sim m_5$ ) were created, and then DRC is used to combine all voting weights to reach a single final decision. Algorithm 3.1 shows the classification procedure.

Algorithm 3.1 Classification with modified DTW and DST.

---

**Input:** Testing model string  $t$  of length  $n_t$  ( $c_1, c_2, \dots, c_n$ )  
**Output:** label --- predicted type of EP

---

```

1  for each training sample  $x$  of length  $n_x$ 
2      for each feature channel  $k=1,2,3,4$ 
3          for each character  $c_i$  in  $t_k$ 
4              for character  $c_j'$  in  $x_k$  between the window
5                  similarity = - LCS( $c_i, c_j'$ )
6                  cost( $i, j$ ) = similarity + min( cost{( $i, j$ ), ( $i+1, j$ ), ( $i, j+1$ )} )
7              end for
8          end for
9          distance( $t, x, k$ ) = cost( $nt+1, nx+1$ )
10     end for
11      $k=5$ 
12     if ( $\sum x_k == 0$  and  $\sum t_k == 0$ ) or ( $\sum x_k > 0$  and  $\sum t_k > 0$ )
13         distance( $t, x, k$ ) = 0.0
14     else
15         distance( $t, x, k$ ) = 1.0
16     end if
17 end for
18 for each feature channel  $k$ 
19     for all training templates  $x$  of label  $y$  in  $S$ 
20          $m_k(y) = \text{mean} \sum_{x \in S} \text{distance}(t, x, k)$ 
21     end for
22 end for
23 label for  $t \leftarrow \text{DRC}(m_1, m_2, m_3, m_4, m_5)$ 

```

---

### 3.2 Designing an Assistance Model

After the commonly-applied exploration procedures (EPs) are learned, designing different exploratory aids based on the learned procedures is the next step to develop an intelligent assisting system. It is defined as the Assistance Model, that can provide appropriate exploratory aids based on the applied EP. To design such a model, understanding the goal behind the detected user behavior, is a crucial aspect. For example, people who are BVI use procedure AP (absolute positioning shown in Table 3.1) to measure the distance between an object and the image boundary.

Different from vision, using haptic-based interaction, distance can only be estimated by the elapse of time, from the point the user left a place till the user reached another place (Hatwell et al., 2003). Due to the lack of sensibility for distance, usually users perform this procedure couple of times between the same object and the same image boundary repetitively, to get a more accurate measurement. To make this process faster, the assistant model can provide users the shortest path between the object and the image boundary. Besides understanding the goal, knowing the limitations of human users is another design criterion that is considered in this work. For example, memorizing the positions of explored objects on an image can be cognitively heavy without vision, and thus lead to unnecessary movements to relocate and repeat the procedures that have been performed on an explored object (Hatwell et al., 2003).

In this section, goals behind each EP was firstly revealed through a “think-out-loud” experiment. Then, exploratory aids were developed based on the two design criteria discussed above.

### **3.2.1 The goals of exploration procedures**

When individuals who are blind explore images, their two main goals are to discriminate what objects are on the image and where these objects are located relative to each other and within the whole picture for understanding image content. These two goals can be formally defined as - identifying objects and locating objects. To understand the relationship between performing each exploration procedure and the ultimate goal of understanding the image, experiments were conducted with users doing a think-out-loud image analysis, requiring users to explain why they were performing a certain exploration procedure using our multimodal image perception system while they were actively exploring parts of the image. The details of the experiment are explained in 4.2. It is observed from experiments that the two major goals are broken into specific pieces. For example, users identify an object by measuring its shape, size and texture. “Contour following” is applied to understand the shape and size, while “surface sweeping” is used to obtain the texture. Therefore, two final goals are broken into smaller goals, that are defined as intermediate goals and summarized in Table 3.3.

Table 3.3 Goals of exploration procedures.

EP	Goals
FF	<ul style="list-style-type: none"> <li>• Measure the size of the whole image</li> <li>• Identify the side of the image</li> <li>• Locate explored objects</li> </ul>
CF	<ul style="list-style-type: none"> <li>• Understand the shape and size of an object</li> <li>• Locate the compass points on an object</li> <li>• Locate explored neighbors</li> </ul>
SS	<ul style="list-style-type: none"> <li>• Understand the texture of an object</li> </ul>
RP	<ul style="list-style-type: none"> <li>• Measure the location (distance/direction) of an object relative to its neighbors</li> </ul>
AP	<ul style="list-style-type: none"> <li>• Measure the distance of an object relative to one image boundary</li> </ul>

From Table 3.3, we can observe that participants used the same EP for different goals. For example, there are three goals for CF. Besides understanding the shape and size of an object, CF was also performed to locate the compass points on an object or locate the explored neighbors. The compass points of an object are defined as the most outreached positions on its boundary in the four directions, North, South, East, and West. While the neighbors of an object are the objects surrounding it.

Among these goals, some of them were pursued frequently, while others were less frequent. A normalized histogram was computed to show the frequency of each goal during the exploration of one image. To identify whether the distribution of used goals is common among different users, Bhattacharyya coefficient is then calculated to measure the similarities between the histograms, following (3-8).  $\mathbf{P}$  and  $\mathbf{Q}$  are the two histograms with  $n$  partitions.  $p_i$  and  $q_i$  are the frequencies of the  $i$ th partition in  $\mathbf{P}$  and  $\mathbf{Q}$ .

$$BC(\mathbf{P}, \mathbf{Q}) = \sum_{i=1}^n \sqrt{p_i q_i} \quad (3-8)$$

The data collected from experiment 4.2 shows a coefficient of 0.954 among 6 participants, which indicates a shared behavior among users. The histogram containing the data of all the participants are shown in Figure 3.7.

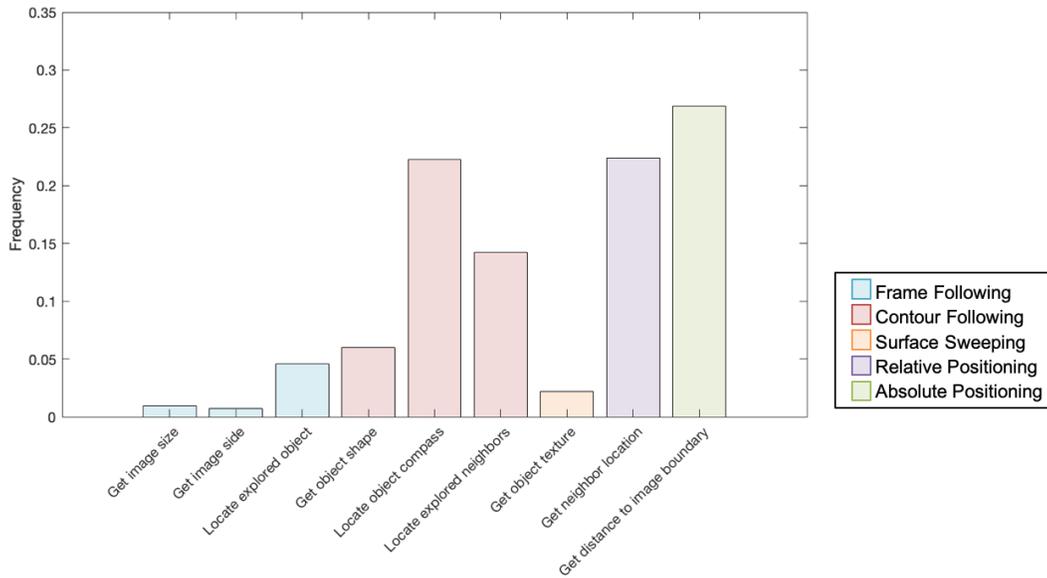


Figure 3.7 Histogram of the goals for exploration procedures.

### 3.2.2 Exploratory Aids

External assistance is necessary when current techniques are not adequate. Therefore, exploratory aids are designed for challenging goals. From Figure 3.7, it is observed that “get distance to image boundary”, “get neighbor location” and “locate object compass” are the top three goals found from experiments. Exploratory aids were firstly designed for these goals, including Contour Compass (CC) and Path Shortcut (PS). Another observation from this study is that users are required to memorize or create a detailed mental picture while exploring images. “Locate explored objects” and “Locate explored neighbors” were the two goals requiring the most memorization. To alleviate the requirement on memory/cognitive load, two other exploratory aids were designed: Border Projection (BP) and Contour Neighbors (CN). The design of these exploratory aids is explained in the following sections.

#### *Contour Compass*

Observed from Figure 3.7, the participants trace the contour of the object (CF) to locate the compass points of an object, so that they can more accurately measure the absolute location of the object. Contour Compass (CC) is designed to facilitate this procedure. The locations of an object’s compass points are provided when the user is moving along the contour of an object. The

compass points of an object are defined as the most outreached positions on its boundary in the four directions, North, South, East, and West. The visualization for CC in Table 3.4 shows an example, where the gray dotted lines indicate the bounding box and the black dots indicate the four compass points. To calculate these positions, the pixels on the contour are firstly identified. The pixels that have the smallest or largest x- or y-positions are annotated as the four compass points.

### ***Path Shortcut***

Path Shortcut (PS) is developed to facilitate the measurement of distance and direction between an object and its surroundings, addressing the other two most common goals observed from Figure 3.7 last two columns. When RP is detected, the path connecting the centers of the two neighboring objects are calculated and rendered to the users. Neighbors of an object are defined as the objects that can be reached following a straight line without interfering with another object. To automatically extract the straight paths between every pair of neighbors, lines connecting the centroids of every two objects are created. Edge detection (Bowyer, Kranenburg, & Dougherty, 2001) was used to extract the contour of an object. The intersection point of the line and the object contour was utilized as the correspondence points (shown in Figure 3.8(a)) for object 1 and 2. The path shortcut is then defined as the straight line connecting the pair of correspondence points. When obstacles were located over any part of the line, the two objects were not considered as neighbors, thus no cues were created (shown in Figure 3.8(b) for object 1 and 3).

When AP is detected, the shortest path between an object and the related image border was provided to expedite the acquisition of distance. The shortest path between an object and an image border is defined as the projected line from the compass points to the image borders. Similarly to the shortcut of neighboring objects, no hints were created if there are obstacles between the object and the image border (Figure 3.8(c)).

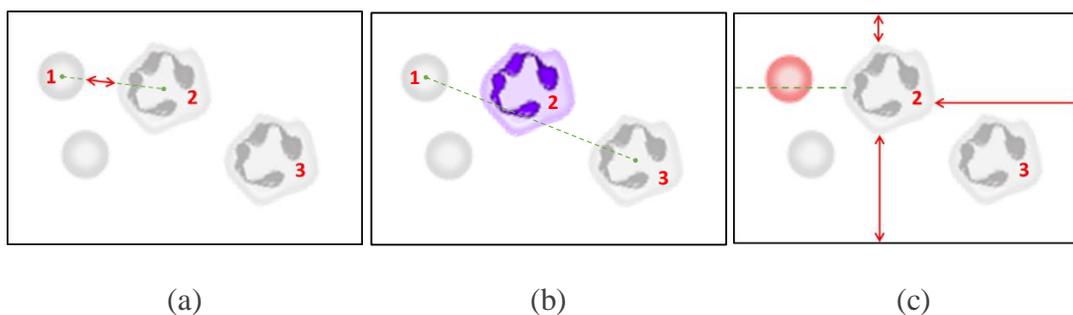


Figure 3.8 Examples of path shortcut. (a) Shortcut between two neighbors; (b) No shortcut for object 1 and 3; (c) Shortcut from object 2 to the top, right and bottom image border and no shortcut from object 2 to the left image border.

### ***Border Projection***

Limited memory resource is another shared problem among BVI users. To alleviate the memory load, Border Projection (BP) was provided by indicating the projected position of objects when the user was moving along the image borders (FF). It's also the most popular goal for FF, observed from Figure 3.7. To construct the aids BP, the bounding box for each object on the image were extracted by applying blob detection techniques (Danker & Rosenfeld, 1981). The regions to the left, right, top and bottom of the object were examined. If obstacles were detected in a region, no aids would be generated for the border associated with the region. In Figure 3.9(a), the right and bottom region of the object 1 has obstacles, therefore, no aids were generated on the right and bottom image border for the object. Among the remaining regions, the distance between the bounding box and the image border,  $dk(i)$ , was calculated, where  $k \in \{\text{left, right, top, bottom}\}$ , indicates the image border, and  $i$  is the unique id of an object ( $i=1$ ). The aid is then generated on the border with shortest distance. If an object was not directly adjacent to any image border, no aids were provided as shown in Figure 3.9 (b) for object 2.

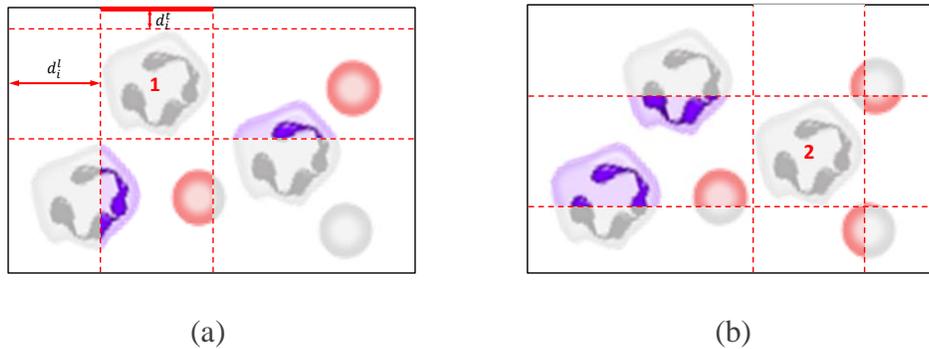


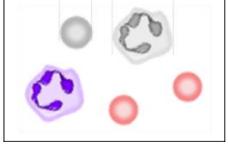
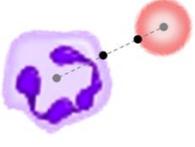
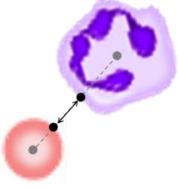
Figure 3.9 Examples of border projection. (a) Aids on the nearest image border; (b) No aids for blocked objects.

### ***Contour Neighbors***

Contour Neighbors (CN) helped to decrease the load on memorization indicating the neighbors when the user was doing CF. It is also the second popular goal for CF. The pair of correspondence points obtained from strategy PS was used to indicate the approximate locations of neighbors.

Table 3.4 summarizes the exploratory aids discussed above, with their related exploration procedures and visualizations. These aids were further compared through user studies along with the selection of proper user interfaces, and explained in section 4.3.

Table 3.4 Summary of exploratory aids.

Exploratory Aids	Description	Related Exploration Procedure	Visualization
Border Projection (BP)	Projects the location of the objects onto the image boundaries.	Frame Following	
Contour Compass (CC)	Indicate the locations of the compass points.	Contour Following	
Contour Neighbor (CN)	Indicate the locations of explored neighbors.		
Path Shortcut (PS)	Calculate the shortest path between the object and <ul style="list-style-type: none"> <li>• its neighbor (RP)</li> <li>• the image boundary (AP)</li> </ul>	Relative Positioning	
		Absolute Positioning	

### 3.3 Sensory-substituted User Interface

Different sensory modalities were utilized to deliver the exploratory aids,. Tactile and auditory sensory substitution are the most commonly used approaches to replace vision. Several studies and applications were reviewed focusing on the performance of sensory substitution. It was found that the optimal feedback modality is rather case-specific than universal. Experiments were conducted in this research to measure the performance of each exploratory aid with vibration, sound and magnetic attraction. The optimal modality for each exploratory aid was determined based on the performance, which included the task completion time and accuracy.

The multimodal interface tested in this study consists of a computer, a haptic controller, and two “Tactors” that generates vibrational feedback (shown in Figure 3.10). The computer runs a program that extracted image features using image processing techniques (Zhang et al., 2017). Then, these features were rendered as other sensory modalities and delivered to the user through the computer speaker, the haptic controller, and the vibrational Tactor. In this study, the multimodal image interface rendered the contour of objects through haptics feedback and expressed color intensity through vibration. One Tactor was attached to the user’s non-dominant hand to receive vibrational feedback inside the object. Besides feedback rendering, the haptic controller works as the input device that is similar to a computer mouse. That is, a user can hold it as a stylus to explore an image.

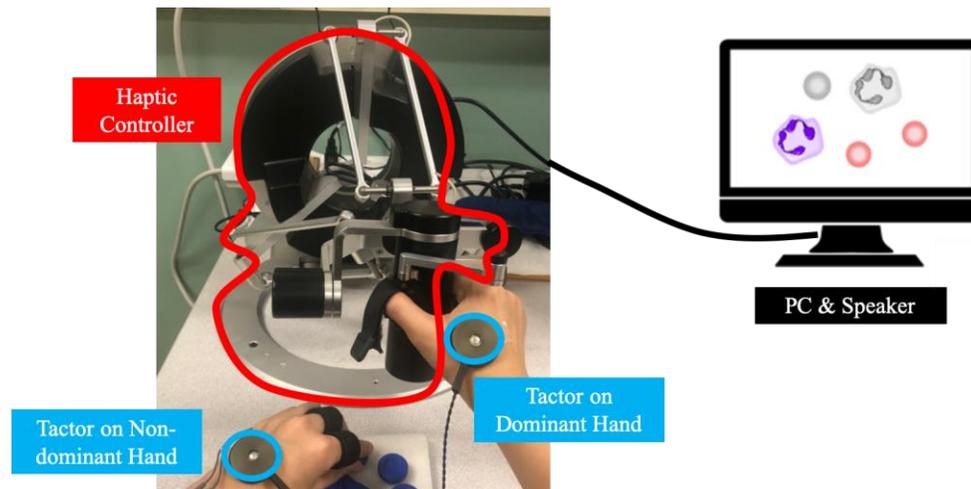


Figure 3.10 Sensory substituted interface.

To render the exploratory aids, vibrational feedback was delivered through a Tactor attached to the dominant hand of the user, magnetic feedback was provided through the haptic controller and sound feedback was delivered through the computer speaker. The users would feel the vibrational or auditory feedback during the time that they were in the aids operational area, while they would feel the magnetic attraction if they were getting close or moving away from the area of the hints.

Section 4.3 explains the design of experiments and the performance metrics to evaluate each exploratory aid delivered through different modalities.

### 3.4 Performance measurement

After the validated exploratory aids were integrated into the multimodal image exploration system, an evaluation of the system was performed. To evaluate the image exploration system with exploratory aids, the system was compared with two other interfaces, including the system without intelligent assistants and the system with human assistants. Multiple evaluation metrics were utilized to measure the system in different aspects, including task performance, and workload.

#### 3.4.1 Task performance

For an image exploration system, the time spent to complete the task and the accuracy of image understanding are two crucial aspects to assess user performance. The intelligent assistant was developed to enhance the understanding of an image in a timely manner.

##### *Task completion time*

A maximum of  $m$  minutes was set to limit the time for the exploration of one image.  $m$  equals 10 in this experiment.

##### *Accuracy of image understanding*

To understand the mental image built by the users after their explorations, a common practice requires asking them to replicate the image using real-world objects and measuring the similarities between the original image and the replicated image. To measure the similarities, both images are converted to binary images eliminating the differences introduced by the material of replicated image. For example, the 3D-printed objects on the replicated image has a different color and texture than the objects digitally. Figure 3.11 illustrates the processing pipeline including the replicated and original image, so that they can be compared. As the copy of the replicated image was acquired by taking a photo of it using a camera, the picture would exhibit projective distortions. Therefore, distortion correction was applied first. The image was then cropped to remove the unrelated background. At last, the image was converted to a binary format.

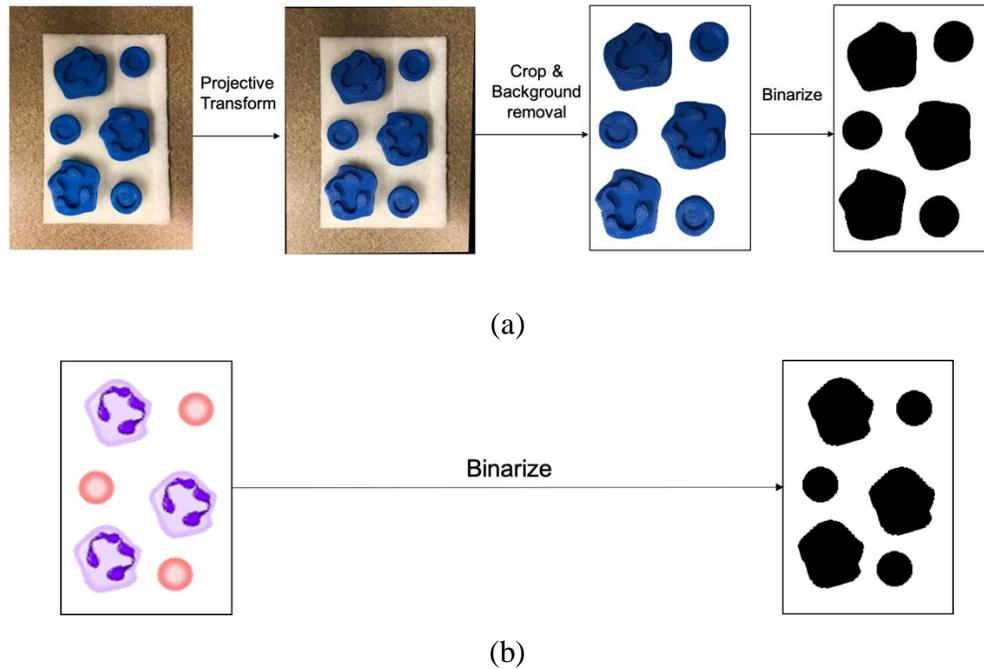


Figure 3.11 Pipeline for image processing (a) replicated image, and (b) original image.

Similarities  $S(A, B)$  between the two binary images, A and B, were calculated following (3-9) and used as the accuracy measurement.

$$\mathcal{P}_A = S(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (3-9)$$

### 3.4.2 Workload Assessment

Upon literature review, it was found that several studies reported that workload is to be related to performance. Workload is often measured using multiple techniques to develop a coherent and complete understanding of workload. In this research, two approaches were used to measure the workload, including subjective rating scales and objective physiological measurements.

#### *Rating scales*

NASA-Task Load Index (NASA-TLX) is a widely used approach to give multidimensional ratings to workload (Hart & Wickens, 1990). A questionnaire containing six ratings was answered

by the participants after each trial, including, mental demand, physical demand, temporal demand, performance, effort and frustration. 21 Scales ranging from 0 to 20 were given for each measurement, with a five-point step. After the participants answered the questionnaire, they were asked to make pairwise comparisons to choose the measurement that is more related to workload for themselves. The number of times a measurement was selected was considered as the weight  $w_i$  for measurement  $i$ . Then the overall work load index  $\mathcal{P}_W^R$  was computed following (3-10), where  $s_i$  is the scale score for measurement  $i$  and  $n$  equals the number of ratings in the questionnaire ( $n=6$ ). The weighted score was further divided by  $m$  ( $m=15$ ) to get the workload index ranging from 0 to 100.

$$\mathcal{P}_W^R = \frac{1}{m} \sum_{i=1}^n w_i s_i \quad (3-10)$$

### ***Physiological measurements***

Physiological measurements including heart rate (HR), hear rate variability (HRV), skin conductance response (SCR), pupil size, and brain activities are commonly used signals to evaluate workload. While pupil size is not applicable in the study for BVI users and brain activities are difficult to measure, HR, HRV and SCR are good measurements. They are sensitive to mental workload (Nourbakhsh, Chen, Wang, & Calvo, 2017). Average heart rate was applied as the metric of HR, average R-R intervals was used for HRV analysis while the frequency of SCR was used as the indicator for workload. Literature indicated that workload can lead to an increasing of average heart rate, decreasing of R-R intervals and increasing of SCR frequency.

## **3.5 Summary**

This chapter explains in detail the framework developed to learn and classify the exploration procedures applied by users who are BVI. A spike-timing neural network based approach was used to encode the collected temporal-spatio data. Classification was then achieved based on distance metrics and Dempster-Shafer decision fusion technique. Being able to understand the exploratory behavior of BVI users, multiple assisting strategies were designed to facilitate their explorations of images. Multiple sensory modalities were then tested to deliver the assisting strategies. At last, performance metrics, including task performance and workload assessment were applied to evaluate the system.

## 4. EXPERIMENTS AND RESULTS

This chapter explains the experimental design and results based on the developed methodology. Regarding RQ1 to learn exploration procedures, experiments were conducted to evaluate its performance. The results obtained from the experiments formed the foundation of this research, as the system is capable of understanding the user behavior. The success of a user model is the key point for further development of the assistant model and the user interface (RQ2). Experiments were then conducted to develop the user model. The results helped the development of exploratory aids discussed in **Error! Reference source not found.** Further experiments were conducted to answer RQ2 that discovered the optimal sensory modalities to deliver each exploratory aid. Final experiments were performed to validate the developed image exploration system with exploratory aids.

### 4.1 Learning Exploration Procedures

To evaluate the proposed framework for EPs learning and classification, our scenario was set to explore blood smear images. Experiments were conducted to collect users' data of observed EPs. These data were then annotated and used for training and testing of the proposed EP learning framework. The results discussed here includes the performance of the proposed learning framework.

#### 4.1.1 Data collection

Participants were asked to explore blood smear images blind-folded using a haptic-based image exploration system (Figure 4.1) and replicate the image by placing 3D printed objects on a board after the exploration. Trajectory data from 10 participants exploring 12 images were collected, including 5 females and 5 males, with age from 18 to 30. Each individual EP was then annotated from the whole exploration, resulting 168 samples for FF, 803 for CF, 207 for SS, 238 for RE and 483 for AB. Figure 4.2 shows an example of captured procedure AP in red. Color intensity indicates the elapsed time of trajectory, where lighter shades occurred earlier than darker shades. Twenty samples per EP were taken out from the dataset for training, and the rest were used as testing samples.

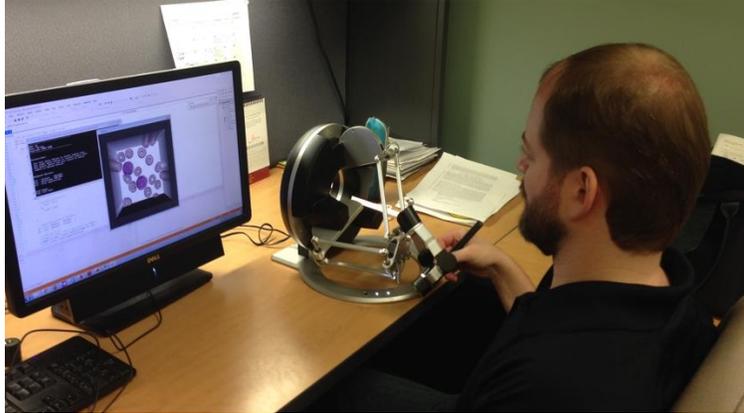


Figure 4.1 A haptic-based image exploration system.

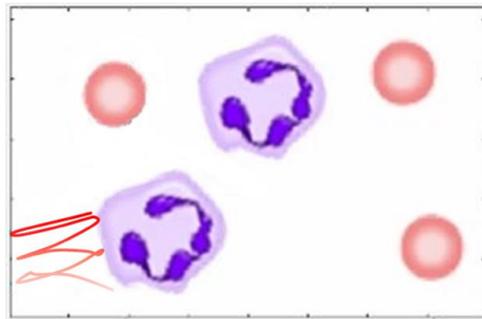


Figure 4.2 Examples of captured EP AP performed by participants.

#### 4.1.2 Evaluation

A 10-fold cross-validation was performed to evaluate the proposed framework, where in each fold, a leave- $n$ -subject-out practice was utilized. The value of  $n$  depends on the number of training samples. For example, when one sample was used for training, the user of the training sample was eliminated from the testing set which resulted in  $n = 9$ . Otherwise, when there were 9 training samples, 9 participants' samples were used for training, so the remaining one subject's data was used for testing ( $n = 1$ ). The proposed framework has an average classification accuracy of 95.89% with 18 training samples for each exploration procedure type. Figure 4.3 shows the confusion matrix, where rows represent ground truth and columns represent predicted labels. The precision rate, recall rate and F1 score were also computed for each class of exploration procedure according to the one-vs-rest basis (**Error! Reference source not found.**).

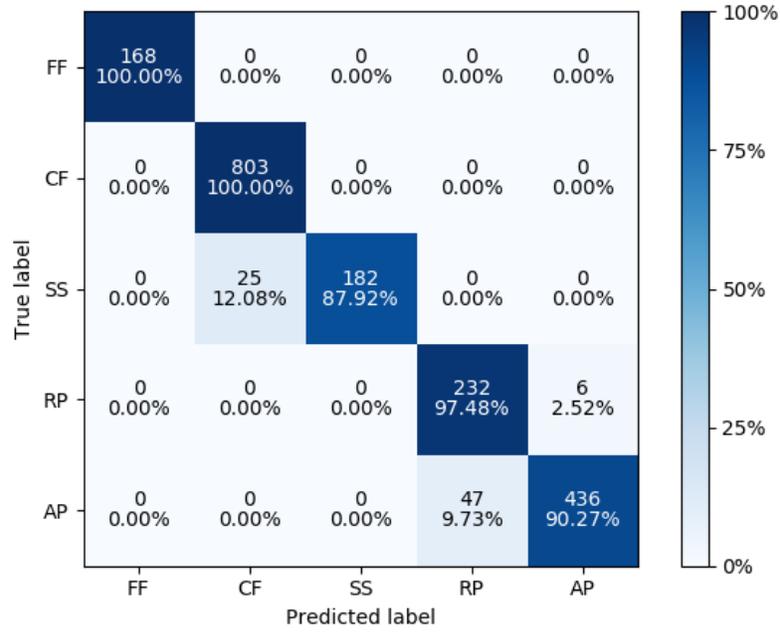


Figure 4.3 Confusion matrix of classification accuracy for the proposed framework with 95.89% accuracy.

Table 4.1 The precision rate, recall rate and f1 score of the results.

	<b>FF</b>	<b>CF</b>	<b>SS</b>	<b>RP</b>	<b>AP</b>
<b>Precision</b>	1.00	0.97	1.00	0.83	0.99
<b>Recall</b>	1.00	1.00	0.88	0.97	0.90
<b>F1 score</b>	1.00	0.98	0.94	0.90	0.94

Further analysis was conducted to understand the effect on classification accuracy when the SNNs were trained with different number of samples. Figure 4.4 shows the average classification accuracy with variance over different number of training samples. The accuracy reaches 94% with 5 training samples.

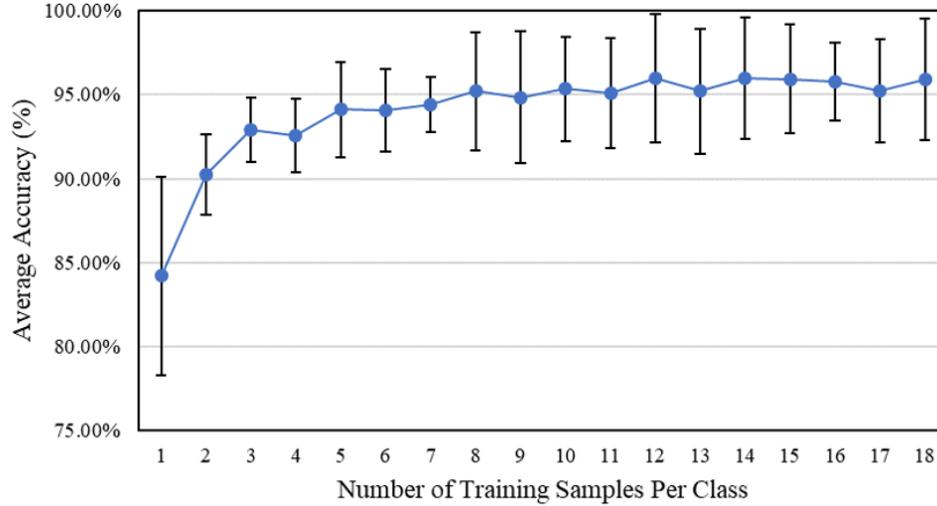


Figure 4.4 Classification accuracy over different number of training samples.

### *Time complexity analysis*

One limitation of the proposed approach is that it takes longer for classification when the number of training samples increases because the testing sample needs to be compared with all training templates. Such an instance-based classification scheme requires  $O(N)$  time for testing, where  $N$  is the number of templates in training set. A different SNN-based classification scheme would tackle this challenge by training Support Vector Machines (SVM) with Normalized Histogram of Neuron Firings (NHNF) as features (Zhou & Wachs, 2018). The NHNF descriptor summarizes the SNN firing patterns over a time window and the SVM training creates discriminative features for classification purposes. As the number of training examples increase, the testing time is held constant as  $O(1)$ , thus making it more suitable for large datasets.

Adapting this approach with the features encoded with/without SNNs, the normalized histogram of feature  $\delta^r$  was also computed and concatenated with the NHNFs of the three SNN-encoded features. Classification was then performed by training this concatenated histogram using a SVM. The 10-fold leave-one-subject-out cross-validation was performed for this comparison. Compared with the proposed approach with 95.89% accuracy, faster classification was achieved by sacrificing the accuracy to 88.68% with the NHNF approach. Figure 4.5 shows its confusion matrix.

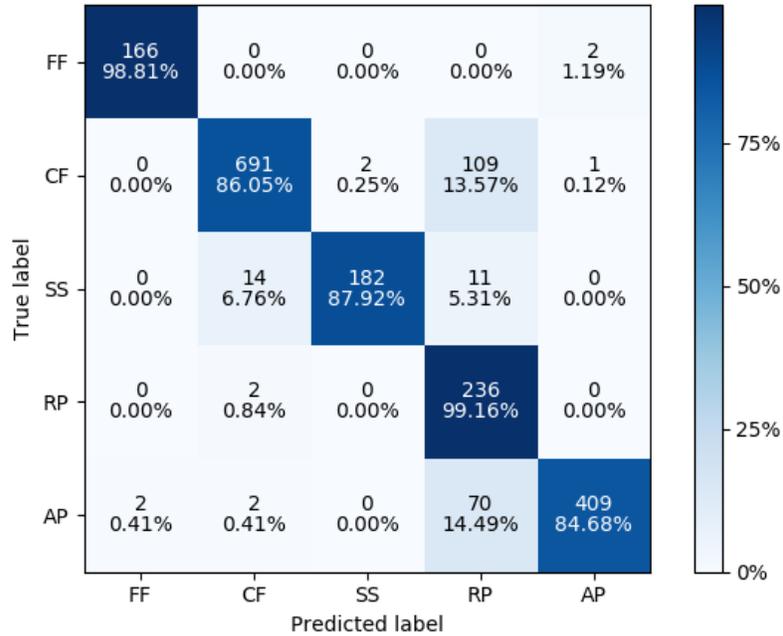


Figure 4.5 Confusion matrix of recognition accuracy for SNN with NHNF descriptor, with accuracy 88.68%.

### Comparisons with DTW and HMM

The proposed approach was also compared with popular recognition algorithms for time-series data, including DTW (“Dynamic Time Warping,” 2007) and HMM (Rabiner, 1989). An overall accuracy of 52.14% was obtained by using the modified DTW proposed in this research, but without the encoding using SNNs. In these comparisons, 18 training samples were used for each type of procedure. The leave-one-subject-out cross-validation was also performed. The confusion matrix is shown in Figure 4.6.

euclidean distance was used to calculate the differences between samples in terms of feature angle since it is a continuous value, while the LCS used in this framework was used as the distance function for feature context index and reference switch. Each testing sample was compared with 18 training samples for each type of procedure. The DTW confusion matrix shows an accuracy of 61.30% (Figure 4.6). It was observed that Frame Following (FF), Contour Following (CF) and Surface Sweeping (SS) were mostly recognized, while DTW failed with the other two types of procedures. Relative Positioning (RP) were likely recognized as CF because both procedures were related to the objects on an image. In contrast, Absolute Positioning (AP) were recognized mostly as FF as both trajectories had contact with the boundary of the image.

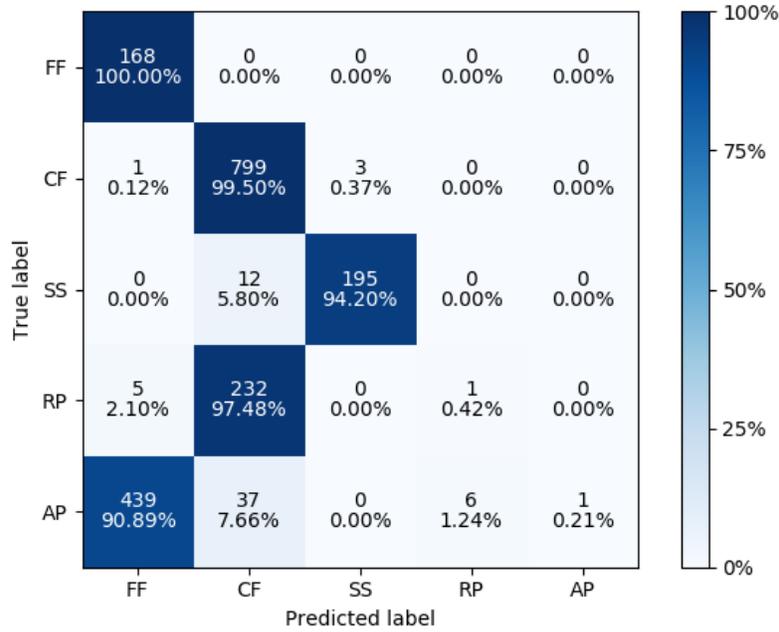


Figure 4.6 Confusion matrix of recognition accuracy for DTW, with accuracy 61.30%.

In terms of HMM, one model was trained for each type of exploration procedure. Therefore, in this experiment, five models were trained for classification. During classification, the testing sample was fed into all five models and the probability that this sample belongs to each model was calculated. The label of model with the highest probability was determined as the predicted label. In this study, every model had five hidden states and k-means clustering was applied to categorize the observations into discrete values as the features contain continuous values. The value of  $k$  was determined empirically as 10 in this experiment. An accuracy of 28.70% was obtained with the confusion matrix shown in Figure 4.7. Except Frame Following (FF), HMM has difficulties distinguishing the rest procedures.

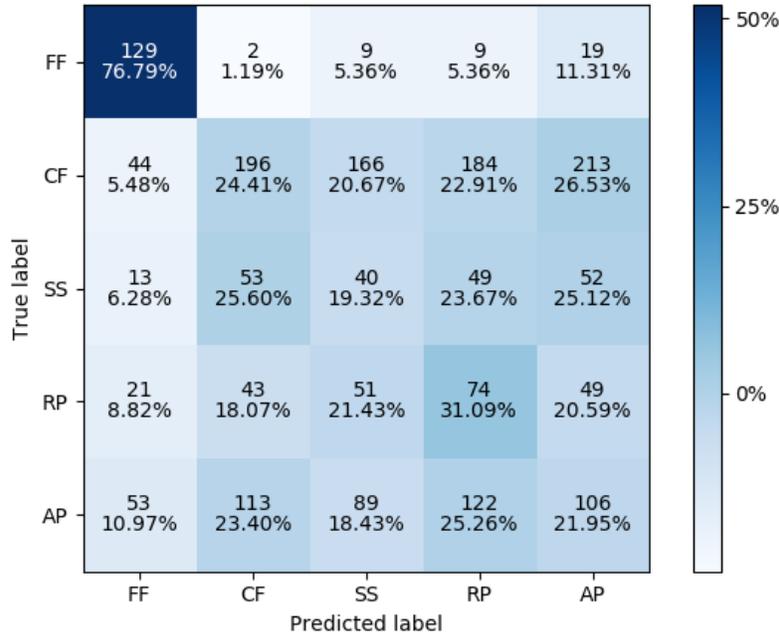


Figure 4.7 Confusion matrix of recognition accuracy for HMM, with accuracy 28.70%.

### ***Recognition of partial EPs***

Our proposed approach can recognize incomplete EPs. The collected data contained a several incomplete EPs. For example, Figure 4.8(a) is a completed CF, while Figure 4.8 (b) – (d) shows several forms of incomplete CF. Similar as Figure 4.2, lighter shade indicates earlier movement. It is a common practice for users to trace only partial contour when they are trying to locate the objects. For instance, to understand the object’s location relative to the top of the image, users would trace the partial contour of the object to determine its top, then leave from there to reach the top of the image. The degree of incompleteness of an EP can vary significantly. Thereby it was unreasonable and time-consuming to collect and train all forms of incomplete EPs. Using this proposed approach, the system would recognize various incomplete EPs while it being trained only with completed EPs.

Recognizing incomplete trajectories also allows early predictions. Having only the first several time frames of an EP is another example of incomplete trajectories. With such an early prediction capability, the system can provide instant assistance to the user, improving efficiency and decreasing the chances of critical errors. For example, the haptic-based system can provide magnetic force feedback to the boundaries of objects, once it detects the CF procedure, and deactivate this effect once the user is leaving the object’s boundary for AP or RP.

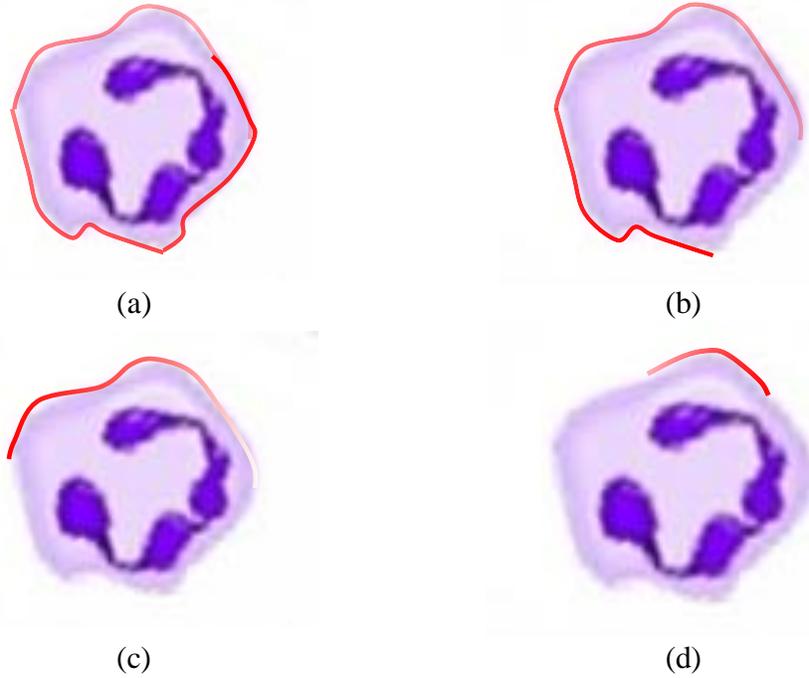


Figure 4.8 Examples of complete and incomplete EPs.

### ***Training SNNs with different properties of neurons***

The spiking neurons used to build the network are model following the approach proposed by Izhikevich (Izhikevich, 2004a). The membrane potential  $v$  and recovery variable  $u$  is updated following (4-1) and (4-2), when the neuron is stimulated with an input current  $I$ .

$$v' = 0.04v + 5v + 140 - u + I \quad (4-1)$$

$$u' = a(bv - u) \quad (4-2)$$

If the neuron spikes ( $v \geq 30mV$ ),  $v$  and  $u$  are reset according to (4-3).

$$\begin{cases} v \leftarrow c \\ u \leftarrow u + d \end{cases} \quad (4-3)$$

The four parameters  $a$ ,  $b$ ,  $c$  and  $d$  determines the property of the neuron thus affect the neuron firing activities. Parameter  $a$  describes how fast the neuron recovers, where smaller values leads to slower recovery. Parameter  $b$  indicates the sensitivity of the recovery variable to fluctuations of membrane's potential.  $c$  is the reset value of a neuron's membrane potential and  $d$  is the reset value of the recovery variable after it is fired.

Fourteen primary types of neurons (Izhikevich, 2004b) that are applicable in this study are examined. Their parameters and properties are explained in Table 4.2, together with the classification accuracy achieved by training SNNs with these types of neurons. Their responses to

different input current are also illustrated in Figure 4.9. The responses over time are shown above the line indicating input current, and the length of bolded horizontal refers 20ms.

Table 4.2 Classification accuracies with different type of neurons.

Type of Neuron	Description	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	Accuracy
<b>(a) Tonic Spiking</b>	<b>Fire trains of spikes with continues input</b>	<b>0.02</b>	<b>0.2</b>	<b>-65</b>	<b>6</b>	<b>94.39%</b>
(b) Phasic Spiking	Fire a single spike at the onset of stimulus	0.02	0.25	-65	6	72.76%
(c) Tonic Bursting	Fire periodic bursts of spikes when stimulated	0.02	0.2	-50	2	82.99%
(d) Phasic Bursting	Fire bursts of spikes when stimulated	0.02	0.25	-55	0.05	75.76%
(e) Mixed Mode	Fire phasic bursts at the onset of stimulus and then fire trains of spikes	0.02	0.2	-55	4	85.55%
(f) Frequency Adaption	Fire trains of spikes with decreasing frequencies	0.01	0.2	-65	8	93.72%
(g) Subthreshold Oscillations	Neurons exhibiting oscillatory potentials	0.05	0.26	-60	0	69.21%
(h) Resonator	respond only to the doublet whose frequency resonates with the frequency of subthreshold oscillations.	0.1	0.26	-60	-1	56.03%
(i) Rebound Spike	Fire a post-inhibitory spike at the onset of an inhibitory input	0.03	0.25	-60	4	61.15%
(j) Rebound Burst	Fire post-inhibitory bursts at the onset of an inhibitory input	0.03	0.25	-52	0	73.21%
(k) Threshold Variability	Variable firing threshold	0.03	0.25	-60	4	46.75%
(l) Bistability	Switching between resting and spiking with a stimulus	1	1.5	-60	0	71.65%
(m) DAP	Depolarizing After-Potentials	1	0.2	-60	-21	76.26%
(n) Accomodation	Less excitable to strong but slowly increasing input	0.02	1	-55	4	18.51%

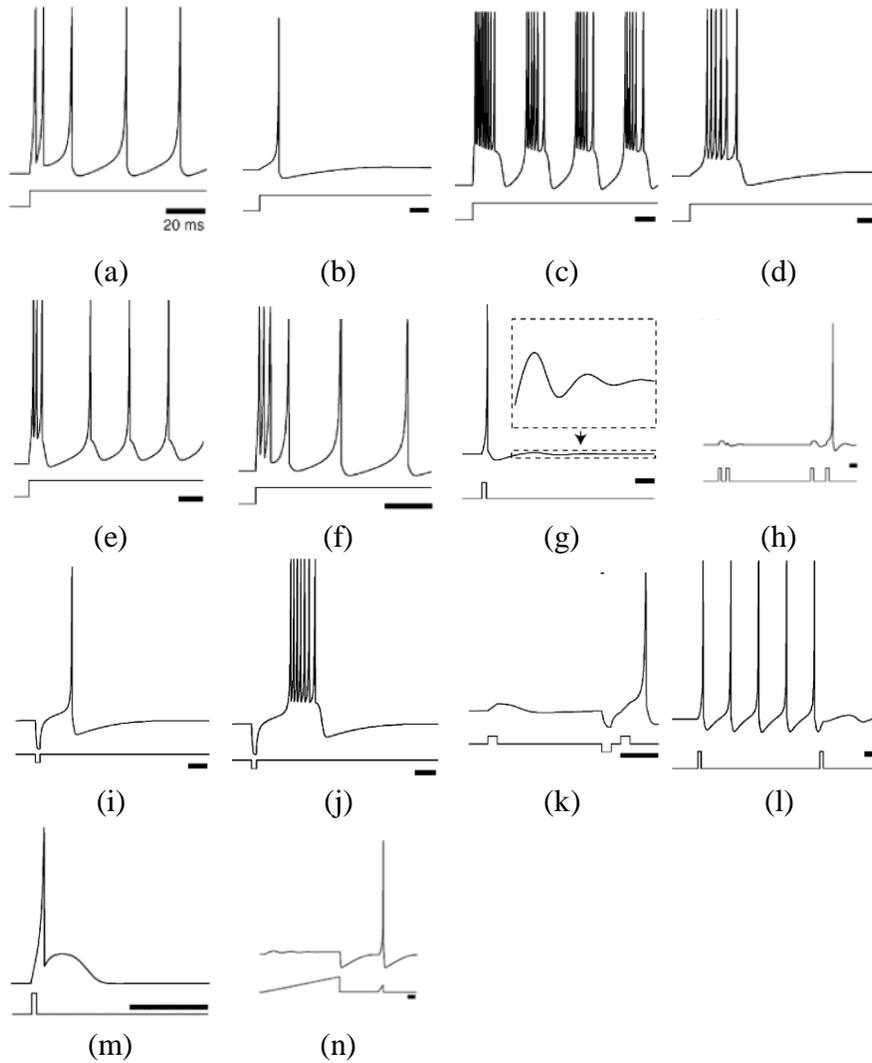


Figure 4.9 The responses to stimulus of different types of neurons.

It was observed that the type of neuron that fired a train of spikes and adapted its spiking frequency over time exhibited the highest the accuracy of 94.39% and 93.72%. These were very close to the higher accuracy of 94.55% achieved by using the regular spiking neurons ( $a=0.02$ ,  $b=0.2$ ,  $c=-65$ ,  $d=8$ ).

## 4.2 Developing an Assistance Model

To develop assisting strategies, experiments were conducted with blind and blind-folded subjects, understanding their goals of performing a particular EP.

### 4.2.1 Data collection

Participants were asked to explore a blood smear image using the same setup illustrated in Figure 4.1. Six blind-folded participants were recruited for this experiment, including 3 females and 3 males, with age ranging from 20 to 30. As opposed to the data collection procedure in the first experiment, this time, participants were doing a think-out-loud image exploration. After they performed one EP, they would need to say loud why they did it. The goals are then summarized by the observer. The number of times a goal appears were recorded for further analysis.

### 4.2.2 Results

The normalized histogram is computed for each user to understand the distribution of goals (Figure 4.10), where each color of bars represent the data of one subject.

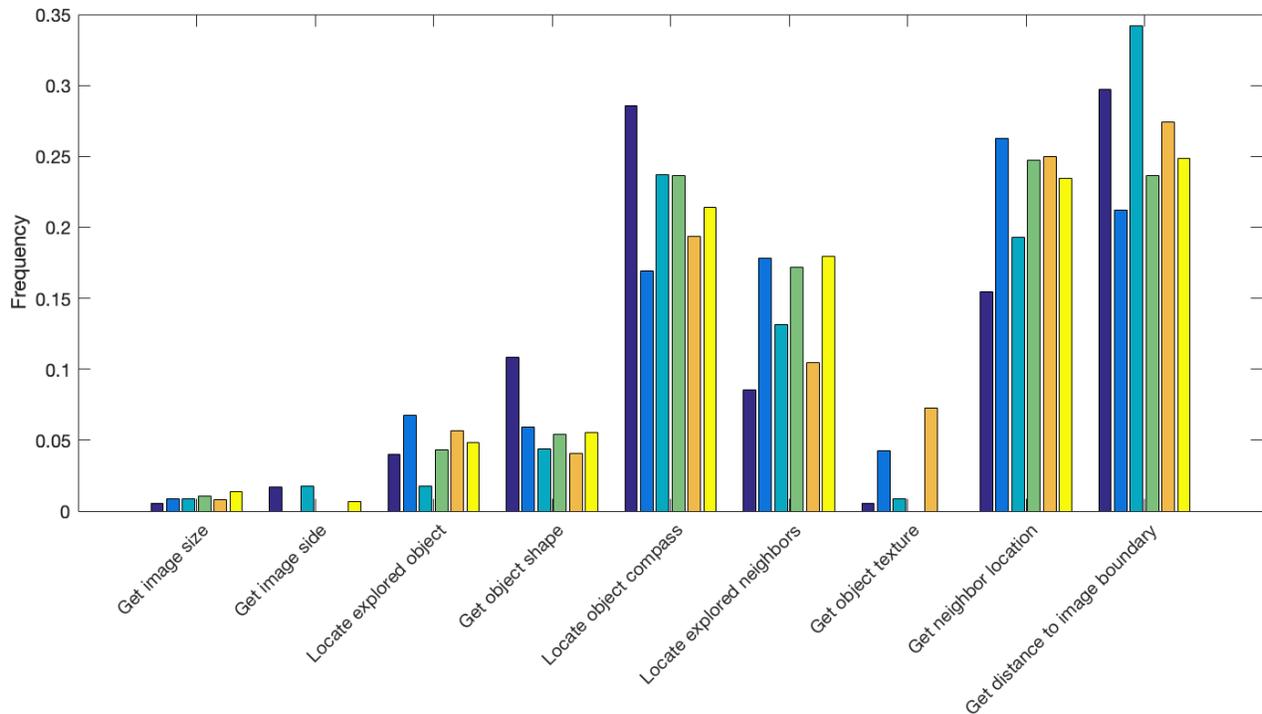


Figure 4.10 Normalized histograms of goals for image exploration.

## 4.3 Sensory Substituted User Interface

Experiments were designed for each assisting strategy to determine the optimal sensory modalities to deliver each exploratory aid. A within-participants experiment setup was utilized to

eliminate the differences introduced by the variations of subjects. Four interfaces were compared in these experiments, including the interface with exploratory aids delivered through vibration, sound, magnetic attraction and the interface without exploratory aids. As a baseline, the interface was compared with versions of the system not providing assistance in order to determine whether a certain aid was necessary.

#### **4.3.1 Participants**

Twenty participants aged from 20 to 50, were recruited for this study, including 2 blind users and 18 blind-folded individuals. This study was approved by Purdue IRB.

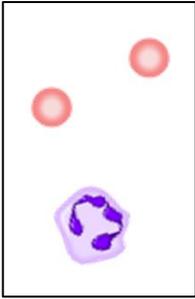
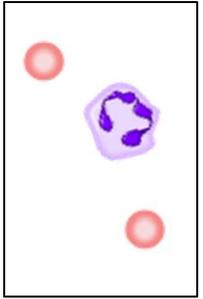
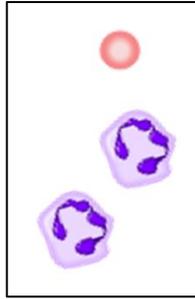
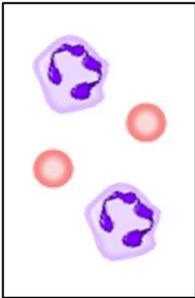
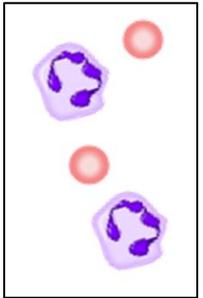
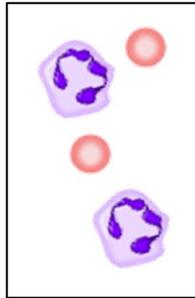
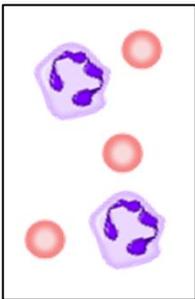
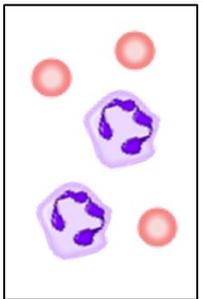
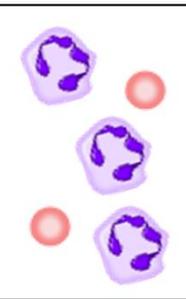
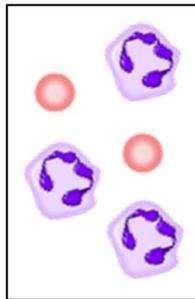
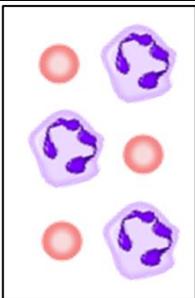
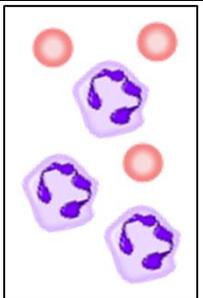
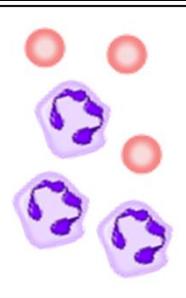
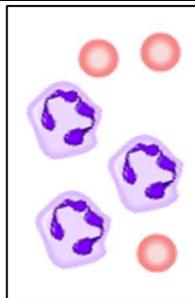
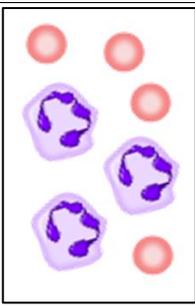
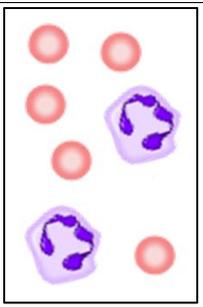
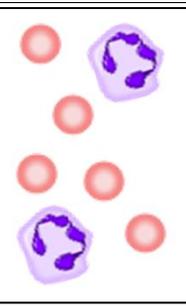
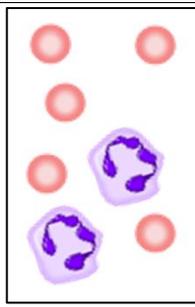
#### **4.3.2 Tasks and Measurements**

Participants performed two tasks in this study, that task 1 was used to evaluate Border Projection (BP) and Contour Neighbors (CN), while users performed task 2 to evaluate Contour Compass (CC) and Path Shortcut (PS). In Task 1, participants were required to explore an image using the multisensory interface and confirm their positions by touching them one by one using the haptic controller; while in Task 2, the users would explore an image using the multisensory interface and replicate the image with 3D printed pieces.

##### ***Task 1: Evaluation of BP and CN***

In each trial of this experiment, the participants explored the whole image using the multisensory interface and build a mental image of their exploration. After they felt confident about their comprehension about the image content, the participants would confirm the positions of the objects by pointing at the object one by one using the haptic controller. Five types of blood smear images were used as examples in this study, with the number of cells on the image ranged from 3 to 7. To eliminate learning effect, different images were used for each testing condition and the sequence of exploring these images were randomized for each participant. Table 4.3 shows all the explored images. The time taken for the whole task and the accuracy of confirmed locations were recorded. The accuracy in this task is defined as the percentage of correctly located objects.

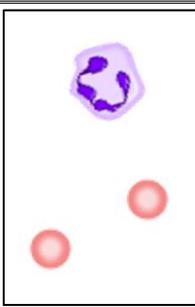
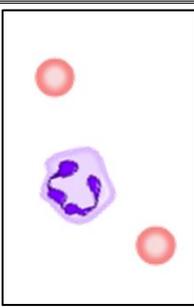
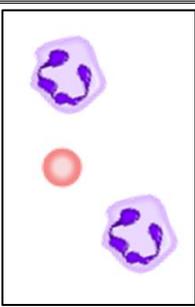
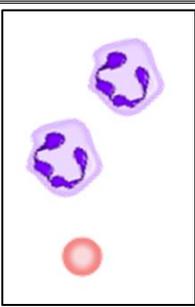
Table 4.3 Testing Images for Task 1.

Number of Objects on the Image	Image			
3				
4				
5				
6				
7				

### ***Task 2: Evaluation of CC and PS***

To understand how CC and PS can help users comprehend image content, participants were asked to explore an image using the multisensory interface and then replicate it using 3D printed blocks. Similarly to task 1, learning effect was eliminated by using different images for testing and the order of these images were randomized for all participants. The testing images were shown in Table 4.4. This task also measured the time used for each trial and how accurate the image was replicated. The accuracy of this task is defined as the similarity between the replicated image and the original image, which was defined in (3-9). Table 4.4 also shows an example of the replicated image for each testing image.

Table 4.4 Testing Images for Task 2.

<b>Testing Image</b>				
<b>Replicated Image</b>				

### ***Subjective Measurements***

Subjective measurements about the usability were also collected through a Likert-scale questionnaire. Participants were asked to rate the amount of helpfulness and intrusiveness (expressed through interruptions) when the exploratory aids were delivered in different types of sensory feedback.

### 4.3.3 Procedure

Practice trials were conducted in order to get the participants familiar with the different interfaces and modalities. After the participants felt comfortable with the system, they performed the task with a random order of experiment conditions. A post-experiment usability survey was filled after all the trials of an exploratory aid were completed.

### 4.3.4 Experimental Results

#### *Border Projection*

Task 1 was used to evaluate BP and it involved two experimental factors, including the number of objects on an image and the type of the sensory feedback to deliver the aid. Therefore, a two-way ANOVA was conducted to analyze the time taken to complete the task. Significant difference was found in both factors: the comparisons between types of sensory feedback has a p-value of  $2e-16$ , and the differences between number of cells has a p-value of  $1.6e-13$ . However, the interaction between the two factors is not significant. Figure 4.11 shows the average time taken for the task using BP.

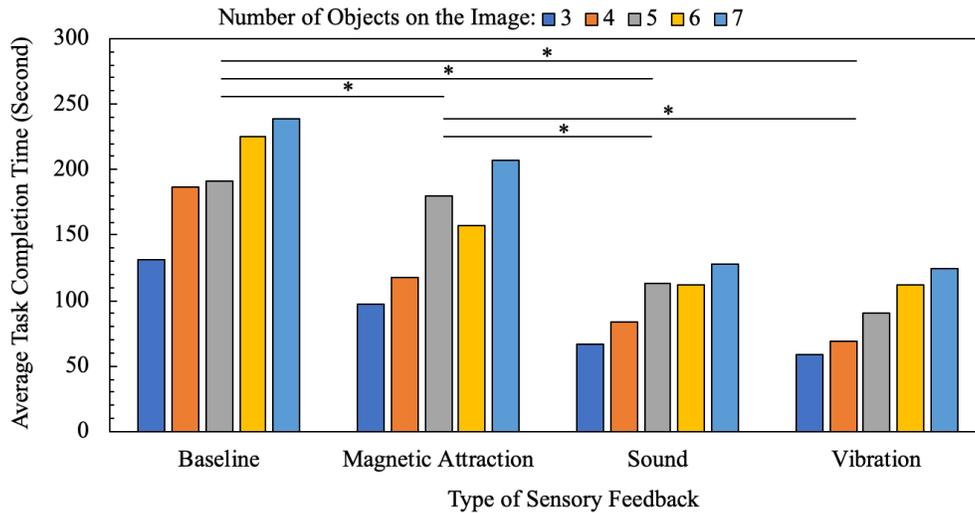


Figure 4.11 Average task completion time for Task 1 with Border Projection.

Post-hoc analysis was performed in terms of the types of sensory feedback. Figure 4.11 shows that Border Projection (BP) helped the participants completing the task significantly faster

than the baseline study. Vibrational and auditory BP has the shortest task completion time of 91.24s and 100.59s without significant difference between themselves. Magnetic feedback has a significant longer task completion time when compared to vibration and sound feedback, with an average of 151.87s, followed by the baseline study with an average of 194.66s. In terms of accuracy, participants correctly confirmed all the locations with the help of BP, while the accuracy for trials without any aids is 87.84%.

Figure 4.12 and Figure 4.13 show the distribution of users' ratings of helpfulness and interruption when BP is provided. It was observed that users' ratings are in accordance with their performance. Sound and vibration were more helpful than magnetic attraction for BP. In terms of interruption, vibration and sound was found to be the least interruptive, while some participants found magnetic attraction was interrupting with their explorations.

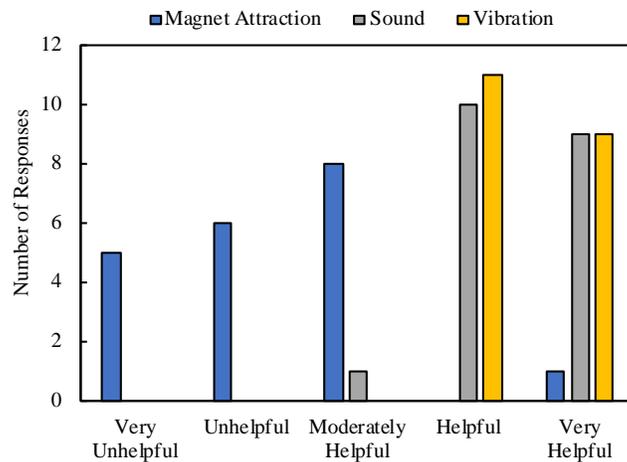


Figure 4.12 Level of helpfulness of Border Projection.

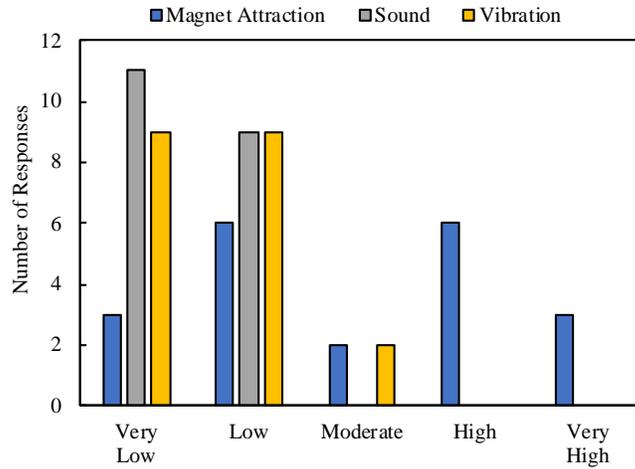


Figure 4.13 Level of interruption of Border Projection.

### Contour Neighbors

A two-way ANOVA was conducted to analyze the time taken with Contour Neighbors (CN), similarly to the analysis performed with BP, including factors: the number of objects on an image and the type of the sensory feedback to deliver the aid. Significant difference was found in both factors: p-value of type of sensory feedback is  $2.66e-15$  and p-value of number of cells is  $1.81e-14$ . However, the interaction between the two factors was not significant ( $p\text{-value}=0.474$ ). Figure 4.14 shows the average time taken for the task using CN.

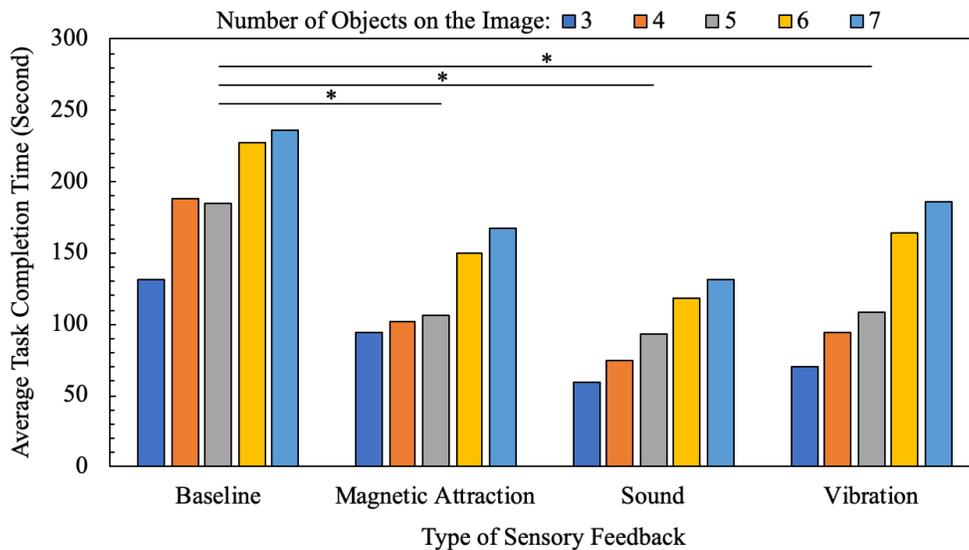


Figure 4.14 Average task completion time for Task 1 with Contour Neighbors.

To understand the difference between different types of sensory feedback, post-hoc analysis was performed. From Figure 4.14 one can see that CN helped the participants completing the task significantly faster than the baseline study. The average task completion time of the baseline study is 193.12s, while the average task completion time of magnetic, auditory and vibrational feedback is 123.72s, 95.42s and 125.03s, respectively, which is 35.95%, 50.60% and 35.27% faster than the trials without CN. However, there was no significance among these three types of sensory feedback. In terms of accuracy, participants also correctly confirmed all the locations with the help of CN.

The distributions of users' ratings of helpfulness and interruption when CN was provided are shown in Figure 4.15 and Figure 4.16. Participants found sound and vibration were more helpful than magnetic attraction. While in terms of interruption, participants didn't indicate strong preference over a certain type of feedback.

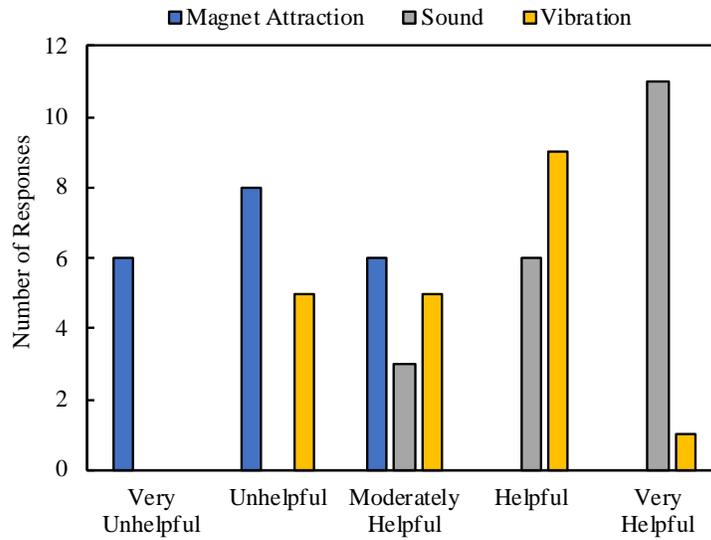


Figure 4.15 Level of helpfulness of Contour Neighbors.

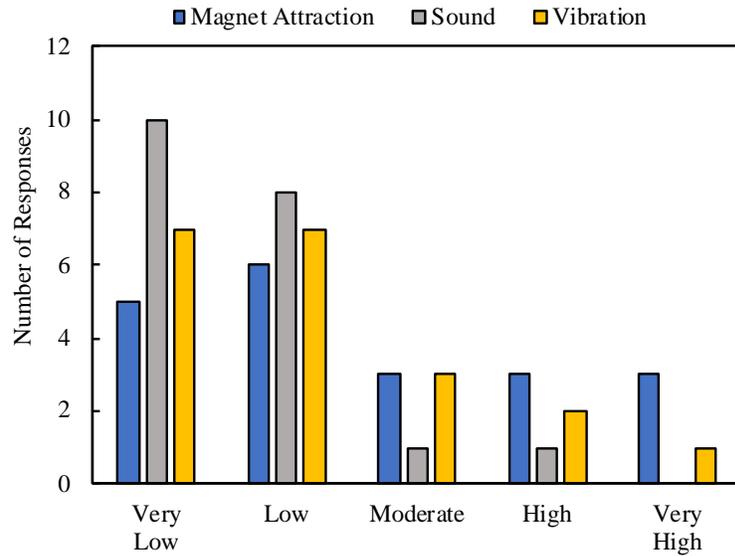


Figure 4.16 Level of interruption of Contour Neighbors.

### ***Contour Compass***

Contour Compass (CC) was evaluated in Task 2 when only one factor was involved, which is the type of the sensory feedback that delivered the aid. Therefore, a one-way ANOVA was conducted to analyze the time taken for the task and the accuracy of image understanding. Figure 4.17 shows the average task completion time and accuracy of image replication comparing the baseline and three types of CC. Significant difference was found among the four testing conditions in both task completion time ( $p\text{-value}=3.94e-05$ ) and replication accuracy ( $p\text{-value}=6.68e-04$ ). Post-hoc analysis indicated that CC delivered through the three types of feedback had significant shorter task completion time compared to the baseline. The average task completion time for baseline, magnetic feedback, auditory feedback and vibrational feedback is 189.23s, 152.03s, 126.94s and 146.16s respectively. Similarly, CC delivered through the three types of feedback had significant higher replication accuracy compared to the baseline. The average replication accuracy is 87.47%, 89.73%, 88.88% and 89.63% for the baseline, magnetic, auditory and vibrational CC. However, there was no significant difference among these three types of sensory feedback in terms of both task completion time and replication accuracy.

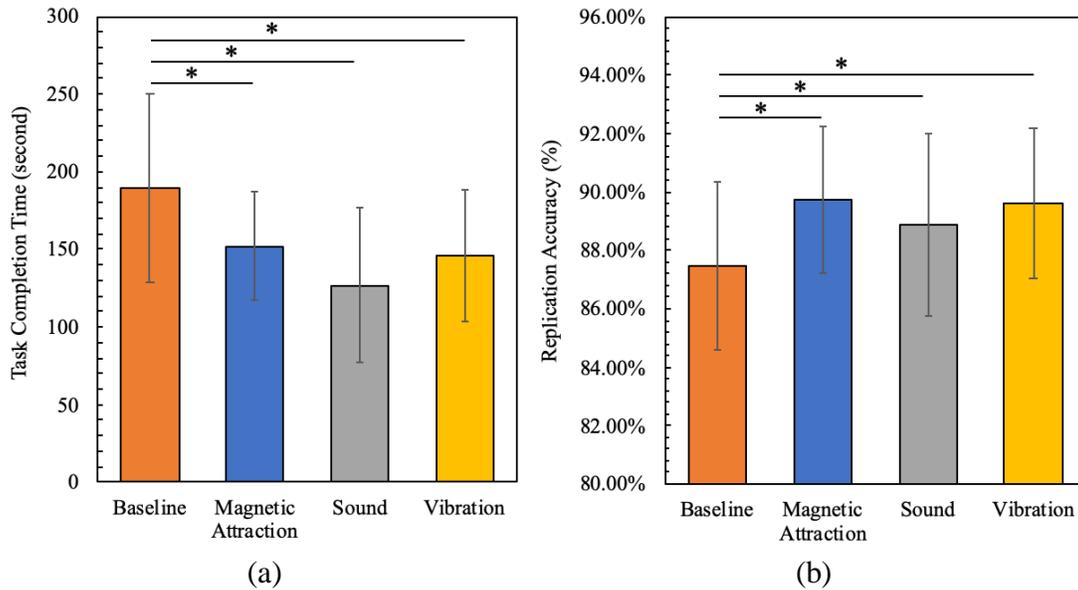


Figure 4.17 Performance comparisons among the baseline and Contour Compass via three types of sensory feedback: (a) Average task completion time; (b) Replication accuracy.

The distributions of users' ratings of helpfulness and interruption are shown in Figure 4.18 and Figure 4.19. Participants found sound and vibration more helpful and less disruptive than magnetic attraction.

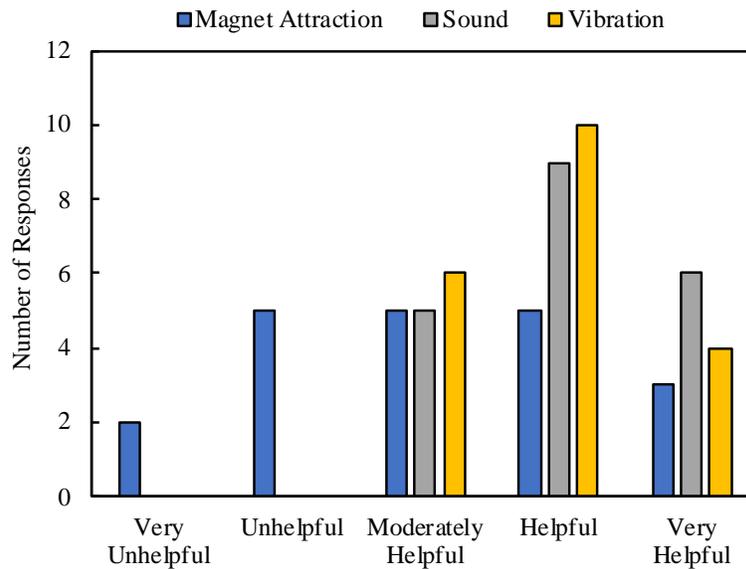


Figure 4.18 Level of helpfulness of Contour Compass.

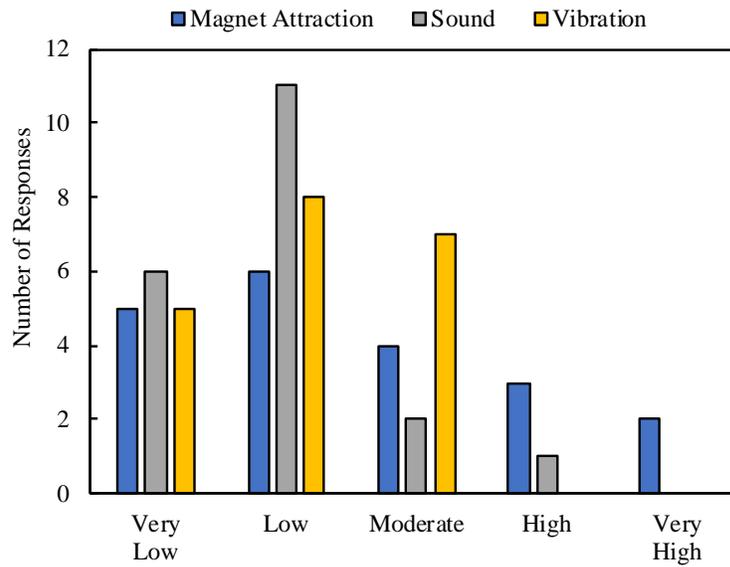


Figure 4.19 Level of interruption of Contour Compass.

### *Path Shortcut*

Task 2 was also performed to assess the Path Shortcut (PS) procedure. One-way ANOVA was conducted to analyze the difference among different types of sensory feedback considering the time taken for the task and the replication accuracy. Figure 4.20 shows the average task completion time and replication accuracy comparing the baseline and the aid PS delivered through magnetic attraction, sound and vibration.

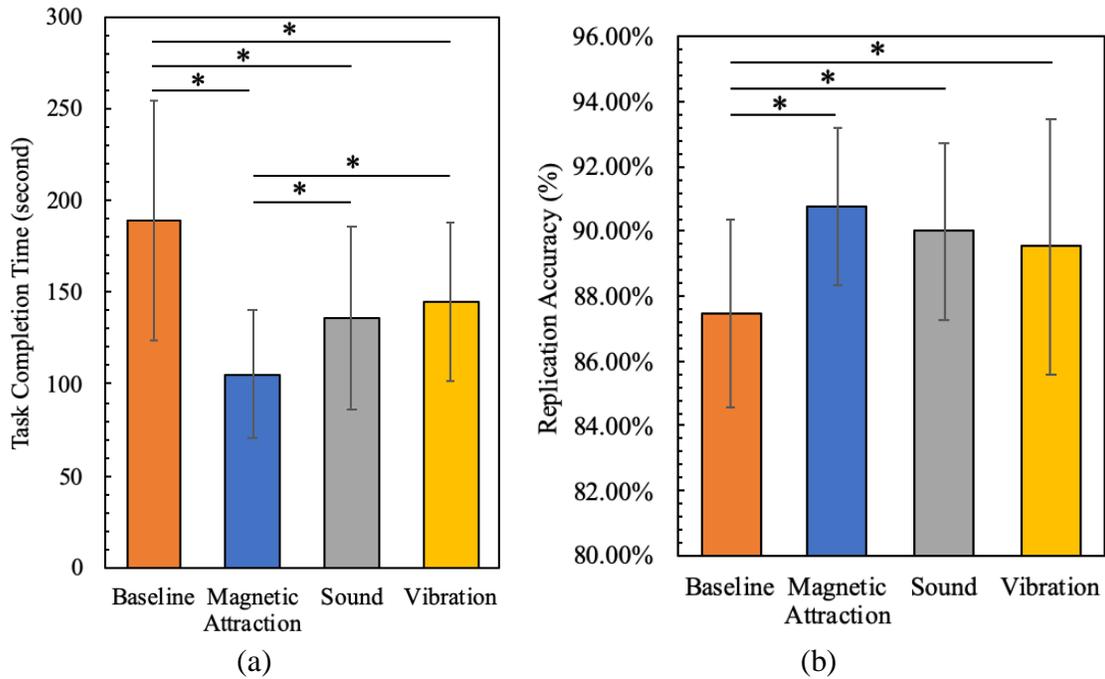


Figure 4.20 Performance comparisons among the baseline and Path Shortcut via three types of sensory feedback: (a) Average task completion time; (b) Replication accuracy.

Significant difference was found among the four testing conditions in both task completion time ( $p\text{-value}=1.09e-10$ ) and replication accuracy ( $p\text{-value}=2.61e-05$ ). Post-hoc analysis indicated that PS delivered through the three types of feedback had significant shorter task completion time compared to the baseline. Among the three types of sensory feedback, magnetic PS has the shortest task completion time ( $t_{mean}=105.08s$ ), which is significant different than the auditory and vibrational PS. However, there was no significant difference between sound and vibration. Exploration using magnetic PS was 44.35% faster than the one without any exploratory aids. Moreover, PS delivered through the three types of feedback had significant higher replication accuracy compared to the baseline. The average replication accuracy was 87.47%, 90.79%, 90.00% and 89.53% for the baseline, magnetic, auditory and vibrational PS. However, there was no significant difference among these three types of sensory feedback in terms of replication accuracy.

The distributions of users' ratings of helpfulness and interruption when PS was provided are shown in Figure 4.21 and Figure 4.22. It was observed that magnetic PS is more helpful than sound and vibration, while the level of interruption is similar among magnetic attraction, sound and vibration.

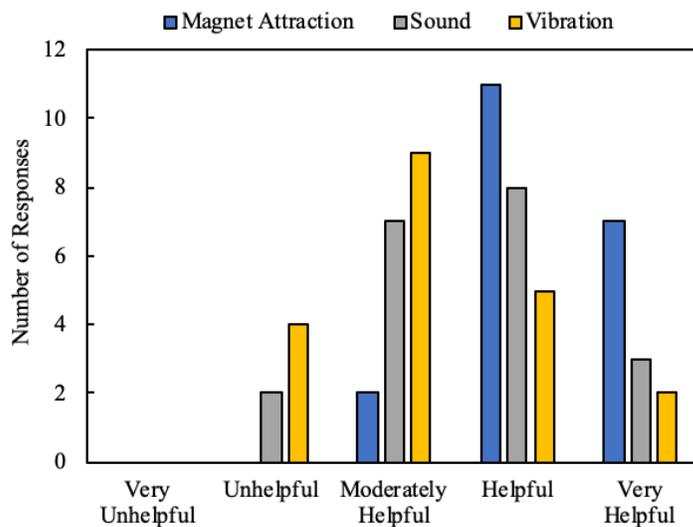


Figure 4.21 Level of helpfulness of Path Shortcut delivered through magnet attraction, sound and vibration.

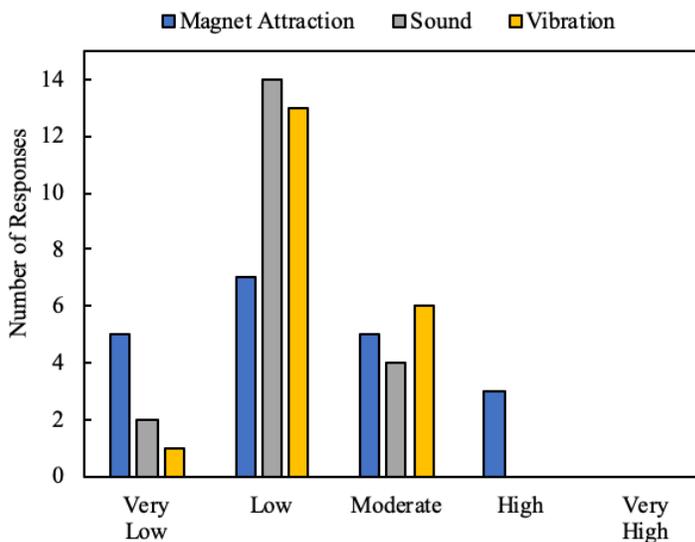


Figure 4.22 Level of interruption with Path Shortcut delivered through magnet attraction, sound and vibration.

#### 4.3.5 Discussion

The above experimental results validated that the use of four exploratory aids improved user performance and interaction experience in multimodal image exploration. The results also indicated the differences between vibrational, auditory and haptics feedback for aids rendering. In

this section, we discussed the differences between the three types of sensory feedback, as well as the differences between the two exploratory aids in terms of task performance and user experience.

### ***Comparisons of feedback***

Auditory feedback was found to be helpful when using Border Projection, Contour Neighbors and Contour Compass strategies. Users found it clear and easy to interpret as opposed to the tactual feedback received from the image exploration interface. Vibration feedback was also found to be helpful for Border Projection; however, this is not the case for the aid Contour Neighbors and Contour Compass. The reason mostly reported was the confusion between different sources of vibration delivered to the users. Besides the vibrational feedback from the dominant hand that indicated the cues for neighbors, participants also perceived vibration from the Tactor on the non-dominant hand that represented color brightness level. This confusion between two sources of vibration did not happen with the aid Border Projection because there was no other vibrational feedback provided at the image borders.

Magnetic attraction was found to be mostly effective to deliver Path Shortcut. Participants found it easy to follow straight lines when the path was delivered through the magnetic modality. Moving along a straight path is essential during exploration when vision is not available to maintain the sense of direction. However, for Border Protection, participants didn't find this type of feedback helpful. As participants moving along the image borders, they could follow straight lines without the help of magnetic attraction.

### ***Comparisons of exploratory aids***

Participants reported that incorporating Border Projection into the image exploration process helped them build more clear and global understanding of the image, as compared to Contour Neighbors. It was also observed from the experiments that Border Projection helped the participants to find all of the unexplored objects faster. However, to relocate the objects, participants took less time with the help of Contour Neighbors because the total distance of moving from object to object was shorter than the distance of moving along the image borders in order to locate objects. Participants had more accurate measurements of the spatial relations between objects when Contour Compass and Path Shortcut were activated.

### ***Exploratory aids integration***

Since all four exploratory aids indicated significant better performance than the interface without any aids, they were integrated into one system. For BP, vibration and sound indicated best performance. Although there is no statistical difference between these two sensory modalities, participants spent least time when BP was delivered through vibration. CN and CC should be delivered through different sensory modalities since they are activated with the same exploration procedure, i.e. when users are following the contour of an object. There was no statistical difference between the three types of sensory feedback for these two exploratory aids. However, vibration was not considered for these two aids because most participants found it confusing. In this case, magnetic attraction was selected for CN, while sound was determined for CC. For PS, participants had best performance when PS was delivered through magnetic attraction with significant advantage over the other sensory modalities. In conclusion, the system integrated with computer assistance had BP delivered through vibration, CC delivered through sound, and CN and PS delivered through magnetic attraction.

## **4.4 Evaluation of the computer-aided image exploration system**

Experiments were designed and performed by blind and blind-folded subjects to evaluate Task 2. Three conditions were compared in these experiments, including image exploration with human assistance, computer assistance and no assistance. Human assistance is the current gold standard helping people who has visual impairments to explore tactile images, while comparisons with a system that does not provide assistance served as the baseline of this study. A within-participants experiment setup was utilized to eliminate the differences introduced by the variations of subjects.

### **4.4.1 Participants**

Twenty participants aged from 20 to 50, were recruited for this study, including 2 blind users and 18 blind-folded individuals. This study was approved by Purdue IRB.

#### 4.4.2 Tasks and Measurements

Participants explored an image using the multisensory interface and replicate the image with 3D printed pieces. The performance metrics mentioned in 3.4 were applied to evaluate and compare the developed systems with the other two conditions mentioned above, including task completion time, replication accuracy and workload.

#### 4.4.3 Experiment Apparatus

The computer-assisted multimodal interface developed in this study consisted of a computer, a haptic controller, and two Tactors that generates vibrational feedback (shown in Figure 4.23). The computer ran a program that extracted image features using image processing techniques (Zhang et al., 2017) and provided exploratory aids according to the actions performed by the user. In this study, the haptic controller acted as the input device that is similar to a computer mouse. The user would hold it as a stylus to explore an image. The multimodal image interface delivered the shape of objects through force feedback and rendered color intensity through vibration. One Tactor was attached to the user's non-dominant hand to receive vibrational feedback for color intensity. To deliver exploratory aids, magnetic attraction was delivered through the haptic controller for CN and PS, sound was delivered by the computer speaker for CC and BP was expressed by vibrational feedback, which was sent to users by the Tactor attached on their dominant hand.

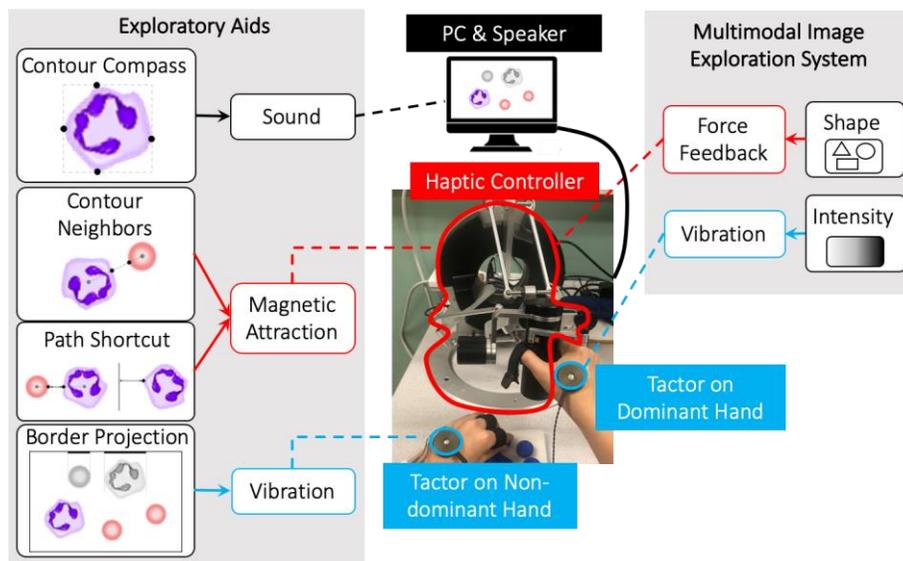


Figure 4.23 Multimodal image exploration system with four exploratory aids.

Participants wore devices to collect physiological data. Shown in Figure 4.24, Consensys Shimmer was used to measure heart rate and skin conductance. The ear clip was used to collect photoplethysmography (PPG) data, which was converted to heart rate. Skin conductivity was measured by two electrodes placed underneath the index and middle finger. Both sensors were placed on the non-dominant side of the participant and connected to the Shimmer, while the Shimmer was worn by the participant on an elastic wrist band on their non-dominant arm.

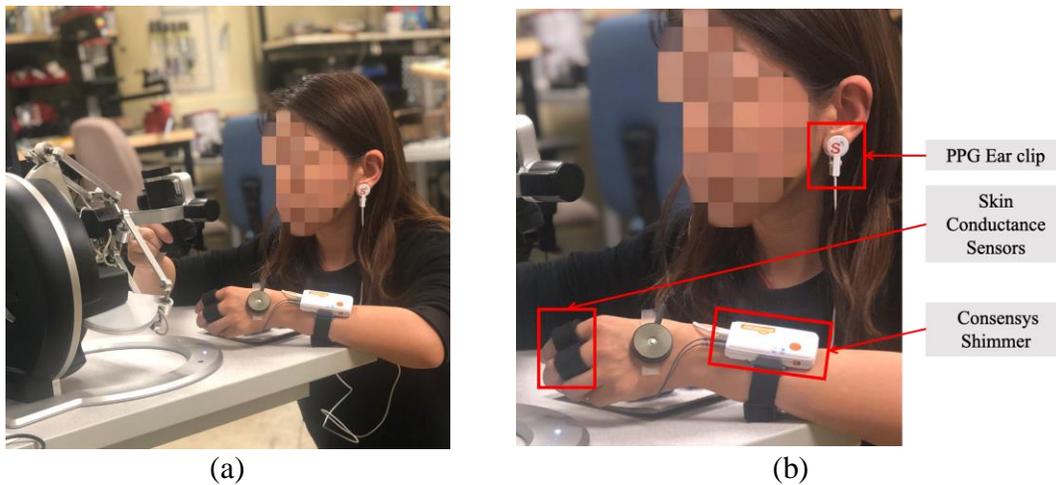


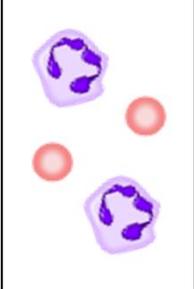
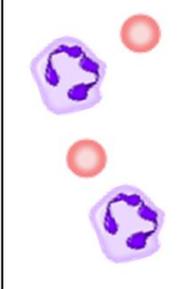
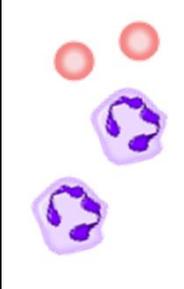
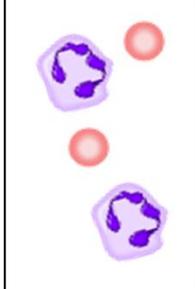
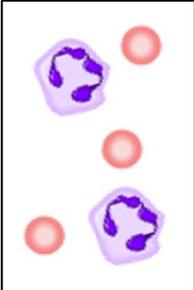
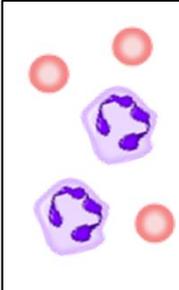
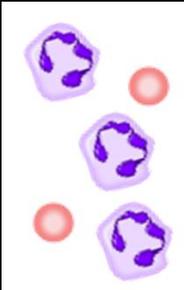
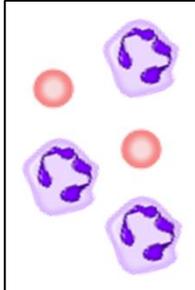
Figure 4.24 Setup for physiological data collection. (a) The participant is wearing a ear clip to collect heart rate data and sensors under the index and middle finger to measure skin conductance; (b) a close-up view of the sensors.

#### 4.4.4 Procedure

Participants were firstly asked to wear the Shimmer and the two Tactors were placed on their hands. Practice trials were then conducted to get them familiar with the different testing conditions. After participants feel comfortable with the system, they were asked to rest for three minutes to collect the baseline data for workload assessment. Participants would then start the trials of the experiment. Through one trial of the experiment, Shimmer was firstly turned on to start streaming the physiological data. Then participants had to explore one blood smear image using one of the testing conditions, and replicate the image using 3D printed objects throughout the exploration. After the participants finished the trial of each image and the Shimmer was turned off, they were asked to answer the NASA LTX workload questionnaire. Two different images with four or five blood cells were tested for each type of system. Table 4.5 shows the testing images. Therefore, each participants performed six trials in total. The order of these trials were randomized

for each participant. Participants were allowed to take a break if they felt tired and another resting session data were collected after each break.

Table 4.5 Test images to evaluate the system with exploratory aids.

Number of Cells	Image			
4				
5				

#### 4.4.5 Experimental Results

The developed computer-aided image exploration system was evaluated by task performance and mental workload. There were two factors involved in this study: types of interface and number of objects on the image, therefore, two-way ANOVA tests were performed on all the experimental data collected.

##### *Task performance*

In this study, task performance measured task completion time and the accuracy of image replication. For task completion time, the two-way ANOVA analysis indicated no significant differences in terms of number of objects on the image, but the type of image exploration system indicated significant difference ( $p\text{-value}=0.00778$ ). Also, there was no significant interaction between these two factors. Post-hoc analysis indicated the task completion time of human assisted

system and computer assisted system was significantly shorter than the time of the baseline system, but there was no significant difference between the interface with human and computer assistant. Figure 4.25 shows the average task completion time for these three types of image exploration system with four and five objects on the image. The average task completion time of the baseline, human assisted and computer assisted system was 266.775s, 211.375s, and 196.325s, respectively.

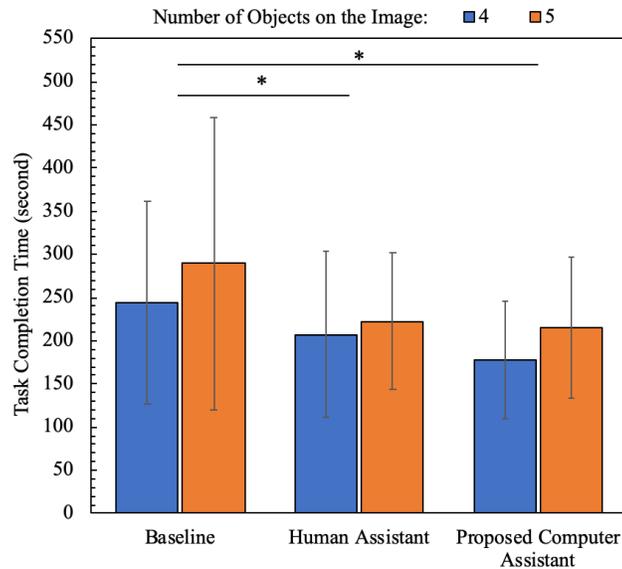


Figure 4.25 Comparisons of average task completion time among the three interfaces: baseline, human assistant and proposed computer assistant.

For replication accuracy, the two-way ANOVA analysis indicated significant differences in terms of both number of objects on the image ( $p\text{-value}=5.75e-08$ ) and the type of image exploration system ( $p\text{-value}=0.0145$ ). But there was no significant interaction between this two factors. Post-hoc analysis of the difference among the three types of system indicated significance between the accuracy of baseline and the system with human assistance, and between the accuracy of baseline and the system with computer assistance. However, there was no significant difference between the interface with human and computer assistant. Figure 4.26 shows the average image replication accuracy for these three types of system. The average accuracy of the baseline, human assisted and computer assisted system is 83.32%, 85.46%, and 85.30%, respectively.

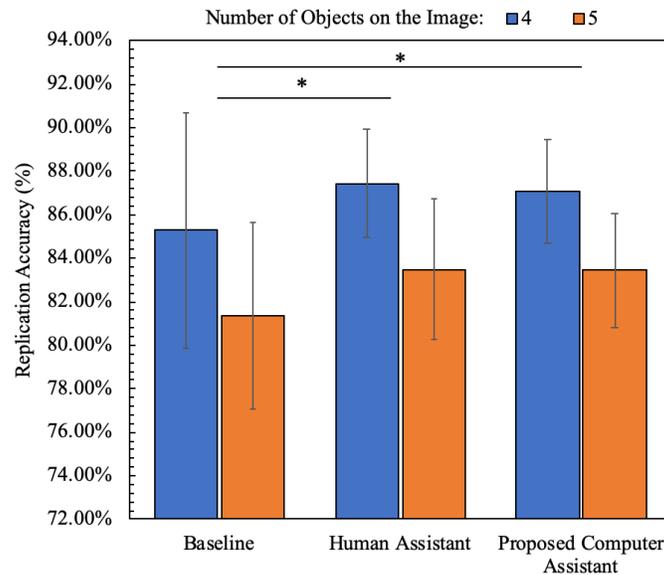


Figure 4.26 Comparisons of average replication accuracy among the three interfaces: baseline, human assistant and proposed computer assistant.

### ***Workload Assessment***

Workload was measured through both subjective self-rating questionnaires and objective physiological metrics. Figure 4.27 shows the results of the NASA TLX workload questionnaire. Significant difference was found among these three types of image exploration system ( $p\text{-value}=0.00126$ ), but the difference in terms of the number of objects on the image was not significant ( $p\text{-value}=0.32542$ ). Also, there was no interaction between these two factors ( $p\text{-value}=0.65914$ ). Post-hoc analysis indicated participants had significant lower workload using the system with human assistant and computer assistant compared to the baseline where tasks were performed without any assistant. However, the difference between the human and computer assistant was not significant. Ranging from 0 to 100, the average workload for the baseline, human assistant and computer assistant was 45.01, 31.79, and 36.32.

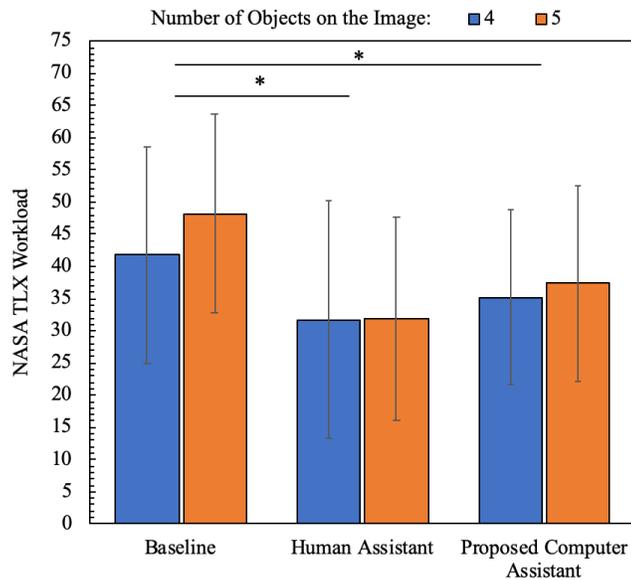


Figure 4.27 Comparisons of average mental workload measured by NASA TLX questionnaire among the three interfaces: baseline, human assistant and proposed computer assistant.

The physiological measurements applied in this study include heart rate, heart rate variability and Skin Conductance Responses (SCR). Average heart rate and average R-R intervals were computed from the PPG data through the software Kubios HRV Premium, while the frequency of SCR was calculated using Ledalab. Physiological data was then normalized in terms of each individual. Previous literature indicated mental workload can cause the elevation of heart rate and SCR frequency, while lead to a decrease of R-R intervals (Haarmann, Boucsein, & Schaefer, 2009; Hjortskov et al., 2004; Novak, Mihelj, & Munih, 2011; Weinger, Reddy, & Slagle, 2004). Therefore, the average heart rate and SCR frequency was normalized by the maximum value of the trials in one testing session. The average R-R intervals were normalized by the minimum value of all the trials performed in one session. Two-way ANOVA didn't show significant difference between these three types of interfaces. Figure 4.28 shows the normalized average heart rate and R-R intervals. Although statistical analysis didn't show significant difference between these three interfaces, participants had higher workload exploring images with five blood cells on it using the baseline interface, compared to the interface with human/computer assistant.

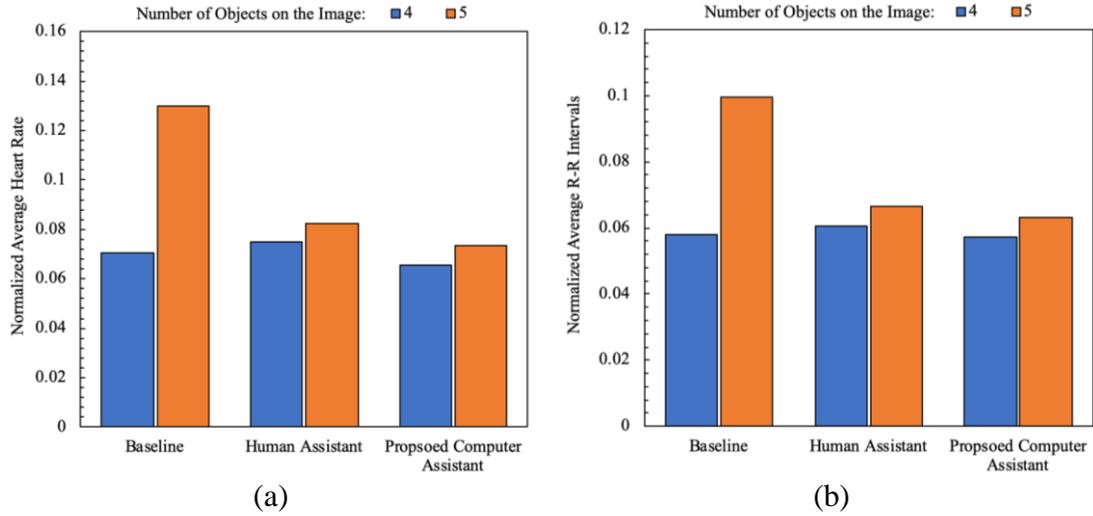


Figure 4.28 Comparisons among the three interfaces: baseline, human assistant and proposed computer assistant in terms of (a) normalized average heart rate and (b) normalized average R-R intervals.

Observed from Figure 4.29, participants had the lowest workload using the system with computer assistant, followed by human assistant, while the baseline system has the highest value.

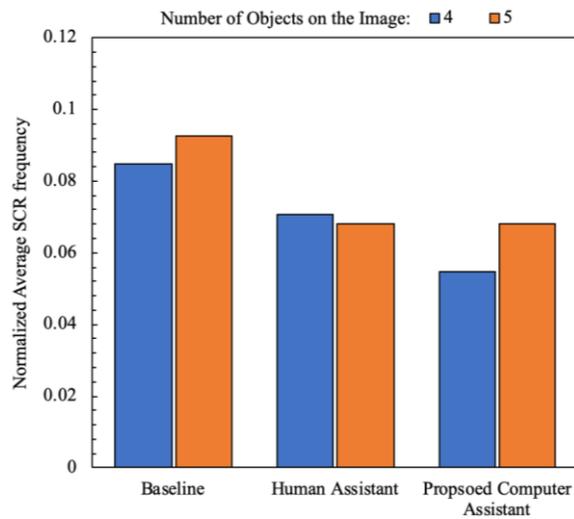


Figure 4.29 Comparisons among the three interfaces: baseline, human assistant and proposed computer assistant in terms of normalized average SCR frequency.

#### 4.4.6 Discussion

The experimental results indicated that the proposed system with exploratory aids improved users' efficiency for multimodal image exploration. The comparisons with the baseline and human assistant are discussed below.

##### *Comparisons with the baseline*

Participants spent an average of 26.41% less time compared to the exploration without any exploratory aids, while maintained slightly higher accuracy of image understanding. Subjective ratings also indicated that exploratory aids decreased the workload of understanding images by 19.31%. Although physiological measurements didn't show statistical significance, the average workload of using system with exploratory aids was lower than the baseline. The benefits of exploratory aids are more crucial when the task becomes more difficult. It is observed that the advantage over the baseline is larger when there are more cells on the image, especially for physiological measurements. From Figure 4.28 it can be seen that the workload index is a lot higher for the baseline with five objects on the image, while it is similar between the two different number of cells for the interface with exploratory aids.

##### *Comparisons with human assistant*

Human assistant is the current gold standard to help users with visual impairments understand images. The experimental results indicated comparable performance between human assistance and computer assistance. Although there was no significant difference between these two interface, participants spent slightly longer time with the human assistant. It is observed from the experiments that participants may spend less time to understand the image but getting instructions and verbal communication from the human assistant can take longer, compared to computer assistant. Participants all found human assistant more helpful when they got lost during the exploration. However, some participants also found the human assistant frustrating and interfering when they prefer to explore the image independently. In real life scenarios, having a human assistant by the side is not always possible, the exploratory aids can be used as a supplement to the human assistant.

## 4.5 Summary

This chapter presented the experiments conducted to validate the proposed approaches. Regarding RQ1, learning exploration procedures applied by BVI users, preliminary results indicated the potential of the proposed research problem, towards an image exploration system with intelligent assistance. RQ2 is answered with the “think-out-loud” experiments to design exploratory aids. The optimal modalities for the exploratory aids were determined by experiments with blind and blindfolded participants. Experiments were also conducted to answer RQ3 that validated the performance of the developed system.

## 5. CONCLUSIONS AND FUTURE WORK

This research led to the development of an intelligent assistant system to help individuals who are BVI explore images in real-time. The system consists of three components: the user model, the assistance model and the user interface. For the user model, a computational framework was developed that classifies different exploratory behaviors of blind users. The exploratory behaviors are summarized as five different exploration procedures. These procedures consist of various spatio-temporal patterns that are uniquely characterized by rotational, translational and scale invariant features. Numerical features representing the angle of movements and context related to the image features were further encoded through the training of multiple SNNs. The logic feature, referred as “reference switch”, was used later for classification without the encoding of a SNN. A distance-based classification scheme was applied in this work to the output of the SNN and the reference switch. We modified the DTW algorithm with a distance function using the length of LCS to compute the differences between model strings. To make the final decision of the predicted label for a sample, DST was integrated in the framework that combined the knowledge obtained from multiple features.

The assistance model was then developed with four exploratory aids. These aids were developed from the investigation on users’ goals during image exploration, including Border Projection, Contour Neighbors, Contour Compass and Path Shortcut. Three types of sensory feedback were compared to render the exploratory aids including vibration, sound and magnetic attraction. The user interface was then constructed with all four aids delivered through the optimal sensory modality. This proposed system improved the efficiency of digital image exploration for users who are BVI to understand images in shorter time and more accurately. Moreover, comparisons with human assistants indicated the developed system can provide almost human-level assistance throughout the exploration.

Nonetheless, the findings of this study have to be seen in light of some limitations. The first limitation concerns the experimental materials. The developed system was tested using blood smear images that are composed of distinct objects. To understand the image, the goal of the user was to find all the objects and investigate the relations between the objects. However, for images like human portraits, users would prefer to identify the person and possibly differentiate between various portraits. Subtle details including size and shape of eyes, face contour and positions of the

facial features are challenging to identify without visual feedback. The exploration behavior would also be different than the exploration of other types of images. Another limitation of this study is related to the computational bottleneck of the classification framework. The exploration procedure performed by participants couldn't be classified in real-time, therefore, the exploratory aids were provided based on users' position on the image, rather than what they were doing. Engineering techniques were developed to minimize the interference between different aids. The findings of this research also lead to insights for better sensory-substituted human-computer interaction.

### **5.1 Adaptive Human Computer Interface**

To provide effective assistance, understanding when the user needs help is another crucial aspect. To improve the current system, it is beneficial to develop an image exploration system that provides assistance adaptively based on users' real-time requirements. Throughout the experiments, participants expressed their frustration when they didn't need help while the human or computer assistant was constantly giving exploratory cues. Computer assistant can be ignored by not following the cues, however, human assistants would insist on giving the same cue if the user was not following their instructions. Therefore, a module that detects when the user needs help can give the user necessary support and at the same time not interfering with the exploration. Understanding when the user needs help is a challenging task that can't be determined by simply recognizing what the user is currently doing, but also needs to consider the actions the user has done and the user's current knowledge of the task. Spiking neural network was applied in this research to recognize user behavior. It can be extended to detect user status by integrating user's knowledge of the image and analyzing the sequence of performed actions.

Another insight from the experiments with human subjects is that users prefer getting real-time feedback on their understanding of the image. Participants not only felt more confident about their exploration, but also spent less time repetitively confirming the same information, which in other words, understood the image more efficiently. Users were placing 3D-printed objects on an attachable pad to present their understanding of the image. Human assistants can give feedback by visually comparing the difference between the replicated and original image. In this case, computer vision techniques can be utilized to provide the information. Depth cameras have been used to monitor the actions of users exploring tactile images by tracking and extracting the positions of the fingertips on the image (Brock et al., 2012). Similar concepts can be applied to extend the

current system. A video camera can be set up looking down at the replicated image and streaming the video data to be processed in real-time. Distortion correction and image processing techniques can be applied to extract the positions of objects places on the pad and compare with the original image.

## **5.2 Virtual Reality for Individuals with Visual Impairments**

As opposed to conventional virtual reality techniques that build virtual scenarios visually, auditory and tactual sensations are used to build a virtual world for individuals who have visual impairments. In this study, tactual sensation including haptics and vibrational feedback served as the major source to render the virtual world, while sound was used to provide exploratory aids. There are also studies investigating how to build a virtual world using different properties of sound. Spatial audio, such as simulated echolocation and distance-dependent hum volume modulation has been utilized to encode the environment around the user that helps with environment exploration (Massiceti, Hicks, & van Rheede, 2018). To extend this study, soundscape can be further explored to deliver the location information to the user more intuitively.

## **5.3 Exploratory Aids for Outdoor Exploration**

Exploring outdoor environments is another application for intelligent user interface with exploratory aids for individuals who are BVI. The system developed in this research could be adapted and extended to help people with visual impairments explore outdoor environment independently. Users who are BVI often rely on audio maps to explore an unknown environment. However, current audio maps can only provide sequential instructions on navigation or points of interests around the user without location information. The haptics-based interface utilized in this research can be adapted to work on mobile devices that helps the user build a mental image of their surrounding environment. Before heading over to the next location, the user could virtually navigate to the destination following the haptic cues. Compared to sequential verbal instructions, participants indicated better sense of orientation and mobility when they reviewed the route following haptic cues (Papadopoulos et al., 2017).

## REFERENCES

- Abboud, S., Hanassy, S., Levy-Tzedek, S., Maidenbaum, S., & Amedi, A. (2014). EyeMusic: Introducing a “visual” colorful experience for the blind using auditory sensory substitution. *Restorative Neurology and Neuroscience*, *32*(2), 247–257. <https://doi.org/10.3233/RNN-130338>
- Abowd, G. D., Atkeson, C. G., Hong, J., Long, S., Kooper, R., & Pinkerton, M. (1997). Cyberguide: A mobile context-aware tour guide. *Wireless Networks*, *3*(5), 421–433. <https://doi.org/10.1023/A:1019194325861>
- Arif, M., Brouard, T., & Vincent, N. (2006). A fusion methodology based on Dempster-Shafer evidence theory for two biometric applications. *Proceedings - International Conference on Pattern Recognition*, *4*, 590–593. <https://doi.org/10.1109/ICPR.2006.68>
- 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. <https://doi.org/10.1068/p281013>
- Bach-Y-Rita, P., Collins, C. C., Saunders, F. A., White, B., & Scadden, L. (1969). Vision substitution by tactile image projection [18]. *Nature*, *221*(5184), 963–964. <https://doi.org/10.1038/221963a0>
- Bach-y-Rita, P., & W. Kercel, S. (2003). Sensory substitution and the human-machine interface. *Trends in Cognitive Sciences*, *7*(12), 541–546. <https://doi.org/10.1016/j.tics.2003.10.013>
- Blenkhorn, P., & Evans, D. G. (1998). Using speech and touch to enable blind people to access schematic diagrams. *Journal of Network and Computer Applications*, *21*(1), 17–29. <https://doi.org/10.1006/JNCA.1998.0060>
- Blum, J. R., Bouchar, M., & Cooperstock, J. R. (2012). *What’s around Me? Spatialized Audio Augmented Reality for Blind Users with a Smartphone*. [https://doi.org/10.1007/978-3-642-30973-1\\_5](https://doi.org/10.1007/978-3-642-30973-1_5)
- Bowyer, K., Kranenburg, C., & Dougherty, S. (2001). Edge detector evaluation using empirical ROC curves. *Computer Vision and Image Understanding*, *84*(1), 77–103. <https://doi.org/10.1006/cviu.2001.0931>

- Brock, A., Lebaz, S., Oriola, B., Picard, D., Jouffrais, C., & Truillet, P. (2012). Kin'touch. *Proceedings of the 2012 ACM Annual Conference Extended Abstracts on Human Factors in Computing Systems Extended Abstracts - CHI EA '12*, 2471. <https://doi.org/10.1145/2212776.2223821>
- 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 Biomedical Engineering*, 45(10), 1279–1293. <https://doi.org/10.1109/10.720206>
- Colwell, C., Petrie, H., Kornbrot, D., Hardwick, A., & Furner, S. (1998). Haptic virtual reality for blind computer users. *Proceedings of the Third International ACM Conference on Assistive Technologies - Assets '98*, 92–99. <https://doi.org/10.1145/274497.274515>
- Crossan, A., & Brewster, S. (2006). Two-handed navigation in a haptic virtual environment. *CHI '06 Extended Abstracts on Human Factors in Computing Systems - CHI EA '06*, 676. <https://doi.org/10.1145/1125451.1125589>
- Csapó, Á., Wersényi, G., Nagy, H., & Stockman, T. (2015). A survey of assistive technologies and applications for blind users on mobile platforms: a review and foundation for research. *Journal on Multimodal User Interfaces*, 9(4), 275–286. <https://doi.org/10.1007/s12193-015-0182-7>
- Danker, A. J., & Rosenfeld, A. (1981). Blob Detection by Relaxation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-3(1), 79–92. <https://doi.org/10.1109/TPAMI.1981.4767053>
- Davidson, P. W. (1972). Haptic judgments of curvature by blind and sighted humans. *Journal of Experimental Psychology*, 93(1), 43–55. <https://doi.org/10.1037/h0032632>
- Dewhurst, D. (2009). Accessing audiotactile images with HFVE silooet. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5763 LNCS, 61–70. [https://doi.org/10.1007/978-3-642-04076-4\\_7](https://doi.org/10.1007/978-3-642-04076-4_7)
- Dynamic Time Warping. (2007). In *Information Retrieval for Music and Motion* (pp. 69–84). [https://doi.org/10.1007/978-3-540-74048-3\\_4](https://doi.org/10.1007/978-3-540-74048-3_4)
- Evans, K. K., & Treisman, A. (2011). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, 10(1), 6–6. <https://doi.org/10.1167/10.1.6>

- Fusco, G., & Morash, V. S. (2015). The Tactile Graphics Helper: Providing Audio Clarification for Tactile Graphics Using Machine Vision. *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility - ASSETS '15*, 97–106. <https://doi.org/10.1145/2700648.2809868>
- Gaunet, F., Martinez, J.-L., & Thinus-Blanc, C. (1997). Early-Blind Subjects' Spatial Representation of Manipulatory Space: Exploratory Strategies and Reaction to Change. *Perception*, 26(3), 345–366. <https://doi.org/10.1068/p260345>
- Golledge, R. G., Klatzky, R. L., Loomis, J. M., Speigle, J., & Tietz, J. (1998). A geographical information system for a GPS based personal guidance system. *International Journal of Geographical Information Science*, 12(7), 727–749. <https://doi.org/10.1080/136588198241635>
- Golledge, R. G., Loomis, J. M., Klatzky, R. L., Flury, A., & Yang, X. L. (1991). Designing a personal guidance system to aid navigation without sight: progress on the GIS component. *International Journal of Geographical Information Systems*, 5(4), 373–395. <https://doi.org/10.1080/02693799108927864>
- Guerreiro, T., Montague, K., Guerreiro, J., Nunes, R., Nicolau, H., & Gonçalves, D. J. V. (2015). Blind People Interacting with Large Touch Surfaces. *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces - ITS '15*, 25–34. <https://doi.org/10.1145/2817721.2817743>
- Gupta, R., Balakrishnan, M., & Rao, P. V. M. (2017). Tactile Diagrams for the Visually Impaired. *IEEE Potentials*, 36(1), 14–18. <https://doi.org/10.1109/MPOT.2016.2614754>
- Haarmann, A., Boucsein, W., & Schaefer, F. (2009). Combining electrodermal responses and cardiovascular measures for probing adaptive automation during simulated flight. *Applied Ergonomics*, 40(6), 1026–1040. <https://doi.org/10.1016/j.apergo.2009.04.011>
- Hart, S. G., & Wickens, C. D. (1990). Workload Assessment and Prediction. In *Manprint* (pp. 257–296). [https://doi.org/10.1007/978-94-009-0437-8\\_9](https://doi.org/10.1007/978-94-009-0437-8_9)
- Hatwell, Y., Streri, A., & Gentaz, E. (2003). *Touching for knowing : cognitive psychology of haptic manual perception*. John Benjamins Pub.
- Hjortskov, N., Rissén, D., Blangsted, A. K., Fallentin, N., Lundberg, U., & Sjøgaard, K. (2004). The effect of mental stress on heart rate variability and blood pressure during computer work. *European Journal of Applied Physiology*, 92(1–2), 84–89.

- Hsu, B., Cheng-Han Hsieh, C.-H., Sung-Nien Yu, S.-N., Ahissar, E., Arieli, A., & Zilbershtain-Kra, Y. (2013). A tactile vision substitution system for the study of active sensing. *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2013*, 3206–3209. <https://doi.org/10.1109/EMBC.2013.6610223>
- Iglesias, R., Casado, S., Gutierrez, T., Barbero, J. I., Avizzano, C. a., Marcheschi, S., & Bergamasco, M. (2004). Computer graphics access for blind people through a haptic and audio virtual environment. *Proceedings. Second International Conference on Creating, Connecting and Collaborating through Computing*, 13–18. <https://doi.org/10.1109/HAVE.2004.1391874>
- Izhikevich, E. M. (2004a). Dynamical Systems in Neuroscience: *Dynamical Systems*, Vol. 25, pp. 227–256. <https://doi.org/10.1017/S0143385704000173>
- Izhikevich, E. M. (2004b). Which Model to Use for Cortical Spiking Neurons? *IEEE TRANSACTIONS ON NEURAL NETWORKS*, 15(5), 1063. <https://doi.org/10.1109/TNN.2004.832719>
- Izhikevich, E. M. (2006). Polychronization: Computation with Spikes. *Neural Computation*, 18(2), 245–282. <https://doi.org/10.1162/089976606775093882>
- Johnson, L. A., & Higgins, C. M. (2006). A navigation aid for the blind using tactile-visual sensory substitution. *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, 6289–6292. <https://doi.org/10.1109/IEMBS.2006.259473>
- Kajimoto, H., Suzuki, M., Kanno, Y., Kajimoto, H., Suzuki, M., & Kanno, Y. (2014). HamsaTouch. *Proceedings of the Extended Abstracts of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI EA '14*, 1273–1278. <https://doi.org/10.1145/2559206.2581164>
- Kaklanis, N., Votis, K., & Tzovaras, D. (2013). Open Touch/Sound Maps: A system to convey street data through haptic and auditory feedback. *Computers & Geosciences*, 57, 59–67. <https://doi.org/10.1016/J.CAGEO.2013.03.005>
- Kane, S. K., Morris, M. R., Perkins, A. Z., Wigdor, D., Ladner, R. E., & Wobbrock, J. O. (2011). Access overlays: Improving Non-Visual Access to Large Touch Screens for Blind Users. *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology - UIST '11*, 273. <https://doi.org/10.1145/2047196.2047232>

- Kaneko, T., & Ooduchi, S. (2010). Tactile graphics in Braille textbooks: the development of a tactile graphics creation manual. *NISE Bulletin*, *10*, 13–28. Retrieved from [http://www.nise.go.jp/cms/resources/content/399/nise\\_a-10\\_2.pdf](http://www.nise.go.jp/cms/resources/content/399/nise_a-10_2.pdf)
- Kouroupetroglou, G., Martos, A., Papandreou, N., Papadopoulos, K., Argyropoulos, V., & Sideridis, G. D. (2016). *Tactile Identification of Embossed Raised Lines and Raised Squares with Variable Dot Elevation by Persons Who Are Blind*. [https://doi.org/10.1007/978-3-319-41267-2\\_11](https://doi.org/10.1007/978-3-319-41267-2_11)
- Kurze, M. A. (1999). TGuide: A guidance system for tactile image exploration. *Behaviour & Information Technology*, *18*(1), 11–17. <https://doi.org/10.1080/014492999119200>
- Lahav, O., Gedalevitz, H., Battersby, S., Brown, D., Evett, L., & Merritt, P. (2018). Virtual environment navigation with look-around mode to explore new real spaces by people who are blind. *Disability and Rehabilitation*, *40*(9), 1072–1084. <https://doi.org/10.1080/09638288.2017.1286391>
- Lederman, S. J., & Klatzky, R. L. (1987). Hand movements: A window into haptic object recognition. *Cognitive Psychology*, *19*(3), 342–368. [https://doi.org/10.1016/0010-0285\(87\)90008-9](https://doi.org/10.1016/0010-0285(87)90008-9)
- Lederman, S. J., & Klatzky, R. L. (1990). Haptic classification of common objects: Knowledge-driven exploration. *Cognitive Psychology*, *22*(4), 421–459. [https://doi.org/10.1016/0010-0285\(90\)90009-S](https://doi.org/10.1016/0010-0285(90)90009-S)
- Lederman, S. J., & Klatzky, R. L. (1993). Extracting object properties through haptic exploration. *Acta Psychologica*, *84*(1), 29–40. [https://doi.org/10.1016/0001-6918\(93\)90070-8](https://doi.org/10.1016/0001-6918(93)90070-8)
- Li, X., Dick, A., Shen, C., Zhang, Z., van den Hengel, A., & Wang, H. (2013). Visual tracking with spatio-temporal Dempster–Shafer information fusion. *IEEE Transactions on Image Processing*, *22*(8), 3028–3040.
- Löttsch, J. (1994). *Computer-aided access to tactile graphics for the blind*. [https://doi.org/10.1007/3-540-58476-5\\_188](https://doi.org/10.1007/3-540-58476-5_188)
- 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. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2958587>

- Massiceti, D., Hicks, S. L., & van Rheede, J. J. (2018). Stereosonic vision: Exploring visual-to-auditory sensory substitution mappings in an immersive virtual reality navigation paradigm. *PLOS ONE*, *13*(7), e0199389. <https://doi.org/10.1371/journal.pone.0199389>
- Meers, S., & Ward, K. (2004). A vision system for providing 3D perception of the environment via transcutaneous electro-neural stimulation. *Eighth International Conference on Information Visualisation*, 546–552. <https://doi.org/10.1109/IV.2004.1320198>
- Meijer, P. B. L. (1992). An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering*, *39*(2), 112–121. <https://doi.org/10.1109/10.121642>
- 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. *2013 International Conference on Advanced Technologies for Communications (ATC 2013)*, 586–591. <https://doi.org/10.1109/ATC.2013.6698183>
- Nourbakhsh, N., Chen, F., Wang, Y., & Calvo, R. A. (2017). Detecting Users' Cognitive Load by Galvanic Skin Response with Affective Interference. *ACM Transactions on Interactive Intelligent Systems*, *7*(3), 1–20. <https://doi.org/10.1145/2960413>
- Novak, D., Mihelj, M., & Munih, M. (2011). Psychophysiological responses to different levels of cognitive and physical workload in haptic interaction. *Robotica*, *29*(3), 367–374. <https://doi.org/10.1017/S0263574710000184>
- Papadopoulos, K., Koustriava, E., Koukourikos, P., Kartasidou, L., Barouti, M., Varveris, A., ... Anastasiadis, T. (2017). Comparison of three orientation and mobility aids for individuals with blindness: Verbal description, audio-tactile map and audio-haptic map. *Assistive Technology*, *29*(1), 1–7. <https://doi.org/10.1080/10400435.2016.1171809>
- Parkes, D. (1988). "Nomad": an audio-tactile tool for the acquisition, use and management of spatially distributed information by visually impaired people. *Proceedings of the Second International Symposium on Maps and Graphics for Visually Handicapped People*. London.
- Picinali, L., Afonso, A., Denis, M., & Katz, B. F. G. (2014). Exploration of architectural spaces by blind people using auditory virtual reality for the construction of spatial knowledge. *Journal of Human Computer Studies*, *72*, 393–407. <https://doi.org/10.1016/j.ijhcs.2013.12.008>

- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257–286. <https://doi.org/10.1109/5.18626>
- Rekabdar, B., Nicolescu, M., Kelley, R., & Nicolescu, M. (2014). Unsupervised learning of spatio-temporal patterns using spike timing dependent plasticity. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8598 LNAI, 254–257. [https://doi.org/10.1007/978-3-319-09274-4\\_28](https://doi.org/10.1007/978-3-319-09274-4_28)
- Rekabdar, B., Nicolescu, M. M., Nicolescu, M. M., & Louis, S. (2017). Using patterns of firing neurons in spiking neural networks for learning and early recognition of spatio-temporal patterns. *Neural Computing and Applications*, 28(5), 881–897. <https://doi.org/10.1007/s00521-016-2283-y>
- Rekabdar, B., Nicolescu, M., Nicolescu, M., Saffar, M. T., & Kelley, R. (2016). A Scale and Translation Invariant Approach for Early Classification of Spatio-Temporal Patterns Using Spiking Neural Networks. *Neural Processing Letters*, 43(2), 327–343. <https://doi.org/10.1007/s11063-015-9436-3>
- Rothwell, J. (2013). *The Hidden STEM Economy*. Retrieved from [https://meaganpollock.com/wp-content/uploads/2013/11/brookings\\_hidden\\_stem\\_economy.pdf](https://meaganpollock.com/wp-content/uploads/2013/11/brookings_hidden_stem_economy.pdf)
- Rottensteiner, F., Trinder, J., Clode, S., Kubik, K., & Lovell, B. (2004). Building detection by dempster-shafer fusion of LIDAR data and multispectral aerial imagery. *Proceedings - International Conference on Pattern Recognition*, 2, 339–342. <https://doi.org/10.1109/ICPR.2004.1334203>
- Sheppard, L., & Aldrich, F. K. (2001). Tactile graphics in school education: perspectives from teachers. *British Journal of Visual Impairment*, 19(3), 93–97. <https://doi.org/10.1177/026461960101900303>
- Sjöström, C., Danielsson, H., Magnusson, C., & Rasmus-Gröhn, K. (2003). Phantom-based haptic line graphics for blind persons. *Visual Impairment Research*, 5(1), 13–32. <https://doi.org/10.1076/vimr.5.1.13.15972>
- Smith, L. B., & Sera, M. D. (1992). A developmental analysis of the polar structure of dimensions. *Cognitive Psychology*, 24(1), 99–142. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1537233>

- Stangl, A., Kim, J., & Yeh, T. (2014). 3D printed tactile picture books for children with visual impairments. *Proceedings of the 2014 Conference on Interaction Design and Children - IDC '14*, 321–324. <https://doi.org/10.1145/2593968.2610482>
- Stewart, J., Bauman, S., Escobar, M., Hilden, J., Bihani, K., & Newman, M. W. (2008). *Accessible Contextual Information for Urban Orientation*. Retrieved from <http://cmusphinx.sourceforge.net/html/cmusphinx.php>
- Suzuki, R., Stangl, A., Gross, M. D., & Yeh, T. FluxMarker: Enhancing Tactile Graphics with Dynamic Tactile Markers. , *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '17* § (2017).
- Tuominen, E., Kangassalo, M., Hietala, P., Raisamo, R., & Peltola, K. (2008). Proactive Agents to Assist Multimodal Explorative Learning of Astronomical Phenomena. *Advances in Human-Computer Interaction, 2008*, 1–13. <https://doi.org/10.1155/2008/387076>
- Ungar, S., Blades, M., & Spencer, C. (1995). Mental rotation of a tactile layout by young visually impaired children. In *Perception* (Vol. 24). Retrieved from <http://journals.sagepub.com/doi/pdf/10.1068/p240891>
- Vinter, A., Fernandes, V., Orlandi, O., & Morgan, P. (2012). Exploratory procedures of tactile images in visually impaired and blindfolded sighted children: How they relate to their consequent performance in drawing. *Research in Developmental Disabilities, 33*(6), 1819–1831. <https://doi.org/10.1016/j.ridd.2012.05.001>
- Way, T. P., & Barner, K. E. (1997). Automatic visual to tactile translation - Part II: Evaluation of the TACTile image creation system. *IEEE Transactions on Rehabilitation Engineering, 5*(1), 95–105. <https://doi.org/10.1109/86.559354>
- Weinger, M. B., Reddy, S. B., & Slagle, J. M. (2004). Multiple measures of anesthesia workload during teaching and nonteaching cases. *Anesthesia and Analgesia, 98*(5), 1419–1425, table of contents. <https://doi.org/10.1213/01.ane.0000106838.66901.d2>
- Wies, E. F., O'Modhrain, M. S., Hasser, C. J., Gardner, J. A., & Bulatov, V. L. (2001). *Web-based touch display for accessible science education*. [https://doi.org/10.1007/3-540-44589-7\\_6](https://doi.org/10.1007/3-540-44589-7_6)
- Williams, G. J., Zhang, T., Lo, A., Gonzales, A., Baluch, D. P., & Duerstock, B. S. (2014). 3D Printing Tactile Graphics for the Blind: Application to Histology. *Annual Rehabilitation Engineering Society of North America Conference*. Retrieved from <https://pdfs.semanticscholar.org/2e5f/7f5eb72d7beafd6b7eb3e689d8256a5c4d8c.pdf>

- Xu, C., Israr, A., Poupyrev, I., Bau, O., & Harrison, C. (2011). Tactile Display for the Visually Impaired Using TeslaTouch. *Proceedings of the 2011 Annual Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '11*, 317.
- Yu, W., & Brewster, S. (2002). Multimodal virtual reality versus printed medium in visualization for blind people. *Proceedings of the Fifth International ACM Conference on Assistive Technologies - Assets '02*, 57. <https://doi.org/10.1145/638249.638261>
- Zhang, T., Duerstock, B. S., & Wachs, J. P. (2017). Multimodal Perception of Histological Images for Persons Who Are Blind or Visually Impaired. *ACM Transactions on Accessible Computing*, 9(3), 1–27. <https://doi.org/10.1145/3026794>
- Zhang, T., Zhou, T., Duerstock, B. S., & Wachs, J. P. (2018). Image Exploration Procedure Classification with Spike-timing Neural Network for the Blind. *2018 24th International Conference on Pattern Recognition (ICPR)*, 3256–3261. <https://doi.org/10.1109/ICPR.2018.8545312>
- Zhou, T., & Wachs, J. P. (2018). Early Turn-taking Prediction with Spiking Neural Networks for Human Robot Collaboration. *2018 IEEE International Conference on Robotics and Automation (ICRA)*, (in press). IEEE.

# VITA

Ting Zhang

School of Industrial Engineering, Purdue University

## **Education**

B.S., Software Engineering, 2012, Zhejiang University, Hangzhou, China.

Topic: A Visual Analysis Approach for Context-Aware Community Detection of Mobile Social Networks

M.S., Industrial Engineering, 2014, Purdue University, West Lafayette, IN US.

Topic: A Multimodal Image Perception System for Visually Impaired People

PhD, Industrial Engineering, 2020, Purdue University, West Lafayette, IN US

Topic: Multimodal Digital Image Exploration with Synchronous Intelligent Assistance for the Blind

## **Research Interests**

Human-Machine Interaction, Machine Learning, Assistive Technologies

## PUBLICATIONS

### Journals

1. Zhang, T., Duerstock B.S., Wachs, J.P. (Accepted 2019). Classification of Image Exploration Procedures with Spike-timing Neural Network for the Blind. *IEEE Transactions on Neural Systems and Rehabilitation Engineering (TNSRE)*.
2. Zhang, T., Duerstock B.S., Wachs, J.P. (2017). Multimodal Perception of Histological Images for Persons Who Are Blind or Visually Impaired. *ACM Transactions on Accessible Computing (TACCESS)*, 9(3), 1-27.
3. Zhang, T., Li, Y.T., Wachs, J.P. (2016). The Effect of Embodied Interaction in Visual-Spatial Navigation. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 7(1), 3.
4. Jiang H., Zhang, T., Wachs, J.P., Duerstock B.S. (2016). Enhanced control of a wheelchair-mounted robotic manipulator using 3-D vision and multimodal interaction. *Computer Vision and Image Understanding (CVIU)*, 149, 21-31.

### Conferences

1. Zhang T., Wachs, J.P., Duerstock B.S. (2019) “Understanding goals of image exploration procedures by blind users.”, *RehabWeek 2019*, Toronto, Canada, June 24-28, 2019.
2. Palaniappan S. M., Zhang T., Duerstock B.S. (2019) “Identifying Comfort Areas in 3D Space for Persons with Upper Extremity Mobility Impairments Using Virtual Reality.”, *The 21st International ACM SIGACCESS Conference on Computers and Accessibility (Assets 2019)*, Pittsburgh, PA, October 28-30, 2019.
3. Zhang, T., Zhou, T., Duerstock B.S., Wachs, J.P. (2018). Image Exploration Procedure Classification with Spike-timing Neural Network for the Blind. *2018 International Conference on Pattern Recognition (ICPR2018)*, Beijing China, Aug 20-24.
4. Zhang, T., Duerstock B.S., Wachs, J.P. (2017). Bimanual Multimodal Image Substitution Perception: A Comparison Study. *RESNA 2017*, New Orleans.
5. Zhang, T., Duerstock B.S., Wachs, J.P. (2015). A Computational Framework for Attention Inference Using a Bayesian Approach. *ISACS 2015*.
6. Zhang, T., Wachs, J.P., Duerstock B.S. (2015). Multimodal Perception of Histological Images for Persons Blind or Visually Impaired. *Demonstration in Haptics 2015*.

7. Zhang, T., Williams, G., Duerstock B.S., Wachs, J.P. (2014). Multimodal approach to image perception of histology for the blind or visually impaired. 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC). San Diego, CA.
8. Jiang H., Zhang, T., Wachs, J.P., Duerstock B.S. (2014). Autonomous performance of multistep activities with a wheelchair mounted robotic manipulator using body dependent positioning. 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems Workshop on Assistive Robotics for Individuals with Disabilities: HRI Issues and Beyond.
9. 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. (2014). Annual Rehabilitation Engineering Society of North America Conference 2014.

### **Patents**

1. Zhang, T., Williams, G., Duerstock B.S., Wachs, J.P. (2016). Multimodal image perception system and method. US 15/284,505.