Bingyu Li

Personal Information

Status:	**MS Student**
Program:	**Computer Science and Engineering**
School:	**Tandon School of Engineering, New York University**
Website:	**https://www.linkedin.com/in/bingyu-li/**
RA Period:	**From 2018-12 to 2019-06**

Biography

I'm a software engineer at Amazon. Before that, I was a research assistant in NYU Multimedia and Visual Computing Lab, advised by Professor Yi Fang. I am broadly interested in 3D Computer Vision, Pattern Recognition and Deep Learning.

Research Project: Robust Object Detection and Recognition

1 Description

Spatial cognition, as gained through the sense of vision, is one of the most important capabilities of human beings. However, for the visually impaired (VI), lack of this perceptual capability poses great challenges in their life. Therefore, we have designed Point-to-Tell-and-Touch, a wearable system with an ergonomic human-machine interface, for assisting the VI with active environmental exploration, with a particular focus on spatial intelligence and navigation to objects of interest in an alien environment. Our key idea is to link visual signals, as decoded synthetically, to the VI's proprioception for more intelligible guidance, in addition to vision-to-audio assistance, i.e., finger pose, as indicated by pointing, is used as "proprioceptive laser pointer" to target an object in that line of sight. The whole system consists of two features, Point-to-Tell and Point-to-Touch, both of which can work independently or cooperatively. The Poin-tto-Tell feature contains a camera with a novel one-stage neural network tailored for blind-centered object detection and recognition, and a headphone telling the VI the semantic label and distance from the pointed object. the Point-to-Touch, the second feature, leverages a vibrating wrist band to create a haptic feedback tool that supplements the initial vectorial guidance provided by the first stage (hand pose being direction and the distance being the extent, offered through audio cues). Both platform features utilize proprioception or joint position sense. Through hand pose, the VI end user knows where he or she is pointing relative to their egocentric coordinate system and we are able to use this foundation to build spa-

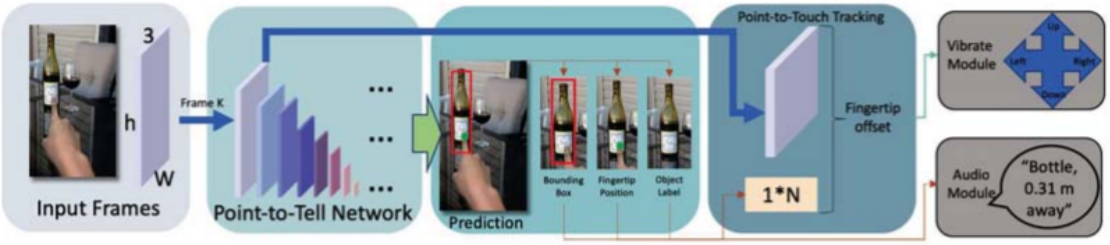


Figure 1: The pipeline of the Point-to-Tell-and-Touch.

tial intelligence. Our successful indoor experiments demonstrate the proposed system to be effective and reliable in helping the VI gain spatial cognition and explore the world in a more intuitive way

2 Method

In this project we have designed Point-to-Tell-and-Touch, a wearable system with an ergonomic human-machine interface, for assisting the VI with active environmental exploration, with a particular focus on spatial intelligence and navigation to objects of interest in an alien environment. As displayed in Figure.1, the frame captured by the left camera will be served as the input to Point-to-Tell net. The result of detection and classification by point-to-Tell net will convey to both the VI via headphone and Point-to-Touch for input. There are two motor control-based feedback loops, as shown in Figure.2, Point-To-Tell and Point-To-Touch. Our platform offers the novel opportunity to close the open loop that vision loss creates by connecting existing sensory channels with computer vision-based spatial intelligence.

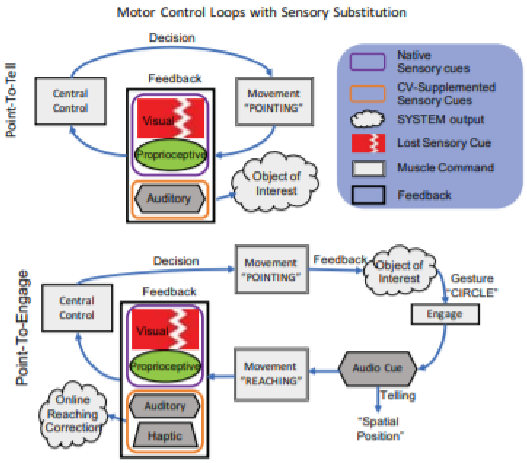
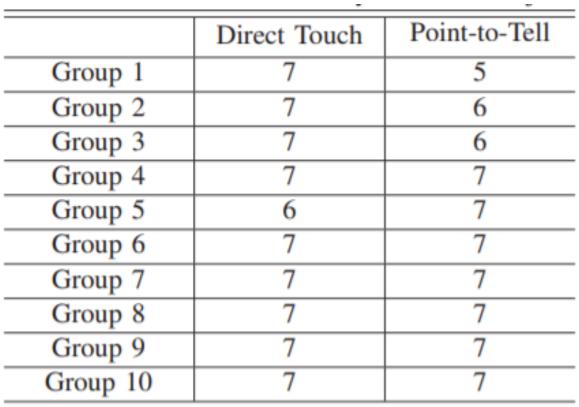


Figure 2: The pipeline of motor control-based feedback loops

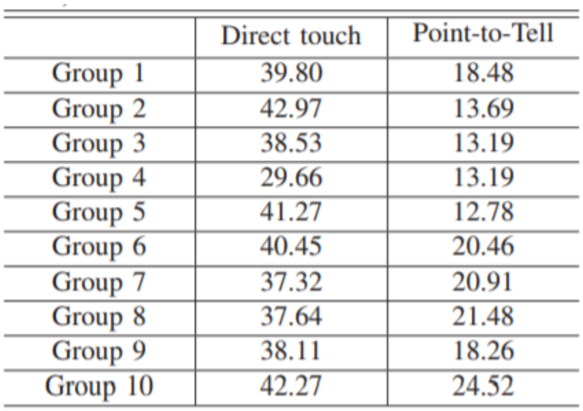
3 Results

In this section, validate our Point-to-Tell function could help visually impaired people efficiently gain spatial information regarding surrounding objects, i.e., in a shorter time with the assistance of Point-to-tell. For the experiment with Point-to-Tell,we supposed a pilot of Point-to-Tell for ADL, i.e the system could be used for shopping assistance. we created a scene with multiple objects in it, a big desk with 7 objects (cellphone, scissors, bottle, laptop, mouse, book, backpack) on it. The distance between them are random, and all the objects are not overlapped. The participants not equipped with the system directly explore the desk in front of them, and the rest using our Point-to-Tell system to identity the objects on the table. In this experiment we will record 1) how many objects can the VI correctly identify during this process and 2) how long does the VI typically need to identify all objects. Among 10 group of blindfolded participants using direct exploration (e.g. their hands), 9 of them correctly identified all objects with one group missed one, meanwhile all of them took a considerably longer time (Table 1). For the blindfolded participants who uses our Point-to-Tell, 1 group missed 2 objects, 2 groups correctly identified 6 objects and the other correctly identified all of them, but only used about half the time (refer to Table 2). This shows that our Point-to-Tell greatly alleviates the cognitive burden when the VI is exploring the scene. As for having more missed out objects when using Point-to-Tell, we would like to note that in this preliminary experiment there is pure object detection running in the backend, which separately detects the objects and the VI's hand through bounding box localization and computes IoU to determine the object that the VI is pointing to, which typically has trouble in detecting small objects. This observation motivates us to design the hand attention guided algorithm pipeline which directly outputs the fingertip and object coordinates that will improve the accuracy. Another possibility is that the previous 5 groups with Point-to-Tell sometimes cannot scan through the whole desk with their fingers, which leads to the missing of some objects in corners. Therefore we suggested to the following 5 groups that they swiped their finger row by row or column by column to cover the whole frame.



	Direct Touch	Point-to-Tell
Group 1	7	5
Group 2	7	6
Group 3	7	6
Group 4	7	7
Group 5	6	7
Group 6	7	7
Group 7	7	7
Group 8	7	7
Group 9	7	7
Group 10	7	7

Table 1. Number of Correctly Identified Objects



	Direct touch	Point-to-Tell
Group 1	39.80	18.48
Group 2	42.97	13.69
Group 3	38.53	13.19
Group 4	29.66	13.19
Group 5	41.27	12.78
Group 6	40.45	20.46
Group 7	37.32	20.91
Group 8	37.64	21.48
Group 9	38.11	18.26
Group 10	42.27	24.52

Table 2. Time to Identify All Objects in the Scene (seconds).