Hanlin Tian

Personal Information

Status:	**MS Student**
Program:	**Computer Science and Engineering**
School:	**Tandon School of Engineering, New York University**
Website:	**https://zachtian.github.io/**
RA Period:	**From 2020-06 to Present**

Biography

I'm a master student at New York University and 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: An Assistive low-vision platform to Enhance Interactive Experience for Visually Impaired: Listen-To-Engage (LTE)

1 Description

Low vision and blindness have given rise to serious inconvenience for the blind and visually impaired (BVI) individuals to conduct daily activities like walking toward a desired object and touch it, which is easy for people with normal vision. This paper presents a novel approach, named Listen-To-Engage (LTE), for the BVI people to help them walk toward and touch the desired object by interactively engaging in the 3D audio. In contrast to traditional methods (e.g. white cane), the LTE system gives full play to the BVI individuals' initiative and assists them in the whole journey. The LTE system will provide the BVI users with 3D audio engagement which indicates the real-world spatial location of the desired object so that to guide them to the destination in a safe and efficient way. Besides, the LTE system is lightweight and portable which only consists of a smartphone and a bone-conduction earphone and does not require any additional equipment. We designed and completed three experiments in the environment generated by Unreal Engine for tracking module, 3D audio module, and engagement respectively to prove the efficiency of our LTE system.

2 Method

In this project, we proposed an interactive approach to guide the BVI users to the desired object through 3D audio engagement, which provides them with interactive auditory experiences. As displayed in Figure.1, the LTE software system consists of two major modules: 1) Object Tracking and Distance Estimation and 2) 3D Audio Engagement for Desired Object. The first input image will be preprocessed with the detection algorithm to find the objects. The images will be sent to the object tracking module to get corresponding point pairs next and then the distance estimation module will use this information to compute the location of the desired object. Finally, The HRTF binaural filter will take both the source 2D audio signal and the estimated location of the desired object to generate 3D spatial audio.

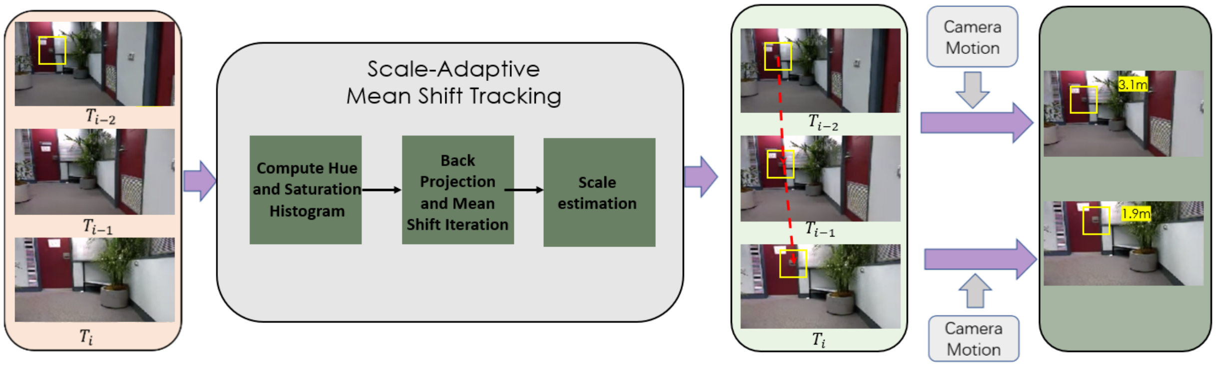

Figure 1: General pipeline to compute of the object distance utilizing IMU-based cameramotion and keypoint tracking.

3 Results

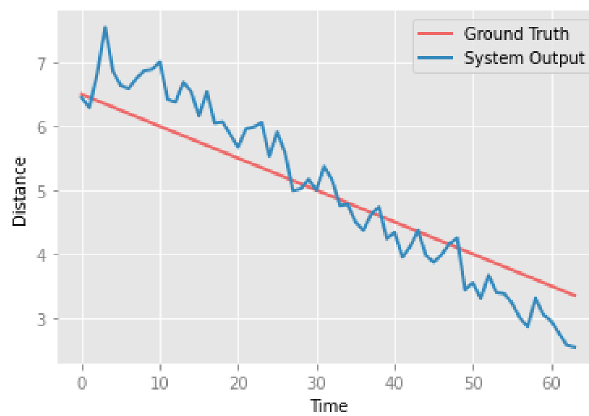

Figure 2: Plot of the calculated distance versus time and the ground truth distance.

In this section, we carry out different experiments for the distance estimation and 3D audio as well as comprehensive tests in the desktop environment. We use the official implementation of the Scale-Adaptive-Mean-shift tracking method as the source program for testing. Considering that the BVI users will follow the guidance of 3D audio in most of the cases but sometimes losing their direction due to external factors so that the desired object disappeared in the capture images for a short time and then appear in the image again, we designed two groups of experiments. As shown in Fig. 3, the algorithm can track the selected object perfectly when the location and orientation of the camera are changing. In the situation where the desired object disappeared in the capture images for a short time, we can see that the tracking algorithm is still running to track other objects, but when the desired object appeared again on the screen, the tracking algorithm will then successfully track the original desired object. In the situation where the desired object always stays on the screen, the tracking algorithm also works well even the desired object is occluded by other objects. Fig.2 illustrate the relationship between estimated distance and time as well as the ground truth distance. We can find that both the estimated distance and ground truth distance

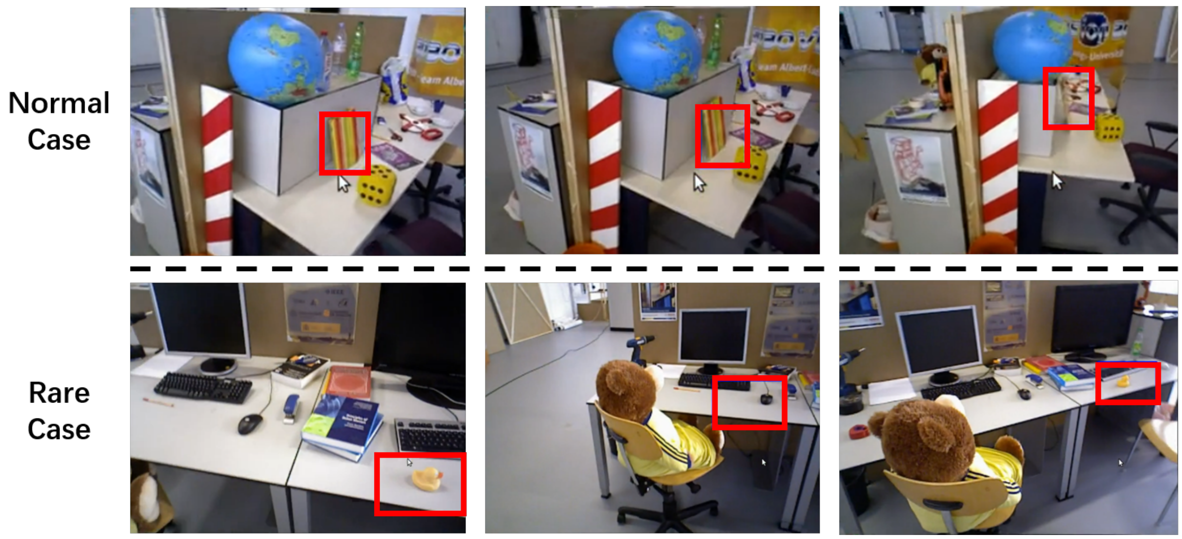

Figure 3: Tracking an object when the camera is moving, the red rectangle is the boundingbox of the selected object.

are decreasing but the estimated distance has a tolerable variance. The overall mean absolute error is 0.4273 and proves the effectiveness of macro engagement.