

Indoor guiding system for the visually impaired

By KHANG, Minsoo & LEE, Lok Yin
Supervised by: Dr. Desmond Y. C. TSOI (CSE)

Motivations

- Over 2.2 billion suffer from visual impairment
- Facing challenges in day to day activities
- Hard to seek and retrieve item indoor

Solution

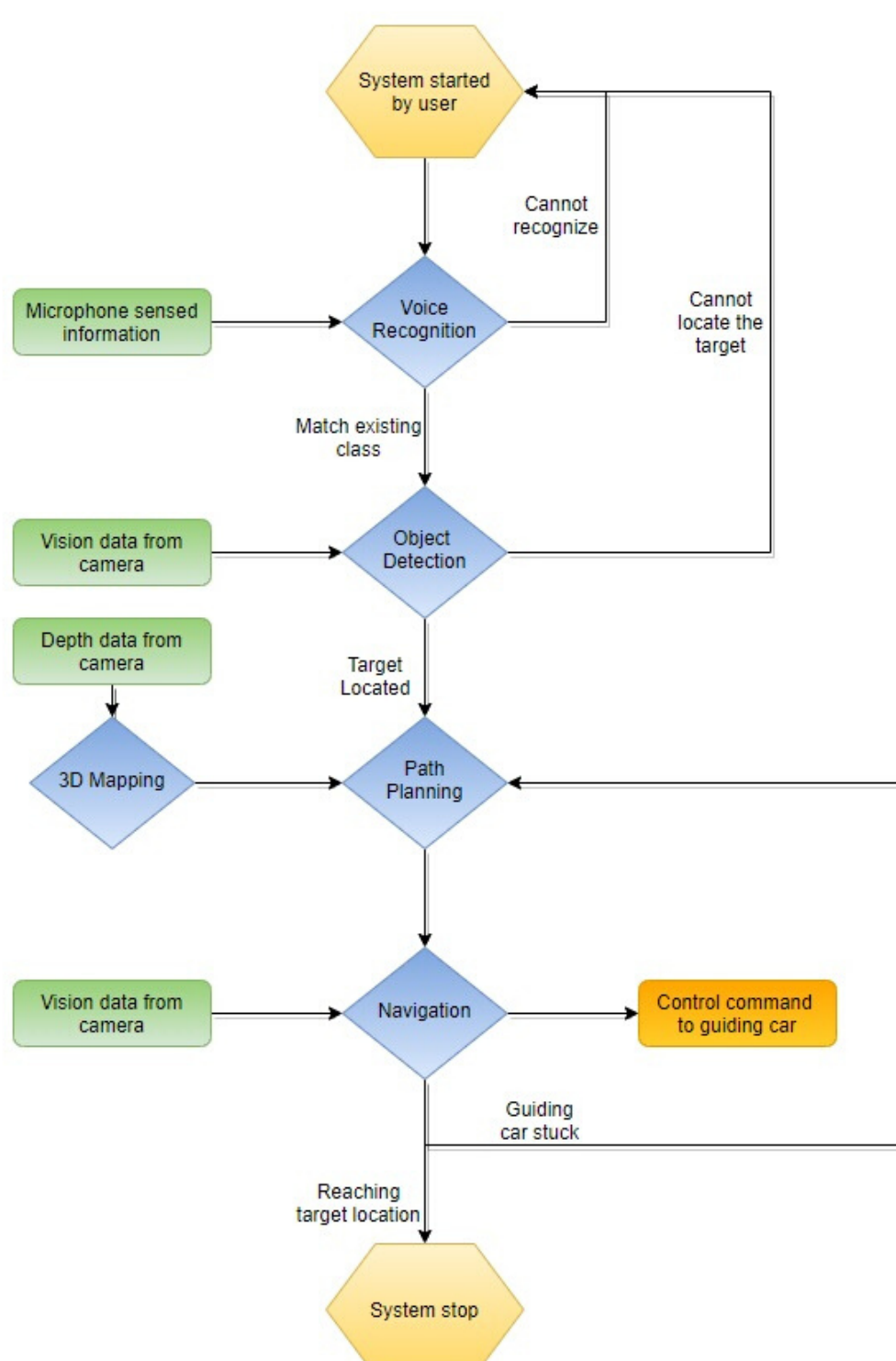
- Object detection and stereo vision to locate items
- Voice interface to receive command from user
- Physical guiding robot generates sound signal to navigate user to target item

Application

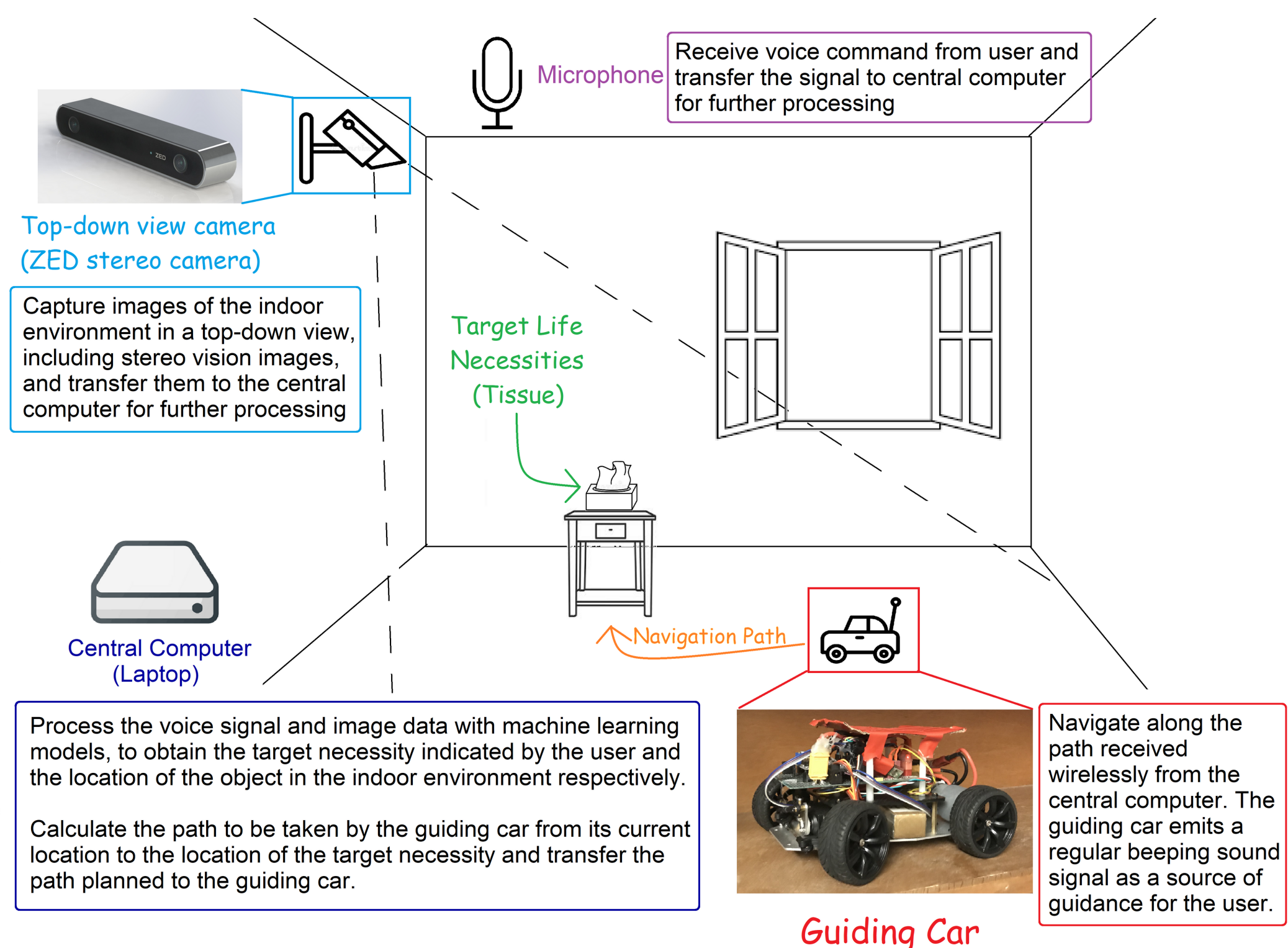
- Customized neural network models to recognize user's voice and detect their necessity items.
- Navigate user to any items indoor with a convenient sound guidance
- Inform user of the item height

System Design

System Flow

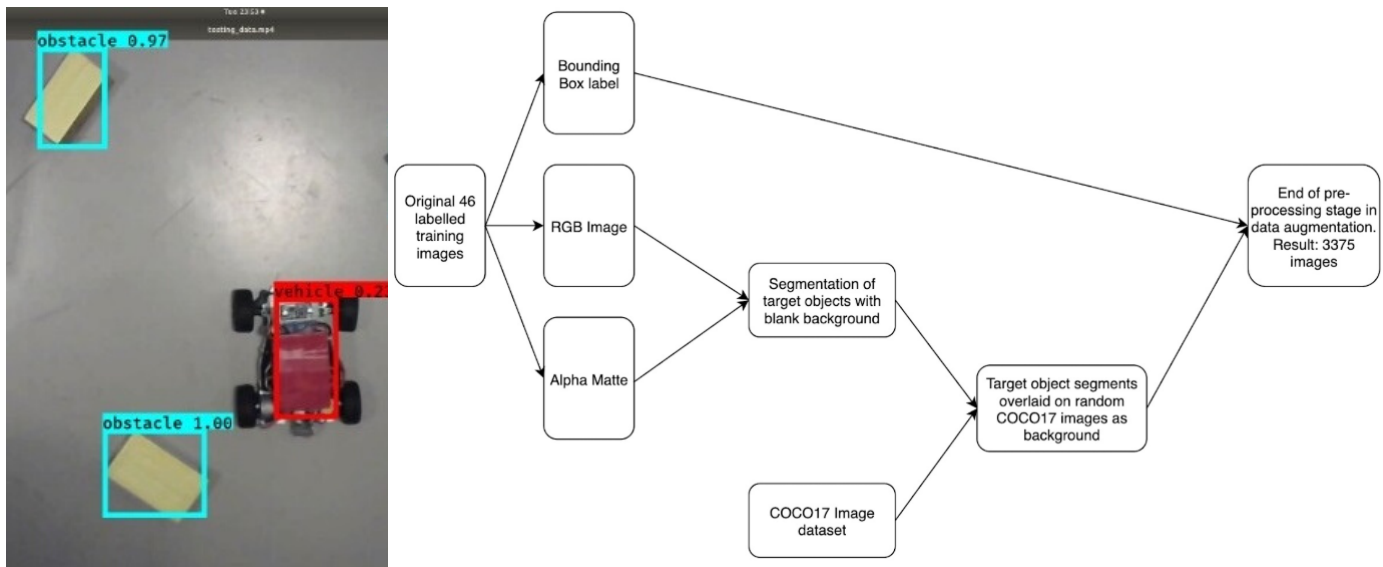


System Setup



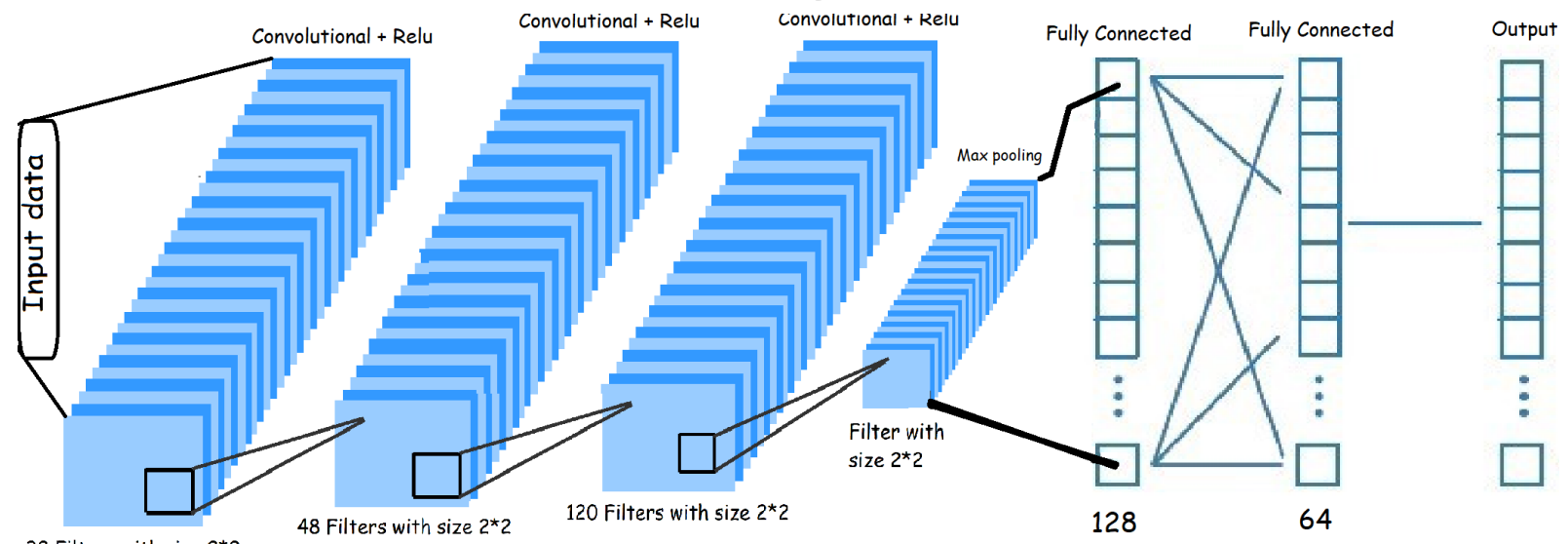
Methodology

Object Detection



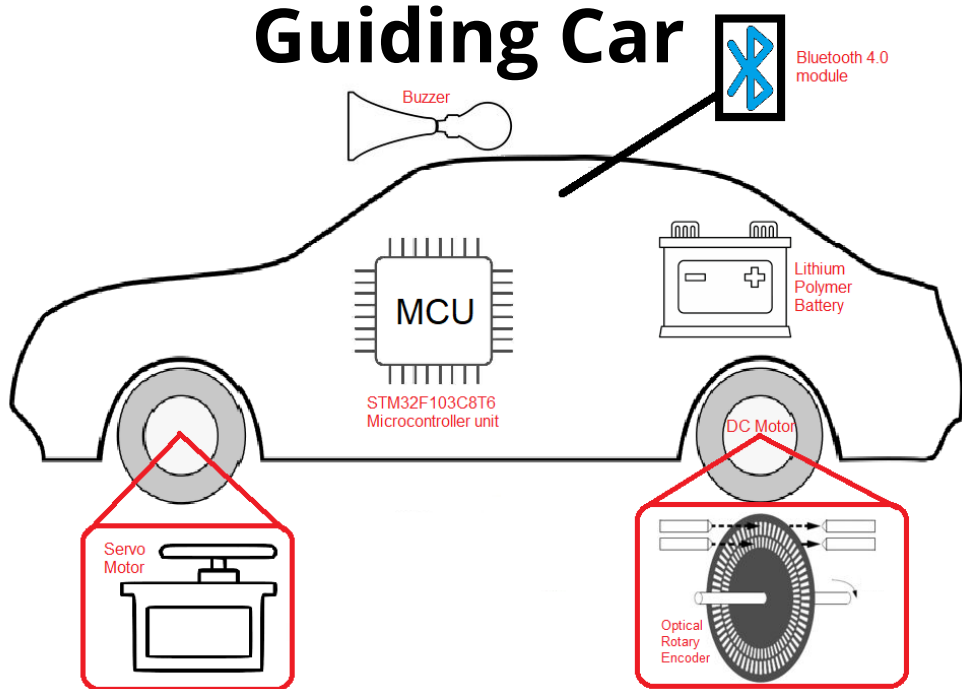
- Implement Retinanet for its high accuracy and lower demand in training images
- Increase training data set size from small user-provided images through data augmentation and segmentation with COCO public image data set

Voice Recognition



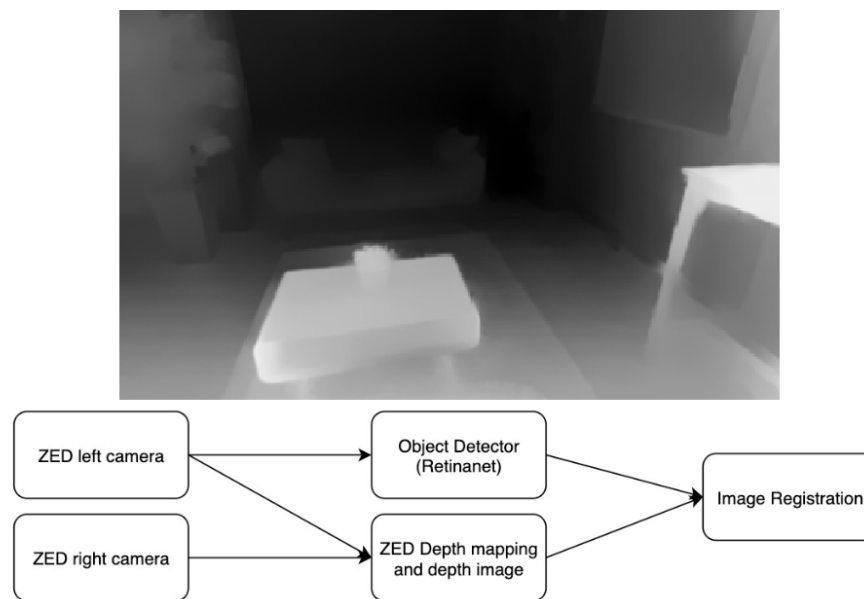
- Develop a pipeline by incorporating an audio convert library LibROSA and convolutional neural network (CNN) by tensorflow keras
- Simplified the training process to achieve a high accuracy regardless of the accent / language (User Customization)

Guiding Car



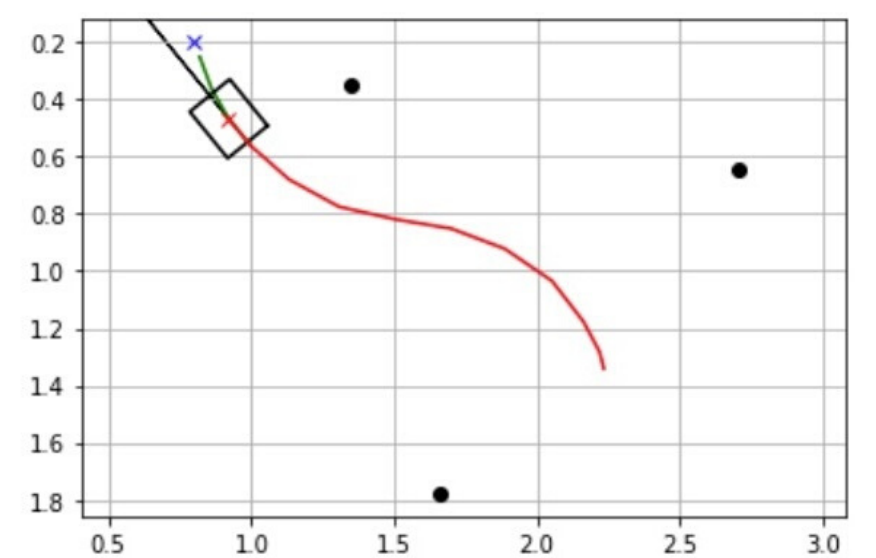
- Design and build a agile mechanical structure with battery system
- Develop a PID positional controller using optical rotary encoder
- Utilize Bluetooth module to form a USART communication with the central computer

Stereo Vision



- Determine the height position of objects after it has been detected by the object detector
- Separate the object and the background to obtain training data automatically (User Customization)

Navigation



- Translate the positional data received from object detection module
- Create a real-time collision avoidance path planning system by Dynamic Window Approach

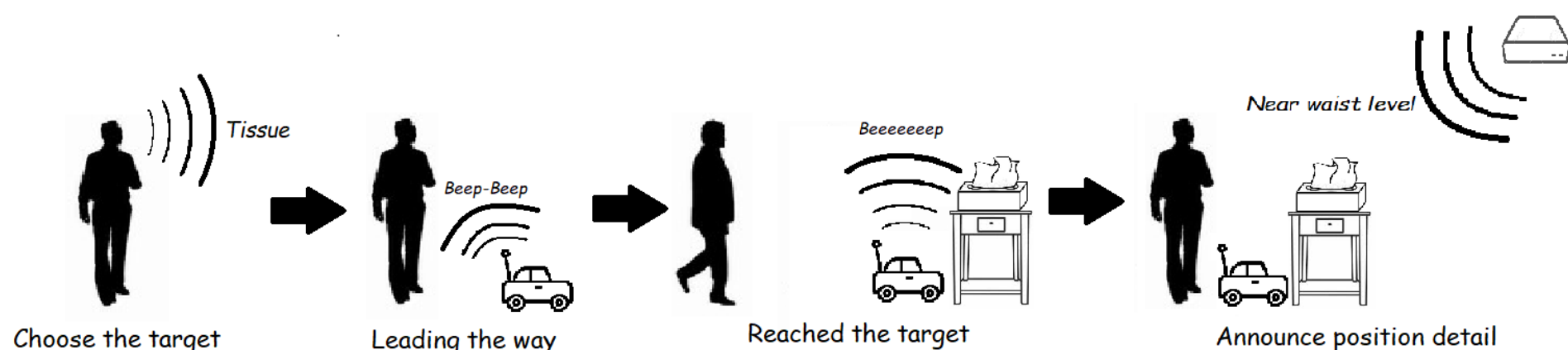
Application

User Customization

Only a small number of images containing target objects provided by the user are needed to generate an effective object detector through our customization pipeline. Through the use of heavy data augmentation and automatic labeling by stereo vision, the user only needs to wait for at most one to two days of training the model prior to its deployment. The final trained model is customized to detect the target objects and fine-tuned for his/her usage.

Voice command can also be customized to different users. By recording and transforming the speech signal into different domains, a small yet effective CNN model can be retrained for each user's voice commands in a short duration of time. Through this customization procedure, the voice recognition model is able to recognize the personalized phrasing, language and ascent of the user, making our system user-friendly and easy to use.

Operation



The implementation of our user-friendly guiding procedure is visualized in the figure above. The user's voice command of a target necessity is recorded by the system's microphone and recognized by our voice recognition model when the volume is above a certain threshold. Subsequently, after the system has located the guiding car and the user, the guiding car will then approach the user followed by emitting the "beep" sound signal. The guiding car then guides the user by navigating itself slowly towards the target necessity while maintaining the "beep" signal. Upon reaching the target destination, it will emit a long "beep" signal to inform the user that the target necessity is in its proximity. Then, the system will inform the user of the approximate height of the target necessity such as "near shoulder level" or "near waist level". The user is now able to retrieve their target necessity without the assistance of another individual.