Problem statement

Build a reliable smart perception system to assist blind and visually impaired people to safely ambulate in a variety of indoor and outdoor environments using an OAK-D sensor.

Introduction

We propose to develop an OAK-D based system that visually impaired people will use to navigate and avoid static and dynamic obstacles. Interviews with visually impaired individuals revealed several specific problems where current approaches (e.g., walking sticks, guide dogs) perform poorly that our system will address. Examples include hanging obstacles, staircases, crosswalks, moving obstacles, reading road signs. The OAK-D camera is ideal for our system because it is a complete package with a sophisticated on-chip AI processing unit, compact, light weight, low power, and reasonably priced.

There are a variety of visual assistance systems for navigation ranging from GPS based voice assisted smartphone apps like Microsoft's Soundscape to camera enabled smart walking stick solutions such as AiServe. While these are useful, they don't capture the visual scene accurately. For example, smartphone apps fail to detect staircases, potholes, hanging obstacles etc. Solutions like AiServe require a walking stick or a guide dog. Microsoft's Seeing AI is an app for general visual assistance. However, it cannot be solely relied upon for navigation as it lacks depth information. Our solution is designed to use OAK-D's compact form factor along with its AI capabilities to overcome these problems.

In this project we solve major navigational challenges faced by visually impaired people. Ranked by priorities based on our interviews and online research, here are cases that we will address:



Moving objects: Spotting and tracking moving objects is a very challenging task especially if the object is moving towards the person. Also, walking behind people with a cane imposes different sets of challenges such as maintaining a gap to avoid collision. This makes navigating in a crowded place much harder. When there is a collision, people's immediate response is negative. However they are fine when a guide dog does the same. Common cases in this category include bicycles on sidewalks, cars driving in and out of driveways. Additionally there is an inherent fear of hurting kids. [Image credits: source, source]


Elevational changes: Spotting elevational changes on a walking surface needs constant effort and are still often missed. Common examples in this category include: Up Curb : entering from road to sidewalk, Down Curb: from sidewalk to road, anomalies on sidewalks like tree roots knocking sidewalk, cracks on the sidewalks. Apart from these examples we will be addressing staircase detection in this category.[Image credits: source, source]


Hanging obstacles are obstacles that are not detected with ground-based tools such as a walking cane or even sometimes by a guide dog. Common examples include tree foliage, bush extending onto sidewalks, cabinet doors left unclosed (indoors). Hanging obstacles are a very common scenario especially in greener states. Failure to spot these obstacles usually results in collision of the upper body such as the face or chest with the obstacle. While few Guide dogs are trained to recognize overhead obstacles, their indication methods are not clear to users and still collide with the obstacles. However the person is unclear on why they stop colliding with obstruction anyway. [Image credits: source, source]

Crosswalk and stop sign detection: Even though crosswalks or stop signs can be detected using sidewalk ramps, detection still relies on car sounds for confirmation. Not all crosswalks are equipped with voice assistance devices, they are commonly found in cities, downtown areas or near government buildings. Also, there are at least 2 kinds of crosswalks: 1) Alternate black and white patterns (also called Zebra crossing); 2). Two long parallel white lines extending from one end to the other.. A common issue with crosswalks is direction alignment to reach the right spot on the other end. Bumps along white paint edges are usually used for alignment. However, these bumps are not always clear. Our solution will provide assistance for accurate alignment as well as indicate presence of nearby automobiles. [Image credits: source]

Road signs detection: Reading traffic sign warnings such as “sidewalk closed” and street names will assist with better perception of the environment. Closed sidewalks are surprisingly common. In most cases an orange cone or traffic sign board is used to mark the closure. In this category we will identify such cases and notify the user. [Image credits: source]

Localization: It is common to lose track of orientation (not physical location) while doing other activities which leads to localization issues. Compass and GPS based apps are useful in this case. However, the apps readjust to new directions instead notifying about changing in orientation. Also, GPS is not accurate in cities, sometimes they are a block off. We aim to provide a customizable waypoint based solution where the users can mark their own locations such as a gym, grocery store or a friend’s apartment with their own labels. This is a complex AI/robotics problem which we may not complete in the course of the competition's one month timeline. Based on our interviews this would be a good nice-to-have feature.

Approach

In this project we will utilize both image based and depth (point cloud) based functionalities of OAK-D along with standard deep learning techniques such as object detection, semantic segmentation etc. to perform inference for visual assistance. We will perform sophisticated scene understanding in 2D and couple that with depth values to obtain accurate 3D information in estimating objects distance from the user. Knowing distance is valuable for a visually impaired person to make decisions ahead while navigating. OAK-D has a critical role in projecting 2D image inference onto the 3D world.

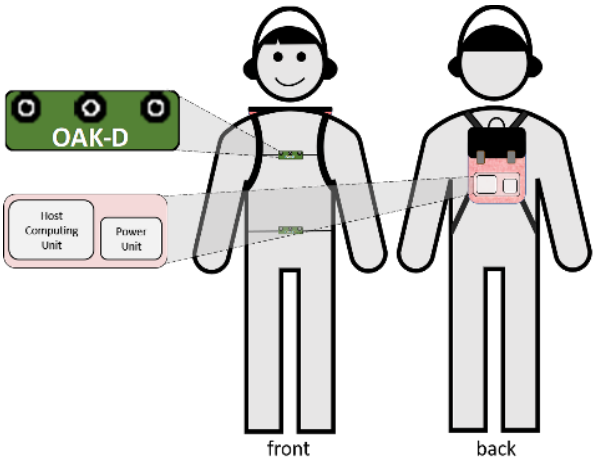
Based on our interviews and research we found that it is more important to have a reliable system that indicates obstacle’s presence accurately while advanced scene understanding can be a valuable additional feature. With that in mind, our perception stack is organized into 3 parts:

1. Primitive perception is a simple yet powerful detector which mainly focuses on detecting if any object/obstacle is within a given distance (more like a binary classifier). This depends heavily on the depth data. This layer provides a minimal scene description. but does so with high accuracy and reliability. An example of the output from this layer would indicate the presence of an obstacle along with the closest cuboid fitted to the obstacle.

2. Advanced perception is a smarter layer of advanced inference that provides a better description of the scene. Examples of outputs from this layer are people, cars, bikes, animals, fences, plants, trees, street lamps, electric poles. Additionally, this layer reports sidewalk width, crosswalk detection, road signs, and staircases. Deep learning techniques such as object detection and semantic segmentation are used in this layer on both images and point clouds.

3. Localizer is another advanced module which performs mapping and localization using images and depth information. Features from advanced perception will be used for this purpose.

Our System



We propose a simple system that doesn't require handheld devices like cane, laser devices or a guide dog. The system would require a small sized backpack to hold a small host computing unit (Raspberry Pi, chrome book or a laptop) and a power unit which is a battery device. OAK-D sensor(s) are attached to bag straps positioned at a particular height facing the front side of the user. The sensor doesn't have to be at the same spot every time as there is no calibration process. The sensor is then connected to the computing unit in the backpack depicted in the figure. The OAK-D sensor along with OAK-D's AI processing unit is triggered by the host unit. The inference data is collected by the host and is updated to the user via a *voice interface* using speaker or earphone. We are aware of bad user experiences in the existing voice assistance devices because of continuous and excess updates to the user. For a better user experience we provide critical updates based on the distance limit set by the user, also non-critical updates are provided upon user's request or a setting adjusted by the user.

Two sensors are likely needed to provide full coverage for a person. Additional sensors could be added--some users might enjoy the opportunity to know what is behind them, in addition to forward sensing-- a sort of "superpower!" Multiple sensors would only require a single computing and battery unit. However, we can complete this project using one sensor and perform testing by changing their placements between 2 positions, the core AI logic is not affected.

Expected Challenges

- Compact and sustaining power unit design
- Overloading OAK-D's on-chip AI computation engine (fix : powerful host or cloud usage)
- Dataset collection and labelling time

Team Capabilities

I am Jagadish Mahendran, senior computer vision / perception engineer with a course work of M.S. in A.I. and extensive experience developing machine learning algorithms for camera-based robotic systems. I have worked for two startups based out of San Francisco : Fellow AI (Fellow Robots) and Chef Robotics.

At Fellow AI, I lead AI projects where I am involved in the design and development of a complete AI pipeline for performing inventory management using deep learning. I have developed deep learning computer vision (CV) models that I deployed and tested on Terabytes of data on a daily basis on cloud platforms (Azure and GCP) as well as edge computing devices such as the Jetson TX2. I am also involved in multiple other AI products at Fellow AI such as SIMPL, a programmable AI automation platform and Canvas, a store space management system using LIDAR point cloud from Velodyne sensor.

At Chef Robotics, I was involved in developing a perception system for cooking robots using image and point cloud data. While at Chef, I developed their perception stack from scratch during which I also worked on performance analysis of Intel's Realsense D435 sensor.

Apart from my industrial experience, I am also involved in AI research activities. I am an inventor on the Fellow Robots' various AI patents. I recently received an excellent reviewer rating on publons for reviewing a COVID-19 related machine learning paper. Here is my Publons profile and Google Scholar page. I have been a part of the panel committee for various AI conferences such as AI4 and Re-Work. I am also mentioned in a few tech blogs such as AltexSoft and TechTarget. I was invited to review a computer vision course in Packt, an online learning platform. I was also invited and participated in Intel OpenVINO toolkit's developers survey in 2018. I was also part of the two member team "Sirius" who won 3rd place in the world wide Animal-AI Olympics competition.

After thoroughly researching the project, I believe that with my strong AI background expertise in computer vision, point cloud systems and sensors, I can complete this project. This is an opportunity for me to make a difference in the community and contribute to the world. Thanks for this fantastic opportunity, looking forward to being in Phase 2.