Smart Navigator for Visually Impaired Person (VIP)

Problem Statement:
A Visually Impaired Person faces a lot of difficulties while moving from one place to another. They usually use red-white canes or guide dogs to move around and sometimes have to rely on other people for directions. Although they are able to get the job done but because of their impairments they are not able to perceive the environment as the normal person does. Also, There is a high reliance on the other people which exposes them to many risks such as traffic and misdirections, etc. We are planning to build a device which will work alongside the traditional modes of navigation to enhance their perception capabilities and make them independent navigators to a certain degree.

Approach:
We propose to give a secondary set of eyes to the blind by leveraging the OAK-D sensor which detects objects/obstacles in the environment and estimates the depth using a stereo or depth sensor. Thinking of this technology in terms of a product, we feel this could be integrated into some sort of a headband/chest-band with the RGB-D sensor mounted in it. The output from the device (like objects in his surroundings, objects in the collision zone, distance to any query object in the scene, direction to the object, information regarding path tracking) will be given to sensory substitution devices which use sensitive areas of tongue, ear or skin for visual perception. We can further enhance the perception capabilities by using deep learning-based models for image captioning and scene understanding. This would enrich the overall experience of the user. Finally, the system could be integrated with GPS to increase the reliability and robustness of outdoor navigation.

Implementation details:

- Depth sensor: The depth camera gives a better perspective of the environment by making the user aware of the surrounding. The Depth module in OAK-D could serve as a primary sensor for navigation in a dynamic environment. Users would be able to get feedback about the presence of an obstacle and this could allow them to make better conscious decisions.
- RGB sensor: The best thing about this sensor module is the native support for the Intel Movidius Myriad X Visual Processing Unit. This can make our device more interactive and user friendly. The data received from the sensors can be post-processed using a neural network to make users more aware of the scene (possibly through concise scene descriptions).
- Sensor fusion: The data from the RGB and depth sensors could be used for distance-based collision warning and to make users aware of the objects around him with the distances to make him spatially aware.

- Sensory substitution device: This device will be the primary interface between the user and the OAK-D sensor. Most of the information received from the OAK-D sensor will be conveyed to users via this device.
- Voice over module: Based on the scene captions generated by the image captioning deep learning module, the data will be fed to a text-to-speech module that could read out the description of the scene on the users request.

Plan of action (Bigger picture):

1. Integrating the neural network for object detection and image captioning.
2. Integrating the Navigation stack with depth sensor stream for navigation.
3. Sensor fusion between depth and RGB camera.
4. Development of Time to collision and warning system modules.
5. Combining steps 3 and step 4 with a sensory substitution device.
6. Integrate the text to speech system with the image captioning module.
7. For reliable outdoor navigation, we can integrate the current sensor suite with GPS

Goals to be achieved in this competition:

1. RGB image processing package
   a. Integration of YOLO for the object detection
   b. Integration of pre-trained deep learning model for image captioning like Show, Attend and Tell: Neural Image Caption Generation with Visual Attention.
   c. The final output of this package will be the annotated image and some text data which describes the scene.
2. Point cloud processing package
   a. First we will process the point cloud data for the ground plane detection using plane segmentation algorithms.
   b. After removing the data associated with the ground plane we will run the clustering algorithm to group the points and for drawing the bounding boxes
   c. The final output of this package will be a segmented point cloud with bounding boxes around the possible object clusters.
3. Sensor fusion package
   a. We will project the processed point cloud data on the image plane and will filter the objects based on the bounding boxes we get from clustering algorithm and object detection algorithm
   b. We will also create a map/dictionary which will hold the objects attributes like object id, object type/name, object distance, and time to collision.
   c. The final output of this package would be a processed depth image with data structure which stores the object specific information for every image frame.

About us:

1. **Suraj Shankar Sapkal (Team member)**
   - Robotics Software engineer (6 River systems)
   - MS in Robotics (GPA 4.0 / 4.0)
     Relevant courses: Deep learning, computer vision, machine learning.
     Udacity Nanodegree: Robotics Software Engineer, Sensor Fusion Engineer
   - I have worked on 3D object tracking, Lidar obstacle detection, Deep learning based Image segmentation, and Object detection related projects.
   - Project practicum on Intel RealSense characterization
     - Derived the confidence matrix and calculated the margin of error for the depth image data based on the experiments conducted for studying the effect of distance, resolution, glare, exposure and motion blur
     - Tuned the environment-specific camera preset parameters, also designed and implemented the spatial and temporal filters to get most of the depth image data without a significant trade-off in the accuracy
   - LinkedIn: https://www.linkedin.com/in/suraj-sapkal/
   - Github: https://github.com/Suraj0712

2. **Bryan Inacio Rathos (Team member)**
   - MS in Robotics Engineering (GPA 3.92/4/0)

   Relevant courses: Computer Vision, Machine Learning, Deep Learning for Robot Perception, Motion Planning.

   Udacity Nanodegree: Self-Driving Car Engineer, Sensor Fusion Engineer
   - Deep Learning Intern, BrainCo Inc.
     - Carried out research in developing deep learning models for EEG signal attention level classification for BrainCo's 2nd generation prototype Focus headband.
     - Worked on EEG signal preprocessing, data cleaning and building models using a combination of Long Short-Term Memory (LSTM) networks and Convolution Neural Networks (CNN) using MXNet framework.
   - I have project experience in 3D object detection, image segmentation, sensor fusion, localization, visual odometry which are relevant to this particular competition.
   - LinkedIn: https://www.linkedin.com/in/bryan-rathos/
   - Github: https://github.com/Bryan-Rathos

3. **James Barabas (Mentor)**
   - Perception Team manager at 6 River Systems