A. James Clark School of Engineering
Department of Electrical & Computer Engineering

408N- NaviGatr

Team NaviGatr:

Using Machine Learning for Foot Travel Assistance

Naitik Gupta, Eliav Hamburger, Micah Wright

Problem Definition

- Worldwide, 2.2 billion people have vision impairments[1]
- Current technologies are limited
  - Ultrasonic Cane: Limited to close proximities, narrow field-view, requires handling[2]
  - Wi-Fi Triangulation: Requires static environment layout, access point cooperation, not publicly viable[2]
  - Camera-based detections: Restricted to motion recognition, not hands free[2]
- Common needs
  - Real-time processing
  - Responsive to dynamic environments
  - Can map new/unknown areas

NaviGatr seeks to provide real-time analysis of a user's surroundings, including the distance, direction, and type of objects nearby and the emotional state of nearby people.

System Architecture

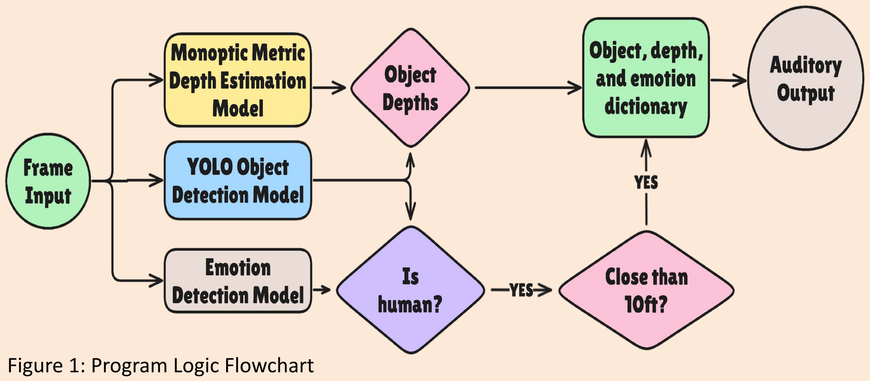

Figure 1: Program Logic Flowchart

The system takes a frame from a video stream and outputs an auditory response to the user which alerts of objects in view, explains the behavior of close persons, and/or provides guidance to navigate in the desired direction of travel.

Design Considerations

- Model response speed
  - To accommodate realistic walking rates, models must be able to produce results in a short, recurring time interval.
- Insightful information gathering
  - Depth sensing can treat all areas as objects once it can isolate the floor area
  - Object locations through labeling is faster than depth sensing and can respond to sudden changes better.
  - Safety of user can be reinforced with behavior tracking. Object model keeps running to alert in case of sudden emergence of an object.
  - Object Detection returns a 2d Array then then refers to the depth array to get the closest point of said object.
- Modeling realistic results
  - Need to simulate practical application of a user wearing smart glasses or similar device
  - Need to be untethered from non-portable computers
  - Machine Learning models need to run on edge devices with limited computational resources.

Prototype

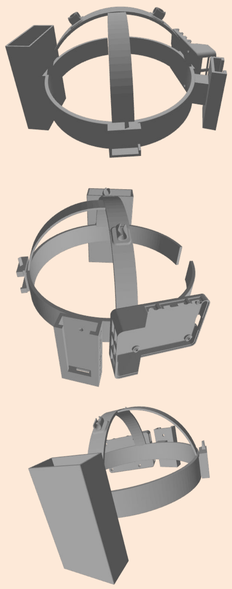

Figure 2: Headband Hardware Front Facing, TPU & RPi Mount, Battery Pack Mount (top to bottom)

Hardware

- Raspberry Pi 5
- Integrated camera
- Coral Tensor Processing Unit (TPU)
- Power bank battery power supply
- 3D Printed head mount

Software

- Apple's Depth Pro for depth sensing[3]
- NanoDet+ for object detection[4]
- EfficentNet for emotion detection[5,6]

Computing

- TPU speedup for closer to real-time edge computation on the Pi.
  - Successfully executed the emotion detection model on the TPU.
- Porting software to Pi
  - NanoDet+ and Depth Pro framework issues
  - SSD MobileNetv2 and YOLOv11n as possibilities[7]
- Cloud computing
  - Depth sensing is a heavy model that may perform better when integrated into a cloud-base environment.

Results

Object Detection

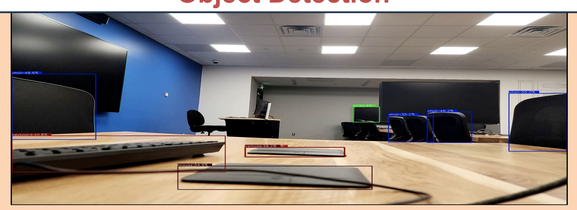

Figure 3: NanoDet+ Bounding Boxes for ENEE408N Classroom

- 5 out of 9 chairs
- Closest proximity objects detected
- Outlet covers mislabeled as remotes

Depth Estimation

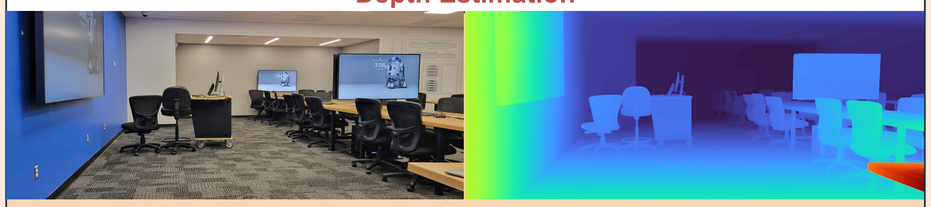

Figure 4: Depth Pro Depth Map of ENEE408N Classroom

- Inference speed ~4 sec (on RTX GPU)
- Depth expressed in color heatmap
- Smooth gradients achieved
- Segments of objects mapped properly (i.e. bottom right desk corner)

Emotion Detection

Figure 5: Sample Emotion Input

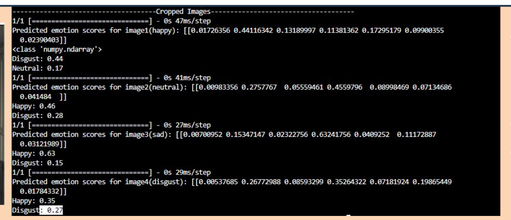

Figure 6: EfficientNet Emotion Model Results

- Our EfficientNet model ~30 fps(RTX3060), a sufficient frame rate to capture a conversation partner's emotional state.
- Takes 4 seconds to run on the Pi CPU.
- Predictions were inaccurate, likely due lack of image cropping. Going forward, we hope to improve results by employing a facial masking model.

References:
[1] World Health Organization, "Blindness and vision impairment," World Health Organization, Aug. 10, 2023. https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment
[2] M. D. Messaoudi, B.-A. J. Menelas, and H. Mcheick, "Review of Navigation Assistive Tools and Technologies for the Visually Impaired," Sensors, vol. 22, no. 20, p. 7888, Oct. 2022, doi: https://doi.org/10.3390/s22207888
[3] A. Bochkovskii et al., "Depth Pro: Sharp Monocular Metric Depth in Less Than a Second," arXiv.org, 2024. https://arxiv.org/abs/2410.02073
[4] RangiLyu. (2021). NanoDet-Plus: Super fast and high accuracy lightweight anchor-free object detection model. [Software]. GitHub. Retrieved from https://github.com/RangiLyu/nanodet
[6] "FER-2013," www.kaggle.com. https://www.kaggle.com/datasets/msambare/fer2013
[7] "TensorFlow | ssd_mobilenet_v2 | Kaggle," Kaggle.com, 2025. https://www.kaggle.com/models/tensorflow/ssd-mobilenet-v2
[5] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," Sep. 2020. Available: https://arxiv.org/pdf/1905.11946