Obstacle Assist for the Visually Impaired

Mihir Dhamankar and Murphy Austin
Carnegie Mellon University, Computer Science Department

Problem Statement

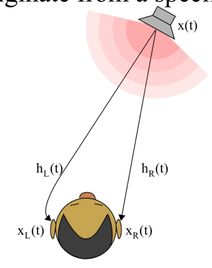
Visually impaired people can effectively use mobility canes to avoid ground-level objects. For overhanging obstacles, we aim to create a similar tool that provides timely warning to users.

Our Obstacle Assist app runs on a Google Glass – a wearable device equipped with a camera, Bluetooth and WiFi connectivity, and a lightweight CPU. The app leverages a nearby cloudlet to process video frames and broadcast actionable audio alerts when an overhanging object is impending.

Existing Work

Spatial Audio using HRTFs (Head Related Transfer Functions) is an existing technology for direction-specific 3D audio. HRTFs utilize the physics of the human ear to create binaural audio that is perceived to originate from a specific point relative the listener.

Virtual Acoustic Displays are the best choice for navigation. Generic HRTFs work well. **White noise** is easier to localize than beeps.

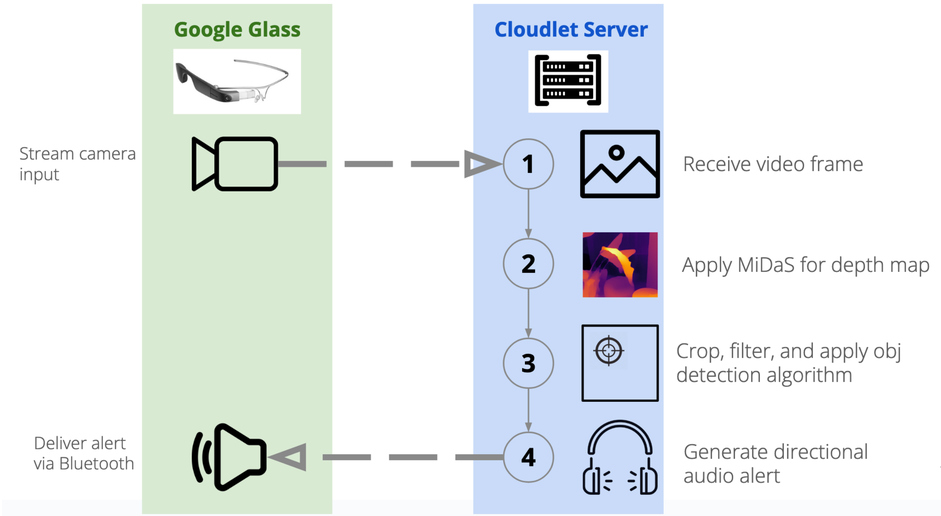


OpenMiDaS is a client/server application for transforming video frames into **relative depth maps.** Frames are streamed from a device to a cloudlet using the **Gabriel** architecture; the cloudlet server uses the **MiDaS ML model** to create a relative depth map and stream back to the deviice.

References:

Ha, K., Chen, Z., Hu, W., Richter, W., Pillai, P., & Satyanarayanan, M. (2014, June). Towards wearable cognitive assistance. In *Proceedings of the 12th annual international conference on Mobile systems, applications, and services* (pp. 68-81).
J. M. Loomis, R. G. Golledge and R. L. Klatzky, "Navigation System for the Blind: Auditory Display Modes and Guidance," in *Presence*, vol. 7, no. 2, pp. 193-203, April 1998, doi: 10.1162/105474698565677.

Solution Design



System processes 2-5 frames per second

Challenges Addressed

Google Glass Programming Environment
- Android 8.1 does not support spatial audio natively (Android 13+)
- On-device audio processing requires NDK/advanced Android knowledge
- Java dependencies have poor documentation

Google Glass resource constraints
- HRTF signal must be convolved with the audio signal, many additions and multiplications in real time
- Google Glass →light client: Stream images to server, decode and play received audio as-is

Binaural audio generation on server
- Sound server and audio interface required– no native audio output on servers
- Emulators installed on host machine; connected to app container via TCP

Relative object detection
- MiDaS approximates relative distance of objects from user, not absolute distance
- Solution: highlight overhanging space through multiple rounds of cropping before and after depth map; filter objects closest to user; procure center point coordinates through blob detection algorithms

Results

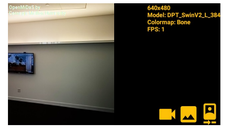
Consistent detection of impending overhanging objects within 5 feet of user

2+ hours of battery life

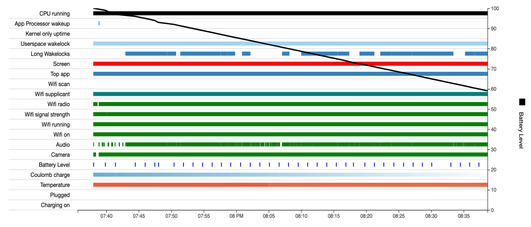
Interactive-time audio alerts with discernible direction

head-height object detected

hip-height object detected



no object detected



CPU Usage By App:

Show 5 entries

Ranking	Name	Uid	User Time / Hr	System Time / Hr
0	CAMERASERVER	1047	33m9s306ms	15m36s444ms
1	edu.cmu.cs.openmidas	10099	10m29s754ms	4m3s423ms
2	AUDIOSERVER	1041	4m30s154ms	7m22s118ms
3	ROOT	0	5s983ms	10m16s408ms

Discussion

- Lower latency audio streaming on client is possible with Oboe
- Processing frames on server is slow (~600ms per frame at worst)
- Better sound localization with absolute depth map and custom HRTFs
  - Absolute depth allows for accurate amplitude control
  - Custom HRTFs limit human localization errors
- Difficult to integrate with existing software used by blind people
- Temporary loss in connectivity may miss an obstacle