ARVI: AR for the Visually Impaired

Final Presentation

Mengshi Feng, James Lin

What is ARVI?

An iOS app for the visually impaired that uses SLAM to place spatialized, user-defined audio cues in a person's surroundings.

Example Use Case

Scanning

Localizing

Background

Literature review

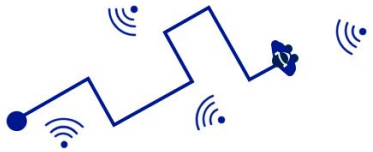
- Read papers, talked to students, and interviewed members from visually impaired communities

LIGHT HOUSE FOR THE BLIND AND VISUALLY IMPAIRED

SKERI
THE SMITH-KETTLEWELL EYE RESEARCH INSTITUTE

Related Works

- Aira
  - Connects user to remote experts through wearable camera
  - Intelligent, but difficult to scale
- Orcam
  - Wearable camera with CV functionality
  - Reads text, recognizes faces, etc.
  - Scalable, but limited in scope
- Soundscape
  - Annotated maps of surroundings through spatial audio
  - GPS is imprecise



Findings

- "Solving" obstacle avoidance is a common trap
  - Existing solutions are well-ingrained, satisfactory, and reliable
- Less solved: learning precise semantic info
- Example: the "**5 meter problem**"
  - GPS is often inaccurate to >5m
  - Difficult to use for precise destinations
    - Doorways, bus stops, store sections, etc.

Methods

SLAM

Requirements

- Persistent environment mapping/tracking
- Robustness > accuracy

Potential Solutions

- ARKit: doesn't support persistent maps natively
- 6D.AI: powerful, but focused on visuals
- *Placenote*: easy-to-use persistence SDK

Spatial Audio

- **Requirements**
  - Binaural Audio: shows depth/direction in sound
  - Varying:
    - Volume
    - Time delays

- **Potential Solutions**
  - *Google Resonance*: spatial sound SDK designed for VR/AR

User Interface

- **Requirements**
  - Blind-friendly
  - Position-invariant input
- **Solution**
  - Gesture recognition (inspired by VoiceOver)
  - Swipes to navigate through available UI elements
  - Tap to hear current element
  - Double-tap to select

Results

Experiment Procedure

- Informal experiment to collect preliminary results
- **Setting**: supermarket aisle, persons *A* and *B*
- **Procedure**:
  1. *A* selects an object and scans its surroundings
     - Drops audio cue on object
  2. *B* localizes to the scan; attempts to find object
  3. Time to find object is recorded

Experiment Results

Blindfolded

Searcher	Object	Time
Mengshi	Coffee Mug	1:40
Mengshi	Protein Powder	1:53
James	Detergent	Failure
James	Cat Litter	Failure

Sighted

Searcher	Object	Time
Mengshi	Baby Wipes	0:20
Mengshi	Peanut Butter	0:25
James	Cat Food	0:29
James	Spaghetti	1:05

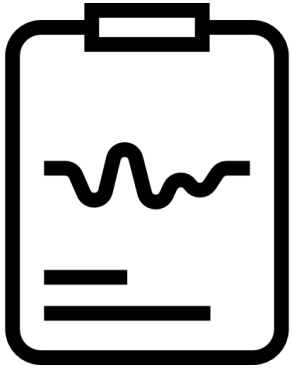
Insights

- Mobile SLAM still fragile/temperamental
  - Can require specific perspectives to localize
  - Hardware differences
- Lack of feedback during localization disorienting
- Spatial audio was very precise
  - < 1ft navigation

Future Work

Next Steps

- **Short Term**
  - User testing + feedback from blind communities
  - UI workflows for sighted users
- **Long Term**
  - Integrating GPS navigation
  - SLAM improvements for robustness



Open Questions

How can you develop a SLAM solution where the goal isn't precision, but robustness and scalability?

What UI interfaces for VR/AR are best suited for the visually impaired?

Thank you!