



# Environmental Text Spotting for the Blind using a Body-worn CPS

Hsueh-Cheng Wang, Rahul Namdev, Chelsea Finn, Peter Yu, and Seth Teller

Robotics, Vision, and Sensor Networks Group

Computer Science and Artificial Intelligence Laboratory (CSAIL), MIT

## Motivation

- Environmental text is important in every-day task, but such information is inaccessible to 285 million blind and visually impaired (BVI) people around the world.
- Fifth Sense Project supported by Andrea Bocelli Foundation



MIT Fifth Sense Project: Providing Key Functions of Vision to the Blind and Visually Impaired.



Italian tenor Andrea Bocelli became blind after a childhood accident.

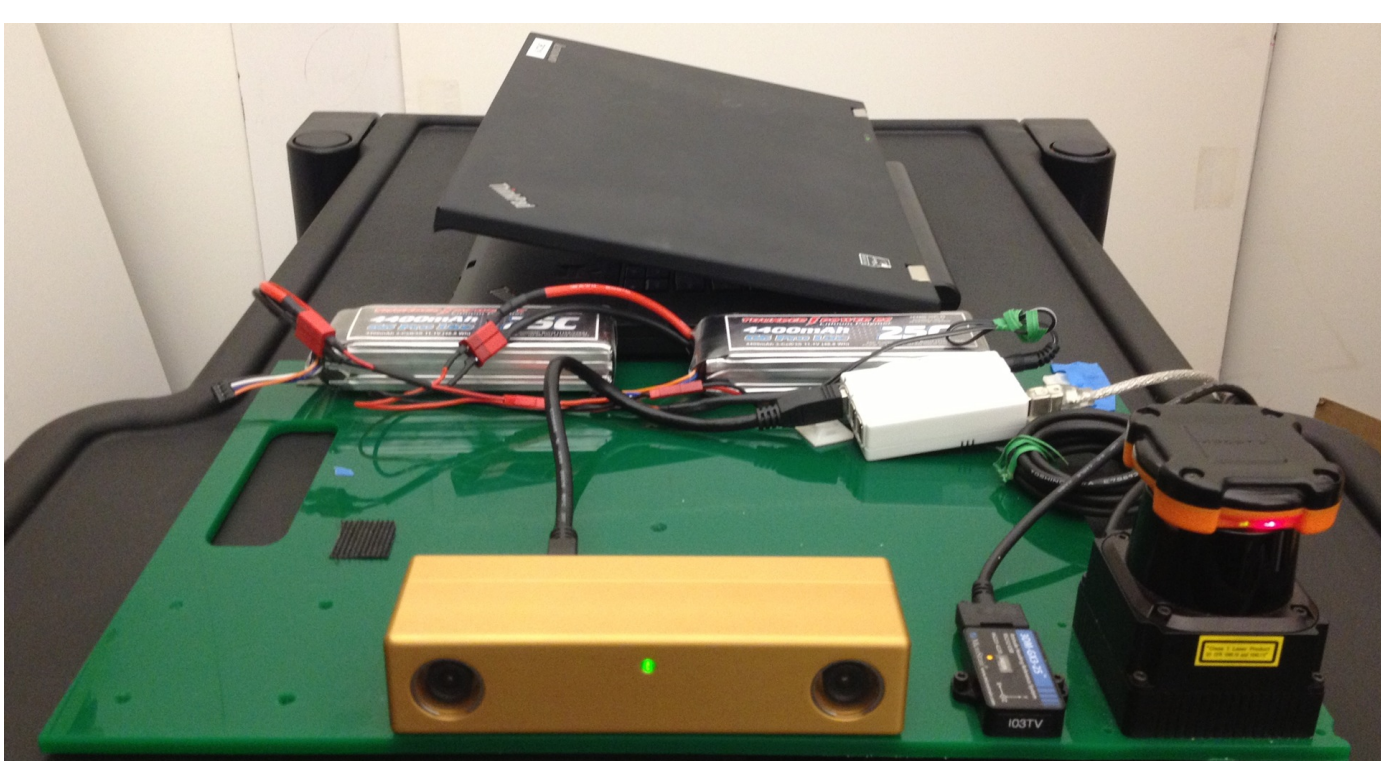
## Challenges

- Unlike scanned documents, scene text only occupies tiny portion of entire field of view (FOV) with high variability
- Decoding is resolution-demanding and computationally intensive
- Similar to classical CPS challenges, a real-time system that allows message passing among computation and physical processes is needed.



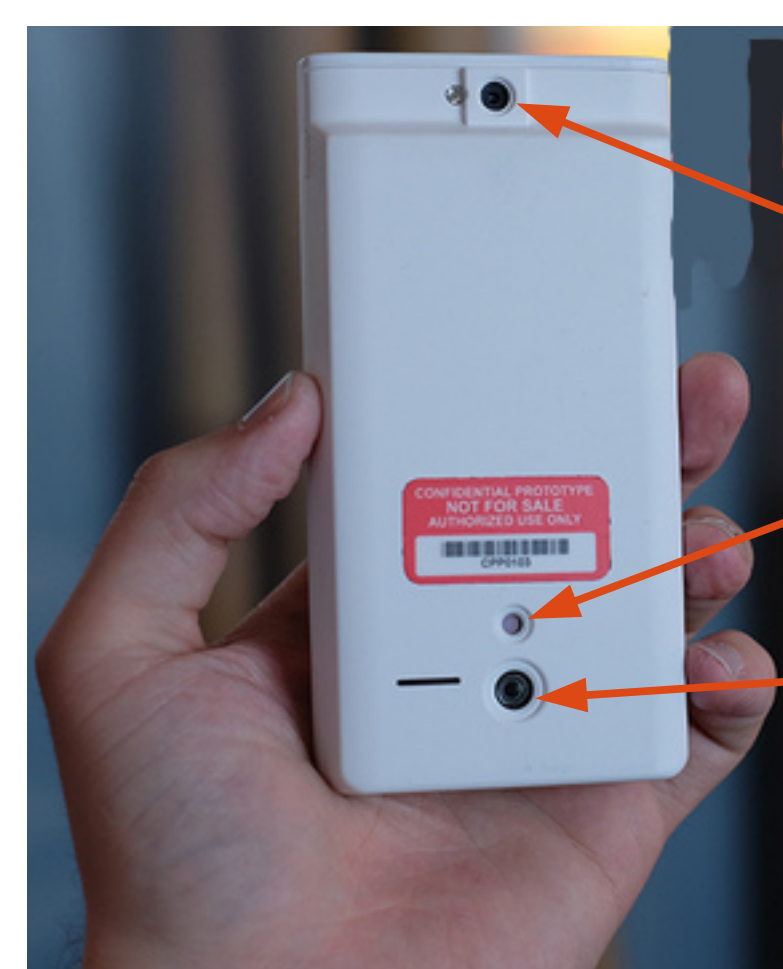
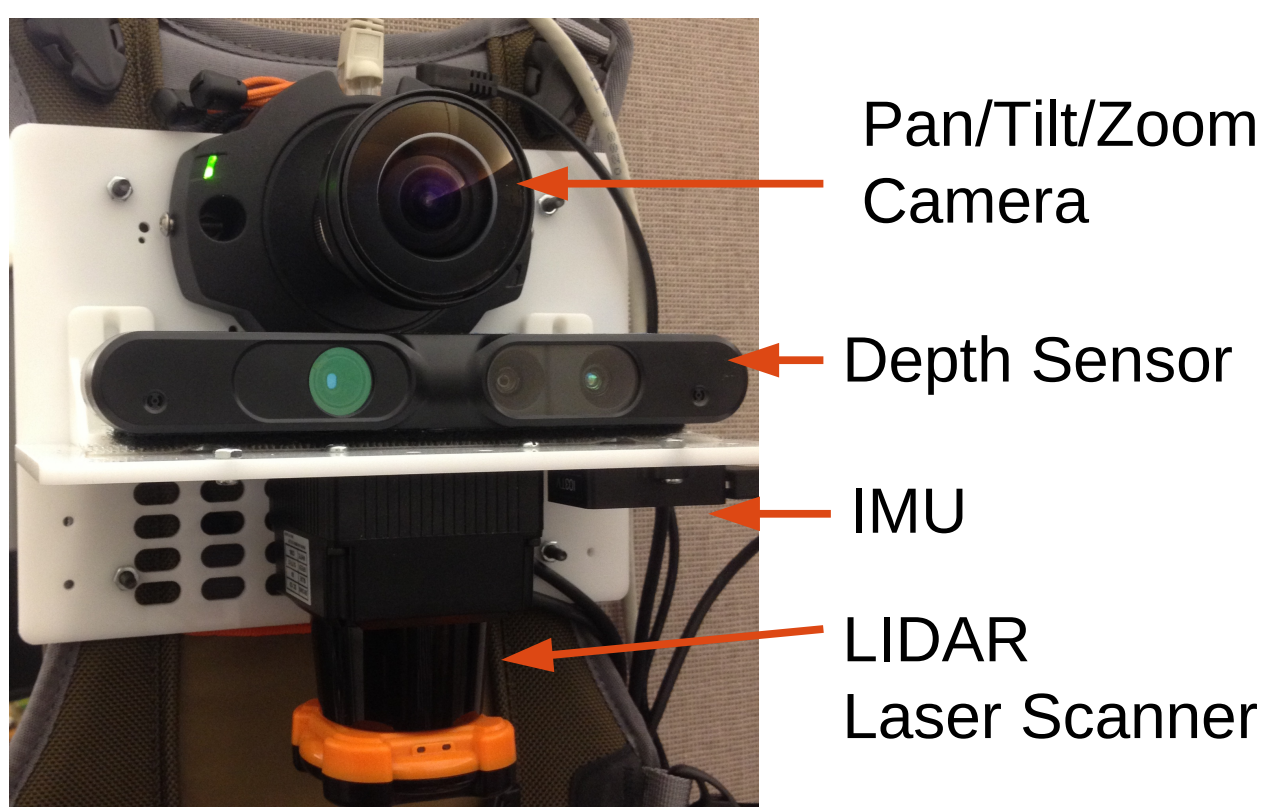
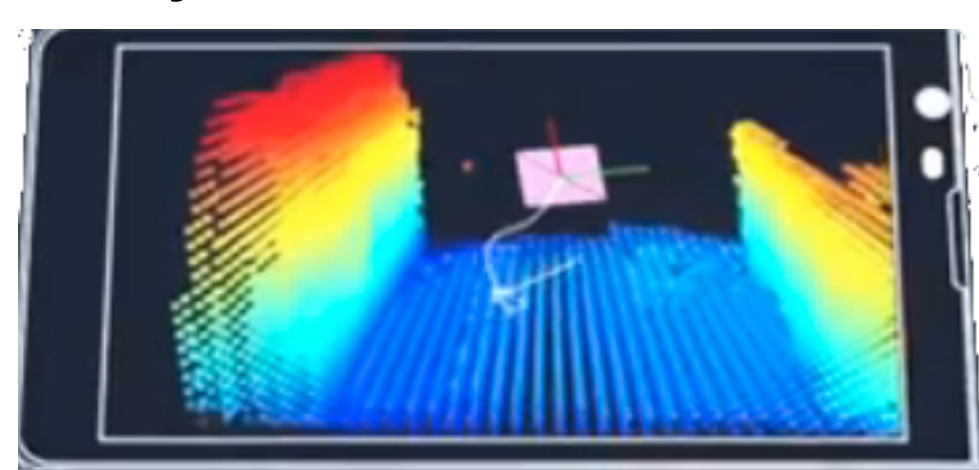
## Body-worn/Mobile CPS

- As a substitute for the eyes to allow communications among sensory devices, algorithms, and BVI users.
- Using frameworks in robotics and sensor networks (LCM and ROS).



June, 2013

Google Tango Project  
March, 2014



Dec., 2013

Pan/Tilt/Zoom Camera

Depth Sensor

IMU

LIDAR Laser Scanner

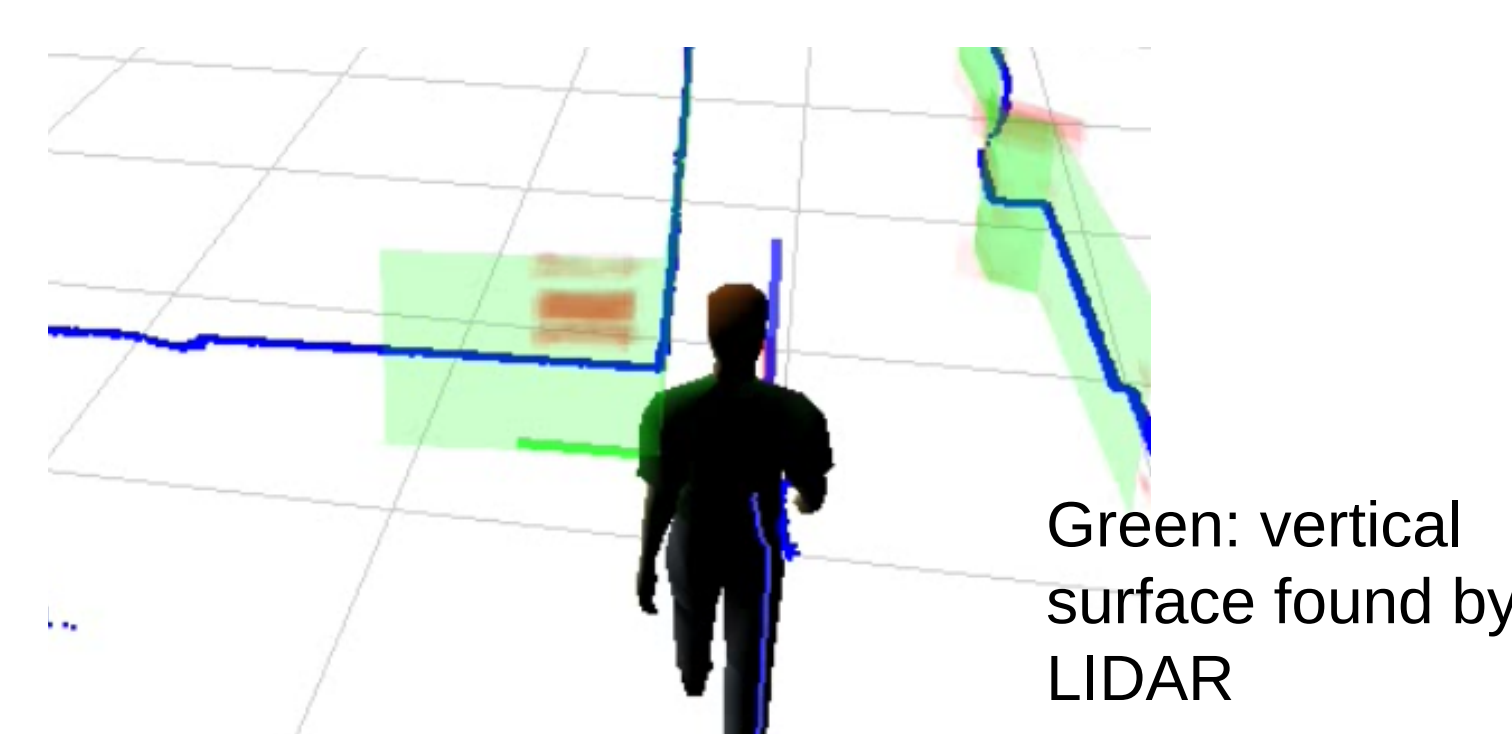
4 MP RGB-IR Camera

Depth Sensor

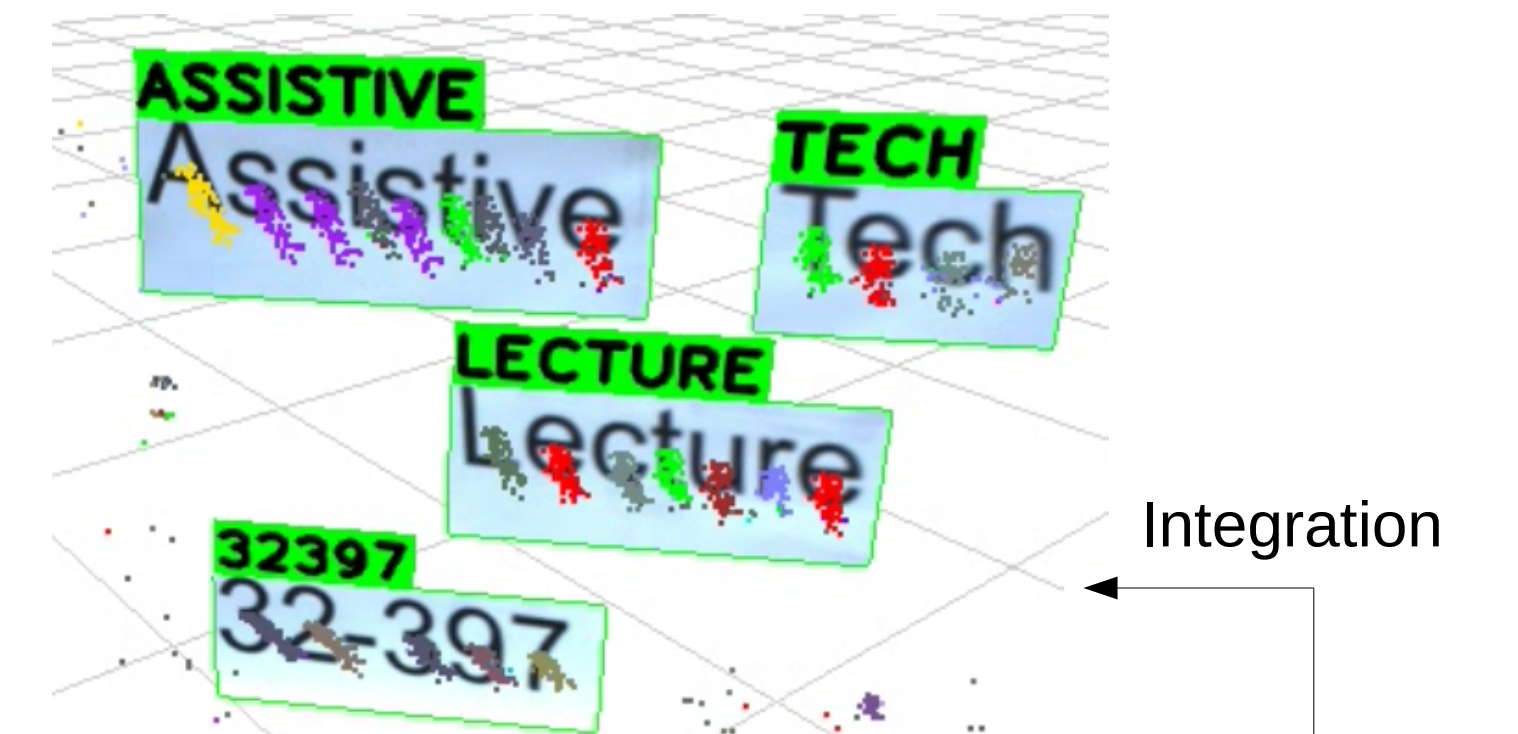
170° Fisheye Motion Tracking Camera

## Text Spotting using SLAM with Feedback Loops

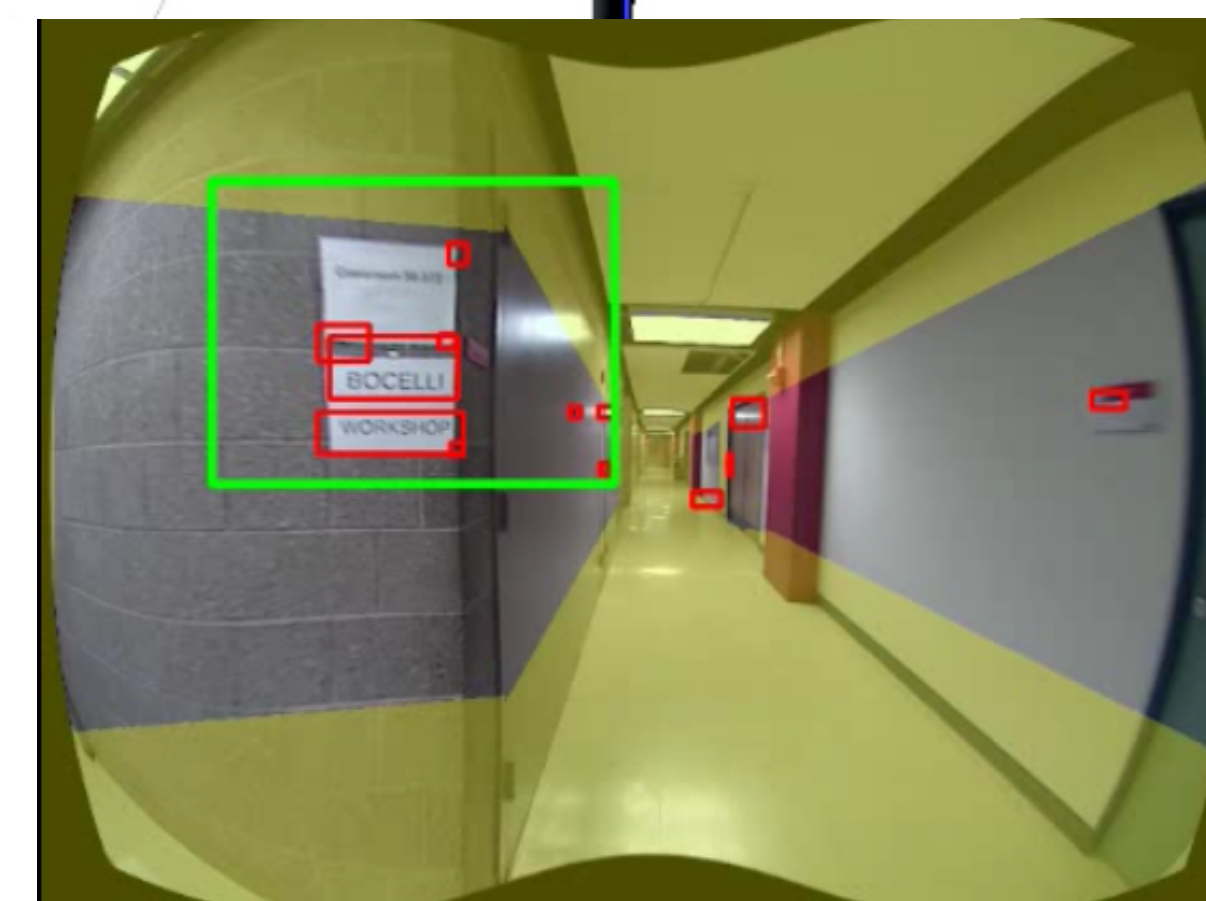
- Incorporate spatial prior on text locations by depth sensors
- Dewarp to remove perspective effects, and Integrate with 3D mapping



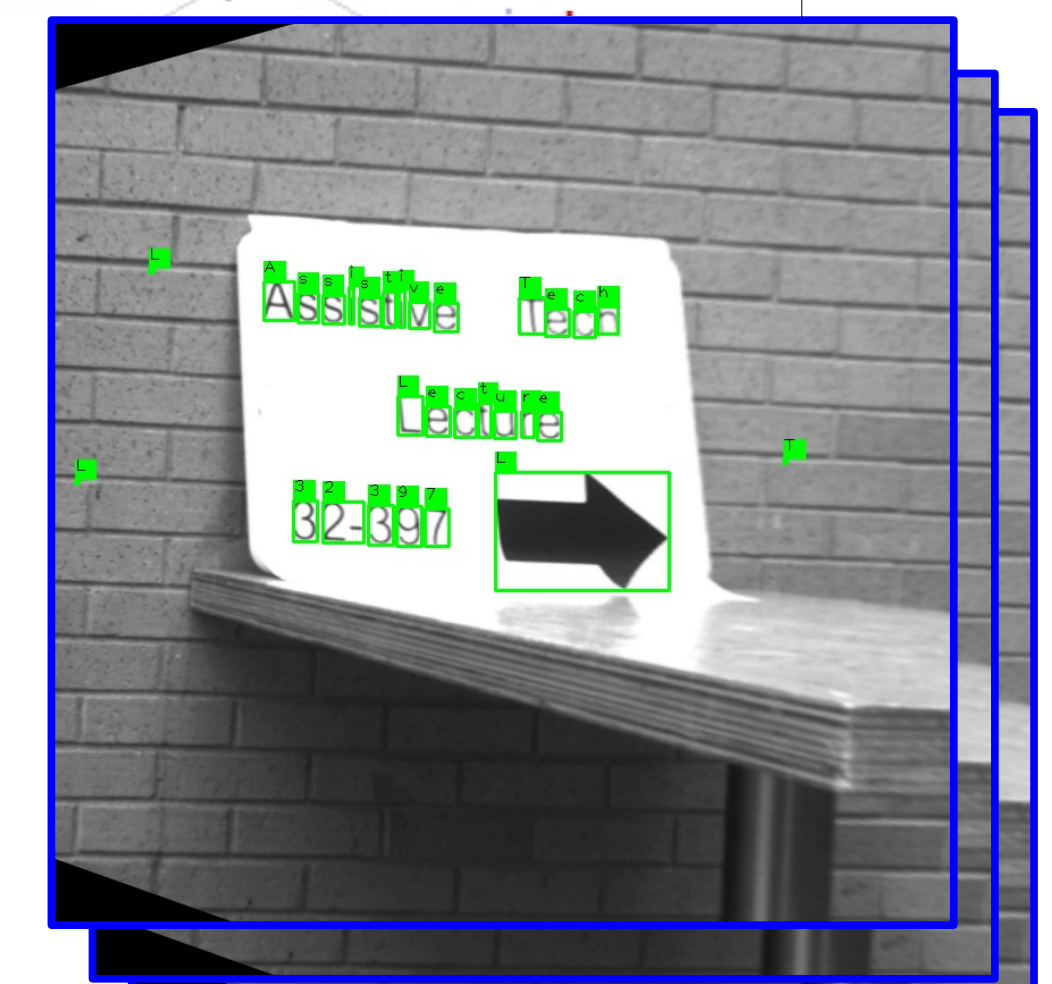
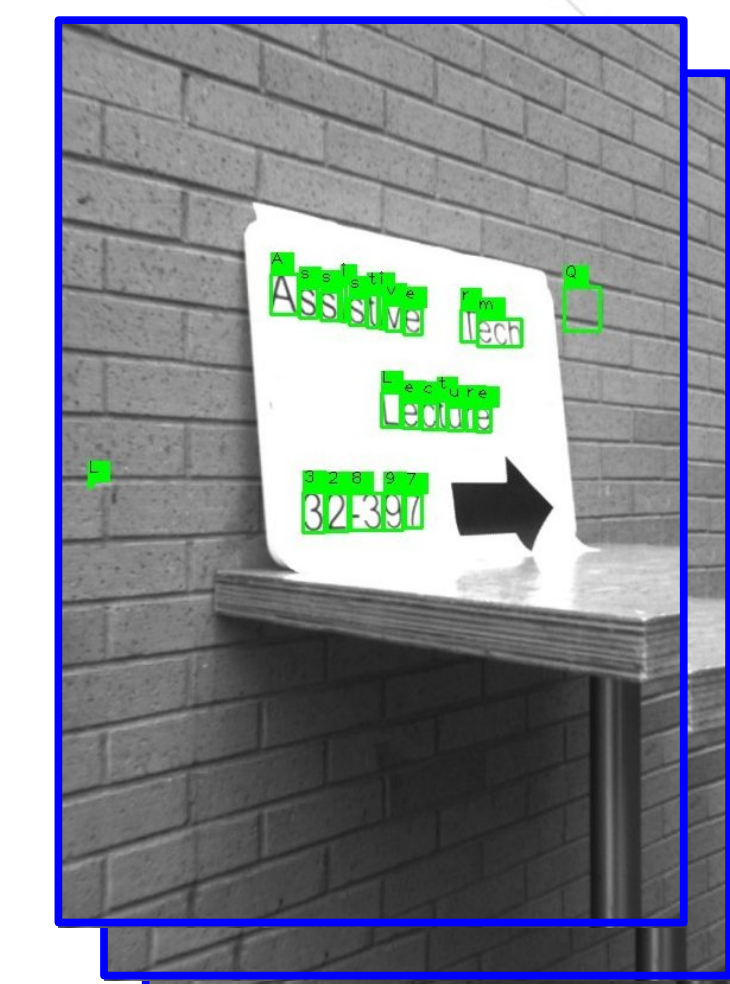
Green: vertical surface found by LIDAR



Integration



Translucent yellow: regions excluded due to prior.



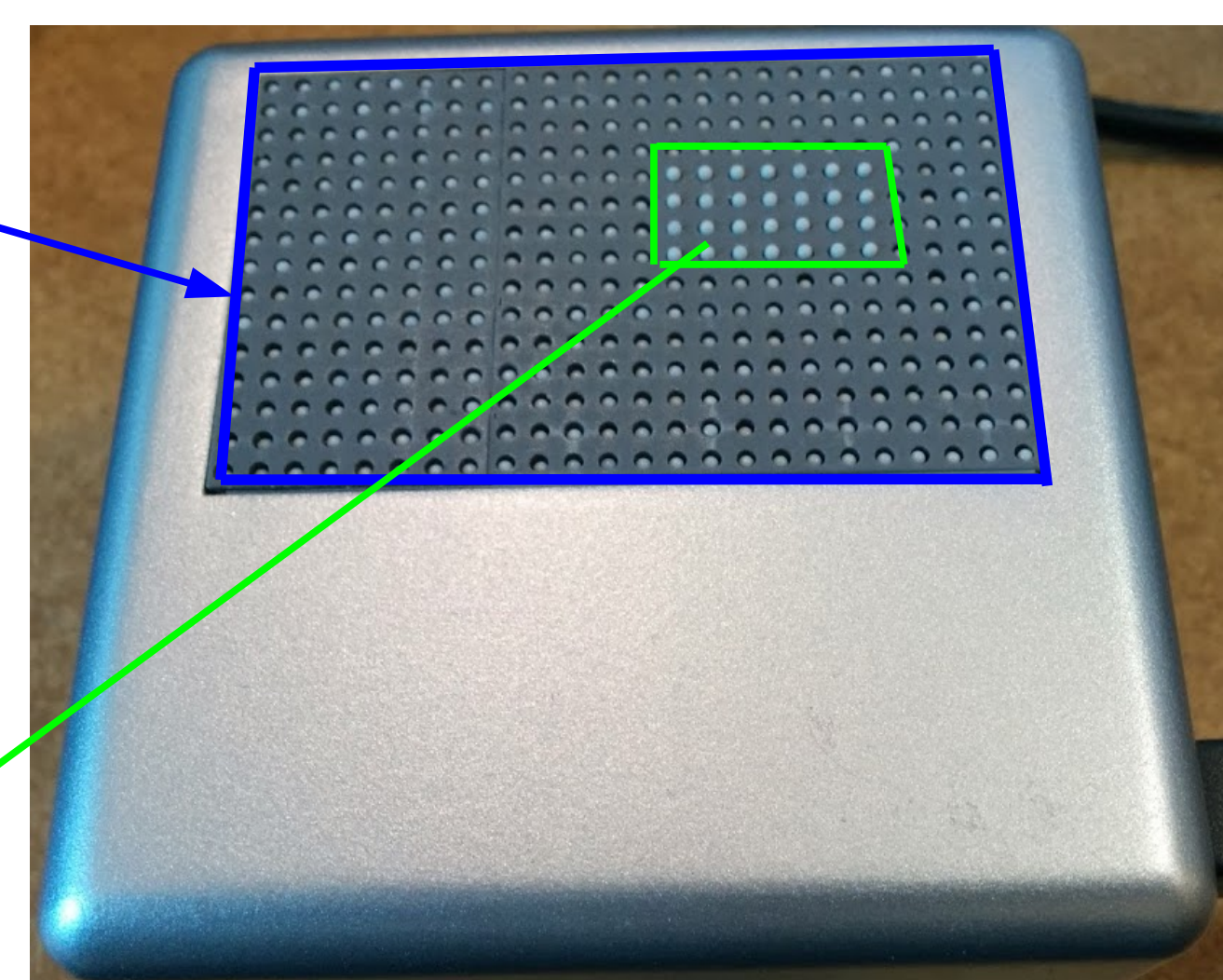
## Human-CPS Interaction

- Using an electronic braille, blind users can not only access where text likely occurs in current field of view, but also control the PTZ cameras to foveate the region of their interests.



Wide FOV

Text detection algorithm displays where text likely occurs



24 x 15 Braille Display

Users press the pins to zoom in



Foveated Imagery

## Potential Impact and Future Work

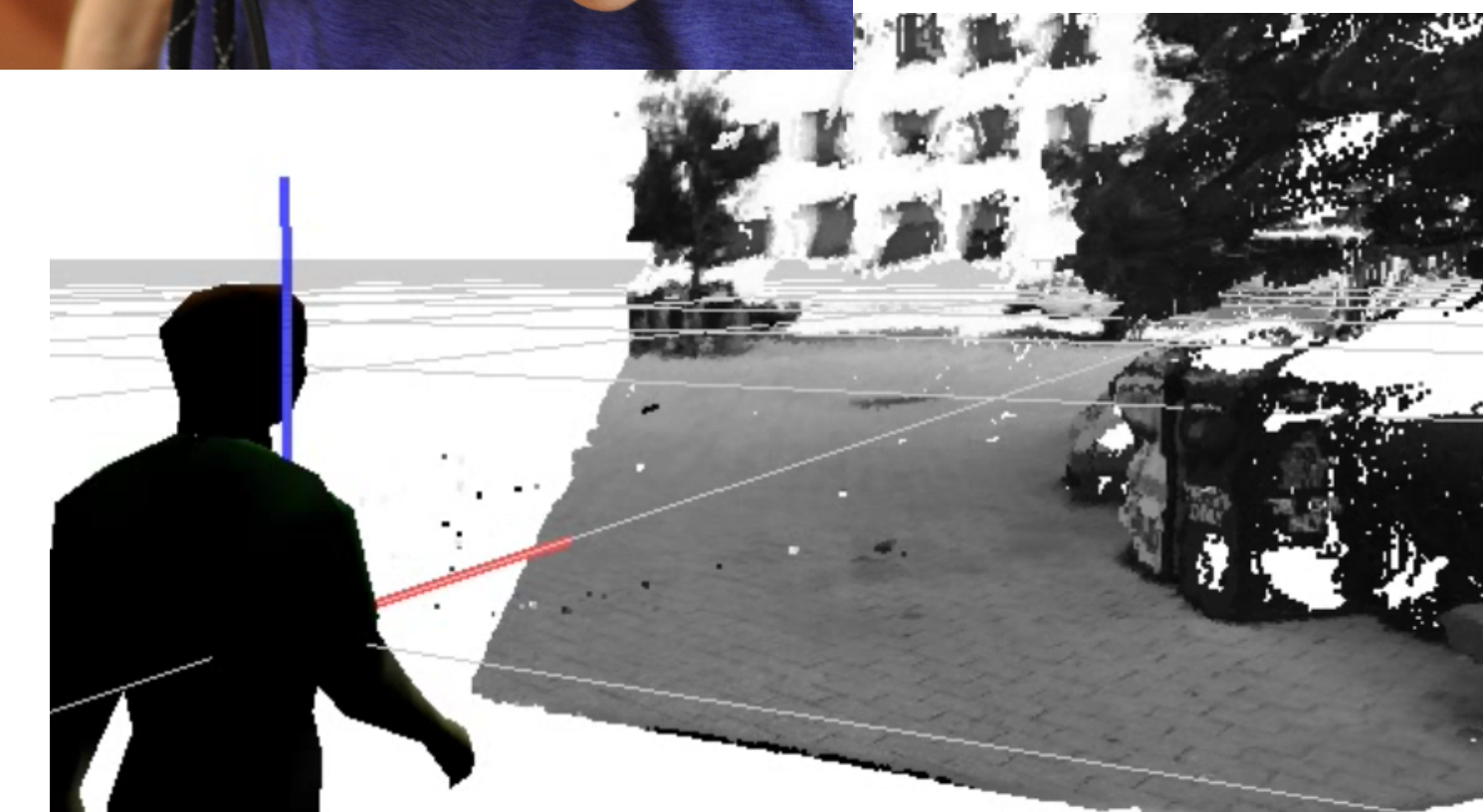
- Our work can lead to many applications, such as health care and augmentation of human capabilities.



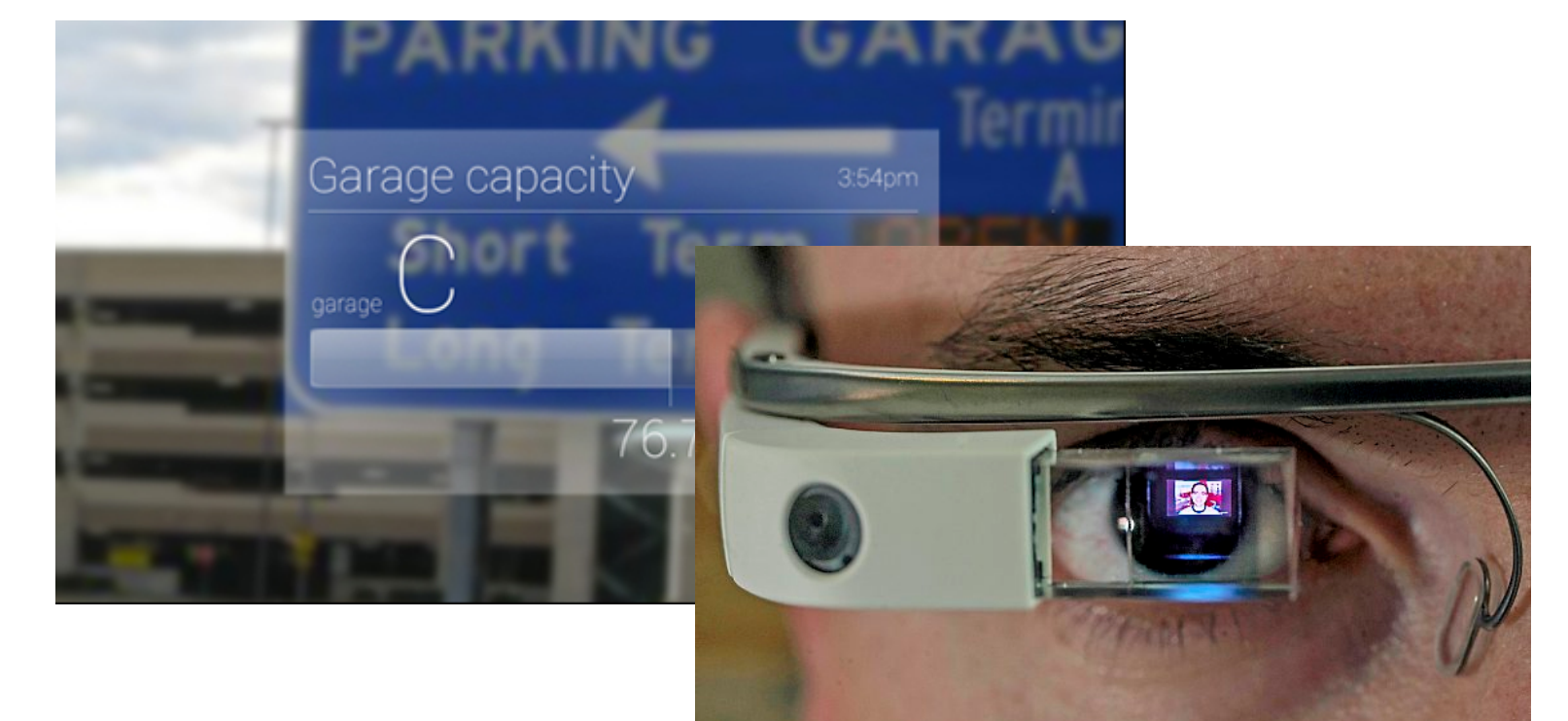
Carnegie Robotics Stereo Sensor



Support decision Making in Supermarket



Outdoor Navigation



Augmented Reality