Environmental Text Reader for the Blind and Visually Impaired

-- Eye of the blind

Problem:

Text is a simple but powerful communication channel for our daily life. We read hundreds of signs with text to obtain crucial information, e.g., room numbers, building directories, warning signs, and name tags. However, there are 285 million blind and visually impaired (BVI) people around the world [2] who can't read text as normally sighted people do. This causes not only inconvenience for them but also safety issues. Moreover, 90% of BVI people live in developing countries, so the solution must be affordable.

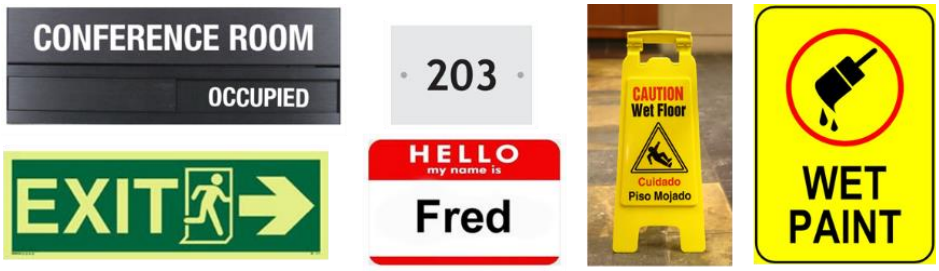


Figure 1. Environmental text

What We Would Build:

Taking advantage of the space and motion awareness technology from Project Tango, we wish to build an App that detects and recognizes environmental text efficiently and accurately, and allow interaction with BVI people.

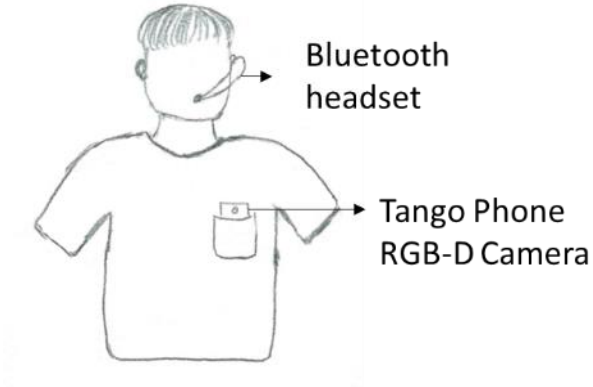
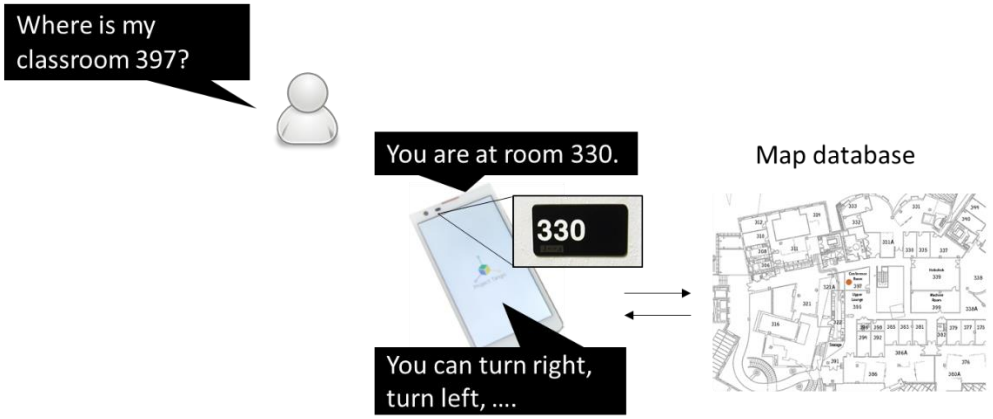


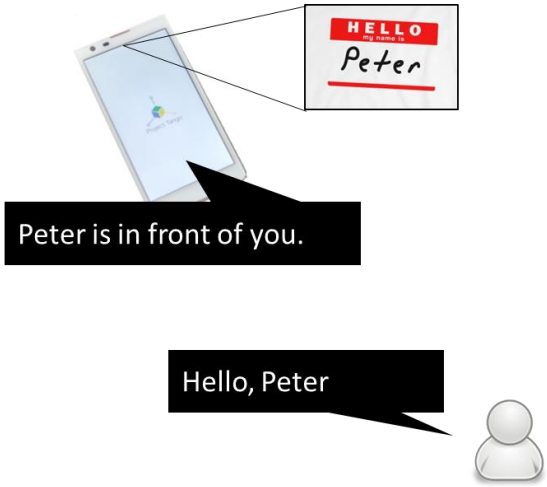
Figure 2. Proposed wearable text reading system

Use cases include:

- Generation of navigation cues from signs and directories.



- Reading name tags accurately in a conference room.



Our work can lead to better intelligent systems to assist people (both with and without disabilities) in everyday tasks.

Why Other Solutions are Insufficient:

- Braille: 1) not available for every signs, e.g. wet floor. 2) need efforts to locate it by touching
- Audio: 1) causes environmental noise. 2) position is inaccurate
- Guide dogs: hard to train them to read text
- Human assistances: expensive, not affordable for everyone

- Existing text-reading mobile Apps: typically require users to fit a text region into a given rectangle in the right scale for successful recognition. Unlike scanned documents with high resolution, environmental text often occurs in small portion of entire field of view, and generally is subject to motion blur or limited depth of field (i.e. lack of focus) of a moving camera. The real-time constraint on mobile devices make it even more challenging to our goal. Therefore, improved methods are required.

Who We Are:

Our team members are domain experts in text detection and simultaneous localization and mapping (SLAM) from Robotics, Vision, and Sensor Network Group (RVSN), Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology (MIT).

- Dr. Nick Wang (http://people.csail.mit.edu/hchengwang/)
- Kuan-Ting Yu, PhD student (http://people.csail.mit.edu/peterkty/)
- Prof. Seth Teller (http://people.csail.mit.edu/teller/)

What We Have Built:

Currently, we have built a prototype text-spotting system using SLAM, and demonstrated that our system can provide more accurate results by integrating multiple observations [1]. The Tango Project fits our previous work well by providing depth information and a map of the environment in real-time, which benefits text spotting as follows:

- Image perspective transformation to improve text recognition
- Spatial constraints computation (vertical surfaces and head/shoulder height)
- Suppression of previous inspected regions to improve detection efficiency

Unlike systems that only rely on RGB information, which requires the user to carefully point the camera at the text. Our system setup is as follows. The user wears a Kinect-style camera, a high resolution RGB camera, an inertial measurement unit (IMU), and a Hokuyo laser in front of the chest (Figure 3 Left). Captured RGB-D images are processed on a laptop. From the RGB image, we will use our text detection algorithm to find out potential text patches. Using depth images, we can extract the plane on which the text is located, and rectify the image patch with text before text recognition (Figure 3 Right). IMU, and laser provide information to register multiple depth images into a local 3D map.

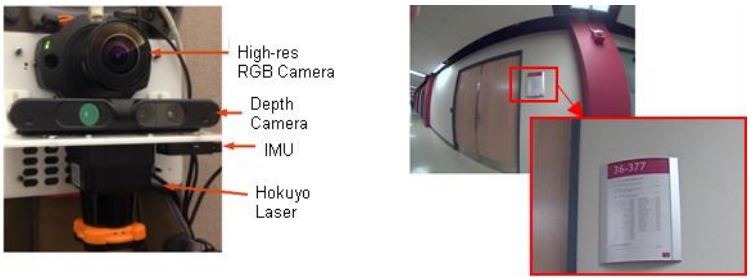


Figure 3. Current system overview

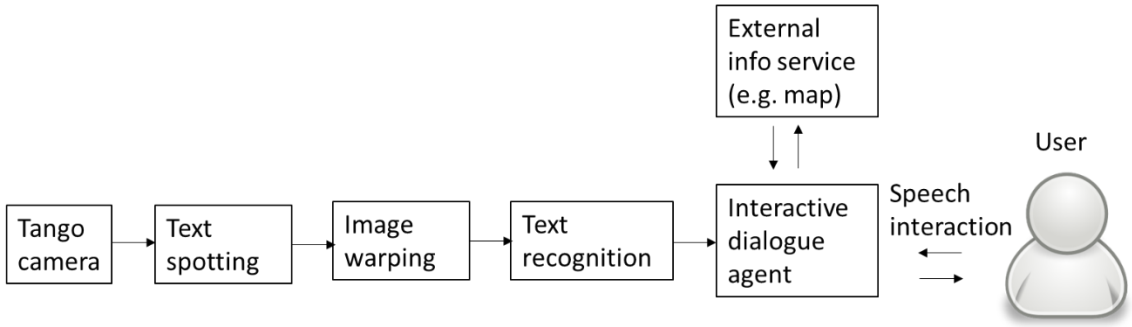
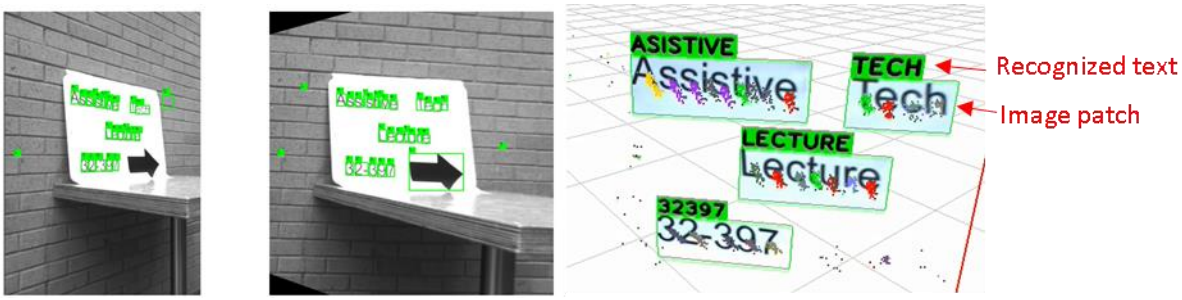


Figure 4. Proposed system flow



(a) Original patch (b) Warped patch (c) Image patches and recognized words

Figure 5. Image processing

Schedule Possible milestones for the next two years include:

- Develop a more robust and fast text detector against common false positive textures, such as window frames or brick walls by May, 2014
- Implement text-spotting capability on Tango by October, 2014
  - The App will be able to find and read the text in the environment
- Apply text-spotting to specific application by December, 2014
  - Link with a map database to provide indoor directions
  - Distinguish name tags from wall signs to provide name tag reading service
- Carry out user evaluations by January, 2015

Reference

[1] Wang, H. C., Landa, Y., Fallon, M., & Teller, S. (2013). Spatially Prioritized and Persistent Text Detection and Decoding. *Fifth International Workshop on Camera-Based Document Analysis and Recognition (CBDAR)*, Washington D. C., USA.

[2] Visual impairment and blindness. World Health Organization. June 2012.