THE JOURNAL ON
TECHNOLOGY AND
PERSONS WITH
DISABILITIES

Indoor Localization for the Visually Impaired Using a 3D Sensor

Feng Hu, Norbu Tsering, Hao Tang, Zhigang Zhu
Department of Computer Science, CUNY Graduate Center
Department of Computer Science, CUNY City College
Dept of Computer Information Systems, CUNY Borough of Manhattan Community College
fhu@gradcenter.cuny.edu, ntsering91@gmail.com,
htang@bmcc.cuny.edu, zzhu@ccny.cuny.edu

Abstract

An indoor localization system offers a significant assistance to the visually impaired in their daily lives by helping them localize themselves and further navigate an indoor environment. RGB-D sensor (e.g. Google Tango Tablet) is able to provide three-dimensional information of the environment around the sensor, which can be used for localization and navigation. In this paper, we propose a system that uses the Tango Tablet to first pre-build an 3D model of an indoor environment, and then utilize the newly captured RGB-D information and an Iterative Closest Point (ICP) algorithm to calculate the device's (i.e., user's) location and orientation corresponding to each RGB-D image. Voice feedback is provided to the users via text-to-speech. The system has three components: an environmental modeling and optimization module, a pose estimation module, and a GUI module. Experiments have been carried out in real indoor environments to test the performance of the system in terms of both time and accuracy.

Keywords

Assistive Technology, Indoor Localization, 3D modeling and matching, Mobile Computing, 3D Model Optimization.

Introduction

An indoor localization system is of significant importance to the visually impaired in their daily lives by helping them localize themselves and further navigate unfamiliar indoor environments. As of August 2014, there are 285 million visually impaired people in the world according to the World Health Organization, among who 39 million are blind. Compared to sighted people, it is much harder for visually impaired people to navigate indoor environments. Nowadays, too many buildings are unfortunately mainly designed and built for sighted people; therefore, navigational tasks and functions that sighted people take for granted could be huge problems to visually impaired people. Despite a large amount of research has been carried out for robot navigation in the robotics community, and several assistive systems are designed for blind people, efficient and effective portable solutions for visually impaired people are not yet available.

Indoor localization in general has been an area that attract researchers to tackle various problems using different sensors; detailed are discussed below in the Related Work section. Previous work in the computer vision community uses 2D images for localization and navigation, which is challenging due to lack of texture in the indoor environments. In this project, we use fully 3D information automatically captured by a tablet with a depth sensor to localize the device itself. In order to facilitate navigation for the visually impaired, we design, implement and evaluate the system to calculate the position and orientation of the device with the following two steps: (1) use a tablet with depth sensor to prebuild a large 3D indoor environment; (2) apply Iterative Closest Point (ICP) algorithm to a newly captured RGB-D (color and depth) image to calculate the user’s new position and orientation. The system includes three components: environmental modeling, pose estimation algorithm, and GUI design. The experiment tested with real model within a university laboratory shows a real time and accurate performance.

Discussion

Related Work

People who are not visually impaired rely almost exclusively on vision to know where they are in a new indoor environment. Since the visually impaired cannot use vision for this task, they need to use alternative sensory tools to collect information to explore the environment. However, the majority of the tools at their disposal are not able to tell them their locations accurately, not even for navigation. For example, a white cane can help them to determine whether an area is walkable or not, but it cannot provide the user their location information. A dog may help to lead the user to walk along a known path, but the user still needs other information to reason his/her location when he/she wants to change the route. GPS is sometimes used for localization in outdoor environments, but it does not function well in an indoor environment because of the significant signal attenuation. Radio Frequency (RF) based methods, e.g. WiFi or Bluetooth, are applied in indoor environments, but they heavily depend on extra devices (wireless routers or iBeacons) pre-installed in the environment. Previous work (Hu, Zhu, and Zhang 600-614; Irschara et al 2599-2606) in the computer vision field explored methods to process images by image matching and estimates the location information. However, image matches are error-prone in the indoor and urban environments with large textureless areas.

Some researchers use Structure from Motion (SfM) to create street 3D models in the outdoor environment, and recognize the places utilizing images from Internet (Torii et al 1808–1817; Zeisl, Sattler and Pollefeys 2704–2712; Sattler et al 2102–2110). Other researchers use Bag of Words (BoW) (Cao, Chen, and Fan 1-19) or ConvNet features (Sunderhauf et al 4297–4304) to represent outdoor environments for localization. However, few researchers focus on the indoor scenarios, especially for an assistive localization purpose. In addition, a practical SfM model heavily relies on the richness and distinguishes of environmental features extracted from the images, which is hard to use in environments where few features are available (such as rooms with white walls).

Mobile and wearable devices are cheap and ubiquitous nowadays, which accelerate the advancement of both general computer vision research and assistive applications. Farinella (Farinella et al 1086–1100) uses Android phones to implement an image classification system with DCT-GIST based scene context classifier. Altwaijry, Moghimi, and Belongie (167-174) apply Google Glass and develop an outdoor university campus tour guide application system by training and recognizing the images captured by Glass camera. Paisios (Paisios 2012), a blind researcher, creates a smart phone app for the Wi-Fi based blind navigation system. Manduchi (Manduchi 9-12) proposes a sign-based way-finding system and tests the blind volunteers with smart phones to find and decode the information embedded in the color marks pasted on the indoor walls. However, there is very few research work on designing user-friendly smart phone apps for helping visually impaired people to localize themselves and navigate through an indoor environment.

System Design and Implementation

We construct the system with three components: an environmental modeling and optimization module, a pose estimation module, and a GUI module; the details of which are discussed in the following subsections. The environmental modeling and optimization module is to utilize the device depth information to create 3D indoor environmental model and use corresponding color information to optimize the model. In the pose estimation module, a newly captured RGB-D image is used to align itself with the pre-built 3D model for calculating the device's location. An easy-to-use GUI module is designed with voice feedback for the visually impaired users. Fig. 1 shows the system diagram.

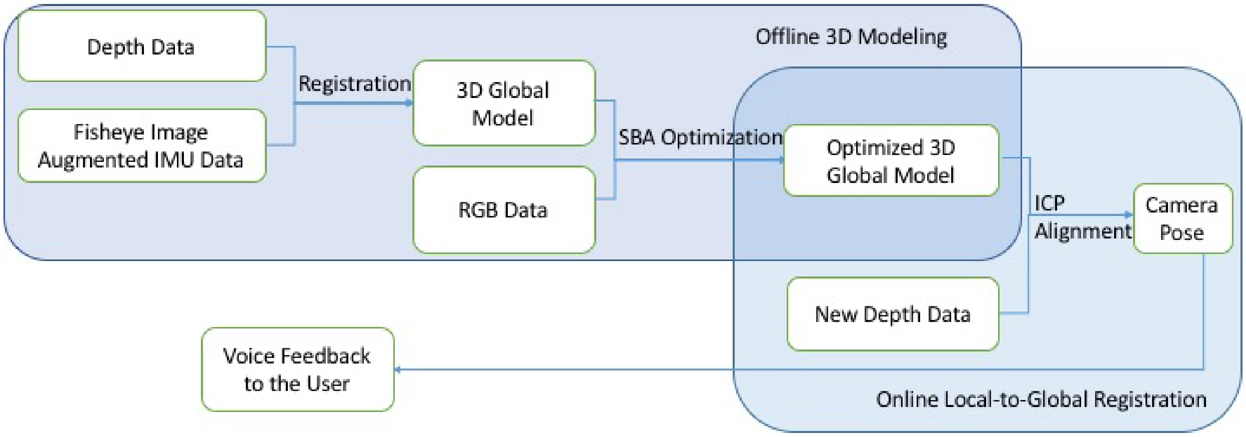


Fig. 1. 3D-sensor-based indoor localization system diagram.

3D modeling and optimization

With recent advance of depth and motion sensors (e.g. Google Tango Tablet, Microsoft Kinect and Structure IO) and development of 3D reconstruction technology (Dryanovski, William, and Xiao 1553-1559), we propose to apply the 3D depth sensor in the localization task. We use both 2D images and 3D depth information captured by a Google Tangle Tablet to improve robustness of location estimation. In this project, a pre-built model of the environment must exist before any localization can take place. The model is created as follows.

1. Capturing piecewise local 3D models of the environment and measuring the corresponding pose information of the Tablet.
2. Fusing the piecewise 3D models by transforming each model to a common coordination system using the given pose information of each local 3D model, and a global optimization framework is applied to increase the accuracy of fusion process (Lourakis and Antonis, 1).

Since Project Tango's Tablet device uses an Inertial Measurement Unit (IMU) for providing pose, augmented using built-in fisheye visual features, there are motion drifts accumulated while the sensor sensing the environment. Thus, loop closure is needed to adjust the generated global model by traversing the same area for a second time, and distribute drift errors along the motion path. In our work, we use Sparse Bundle Adjustment (SBA) (Lourakis and Antonis, 2) for the loop closure. The general idea of SBA is first to find a sequence of pairs of 2D image features and corresponding 3D points, and then project the 3D points back to the 2D images using the known camera intrinsic parameters and unknown extrinsic parameters, which includes the camera's pose information. By minimizing the distance between the projected features and observed features, we estimate the optimized camera motion parameters.

Pose estimation

As we have discussed in Step (1), the major component of the system is the accurate 3D reconstruction of the global model, which has been done offline on a separate machine. In our implementation, we have a reconstructed global model generated by stitching RGB-D data of frames together with the camera pose data of each frame. To localize the Tablet device, a user captures new RGB-D data at a new location by holding the Tablet, and the indoor navigation system then registers the new data with the pre-built global 3D model using the ICP algorithm (Zhang 119-152), in order to calculate the pose where the new data is captured. The registration algorithm will return a transformation matrix whose rotation matrix and translation vector can give us the orientation and position of the camera, respectively.

Graphic User Interface (GUI) design

In order to make the system applicable to the visually impaired, we've designed a GUI on the tablet using audio-tactile feedback, which reminds and guides users to adjust the tablet poses when capturing new RGB-D data for localization, and informs user the estimated location. Since the blind people cannot see the screen and are usually not comfortable with complex GUI, we simplify the interface by adding a big button on the right bottom part of the screen. Once the user presses the button, the tablet will begin to capture surrounding 3D data, and feedback to the user via voice after calculating the location. Fig. 2 shows an image of the GUI layout.

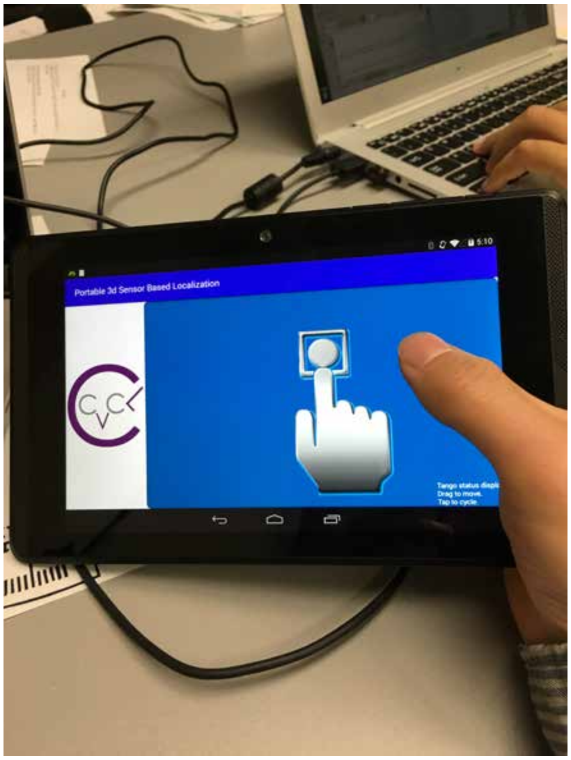
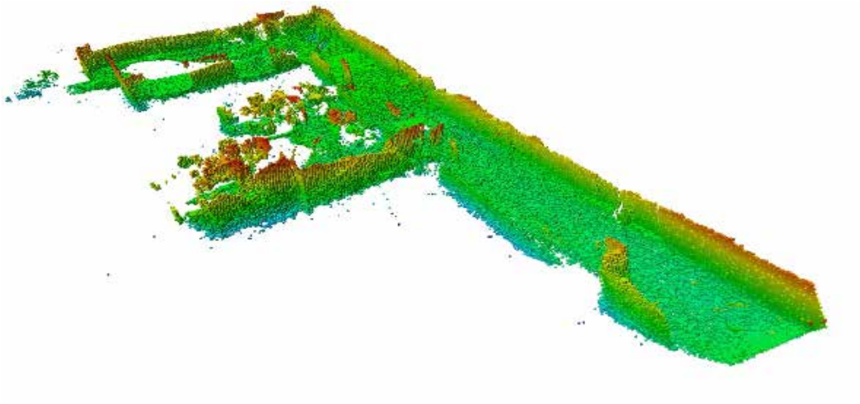


Fig. 2. GUI design. The large bottom right button is for starting localization service, and the rest area can be used for displaying information for administrative purpose.

Experiments

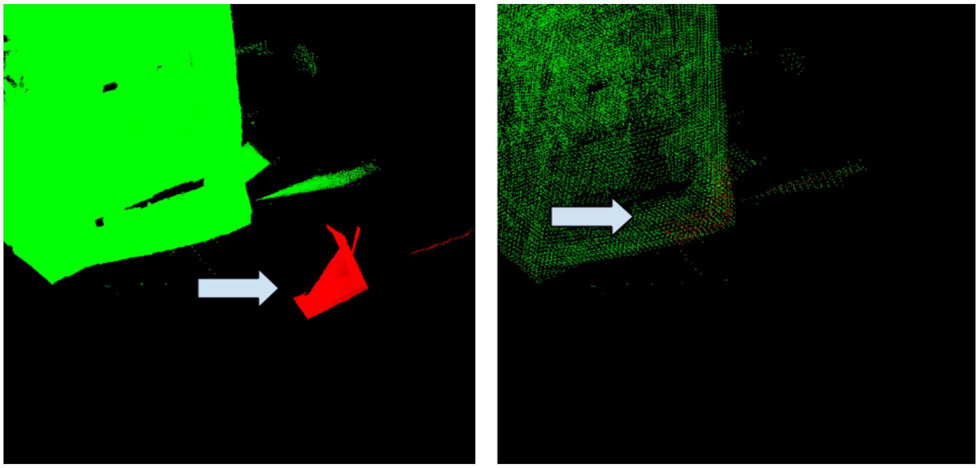
The Google Tango Tablet, which has a 3D sensor onboard, is adopted in our experiments. We have built a few large global 3D models in a campus building, as shown in Fig. 3. In the Fig. 3(a), we build a 3D model for an indoor research lab by scanning the room for 5 times, each with a different tilt angle. A short video of this model can be accessed from this link: https://goo.gl/ILFlTg. In the Fig. 3(b), half of the 8th floor of our NAC building are scanned and visualized. We then captured some RGB-D frames at different locations and apply the aforementioned registration method to estimate where each RGB-D frame is captured, that is the location of the user. The system can estimate correct locations in the experiments.

3(a)



3(b)

Fig. 3. A partial global 3D model includes corridors and rooms.



4(a) 4(b)

Fig. 4. Apply the ICP based registration algorithm for localization. (a) Local model (RGB-D frame, 3 feet by 3 feet, in red) and a partial global model (in green) before registration; (b) Registration result shows newly captured frame is correctly aligned with the global model, which indicates the location is correctly estimated.

Fig. 4 shows some experimental results. The green part represents the partial global 3D model of Fig. 3(b), whereas the small red model (a corner of a lab) indicates a local 3D model, which is a RGB-D frame captured after the global 3D model is built. Fig 4(b) shows the location is correctly estimated. A short video of this process can be accessed via the link: https://goo.gl/VqBZrp.

An optimized 3D model eliminating drifts is critical for an accurate indoor localization. In our work, we utilize the Sparse Bundle Adjustment (SBA) for improving the 3D model built with Tango device. We first extract SIFT features on all color images corresponding to each depth image, and then match them pairwise. After that, we select the qualified SIFT features by setting a threshold to make sure each qualified feature appears at least multiple images. Then we find the corresponding 3D points for each feature from the depth images. The SBA then accepts the features and corresponding 3D points and outputs optimized camera poses. Fig. 5 shows SBA optimization on the synthetic data, where the left part of Fig. 5 shows the camera poses and 3D points before SBA optimization, and the right part of Fig. 5 shows the new poses and points after the optimization.

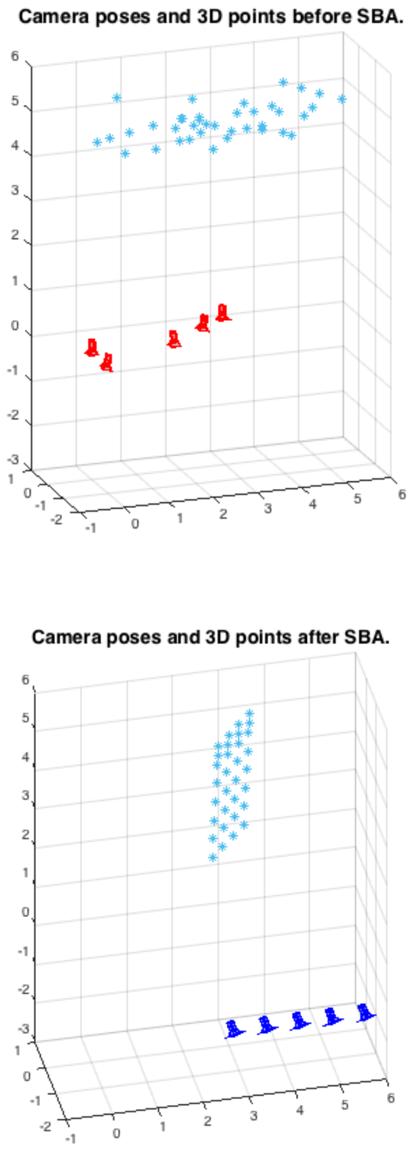


Fig. 5. Camera poses and 3D points before and after SBA optimization.

Conclusions

In this paper, we proposed a new method of localization for visually impaired, using a tablet with depth sensor. We believe the study is important for visually impaired, since portable depth sensors become popular on some tablets and smartphones. As ongoing work, we are expanding the testing database to larger environments, e.g. a whole campus building. In addition, we are going to recruit blind subjects for more formal testing and revise our system design based on their feedbacks.

Acknowledgments

This work is supported by the National Science Foundation under Award # EFRI-1137172, the 2015 NSF EFRI-REM pilot program at the City College of New York, and PSC-CUNY research Award # TRADB-46-121. Also thanks to Prof. Jianting Zhang for providing the Tango Tablet device.

Work cited

Altwaijry, Hani, Moghimi, Mohammad, and Belongie, Serge. "Recognizing locations with google glass: A case study." *In 2014 IEEE Winter Conference on Applications of Computer Vision (WACV)* (2014): 167-174. Print.

Cao, Jiuwen, Chen, Tao, and Fan, Jiayuan. "Landmark recognition with compact bow histogram and ensemble elm". *2015 Multimedia Tools and Applications* (2015): 1-19. Print.

Dryanovski, Ivan, William, Morris, and Jizhong, Xiao. "Multi-volume occupancy grids: An efficient probabilistic 3D mapping model for micro aerial vehicles." *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. (2010): 1553-1559. Print.

Farinella, Giovanni Maria, Ravì, Daniele, Tomaselli, Valeria, Guarnera, Mirko, and Battiato, Sebastiano. "Representing scenes for real-time context classification on mobile devices." *Pattern Recognition* 48:4 (2015): 1086–1100. Print.

Hu, Feng, Zhu, Zhigang, and Zhang, Jianting. "Mobile panoramic vision for assisting the blind via indexing and localization." *In Computer Vision-ECCV 2014 Workshops* (2014): 600-614. Print.

Irschara, Arnold, Zach, Christopher, Frahm, Jan-Michael, and Bischof, Horst. "From structure-from-motion point clouds to fast location recognition." *In Computer Vision and Pattern Recognition, 2009 IEEE Conference on* (2009): 2599-2606. Print.

Lourakis, Manolis IA, Antonis A. Argyros. "SBA: A software package for generic sparse bundle adjustment." *ACM Transactions on Mathematical Software (TOMS).* 36.1 (2009): 2. Print.

Manduchi, Roberto. "Mobile vision as assistive technology for the blind: An experimental study." *The 13th Int. Conf. on Computers Helping People of Special Needs (ICCHP 2012)* (2012): 9-16. Print.

Paisios, Nektarios. "Mobile Accessibility Tools for the Visually Impaired". PhD thesis, New York University, 2012. Print.

Sattler, Torsten, Havlena, Michal, Radenovic, Filip, Schindler, Konrad, and Pollefeys, Marc. "Hyperpoints and fine vocabularies for large-scale location recognition." *In Proceedings of the 2015 IEEE International Conference on Computer Vision* (2015): 2102-2110. Print.

Sunderhauf, Niko, Shirazi, Sareh, Dayoub, Feras, Upcroft, Ben, and Milford, Michael. "On the performance of convnet features for place recognition." *In Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on* (2015): 4297-4304. Print.

Torii, Akihiko, Arandjelovic, Relja, Sivic, Josef, Okutomi,Masatoshi, and Pajdla, Tomas. "24/7 place recognition by view synthesis." *In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition* (2015): 1808-1817. Print.

Zeisl, Bernhard, Sattler, Torsten, and Pollefeys, Marc. "Camera pose voting for large-scale image-based localization." *In Proceedings of the 2015 IEEE International Conference on Computer Vision* (2015): 2704-2712. Print.

Zhang, Zhengyou. "Iterative point matching for registration of free-form curves and surfaces." *International Journal of Computer Vision*. (1994): 119-152. Print.

Journal on Technology and Persons with Disabilities

ISSN 2330-4216

LIBRARY OF CONGRESS * U.S. ISSN CENTER
ISSN Publisher Liaison Section
Library of Congress
101 Independence Avenue SE
Washington, DC 20540-4284
(202) 707-6452 (voice); (202) 707-6333 (fax)
issn@loc.gov (email); www.loc.gov/issn (web page)