RGB-D camera Based Navigation for the Visually Impaired

Young Hoon Lee
Department of Electrical Engineering
University of Southern California
Los Angeles, California 90089
Email: lee126@usc.edu

Gérard Medioni
Department of Computer Science
University of Southern California
Los Angeles, California 90089
Email: medioni@usc.edu

***Abstract*—We present a wearable RGB-D camera based navigation system for the visually impaired. The navigation system is expected to enable the visually impaired to extend the range of their activities compared to that provided by conventional aid devices, such as white cane. Since this design is a successor version of a previous stereo camera based system to overcome a limitation of stereo vision based systems, algorithmic structure of the system is maintained. In order to extract orientational information of the blind users, we incorporate visual odometry and feature based metric-topological (Simultaneous Localization And Mapping) SLAM into our system. We build a vicinity map based on dense 3D data obtained from RGB-D camera, and perform path planning to provide the visually impaired with 3D traversability on the map. The 3D traversability analysis helps subjects steer away from obstacles in the path. A vest-type interface consisting of four microvibration motors delivers queues for real-time navigation with obstacle avoidance. Our system operates at $12-15$Hz, and helps the visually impaired improve the mobility performance in a cluttered environment. The results show that navigation performance in indoor environments with the proposed lightweight and affordable RGB-D camera improves compared to a stereo-vision based system.**

I. Introduction

Visual impairment hinders a person's essential and routine activities [3]. Furthermore, low vision or complete vision loss severely lowers functional independence of the visually impaired. It also highly reduces their ability and willingness to travel independently.

About 109,000 people with vision loss in the U.S. used long canes for navigation and obstacle avoidance purposes [10]. However, the device has limited functionality in crowded public areas. Most importantly, the relatively short range of the device means there is still a high risk of collision, because the visually impaired can avoid obstacles only when they make contact with them.

To provide answers to this problem and improve the mobility of people with vision loss, recent work has proposed utilizing various types of sensors to replace the white cane. These include ultrasonic [1, 5] and laser [20].

Recently, a real-time wearable, stereo-vision based, navigation system for the visually impaired [12] was proposed. The system is known to be the first wearable system. It consists of a head-mounted stereo camera and a vest-type interface device with four tactile feedback effectors. The head-mounted design enables blind users to stand and scan the scene to integrate wide-field view information, whereas waist-mounted or shoulder-mounted systems require body rotation. Furthermore, a head mounted device matches the frame of reference of the person, allowing relative position commands. The system uses a stereo camera as a data acquisition device and implements a realtime SLAM algorithm, an obstacle avoidance algorithm, and a path planning algorithm. Based on the algorithms mentioned above, an appropriate cue is generated and delivered at every frame to the tactile sensor array to alert the user to the presence of obstacles, and provide a blind user with guidance along the generated safe path. The authors carried out experiments to evaluate the mobility performance of blindfolded and truly blind users. The experiments were designed to evaluate the effectiveness of tactile cuing in navigation compared to widely used white canes. The experiment results indicate that the tactile vest system is more efficient at alerting blind users to the presence of obstacles and helping blind subjects avoid collisions than the white cane. The navigation performance was proved successful by showing trajectories generated by the proposed navigation system are the closest to the ideal trajectories from sighted subjects.

Fig. 1. A RGB-D camera on the top left and a tactile vest interface device on the bottom left. RGB-D camera operates with USB interface and tactile interface system communicates via Zigbee wireless network

The main limitation of the system is due to inherent shortcomings of stereo vision systems. For example, depth maps extracted by stereo camera systems in a low textured environment such as white walls are not accurate enough to

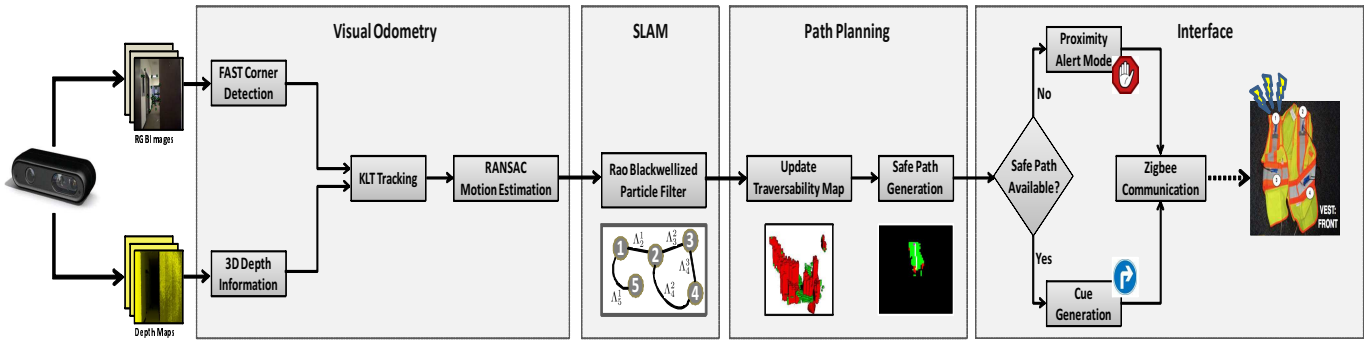

Fig. 2. System Overview

build a consistent 3D map of the environment for successful navigation.

Recently, Primesense introduced an affordable RGB-D camera that provides dense 3-D information registered with an RGB image at a faster rate than that of sensors used in the past and at low cost and power consumption. The main contribution of the paper is to introduce a wearable, mobility aid system for the visually impaired using the RGB-D camera built upon the system in [12] working in more robust way as presented in Section III. Since this system is a successor version of the previous stereo camera based system, the algorithmic structure of the system remains unchanged. The system uses the Primesense RGB-D camera as an input device to run obstacle detection route-planning algorithm. The tactile interface device receive a resulting cue through a Zigbee network for guidance along the safe path generated by the algorithms.

The rest of this paper consists of four sections as follows. Our system with all the building blocks is described in Section II. In Section III, experimental results are presented. And we finish with conclusions and future work in Section IV.

II. System Description

In this section, we provide a system overview as illustrated in Fig. 2. Our system input device is a RGB-D camera that provides dense 320×240 (QVGA) 3D point clouds synchronized with a RGB image at 60 frames per second (fps) through a USB interface.

A. Visual Odometry

Since its introduction in [19], Iterative Closest Point (ICP) algorithm for registration of dense point clouds has become the most popular method. In recent years, there have been researchers proposed variants of ICP algorithm to reduce computation time and to make it suitable for real-time application [13, 16]. Nevertheless, it may still be unsuitable for real-time SLAM applications which runs as fast as 12-15 Hz due to its computational complexity. Instead, we keep the main structure of the system described in [12]. In order to maintain a sense of egocentricity of a blind user, we continuously estimate the camera location and orientation with respect to the world coordinate frame.

FAST is a high-speed corner detector proposed by Rosten and Drummond in [14]. The corner detector is used as a saliency feature for tracking because of its speed and robustness in tracking as shown in [14, 15]. Let P^{t-1} and P^t be the position of the camera at time instance $t-1$ and t, respectively. To find the 6DOF camera motion between P^{t-1} and P^t, we first extract FAST corners in the previous frame, at $t-1$, and in the current frame, at t. The relative 2D motion in RGB image space of the FAST corners between two frames is calculated using KLT tracker[6]. Let F^{t-1} and F^t be a set of FAST corners extracted and matched across two frames at $t-1$ and at t, respectively. Then FAST corners successfully tracked between two frames are passed on to a RANSAC process. The camera motion can be calculated from correspondences between F^{t-1} and F^t along with their corresponding depth information at $t-1$ and t using the well-known 3 point algorithm [2]. Once camera motion is estimated, sparse bundle adjustment(SBA) is applied for refinement [17].

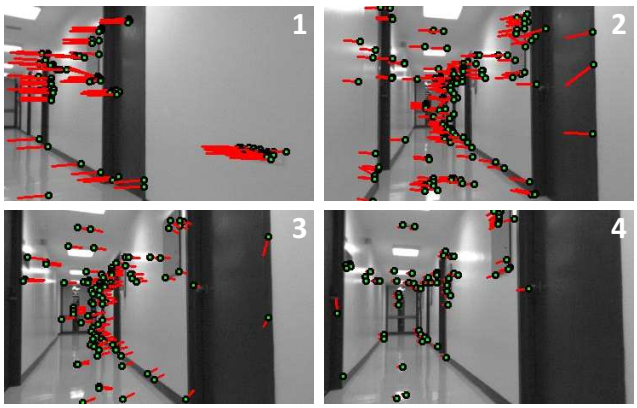

Fig. 3. FAST corners extracted are shown as green dots. KLT tracking estimates optical flow, and estimated camera motion is represented in red bars. Four consecutive image frames as a subject made a left turn and moved forward in a corridor are represented. Numbers in each figure represent the order of the scene. As seen in the top left figure, red bars head left which indicates that the subject is making a left turn. When a subject has finished making a turn and starts moving forward, red bars point toward the front direction as shown in the bottom right figure.

B. Metric-Topological SLAM

Consistent and accurate 3D location of the camera is important to perform robust path planning. SLAM algorithms help to prevents accumulation of errors by continuously estimating and correcting the camera location based on known landmarks and to maintain accurate position of the camera. However, as we expect to navigate a large area, the number of landmarks to be observed and registered would be very large which would slow down the SLAM process. Hence, we adopt the metric-topological SLAM approach as described in [11]. The approach has two different levels to describe an environment, metric and topological. The metric (local) level estimates the 6 dimensional camera location and orientation information s^t and sparse map m_t given a feature observation z^t such as KLT/SIFT or DAISY and camera motion estimation u^t until frame t. This can be represented using a standard Rao-Blackwellized FastSLAM framework [7, 8, 9] as follows.

$$p(s^t, m_t|z^t, u^t) \approx p(s^t|z^t, u^t) \prod p(m_t(i)|s^t, z^t, u^t) \quad (1)$$

$m_t(i)$ is the i^{th} landmarks in the local map with a normal distribution $\sim N(\mu^i, \Sigma^i)$. Every time registered landmarks in the map are observed, the map and their covariance matrix are updated using Extended Kalman Filter(EKF). However, a Rao-Blackwellized structure lets us update only the observed landmarks instead of the whole map. When the state vector of a submap grows up to a certain threshold size, manually chosen, or a 'visually novel' region is detected, a new submap is created. On creating a new submap, the new map is initialized by copying landmarks corresponding to the current observations from the previous submap and transforming to the current submap coordinate. Note that adjacent submaps are conditionally independent because of these common landmarks shared and the following transformation at the initialization step.

In the topological level, the global map contains a collection of submaps and their geometric relations to adjacent maps. The global map can be represented using a graph annotation.

$$G = \left(\{^iM\}_{i\in\sigma_t}, \{\Lambda^a_b\}_{a,b\in\sigma_t}\right) \quad (2)$$

^iM represents the i^{th} submap, σ_t is the set of computed submaps, and Λ^a_b indicates the coordinate transformation from ^aM to ^bM. More detail of the approach is provided in [11].

C. Traversability Map and Path Planning

The traversability map is a key part of this navigation system as the goal of this research is to provide a mobility aid and to help avoid obstacles in the path. By imposing some search distance restriction, we can decrease the amount of computation when detecting obstacles. In order to protect a blind user from collision as much as possible, we set a conservative radius of obstacle detection, 5m, where 3m is known as enough range for successful navigation.

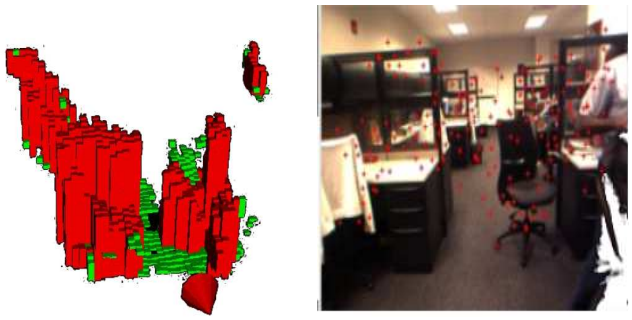

Fig. 4. Vertical (red) and horizontal (green) patches generated by 3D maps from a RGB-D camera for traversability map

We adopt a map representation from [18]. The traversability map is a 2D grid map quantized into 30cm $\times$ 30cm cells to reduce computation complexity. The vertical patches represent occupied regions in the map and have red colors while horizontal patches are free spaces that a subject is safe to navigate through. When 3D points clouds that fall into a cell have a different height range (maximum value - minimum value) exceeding a certain threshold value, the cell is registered as a vertical patch. For each remaining unknown state cell, we search the height of 8-neighbor cells. When the difference in height among 9 cells are in a certain boundary, we register the cell in the middle as a horizontal patch. Fig. 4 shows an example of a traversability map generated by 3D information obtained by a RGB-D camera. Note that the traversability map is not updated at every frame since 3D information are usually very noisy and this simple update strategy would cause an inconsistent map. Instead, we track the update of cells in the search area and those that match the previous state increase their confidence level. Cells with confidence level over a certain threshold are used in the next state.

In order to provide a reliable path to a way point, we maintain the moving heading vector of a blind user by comparing the previous position and current position. The previous position is updated only when there is translation greater than 1m to avoid frequent updates of the generated path and redundant computation. The way point is the farthest accessible (connected through green dots on the traversability map from the current position) point in a triangular region as shown in Fig. 5. On finding the way point in a map, the shortest path is generated using the D* Lite Algorithm as suggested by [4].

D. Cue Generation and Interface Device

This system has 6 different cues for a user as followings:

- Go straight (no tactile sensors on)
- Stop: Proximity Alert mode (all tactile sensors on)
- Step left (bottom-left sensor on)
- Step right (bottom-right sensor on)
- Turn left (top-left sensor on)
- Turn right (top-right sensor on)

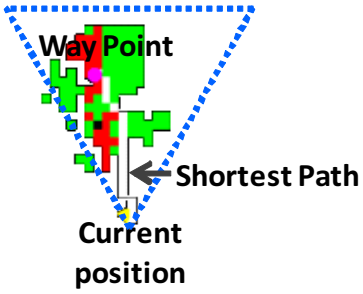

Fig. 5. Free spaces(green areas), occupied spaces(red areas), current position of a blind user(yellow dot), the way point (purple dot), generated path (white line), and search region (inside blue triangle).

The cue is generated simply based on a direction obtained by drawing a line from the current position to the way point. The generated cue is transmitted through the Zigbee wireless transmitter to the vest interface that has four vibration motors as provided in the bottom left corner of Fig. 1. The tactile feedback instead of audio feedback was adopted in order to reduce the amount of cognitive loads from hearing, which most blind users rely on for many other tasks. The wireless communication module has a very light weight of 0.1 pound and small packaging size, 7.5 cm $\times$ 4.5 cm $\times$ 2.5 cm (width$\times$length$\times$height), which is suitable for this wearable system. The operating range of this wireless system is about 100 feet at maximum indoors and 300 feet outdoors.

III. Results

We present the result of the navigation system using the Primesensor in an indoor environment. Our navigation system runs at $12-15$ Hz on the following configuration.

- CPU: Intel(R) Xeon Quad Core @ 3.72GHz
- RAM: 3.00 GB
- OS: Windows XP - 32 bit

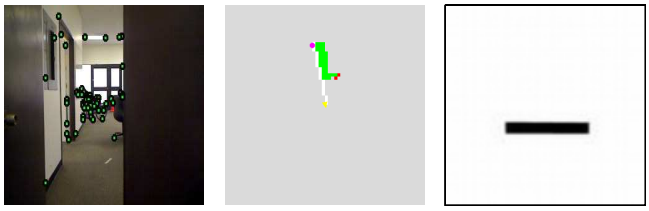

Fig. 6. Visual odometry and SLAM result. Left: RGB-D camera image with feature points extracted. Center: Traversability map with path finding algorithm. White bar show the safe path to the way point. Right: Calculated cue based on traversability map (Go straight).

Fig. 6 illustrates RGB image with feature points and the calculated traversability map. Green areas represent free regions while red areas represent occupied regions. White bar depicts the safe path to the way point. The way point is the farthest point in the traversability map within a certain triangle as described in the previous section.

A. 3D Depth Map Generation in a Low Textured Environment

For successful navigation tasks, it is essential to have an accurate map based on precise sensor information. There is a couple of advantages of a RGB-D sensor based systems over conventional stereo vision based systems, such as processing speed. Accurate 3D depth maps, especially in low texture environments which is very common indoors, are one of the most obvious reasons to operate a navigation system based on a RGB-D camera. As can be seen from Fig. 7, a stereo-vision based navigation system is vulnerable in certain areas where few visually salient features are available. In this section, we would like to provide a simple test result to visually compare the accuracy of 3D depth data obtained by both a stereo based system and a RGB-D based system. The figures in the 2nd column of Fig. 7 are disparity maps generated by a stereo camera system. As seen in Fig. 7, the stereo based system does not provide correct disparity information of low texture scenes such as a solid color door or white walls, which will lead incorrect 3D depth information. As a consequence, a blind subject may receive incorrect feedback information, heading to possible collisions.

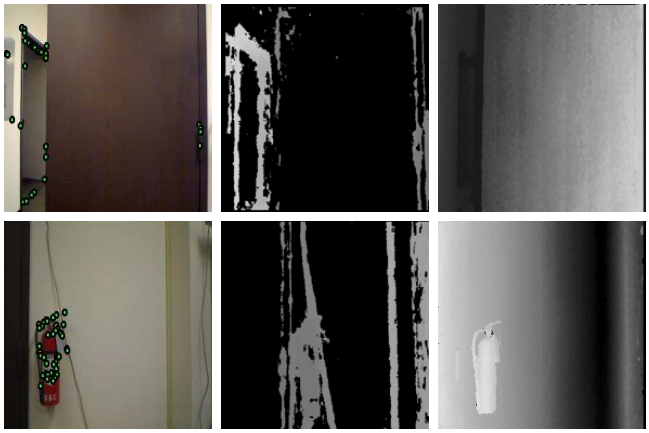

Fig. 7. Disparity map experiments in low textured areas. Left: RGB camera images. Center: Disparity maps generated by a stereo camera. Right: Disparity maps generated by a RGB-D camera.

B. Traversability

The goal of the system is to guide blind users without collision with obstacles on the way to a destination. To verify the performance of the navigation system, experiments were conducted for a simple navigation task. A subject starts from one end of a corridor and aims to reach the other end of the corridor. As shown in Fig. 8 and Fig. 9, the experiment environment has solid walls and reflective floors, which is quite common in indoor environments. Both experiments with RGB-D camera and Stereo Camera involved about 350 frames. Fig. 8 suggests that the 3D depth map from the stereo camera is inaccurate and results in inconsistency of the traversability map in the experiment area. From the 2nd row of Fig. 8 and on, errors in building the traversability map of an environment accumulate. This spurious information of the map causes the

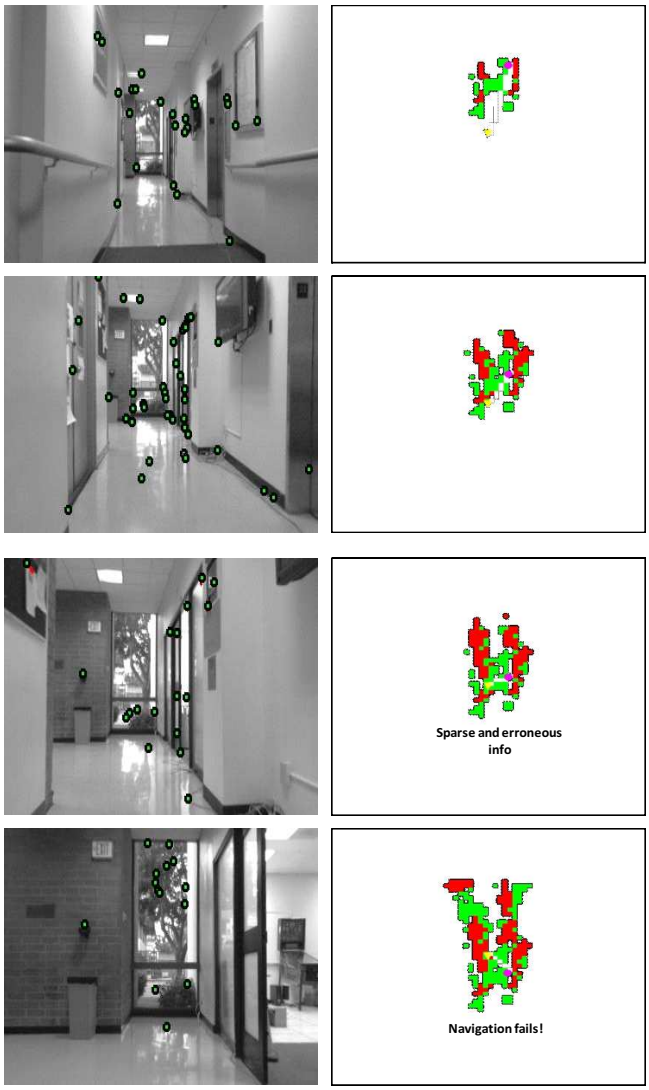

Fig. 8. Four snapshots of traversability map of the system with stereo camera

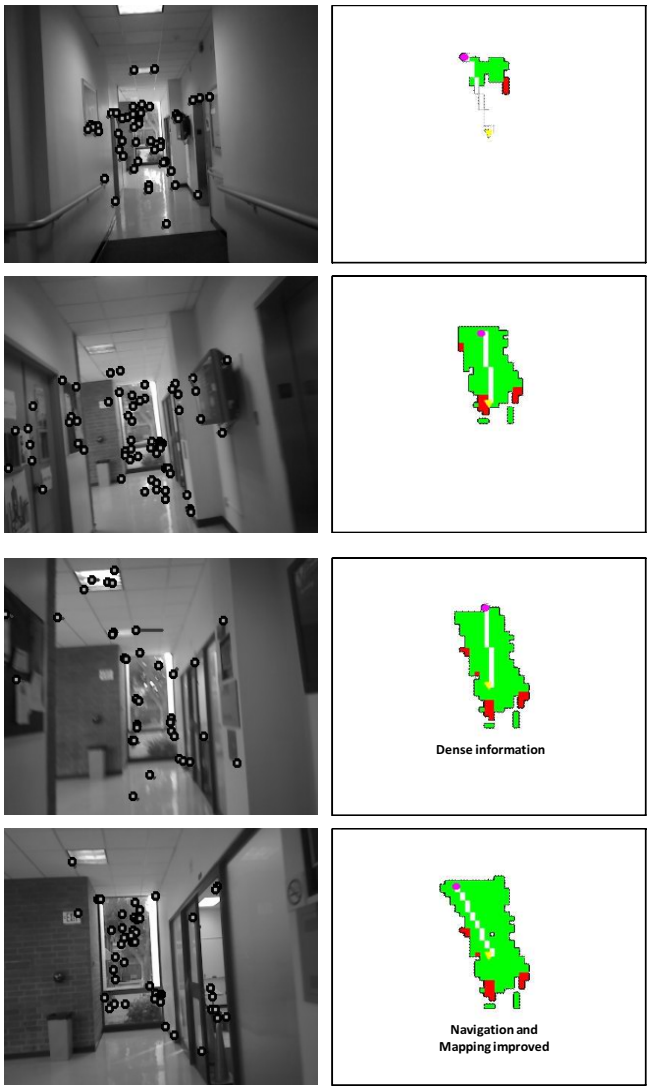

Fig. 9. Four snapshots of traversability map of the system with RGB-D camera

navigation algorithm to fail to find an appropriate way point and a safe path. As depicted in the 4th row of Fig. 8, the stereo camera based navigation system hallucinates obstacles that do not exist, and navigates a subject in the wrong direction. In Fig. 9, we replace the stereo system by a RGB-D camera. The dense traversability map was built and showed more accurate and consistent map with the real corridor environment. It is notable that the low textured experiment area still affects the visual odometry performance of the system since we perform visual odometry on RGB images. However, traversability map and navigation worked much more robustly than the navigation system with a stereo camera pair. Then the performance and reliability of the system improved.

IV. Conclusion and Future work

We have presented an integrated system using RGB-D camera to improve navigation performance of the stereo camera based system proposed in [12] in low textured environments. A RGB-D camera enables a blind subject to navigate through an indoor environment where stereo-vision based algorithm is not working properly such as a corridor with very low texture. As an extension of this work, we aim to carry out a quantitative performance analysis of different type of sensors for the navigation systems. We will also investigate other ways to register dense depth maps without the use of features extracted in RGB images.

V. Acknowledgement

This research was made possible by a cooperative agreement that was awarded and administered by the U.S. Army Medical Research & Materiel Command (USAMRMC) and the Telemedicine & Advanced Technology Research Center (TATRC), at Fort Detrick, MD under Contract Number: W81XWH-10-2-0076

REFERENCES

[1] J. Borenstein and I. Ulrich. The GuideCane - A computerized travel aid for the active guidance of blind pedestrians. In *IEEE Int. Conf. on Robotics and Automation*, pages 1283–1288, 1997.

[2] M. Fischler and R. Bolles. Random Sample Consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[3] R. G. Golledge, J. R. Marston, and C. M. Costanzo. Attituds of visually imparied persons towards the use of public transportation. *Journal of Visual Impairment Blindness*, 91(5):446–459, 1997.

[4] S. Koenig and M. Likhachev. Fast replanning for navigation in unknown terrain. *IEEE Transactions on Robotics*, 3(21):354–363, 2005.

[5] B. Laurent and T. N. A. Christian. A sonar system modeled after spatial hearing and echolocating bats for blind mobility aid. *Int. Journal of Physical Sciences*, 2 (4):104–111, 2007.

[6] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Int. Joint Conf. on Artificial Intelligence*, pages 674–679, 1981.

[7] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit. FastSLAM: A factored solution to the simultaneous localization and mapping problem. In *Proc. of the AAAI National Conf. on Artificial Intelligence*, pages 1151–1156, 2002.

[8] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit. FastSLAM 2.0: An improved particle filtering algorithm for simultaneous localization and mapping that probably converges. In *Int. Joint Conf. on Artificial Intelligence*, pages 1151–1156, 2003.

[9] K. Murphy. Bayesian map learning in dynamic environments. In *Neural Information Processing Systems*, pages 1015–1021, 1999.

[10] JVIB news service. Demographics update: Use of white "long" canes. *Journal of Visual Impairment Blindness*, 88(1):4–5, 1994.

[11] V. Pradeep, G. Medioni, and J. Weiland. Visual loop closing using multi-resolution SIFT grids in metric-topological SLAM. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 1438–1445, 2009.

[12] V. Pradeep, G. Medioni, and J. Weiland. Robot vision for the visually impaired. In *Computer Vision Applications for the Visually Impaired*, pages 15–22, 2010.

[13] D. Qiuand, S. May, and A. Nüchter. GPU-Accelerated nearest neighbor search for 3d registration. In *Proc. of the 7th Int. Conf. on Computer Vision Systems: Computer Vision Systems*, pages 194–203, 2009.

[14] E. Rosten and T. Drummond. Fusing points and lines for high performance tracking. In *IEEE Int. Conf. on Computer Vision*, pages 1508–1511, 2005.

[15] E. Rosten and T. Drummond. Machine learning for high-speed corner detection. In *European Conf. on Computer Vision*, pages 430–443, 2006.

[16] S. Rusinkiewicz and M. Levoy. Efficient variants of the ICP algorithm. In *Proc. of 3rd Int. 3D Digital Imaging and Modeling Conf.*, pages 145–152, 2001.

[17] N. Sunderhauf, K. Konolige, S. Lacroix, and P. Protzel. Visual odometry using sparse bundle adjustment on an autonomous outdoor vehicle. In *Tagungsband Autonome Mobile Systeme*, pages 157–163, 2005.

[18] T. Triebel, P. Pfaff, and W. Burgard. Multi-level surface maps for outdoor terrain mapping and loop-closing. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems.*, pages 2276–2282, 2006.

[19] C. Yang and G. Medioni. Object modelling by registration of multiple range images. *Image and Vision Computing*, 3(10):145 155, 1992.

[20] D. Yuan and R. Manduchi. Dynamic environment exploration using a virtual white cane. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 243–249, 2005.