Robot Vision for the Visually Impaired

Vivek Pradeep, Gerard Medioni, James Weiland
University of Southern California, Los Angeles, California, USA
{vivekpra, medioni, jweiland}@usc.edu

Abstract

We present a head-mounted, stereo-vision based navigational assistance device for the visually impaired. The head-mounted design enables our subjects to stand and scan the scene for integrating wide-field information, compared to shoulder or waist-mounted designs in literature which require body rotations. In order to extract and maintain orientation information for creating a sense of egocentricity in blind users, we incorporate visual odometry and feature based metric-topological SLAM into our system. Using camera pose estimates with dense 3D data obtained from stereo triangulation, we build a vicinity map of the user's environment. On this map, we perform 3D traversability analysis to steer subjects away from obstacles in the path. A tactile interface consisting of microvibration motors provides cues for taking evasive action, as determined by our vision processing algorithms. We report experimental results of our system (running at 10 Hz) and conduct mobility tests with blindfolded subjects to demonstrate the usefulness of our approach over conventional navigational aids like the white cane.

1. Introduction

Visual impairment leads to loss of independence in performing several routine and life-enabling tasks. According to a survey carried out in [14], the task of route planning and unforseen obstacles can severely impede the independent travel of sightless individuals and will reduce their willingness to travel, despite having access to the long cane or guide dog. Genensky et al. [13] reported difficulties in detecting small obstacles in the path of travel, uneven walking surfaces and horizontal objects at eye level or lower for a majority of 94 visually impaired subjects interviewed. It has been estimated that more than 30% of the blind population do not ambulate autonomously outdoors [7]. All these factors, therefore, underscore the necessity of a device that aids in detecting obstacles (structured or unstructured) and helps perform wayfinding.

Pending the availability of more current information, approximately 109,000 people with vision loss in the US used long canes to get around in 1990 [1] and just over 7000 used dog guides. It is estimated that annually, only about 1,500 individuals graduate from a dog-guide user program [2]. While these navigational aids are currently the most popular, they also have limited usability in crowded regions, social situations and definitely do not resolve the primary issues of independent mobility and non-reactive navigation (that is, avoiding an obstacle without actually having to touch it). Electronic travel aids (ETAs), leveraging ultrasonic [6, 10, 18], laser [5, 33] or vision sensors, are designed to further enhance user confidence about mobility. Vision offers a cheap, passive and low power modality that provides high information bandwidth that can be exploited for developing sophisticated applications towards a real-time mobility aid.

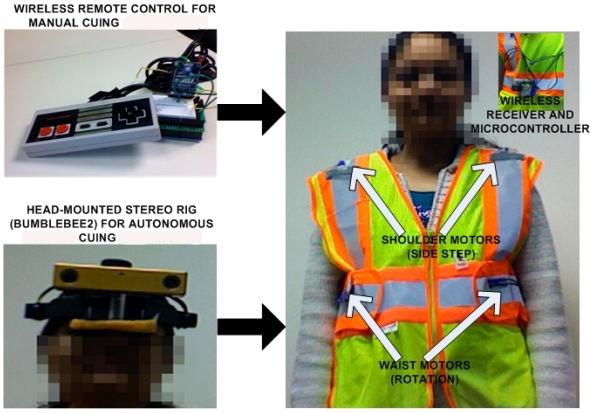


Figure 1. Tactile cuing system with headmounted stereo camera for a wearable mobility aid. Vibration motors located on the shoulder and waist guide the subject around obstacles. There are two modes of operation - the vibrations motors on the vest can be activated by manual control for initial validation experiments. In the final experiments, an online SLAM algorithm with obstacle detection autonomously communicates via wireless interface.

In this paper, a novel, wearable, low-vision aid is proposed that utilizes computer vision algorithms for providing navigational assistance to patients with impaired vision[1].

[1]This work was supported in part by the National Science Foundation under Grant No. EEC-0310723

978-1-4244-7030-3/10/$26.00 ©2010 IEEE

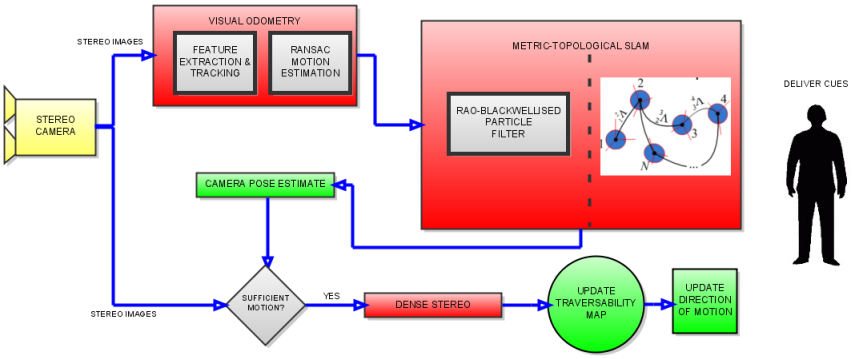


Figure 2. System Overview

Our design consists of a head-mounted stereo camera that provides data for running an online SLAM algorithm along with obstacle detection (see figure 1). Based on the processed output, a wearable array of microvibration motors provides tactile cues to alert about the presence of obstacles and guide the user along a computed, safe path. To the best of our knowledge, our device is the first instance of a completely developed system for such a wearable application. Furthermore, we also present results of experiments carried out with the system to evaluate mobility performance for blindfolded subjects. The rest of this paper is divided as follows. We cover vision based mobility devices and SLAM applications in Section 2. Beginning with an overview of our system, we describe the various building blocks in Section 3. Experimental and evaluation results are presented in Section 4 and we finish with conclusions and plans for future work in Section 5.

2. Literature Review

A stereo camera configuration is described in [8], that conveys depth information by means of a tactile interface and requires the user, over time, to learn to associate patterns of tactile stimuli with corresponding objects. This approach leverages the concept of sensory substitution [3], which is a mechanism by which characteristics of one sensory modality (in this case, vision) are transformed into stimuli of another sensory modality (touch). However, the system proposed in [8] can be expected to impose significant training time and cognitive load on the user due to the non-semantic nature of the information conveyed. The ASMONC system described in [22] integrates stereo vision with sonar sensors to derive high level information that requires less interpretation and is more meaningful for navigation. As all the sensors are fixed on the body, the subject has to perform bodily rotations to integrate wide field scene information. A head-mounted setup, on the other hand, enables a subject to stand and scan, thereby allowing the system to gather more information about the scene to plan the safest and most efficient route. This has significance for those who are completely blind as they might not always be oriented in the correct direction. Orientation information is particularly useful as determining how the user is spatially located with respect to obstacles helps in generating correct cues for taking evasive action.

Simultaneous Localization and Mapping (SLAM) incorporates noisy sensor data and motion model to reliably compute camera trajectory and scene information and can be used to maintain reliable estimates of pose. This approach has been proven to be quite successful in navigation for autonomous robotic systems [24, 21] and urban/landscape modeling [26]. Vision based SLAM systems have also been proposed for wearable applications such as augmented reality [17]. Implementing SLAM systems intended for wearable applications, however, is a significantly different problem from that in robotics as there is no reliable way to predict user motion.

While visual SLAM systems with single cameras [9] show good performance, we are interested in reconstructing metric maps for obstacle avoidance and therefore, employ stereo cameras. The idea of a SLAM-based mobility aid is not new. In [29] and later, in [28], the authors describe a stereo vision based setup that utilizes estimates of camera trajectory to predict user motion and build 3D maps of the vicinity. The described system, however, detects only overhead obstacles and being shoulder-mounted, does not have the advantages just described.

3. System Description

Figure 2 provides an overview of the main components of our system. We always maintain information about the current camera pose and a 'traversability map' of the region surrounding the user's location. Camera motion is tracked using a visual odometry pipeline along with a SLAM module to obtain consistent results. The SLAM implementation is a Rao-Blackwellised particle filter (RBPF) [25] in a FastSLAM [24, 23] framework using a combination of

KLT [20] and SIFT [19] tracking to solve for data association. Each observed feature i yields a corresponding 3D point (from stereo triangulation) μ^i and an associated covariance Σ^i arising from errors in feature localization and stereo reconstruction. This landmark-based scheme is inspired from the approach described in [30]. An important point to note is that the SLAM map and traversability map are two different objects. The SLAM map is collection of sparse landmarks that is propagated every frame to yield consistent camera pose estimates. It is only a by-product of the SLAM algorithm and is not employed for any other purpose. On the other hand, the traversability map is created by registering and processing dense 3D point-clouds obtained by stereo triangulation. However, we do not compute a dense 3D cloud at every frame. The traversability map is only updated when camera motion exceeds a certain threshold such that there is significantly new 3D data available. This helps to keep computational and memory costs low and not perform redundant computations. The current orientation and traversability map are used to generate tactile cues to keep the subject on a 'safe' path. We have adopted tactile cues instead of audio since the latter impose greater cognitive load on the subject and blind users rely on hearing to perform a wide variety of other tasks.

3.1. Stereo Vision Odometry

Camera motion can be estimated by matching features across two stereo views. Let P_L^{t-1}, P_R^{t-1} represent the left and right cameras before the motion and P_L^t, P_R^t be the pair after. Matched correspondences across $(P_L^{t-1}, P_R^{t-1}, P_L^t)$ or $(P_L^{t-1}, P_R^{t-1}, P_R^t)$ can be used to compute the camera motion using the three-point algorithm [15] in a RANSAC [12] setting. For added robustness, however, we match features and measure reprojection errors across all four views. Furthermore, since the stereo geometry is known apriori, the associated fundamental matrix F is used to refine matches by minimizing the following constraint

$$x_R^\top F x_L = 0 \tag{1}$$

(x_L, x_R) represents the feature correspondence set in either of the stereo pairs. This minimization can be achieved in closed form using the technique described in [16]. Finally, once the motion is estimated and corresponding inliers found, sparse bundle adjustment [31] is applied to get a more consistent solution. The feature covariances can be propagated to get the motion uncertainty for use in the SLAM filter.

3.2. Metric-Topological SLAM

Since we expect the system to be used for traversing and mapping large areas that could potentially lead to several thousands of landmarks, we adopt the metric-topological approach described in [27]. This has two levels of environment representation: local, metric (submap) and global, topological. The local submap level estimates state information corresponding to the six dimensional camera trajectory s^t and sparse map m_t, given feature observations (KLT/SIFT) z^t and camera motion estimates u^t collected until the current time t. In the standard RBPF formulation, an approximate but highly efficient solution can be obtained by the following factorization:

$$p(s^t, m_t | z^t, u^t) \approx p(s^t | z^t, u^t) \prod_i p(m_t(i) | s^t, z^t, u^t) \tag{2}$$

$m_t(i)$ is the i^{th} landmark in the map, represented by a normal distribution $\sim N(\mu^i, \Sigma^i)$. Each time a feature is reobserved, the corresponding landmark is updated using the extended Kalman filter. Rao-Blackwellisation enables us to only update the observed landmark instead of the whole map. At the global level, the map is represented as a collection of 'submaps' using an annotated graph:

$$G = \left(\left\{ {}^iM \right\}_{i \in \Omega_t}, \left\{ {}^b_a\Lambda \right\}_{a,b \in \Omega_t} \right) \tag{3}$$

${}^iM$ are the metric submaps, Ω_t is the set of computed submaps and ${}^b_a\Lambda$ are the coordinate transformations between adjacent maps. Each local submap, built using equation 2 in the time interval $t-\tau$ to t, encapsulates N samples of the camera pose trajectory and per-particle maps. A new submap is created when the per-particle map size exceeds a certain threshold (hard constraint) or a 'visually novel' region is detected (soft constraint). The latter determination is achieved by keeping track of the ratio of new feature observations to older feature tracks. The new submap is initialized by copying the landmarks in the previous submap corresponding to the current observations and transforming them into the new coordinate frame. Adjacent submaps are conditionally independent due to these shared landmarks and coordinate transformation. A formal proof of this conditional independence can be found in [11].

3.3. Traversability Map

The ultimate goal of the system is to detect obstacles in the path and warn subjects about their presence. Obstacle detection on point cloud data obtained from stereo processing is challenging due to the high noise content, irregular structure of the operating environment, unknown shape of the ground surface, changing camera pose and real-time requirements. The typical approach of imposing a horizontal plane in the cloud and computing distance to individual points in infeasible. We can minimize the computation load by limiting obstacle detection to only the current vicinity of the subject. Mobility studies in visually impaired subjects have found that successful navigation requires information only upto 3 meters ahead [4]. Our conservative strategy, therfore, is to extract a sphere of data around the subject

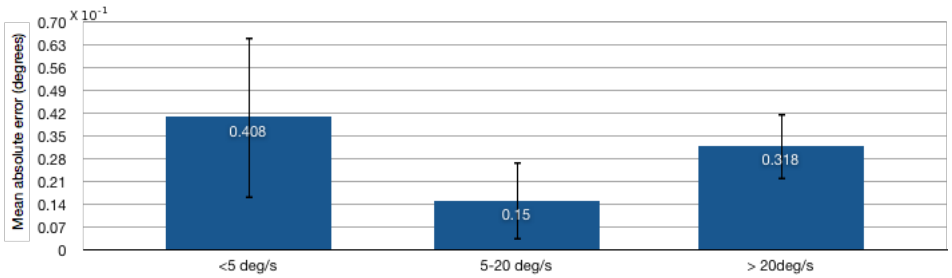
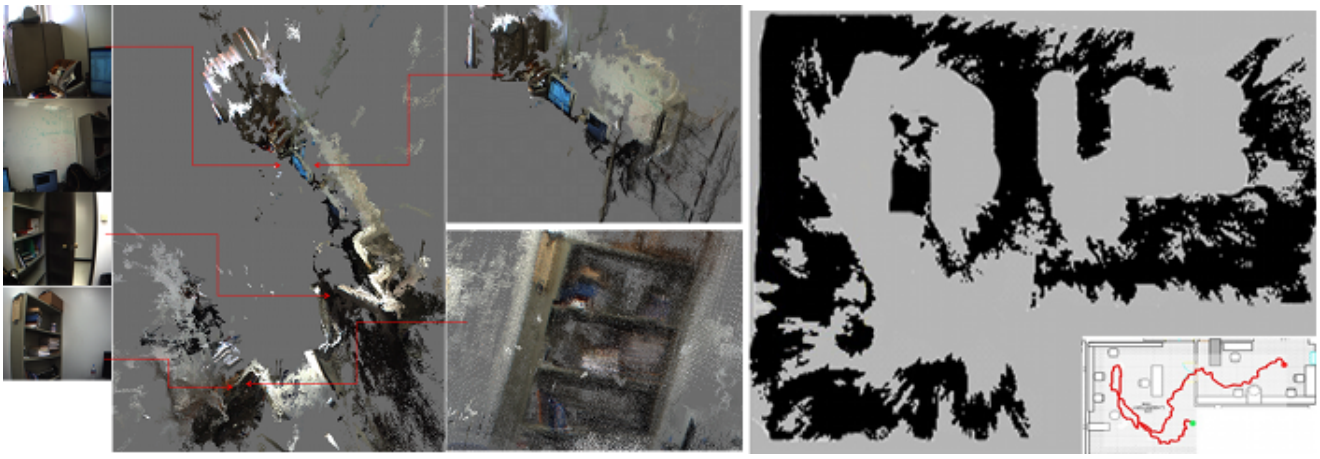


Figure 3. Mean absolute errors in estimation of camera yaw for three levels of angular velocity



(a) Mapping a room (b) Mapping a larger environment

Figure 4. Mapping results for small and large indoor environments. (a) shows a dense 3D, textured map with selected regions magnified with the corresponding images and (b) shows an untextured map projected onto the ground plane for a much larger environment. Inset displays the floorplan corresponding to the latter with the camera trajectory (30 meters long) superimposed in red.

with a radius of 5 meters and perform scene interpretation in this region.

We adopt the technique described in [32] for handling large point-clouds by constructing a multi-surface elevation map. The point cloud region is quantized into a 2D grid, and for each bin, the heights of the points falling within it are stored. A brief review of the algorithm, modified for our application, follows:

- *Height intervals* are computed for each bin from the stored values. Two consecutive heights belong to the same interval if the values are closer than a given *gap size*. The gap size can be equal to or greater than the height of the person to enable him/her to pass through. This implies that each bin can have more than one height interval. For instance, if there is an open door, points belonging to the floor will fall in one interval and those belonging to the upper section of the wall will fall in another.

- Each interval is classified as a horizontal or vertical patch, based on the height of the interval. If the height exceeds a certain threshold thickness, then it is considered as vertical, otherwise it is horizontal.

- We label any vertical patch as untraversable and if any new data point during update falls in it, it is ignored.

- If a horizontal patch has atleast 5 neighbors with horizontal patches, all under a certain height difference, then it is labeled traversable. To compute this neighborhood efficiently, we scan the 2D grid from left to right, top to bottom. If the bin has a horizontal patch, it is stored as a node in a graph. Next, the upper three and left grid cells in the 3×3 neighborhood are checked and edges are added to corresponding nodes if they exist and height differences are below the threshold. A second pass checking for the number of edges for each node determines the final traversability map.

3.4. Predicting Motion and Cue Generation

As mentioned before, our head-mounted design is motivated by the potential for subjects to stand and scan, integrating wide field information. To perform obstacle avoidance, however, it makes more sense to convey information relative to the body position. To this end, we keep track of the magnitude and direction of motion. At any instant, if the magnitude of translation with respect to the previous

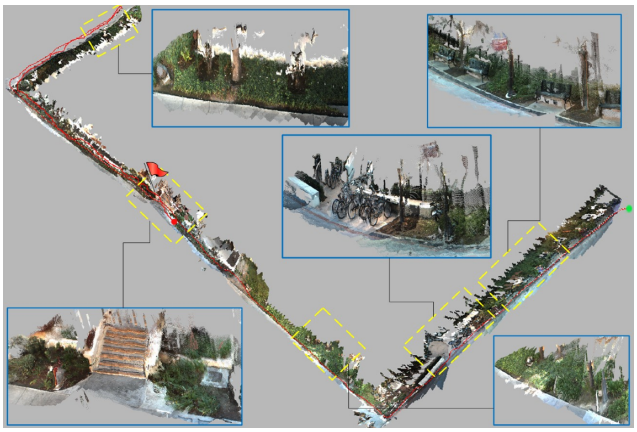


Figure 5. 3D reconstructed map with camera trajectory (red, 232 meters) for the outdoor test environment. Selected regions have been magnified for visualization.

(reference) position exceeds a certain threshold, then the direction of motion and reference position are updated. The direction is obtained by the line joining the current position to the previous position. If there is very little translation, however, no position updates are performed and it is assumed that the subject is standing in the same location (see figure 7).

Currently, the approach for cue generation is very simple. For instance, if the user is at a location such that the nearest and most continuous traversable path is to the left, then the vibration motor corresponding to that direction is activated. This can obviously lead to trap situations, and a more sophisticated route planning algorithm is planned for future work (see Section 5).

4. Results

We present results for the various modules of our system in this section. The Bumblebee2 stereo camera sold by Point Grey Research was used for all our experiments. Results of human mobility testing with normal sighted subjects under blindfold are also provided.

4.1. Visual Odometry and SLAM

Figure 3 shows the errors obtained for camera heading (yaw) when compared with readings from a commercially available Inertial Measurement Unit (IMU). The IMU, sold and manufactured by Intersense Inc under the brand name InertiaCube3, was mounted on the camera to provide ground truth data. Visual odometry and IMU are both prone to drift and so, the error estimates compare frame-to-frame motion rather than accumulated motion. The yaw angle was selected for analysis as it undergoes the most change during head scanning motions. Performance for three different conditions are evaluated: slow camera rotation (less than 5 degrees/second), medium camera motion (5-20 degrees/second) and fast camera motion

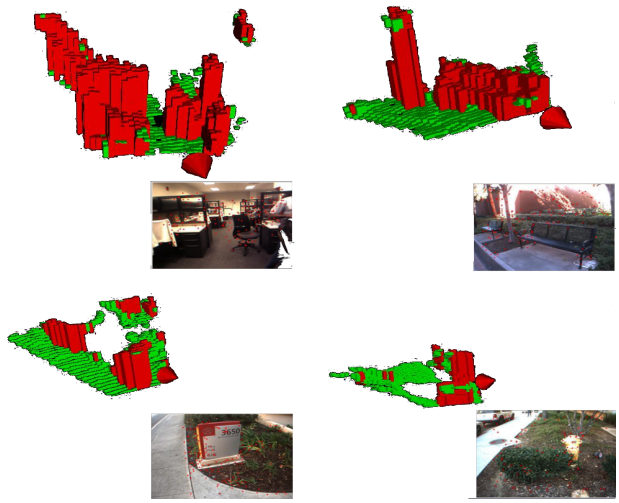
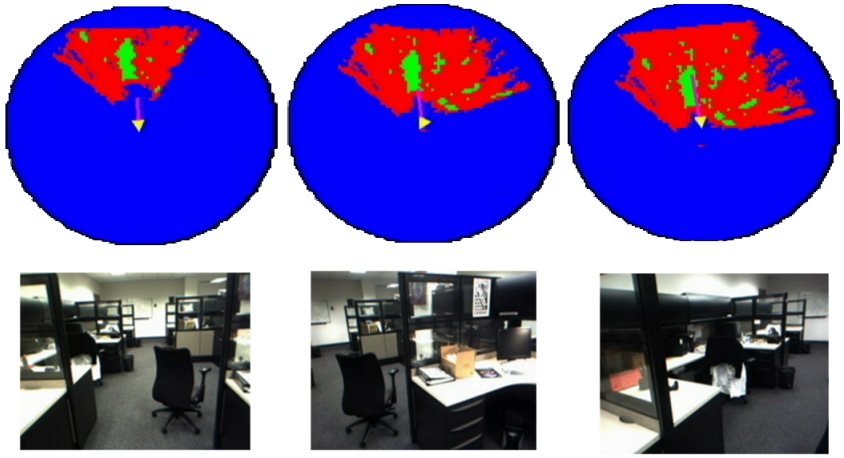


Figure 6. Multi-level surface patch models of 3D point clouds. Red regions show vertical surfaces and green regions indicate horizontal surfaces. The cone represents the current camera orientation.

(20-30 degrees/second). Faster camera motions are not explored as these lead to significant motion blur and small to no image overlap between two consecutive frames resulting in tracking failure. During SLAM execution, such cases are handled by resetting the tracking system and reinitializing the odometry. All the errors are of subdegree magnitude. The error and variance is largest for small rotations and this is because of the increased sensitivity to outliers and numerical inaccuracies. Medium scale motion represents the optimum trade-off between this scenario and that caused by mismatches and motion blur during faster rotations. Figure 4 shows the result for indoor mapping and figure 5 shows SLAM results for a large, outdoor environment. Over 2206 frames were processed for the result in figure 4(b), while the outdoor result involved 8000 frames. The SLAM and visual odometry modules, encompassing feature extraction, matching, motion estimation and RBPF FastSLAM with metric-topological representation currently operate at 10 Hz on a 3.39 GHz, Pentium dual core PC equipped with 3 GB of RAM. The dense 3D, textured maps are shown here only for providing a qualitative measure of the reconstruction accuracy. In actual operation, storing such large amounts of data is not recommended. Instead, as described earlier in Section 3.3, 2D elevation grids are constructed for the regions around the user's current location.

4.2. Traversability

Figure 6 shows multi-level planar matches for the 2D elevation grids constructed from dense 3D point clouds. The gap interval is set at 2 meters and any patch with a thickness greater than 30 cm is labeled as vertical. These results correspond to data acquired for one frame only. In figure 7, snapshots of system operation during different time instants



(a) (b) (c)

Figure 7. (Top) Top-down view of the traversability maps corresponding to different time instants (below) of system operation. Blue regions indicate unmapped areas in the subject's neighborhood, green regions are traversable and red regions are not. The yellow cone shows the current user position and head orientation. The magenta line attached to the cone is the direction of motion and predicted body orientation.

are shown. Note that between figure 7(a) and 7(b), there is not much translation motion as only scanning is being performed. Consequently, the estimated body orientation remains the same and therefore, the magenta line (motion direction) divides the scene accordingly into left and right halves. In (c), however, the motion direction has changed as the subject moves along that particular direction. The cuing system will now identify between left and right based on this updated line. This approach is able to integrate body and head motions into an accurate sense of ego-location, thereby providing the system with a sense of the user's orientation with respect to the scene.

4.3. Mobility Experiments[2]

We conducted some pilot experiments to evaluate the performance of our system in providing navigational assistance. Since we are still in the early experimental stages, we recruited normal sighted subjects and completely blindfolded them to simulate blindness. Our experiments were designed to study the effectiveness of tactile cuing in guiding subjects across a simple, indoor mobility course and compare performance with and without the popularly used white cane. To this end, we divided our sixteen subjects into four groups. Group 1 subjects were not blindfolded and asked to walk through the obstacle course to establish the most optimum trajectory and time performance. Remaining three groups had blindfolded participants, with group 2 allowed free use of both hands but no mobility aids, group 3 with a white cane only and group 4 with the tactile vest. None of the subjects had ever seen the obstacle course and each one went through 10 trials in a single session. An overhead surveillance camera system recorded the path taken

[2]Videos of the mobility sessions can be viewed at `http://www-scf.usc.edu/~vivekpra/`

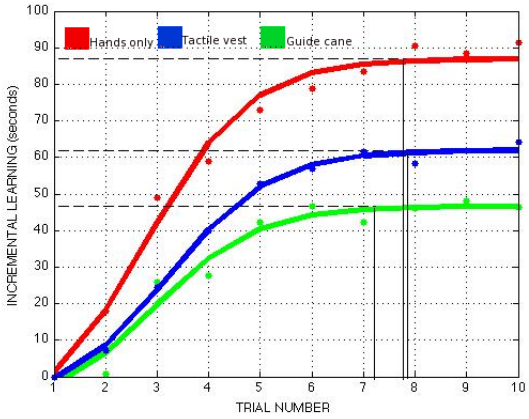


Figure 8. Plots of incremental learning over trials for Groups 2,3 and 4 with sigmoid fits. The horizontal asymptote indicates learning is achieved between the seventh and eight trials

and a custom software tracked the position of the subject at each time instant. Only the shoulder-mounted vibration motors were employed for these experiments, with subjects instructed to step sideways in the direction corresponding to the vibration. We performed two sets of experiments - one with manual guidance to establish a baseline and another with the computer vision system autonomously generating guidance cues.

In experiment 1, we manually generated the cues using a wireless remote control (see figure 1). The motivation here was to learn whether tactile stimuli can be interpreted for avoiding obstacles. Motors were activated in a manner such that the subject moved away from the nearest obstacle lying within 3 meters of his/her path. We expect our system to be able to autonomously generate such cues (as validated later in experiment 2). All our blindfolded subjects showed incremental improvement in time performance, indicated by

the typical sigmoid curves in figure 8. Subjects with full vision showed no learning and as expected, performed better by several orders of magnitude. Data from position tracking was converted into heatmaps to show distribution of time spent at various locations along the trajectory typical to each group (figure 9). Group 2 subjects performed the second best, as the strategy of wall-following enabled them to correlate wall texture (such as a window frame or junction) with obstacle location. This performance, however, came at the cost of a lot of collisions. Group 3 subjects fared the worst as the limited information provided by the walking cane lured them into trap situations, particularly around wider obstacles. Subjects with the tactile vest did significantly better than this group (ANOVA tests yielded a p-value of 0.0002) and their trajectory was closest to the optimal set by the sighted subjects. These experiments, therfore, established the need for a device that complements the guide cane and proved that effective navigation can be achieved by our strategy of obstacle avoidance.

In experiment 2, the computer vision system autonomously generated the cues and no manual intervention was attempted. The system built a traversability map and computed current user orientation and direction of motion (as shown in figure 7). The subjects were given cues to move such that the most continuous, traversable area from the current position lay in the path directly up ahead. The stereo camera was tethered to a laptop placed on a cart and owing to difficulties in moving the cart around obstacles, navigation tests were limited to only a small section of the whole course. However, we chose the section that posed the most difficulty for all groups. The goal here was to evaluate whether autonomous cuing by the algorithm can guide subjects along a safe path as obtained by manual cuing. Ideally, the system should generate cues such that subjects follow the path taken by those under manual guidance. The heatmap for this test is shown in figure 10. The realised trajectory is close to that in figure 9 for group 4 and demonstrates desired performance of the system. The accompanying video file has clips showing navigation by blindfolded subjects using autonomously generated cues by our system.

5. Conclusion and Future Work

We have presented an integrated framework which uses advanced computer vision techniques like SLAM and obstacle detection for application in a mobility assistance device for the visually impaired. Furthermore, we have also shown that coupled with a tactile cuing system, an intelligent set of algorithms can guide vision-deficient subjects through an obstacle course with minimal cognitive load. However, the navigation strategy adopted here can be considered too simplistic for real-world scenarios. Directing users towards the nearest open space might not always work as there might be an obstacle further down the path. We are working on developing an automated route-planning algorithm that is not just reactive, and in future, plan to implement a cuing strategy that tries to keep the user on such a computed path. The Bumblebee2 stereo camera is also very cumbersome to use and a tethered firewire connection imposes constraints during mobility experiments. Our future plans include incorporating a lightweight, wireless stereo camera that streams data to a server running our algorithms. This server will include Graphics Processing Units (GPU's) to further boost the frame rate of our system and free up CPU cycles for other processes. Finally, validation experiments with visually impaired subjects in challenging environments are planned. Apart from evaluating the performance, we hope to gain valuable insight into specific requirements for subjects who have been coping with visual dysfunction for a long period of time, that might not be obvious from data collected on normal sighted subjects who have had simulated vision loss for only a brief period.

References

[1] JVIB news service. demographics update: Use of white ("long") canes. *J. Vis. Impairment Blindness*, 88(1):4–5, 1994.

[2] JVIB news service. demographics update: Alternate estimate of the number of guide dog users. *J. Vis. Impairment Blindness*, 89(2):4–6, 1995.

[3] P. Arno and et al. Auditory coding of visual patterns for the blind. *Perception*, 28(8):1013–1029, 1999.

[4] J. L. Bath. The effect of preview constraint on perceptual motor behaviour and stress level in a mobility task. *Unpublished Doctoral Dissertation, University of Louisville*, 1979.

[5] J. M. Benjamin, N. A. Ali, and A. F. Schepis. A laser cane for the blind. *Proceedings of the San Diego Biomedical Symposium*, 12(53–57), 1973.

[6] J. Borenstein and I. Ulrich. The GuideCane-a computerized travel aid for the active guidance of blind pedestrians. *ICRA*, pages 1283–1288, 1997.

[7] D. D. Clark-Carter, A. D. Heyes, and C. I. Howarth. The effect of non-visual preview upon the walking speed of visually impaired people. *Ergonomics*, 29(12):1575–1581, 1986.

[8] G. Costa, A. Gusberti, J. P. Graffigna, M. Guzzo, and O. Nasisi. Mobility and orientation aid for blind persons using artificial vision. *J. Phys*, 2007.

[9] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse. Monoslam: Real-time single camera slam. *IEEE PAMI*, 29(6):1052–1067, 2007.

[10] A. Dodds, , D. Clark-Carter, and C. Howarth. The sonic path finder:an evaluation. *J. Vis. Impairment Blindness*, 78(5):206–207, 1984.

[11] C. Estrada, J. Neira, and J. D. Tardos. Hierarchical slam: Real-time accurate mapping of large environments. *IEEE Transactions on Robotics*, 21(4):588–596, 2005.

[12] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *IJCV*, 22(2):125–140, 1997.

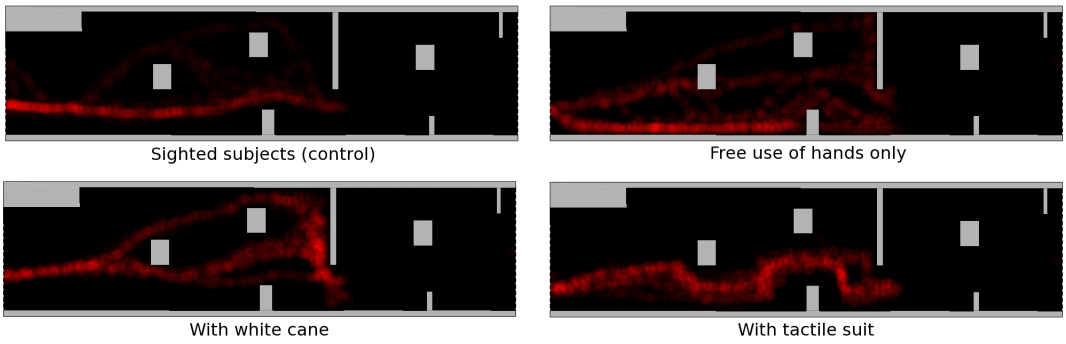


Figure 9. Experiment 1: Heatmaps displaying the averaged out trajectories for each group. Brighter red signifies more time spend in that location. Grey regions indicate obstacles/walls in the mobility course. Tracking data was not available for the complete obstacle course due to the limited field of view of the recording camera.

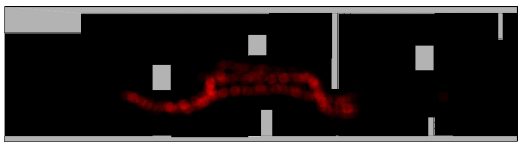


Figure 10. Experiment 2: Heatmap for autonomous steering

[13] S. M. Genensky, S. H. Berry, and T. H. Bikson. Visual environment adaptation problems of the partially sighted: Final report (cps-100-hew). *Center for the Partially Sighted, Santa Monica Medical Center*, 1979.

[14] R. G. Golledge, J. R. Marston, and C. M. Costanzo. Attitudes of visually impaired persons towards the use of public transportation. *J. Vis. Impairment Blindness*, (90):446–459, 1997.

[15] R. Haralick, C. Lee, K. Ottenberg, and M. Nolle. Analysis and solutions of the three point perspective pose estimation problem. *CVPR*, 1991.

[16] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[17] G. Klein and D. Murray. Parallel tracking and mapping for small AR workplaces. *ISMAR*, November 2007.

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

[19] D. Lowe. Object recognition from local scale-invariant features. *ICCV*, pages 1150–1157, 1999.

[20] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *IJCAI*, pages 1151–1156, 1981.

[21] L. Maohai, H. Bingrong, and L. Ronghua. Novel mobile robot Simultaneous Localization and Mapping using Rao-Blackwellised Particle Filter. *International journal of Advanced Robotics Systems*, 3(3):231–238, 2006.

[22] N. Molton and et al. Robotic sensing for the partially sighted. *Robotics and Autonomous Systems*, 26:185–201, 1999.

[23] M. Montemerlo, S.Thrun, D.Koller, and B.Wegbreit. FastSLAM: A factored solution to the simultaneous localization and mapping problem. *Proceedings of the AAAI National Conference on Artificial Intelligence*, pages 593–598, 2002.

[24] M. Montemerlo, S. Thrun, D. Koller, and B. Weigbreit. FastSLAM 2.0: An improved particle filtering algorithm for simultaneous localization and mapping that provably converges. *IJCAI*, pages 1151–1156, 2003.

[25] K. Murphy. Bayesian map learning in dynamic environments. *NIPS*, pages 1015–1021, 1999.

[26] M. Pollefeys, D. Nister, and et al. Detailed real-time urban 3d reconstruction from video. *IJCV*, 2007.

[27] V. Pradeep, G. Medioni, and J. D. Weiland. Visual loop closing using multi-resolution sift grids in metric-topological slam. *CVPR*, pages 1438–1445, 2009.

[28] J. M. Saez and F. Escolano. Stereo-based aerial obstacle detection for the visually impaired. *CVAVI, ECCV*, 2008.

[29] J. M. Saez, F. Escolano, and A. Penalver. First steps towards stereo-based 6DOF SLAM for the visually impaired. *CVPR*, 2:23, 2005.

[30] R. Sim, P. Elinas, and J. J. Little. A study of Rao-Blackwellised Particle Filter for Efficient and Accurate Vision-Based SLAM. *IJCV*, 74(3):303–318, January 2007.

[31] N. Sunderhauf, K. Konolige, and S. Lacroix. *Visual odometry using sparse bundle adjustment on an autonomous outdoor vehicle*. Springer-Verlag, Jan 2005.

[32] T. Triebel, P. Pfaff, and W. Burgard. Multi-level surface maps for outdoor terrain mapping and loop-closing. *IROS*, (2276–2282), 2006.

[33] D. Yuan and R. Manduchi. Dynamic Environment Exploration Using a Virtual White Cane. *CVPR*, 2005.