2017 IEEE International Conference on Robotics and Automation (ICRA)
Singapore, May 29 - June 3, 2017

Enabling Independent Navigation for Visually Impaired People through a Wearable Vision-Based Feedback System

Hsueh-Cheng Wang*,1,2, Robert K. Katzschmann*,1, Santani Teng[1], Brandon Araki[1], Laura Giarré[3], Daniela Rus[1]

Abstract— **This work introduces a wearable system to provide situational awareness for blind and visually impaired people. The system includes a camera, an embedded computer and a haptic device to provide feedback when an obstacle is detected. The system uses techniques from computer vision and motion planning to (1) identify walkable space; (2) plan step-by-step a safe motion trajectory in the space, and (3) recognize and locate certain types of objects, for example the location of an empty chair. These descriptions are communicated to the person wearing the device through vibrations. We present results from user studies with low- and high-level tasks, including walking through a maze without collisions, locating a chair, and walking through a crowded environment while avoiding people.**

I. Introduction

We wish to design a system that assists local navigation for blind and visually impaired people by providing (1) generic feedback on obstacles and (2) descriptions of identified useful objects.

According to the World Health Organization, approximately 285 million people worldwide are blind or visually impaired (BVI) [1], with one person losing their vision every minute. Visual impairment poses a number of challenges related to mobility: BVI people find it difficult to map the environment and move collision-free within it, even with the help of canes or guide dogs. BVI individuals are thus significantly less willing to travel independently [2]. Independence in daily activities for BVI people requires tools to enable safe navigation. Cane users naturally sample their immediate environment with physical contact, but in some situations it can be desirable to avoid contact if possible, e.g. when navigating among pedestrians or in quiet environments.

Computer vision can be used to enable *purposeful navigation* and *object identification*. Purposeful navigation can be defined as guided motion through space toward a desired target while avoiding obstacles. The challenge is to robustly process the sensor feedback from a wearable system and to intuitively map the feedback to directions and semantic descriptions of the environment that meet the needs and goals of a BVI person. A natural option would be to produce an audio stream that describes the space. However, audio feedback becomes indistinguishable in noisy environments and can interfere with the perception of auditory environmental cues on which BVI people depend for situational awareness. Previous work on blind navigation proposed systems that used backpacks to carry the computing units and relied primarily on audible feedback. BVI users considered the size and design of these solutions as too bulky and obtrusive and the audio feedback as undesirable and restrictive [3]–[5]. BVI users have expressed a strong preference for a miniaturized design packaged in an unobtrusive solution. This limits the choices in sensing, computation, and feedback mechanisms.

Fig. 1: The wearable system enables a blind user to navigate through a maze while avoiding collisions. Top left: Depth camera and embedded computer with battery. Top right: A blind user navigates a maze using the haptic feedback from vibration motors; the walking trajectory is recorded by a motion capture system. Bottom left: on-board RGB image with an obstacle on the right-hand side. Bottom right: the five vibration motors are visualized as blue (vibration off) and red (vibration on) dots.

In this paper we develop an end-to-end wearable solution to enable purposeful navigation that uses a portable depth-camera to extract information about the local state of the world, such as the range and direction of obstacles, the direction of free space, and the identity and location of target objects within a space (Fig. 1). The object detection and recognition algorithm of the system assists the user in navigating to objects like an empty chair. The solution provides navigational cues through haptic feedback from vibration motors worn around the torso.

Specifically, this paper contributes:

1) A wearable and unobtrusive vision-based system providing haptic feedback to enable purposeful navigation for BVI people;
2) A real-time algorithm for segmenting the free space

*H. Wang and R. Katzschmann contributed equally to this work. The authors are with the [1]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 32 Vassar St. Cambridge, MA 02139, USA, [2]Department of Electrical and Computer Engineering, National Chiao Tung University, Taiwan, and [3]DEIM, Università di Palermo, Italy. {hchengwang, rkk, santani, araki, giarre, rus}@csail.mit.edu

and mapping it to free-space motion instructions;
3) User studies demonstrating the system's capabilities of providing guided navigation through free space, obstacle avoidance, and guidance to specific types of target objects such as empty chairs.

II. Related Work

Several recent papers have applied robotic techniques to assistive safe navigation for people. The systems typically include sensors, computation, and feedback components, and provide the functions of either global frame localization [6], local frame obstacle detection/avoidance [7], or both [8]–[10].

A cane-mounted prototype estimates the pose of a user walking inside a building with a known map [6]. The sensing package includes a 3-axis gyroscope, a 2D laser scanner, and a foot-mounted pedometer. The sensing data is transferred wirelessly to a portable laptop computer for processing. Although the system provides no user feedback, it is a step towards indoor global frame localization in combination with local obstacle detection through a white cane. In a different study, the data from a stationary laser scanner steers a user through a room using vibration motors [7]. A planar laser scanner is used to detect obstacles and trace moving pedestrians in the environment. The data is processed on a controller PC and mapped to eight vibration motors through various actuation patterns. The system focuses on local obstacle avoidance and takes moving objects into account. Users in that study could respond to the vibrations within 1.9 s to avoid obstacles in the environment. Another system [8], [11] considers both global frame localization and local frame navigation and uses a tactile feedback vest to deliver navigation cues. An RGB-D camera provides a 3D point cloud and an IMU provides initial orientation. The point cloud is downsampled into a representation of a 3D voxel grid map and added to the global frame using the estimated visual odometry algorithm [12]. An occupancy map is established, and the D^*-Lite planning algorithm [13] is used to generate four navigation cues, including "straight," "stop and scan," "turn left," and "turn right." All computations were done on a laptop computer carried in a backpack. In addition to wearable systems, robotic walkers [9], [14] have been developed to provide walking assistance to blind users. A robotic walker [9] estimates egomotion and detects obstacles such as curbs, staircases and holes by two planar laser range finders, continuously tilted by a servo motor to obtain the 3D point cloud for indoor and outdoor environments. A controller module issues directions such as "straight", "left", "right." The human users' reaction times to four different vibration signals are $0.87\,\mathrm{s}$ on average. The authors suggest that their solution requires less traveling time and distance to reach the goal. The perception and computation components for self-driving cars is also used to develop a driver assistance system for blind people to independently operate an automobile [10]. The system includes passive and adaptive controlling software and non-visual vibrotactile motors positioned in gloves, beneath the thighs, and on the back of the driver. There have been other attempts of helping BVI communities using RGB-D sensors [15]–[17]. We refer the reader to [4], [18], [19] for other aspects of safe navigation.

Our system builds on the work of Stixel World [20]–[23]. Stixel World represents the scene using a few upright objects on the ground plane. The computation includes dividing the input image into a set of vertical columns called 'sticks', and searching for regions that can be immediately reached without collision. There have been a few variations, including allowing multiple segments along columns and combining nearby segments for better and more meaningful segmentation [22]. Recently [23] proposed a 4-layer compact representation for street view, encoding semantic classes including ground, pedestrians, vehicles, buildings, and sky.

In indoor environments, useful objects such as chairs and tables can be detected by their 3D geometry. Under the Manhattan world assumption [24], detection of the ground plane from Stixel World allows the efficient localization of knee-or waist-height objects, enabling real-time detection of a certain category of objects on mobile platforms.

Our system builds on this prior work and contributes:
1) A miniaturized device that estimates the local state of the world and divides point cloud data into free space and obstacles. This enables a blind user to walk through complex environments such as mazes or hallways without collisions. To our knowledge, no previously developed wearable system achieves this functionality.
2) A system that uses depth information from a moving camera to provide on-board object detection in real-time. This can be used to localize a target object, for example an empty chair. Prior work, such as Sliding Shape [25], assumed datasets with static depth and unconstrained computational resources.
3) Unobtrusive haptic feedback is given to the user through a belt with vibration motors while not overloading the user's sensory capabilities.
4) A user study showing BVI people able to complete maze-navigation, chair-finding and path-following tasks. Most experiments carried out in previous work were only evaluated using blindfolded sighted users.

III. System Description

The system is designed in a serial architecture consisting of three stages: perception, planning, and human-robot interaction (Fig. 2). In the following, we first introduce the system, including the requirements and assumptions made. Next, we detail each stage of the system architecture and its subcomponents.

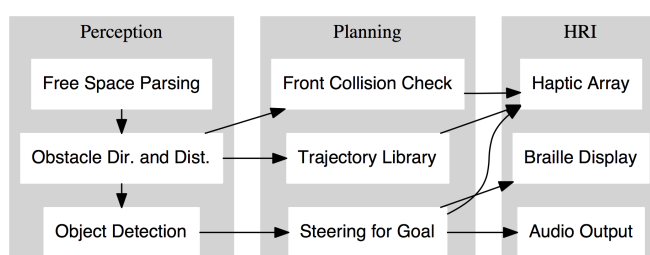


Fig. 2: System overview.

A. System Requirements

The user requirements are:

1) A miniaturized and compact package that is portable and wearable;
2) Socially unobtrusive form;
3) Low cost;
4) Long battery life through low power consumption;
5) Intuitive feedback to the user, requiring minimum training time.

The functional requirements are

1) High frame rate, low latency, and high reliability;
2) Efficient feature and environment recognition;
3) On-board computation over video streams.

These functional requirements can be achieved with many sensors and extensive computation. However, there are trade-offs between sensing, computation, and system usability. While users require compactness and acceptable battery life, minimizing cost and power consumption at the cost of increased latency and less sophisticated recognition schemes is detrimental to a full-fledged user experience. Thus, in this work, we propose a system that balances these requirements.

B. Assumptions

We assume an indoor environment, where the main scene surfaces, i.e. the ground and the walls, follow the Manhattan world assumption [24]. The objects of interest, such as chairs, tables and other furniture, are adjacent to the ground plane. We use pre-trained models for the objects we want to recognize, based on their distinct geometry. The object classes include "chair," "table," "stair up," "stair down," and "wall". We assume that the ground plane can initially be observed and that the height and rotation of the sensor remain constant when the ground plane cannot be seen. All computation is performed on board with real-time detection and feedback.

C. System Overview

The system architecture and key capabilities of the system are shown in Fig. 2. Environment sensing is achieved through a structured light camera, which provides a point cloud that represents the measured depth of the field of view. This technology works best indoors and on non-reflective, non-absorptive surfaces, but it suffers from exposure to direct sunlight when facing windows or other transparent structures. This depth sensor technology is best suited for detecting unstructured walls without textures. The depth-sensor and the embedded computer are disguised as a fake SLR camera within a leather case.

The hardware overview of all the system components is given in Fig. 3. The components communicate using the LCM (Lightweight Communications and Marshaling) package [26]. The algorithms are implemented in C++ with OpenCV and the Point Cloud Library (PCL).

D. Perception

Independent navigation for a BVI person requires the indication of safe walking regions, drop-offs, ascents or descents, trip hazards, obstacles, overhangs, or boundaries, whereas semantic descriptions improve situational awareness

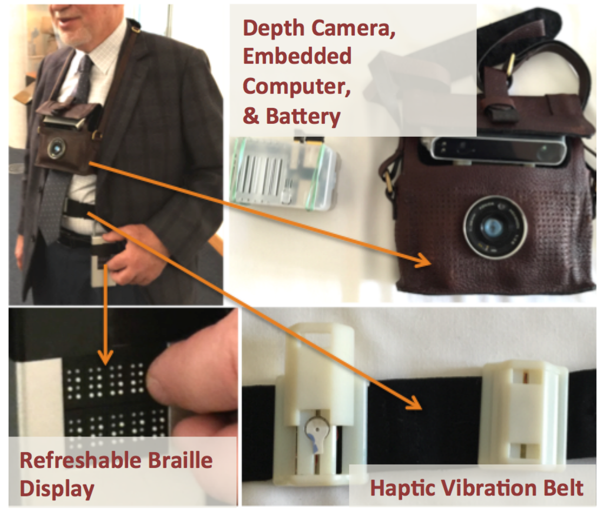


Fig. 3: Hardware overview.

of the surroundings for everyday activities. This perceptual information is sufficient for basic planning such as avoiding frontal collisions and providing simple motion trajectories. Object recognition is needed for steering the user toward a target object.

The perception capabilities of our system consist of two components: (1) finding the free space, detecting obstacles and corresponding distances, and (2) recognizing the type of object.

The following describes our solutions to these two aspects of perception:

1) Free Space Parsing:

a) Challenges: The inputs from depth sensors and cameras are usually subject to pitch and roll in a walking scenario. This is undesired for detecting the ground height and walkable free space. The typical solution is to integrate IMU observations along with vision inputs. This solution fails to obtain correct ground height and ground-to-image transformation when a user stands on a ramp, in front of a staircase, or next to a drop-off.

b) Algorithm: We use a variant of the Stixel World [20] to compute the free space. The parsing of the free space is shown in Algorithm 1.

The surface normal step of Algorithm 1 is estimated using the integral images implementation in the PCL. The system

Algorithm 1 Free Space Parsing

procedure FREESPACEPARSE(**C**)
Input: Point cloud **C**
Estimate surface normal N
Find ground plane G from the Stixel World, and estimate normal vector n_g of the ground plane
Rotate **C** based on n_g, obtain ground height h_g to translate the point cloud **C**
Compute ground-to-image frame transformation T_g
Find the occupancy grid, and extract the free space and distances d_o to objects/obstacles
Output: Free, walkable space

can trace ground height changes when the user approaches a stair or an object such as a desk. The point cloud $\mathbf{C}$ is then transformed according to the estimated ground and an occupancy map is obtained (see Fig. 4).

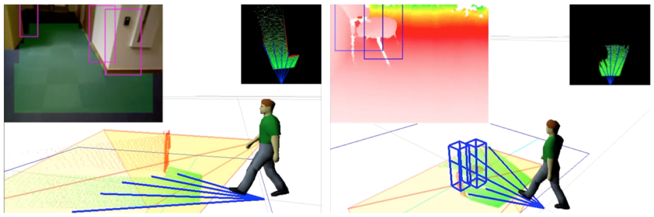


Fig. 4: Left: free space parsing. Right: detecting chairs. For each sub-figure, we show the point cloud of ground plane as green regions and each blue line representing the free space in each direction. Top-left: RGB or depth visualization, where pink rectangles are detected obstacles, and blue rectangles are region proposals of objects. Top-right shows the projected occupancy grid map.

c) Implementation: We choose to use a resolution of 320×240 at 10 frames per second for the point cloud $\mathbf{C}$. This resolution allows effective ranges of about 3 meters when the camera is positioned on the user's chest facing forward and angled about $45°$ downward from the viewpoint of the user. The obstacle directions and distances can be used in full scale sliding window search to achieve efficient computation for object recognition, which is described in the next section.

2) Object Recognition:

a) Challenges: On the hardware side, the sensor resolution is limited by the requirement for on-board computation on the wearable platform. Furthermore, the images observed by the moving camera will generally be subject to motion or out of focus blur, making it challenging for object recognition from RGB imagery. It is also known that sliding window classification is too computationally expensive for real-time implementation due to the large search space.

b) Algorithm: We use a depth-based method that detects the ground plane and accurately transforms the point cloud. We search for the set of "known" objects that contain horizontal surfaces. Algorithm 2 shows the object detection algorithm.

Algorithm 2 Object Recognition using Depth

procedure OBJECTRECOGNITION(W, b)
Input: Outputs of Algorithm 1, weights matrix W, and bias vector b
  From T_g and d_o, generate region proposals R using bounding cubes
  for each region proposal r_i in R **do**
    Remove ground surfaces
    Obtain and normalize feature vector x_i
    Compute score function by $Wx_i + b$
    Obtain the object class with the highest score
Output: Extract object class from all region proposals R

c) Implementation: An object region proposal is generated according to the ground-to-image frame transformation T_g. For example, the region proposal for a chair is a bounding cube of size $0.3 \times 0.3 \times 1.0\,\text{m}$, see right image in Fig. 4. The cropped depth map or RGB image patches are further processed for object recognition. We use a feature vector of 2 (vertical or horizontal) $\times$ 18 (height) and a linear classifier for the object classes. The weights of matrix W and bias vector b are pre-trained from training data, and the dimensions are 36×5 and 1×5, respectively.

d) Object Scope: The target object classes include common objects that appear in engineered indoor environments such as tables, chairs, lounges, and other types of furniture. Although the current implementation does not recognize smaller objects such as cups and pencils, adding them to the system is a possible extension. For the purpose of navigation, the key is to obtain from the point cloud data of a moving camera the ground surfaces so that the physical heights of vertical and horizontal surfaces result in reliable feature vectors.

E. Planning

Purposeful navigation requires task-relevant feedback from free-space parsing and object recognition processes. Planning includes the following behaviors [27]: (1) front collision checks, (2) a trajectory library for possible walkable paths, and (3) steering toward a goal. Multiple behaviors are activated simultaneously in various tasks: the task of finding a chair applies to behaviors (1) and (3), and the task of tracing a tiled path uses behaviors (1) and (2) for planning free, walkable space momentarily while avoiding obstacles. The execution of each behavior is specific to different human-robot interfaces, which are described in detail in the following.

F. Human Interaction

Human interaction with our system can either occur through a haptic array that vibrates at a suitable position on the user's body or through a braille display that allows the user to feel the occupancy grid or object description with his or her fingers. The feedback should avoid overwhelming the users' sensory and cognitive capacities, but provide enough information to steer the user around obstacles or to a goal.

The communication between the computing unit and the receiving unit occurs via wireless transmission. We implemented it using bluetooth communication. The signal receiver uses a microcontroller to parse the information and actuate the haptic motors or braille display. The audio output is synthesized by a text-to-speech engine.

1) Haptic Array: The haptic array is an elastic belt that can be worn around the abdomen. It consists of five linear resonating motors mounted within clips attached to the belt (see Fig 3). The motors deliver feedback to the user through pulses of varying strength and frequency. The haptic feedback is tuned to ensure that the user can feel it: the motors are placed with wide spacing of more than 10cm at sensitive positions around the abdomen and pulsed with high strength and frequency. The design of the haptic feedback for an intuitive user experience is not the focus of this paper, but the subject of an unpublished study.

For both experimental tasks (maze navigation and chair-finding), three of the five motors (middle, left, and right) were activated. Users were instructed to interpret a vibration as the presence of a detected obstacle. In the maze-navigation task, the motors signaled when an obstacle was detected within 1 m; the left, middle, and right motors corresponded to obstacles positioned $-30°$, $0°$, and $30°$ from the sagittal plane of the user, respectively. A lack of vibration indicated free space.

In the chair-finding task, the vibration of the front motor indicated the proximity of the chair or another obstacle. The left or right motors vibrated only when an empty chair was detected. A vibration by the left motor indicates that the chair is towards the left side of the user, and vice versa. The maximum amount of information sent to a user was 15 bits per second: three signaled directions triggered at a rate of five frames per second.

2) Braille Display: The braille display by Metec AG is a handheld module with a 10-cell braille display on the top surface. Each cell contains 4 rows and 2 columns of single pins. The arrangement of the cells leads to a total of 8 rows and 10 columns of single pins (see Fig. 3).

In the mode of obstacle distance description, each of the 10 columns encodes a different direction away from the user towards the obstacles. The rows of each column indicate how far away an obstacle is along that particular direction. The range distances are matched to the settings in the haptic belt. For object recognition, we encode four different object types using one braille symbol per object type: *o* for obstacle, *c* for chair, *t* for table, and *a* space for free space. The first row of cells encodes long distance ($>$ 1 m), and the second row encodes short distance.

IV. User Studies

Since a white cane is the most commonly used mobility aid for obstacle detection and local navigation, it is important to identify the practical relevance of the proposed system when used instead of or in concert with a cane. To this end, we surveyed respondents on the importance of some common mobility tasks, and the utility of a cane in performing them. All 15 respondents were blind (9 congenitally) and used a cane as a daily mobility aid.

The survey results (Table I) indicate that a cane can serve well in tracing along walls, curbs, tactile tiles, or passing entrances. Nevertheless, canes were rated of less utility in finding an empty chair or avoiding contact with pedestrians in crowded sidewalks and corridors or in other crowded environments. Some users also reported being less willing to enter crowded environments or social events with a cane alone. We further observed during experiments that sighted people rather than the cane users avoid the contact during encounters.

We measured the effectiveness of the system for (a) purposeful navigation of simple and complex paths, and (b) target object localization in a cluttered environment. We have conducted a total of 100 hours of experiments with the system and traversed in aggregation several kilometers of indoor terrain using this system. In order to evaluate the utility of the system in various scenarios for the BVI community, we conducted a user study with 11 blind users who had no prior exposure to the system. Each session lasted 2 hours, with some users returning for multiple sessions. Each participant was fitted with an adjustable "pendant" hung around the neck to hold the depth sensor and embedded computer, and an adjustable haptic array belt around a comfortable location on the upper abdomen. Reflective markers mounted on a helmet were tracked by a ceiling-mounted motion capture system (Vicon) to monitor time to completion, average walking speed, collisions, and any failures. We also assessed the users' subjective experience with a questionnaire.

TABLE I: Survey of common mobility tasks based on two ratings: Importance (Imp.) of task for daily life (5 = very important, 1 = not important) and utility (Uti.) of a cane to perform the task (5 = greatest utility, 1 = no utility). All survey results are gathered from 15 respondents except task 5 (10 respondents).

Tasks	Importance	Utility
1. Tracing along walls, curbs, stairs, and other clues indoors and outdoors	4.67	4.87
2. Being able to use tactile tiles	3.8	4.93
3. Passing entrance, door, or narrow gates	4.6	4.6
4. Finding empty chairs in a crowded public area	5	1.29
5. Avoiding contacts with other pedestrians in crowded sidewalks and corridors	4.9	2.8
6. Avoiding contacts with other pedestrians in crowded environments	4.8	2.53

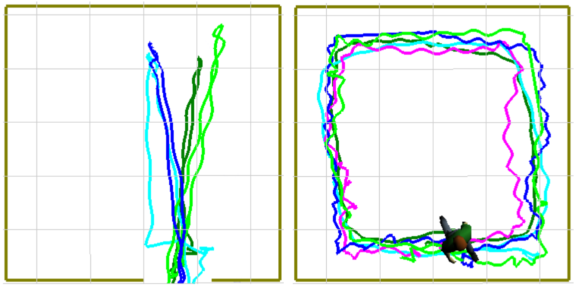
The users' first task was to detect an obstacle in their path and to shoreline along a wall. This task served as the user training. The second task was the navigation test, conducted in a series of three challenging maze-like enclosures, containing 8-11 turns, and suitable for tracking with a motion-capture system. The users did not know the shape of the maze and had to find their way through the environment using the system or their white cane. In the third task, we placed several chairs in an open space defined by the walls of the enclosure. The space contained empty chairs, occupied chairs, and recycling bins as clutter objects. The task was to identify and navigate to an empty chair. Finally, we tested system performance with blind users following a tiled path in a crowded real-world environment.

A. Obstacle Detection and Shorelining

Users walked forward into an open space until the system signaled an upcoming obstacle, then reached forward to physically "calibrate" their sense of distance from the wall to the output of the system before returning to their point of origin, shown in Fig. 5a. Next, users shorelined along the inner perimeter of the enclosure, maintaining constant distance and orientation with respect to the wall, shown in Fig. 5b. These two tasks required that the users utilize both the range and azimuth information provided by the system. No white cane was used. Both trajectory visualizations show that the users completed the tasks successfully without collisions.

B. Maze-Navigation Task

The practical relevance of the maze trials is to enable independent navigation in crowded environments, where cane contacts are undesired. Such environments include



(a) Obstacle Detection (b) Shorelining

Fig. 5: Initial BVI user training tasks of (a) obstacle detection and (b) shorelining. The walking trajectories are recorded by a motion capture system and color-coded for each user.

restaurants and other social events, in which blind users often rely on human assistance rather than using a cane independently.

The maze environments were $5.2\,\mathrm{m} \times 5.2\,\mathrm{m}$ in size, with a wall height of $124\,\mathrm{cm}$ (see Fig. 1). The paths were designed to test safe traversal of an unpredictable path and consisted of 8 to 11 turns, depending on the maze configuration. There were no configurations with T-junctions. Each of the five users traversed three maze configurations with similar topology during each session (Fig. 6). For each maze type, we recorded a reference trajectory traversed by a sighted user. In each session, users traversed each maze configuration using a white cane and using our system. Users were asked to traverse each maze path twice, once in each direction, not repeating any path/direction combination. Since some users were invited for multiple sessions, we conducted a total of 9 sessions. This amounted to a total of 32 maze traversals. For each trial, we measured the time to completion and the number of minor and major collisions. Table II shows the results.

TABLE II: User performance in a maze navigation task. The speed was calculated by the trajectory length divided by the time to completion. For comparision, a sighted user walked each maze at an average speed of $0.7\,\mathrm{m/s}$. Major collisions were defined as wall collisions during forward movements without the user responding to the vibration signals. Minor collisions were defined as other contacts that did not affect users' forward movements, typically due to the limited field of view of the depth camera. Both the speed and number of collisions were averaged across trials.

User #	B1	B2	B3	B4	B5
White Cane					
No. of Trials	1	2	1	2	-
Speed (m/s)	0.23	0.30	0.34	0.26	-
System: 1st Session					
No. of Trials	4	3	6	5	2
Speed (m/s)	0.15	0.09	0.23	0.13	0.10
Major Collision	0.50	0.33	0.33	0.40	0.00
Minor Collision	1.25	1.33	0.33	0.20	0.50
System: 2+ Sessions					
No. of Trials	6	6	-	-	-
Speed (m/s)	0.17	0.12	-	-	-
Major Collision	0.17	0.00	-	-	-
Minor Collision	1.83	0.33	-	-	-

Users B1 and B2, who participated in multiple sessions, showed an increase in average speed and decrease in major collisions. The walking paths of the BVI users using the system contained moderate serrations, because the users stepped and turned to better understand the topology of the space around them. Users walked faster when using a cane than when using our system, due to the users' greater familiarity with the cane and the system's requirement to turn one's body to assess a walkable free space. Nevertheless, our system provides rich information enabling collision-free navigation independently of a cane.

C. Chair-Finding Task

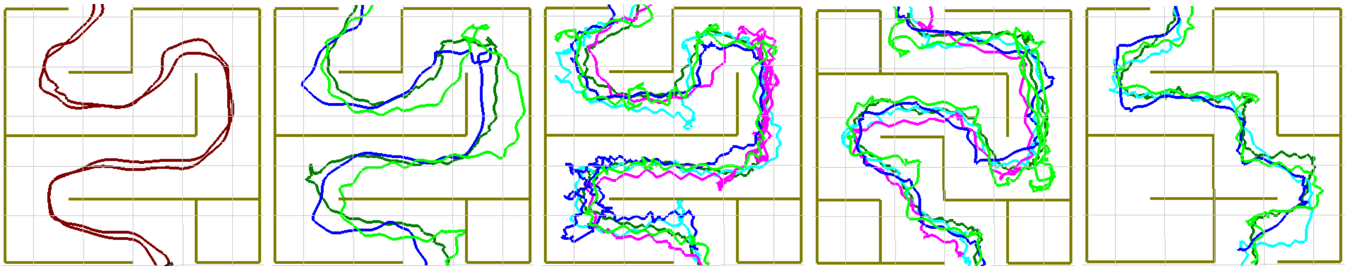
The chair-finding task was intended to measure the system's ability to help a user find a specific target in an environment. In this case, the target was an empty chair, with non-target distractors consisting of an occupied chair or a tall recycling bin. The task serves as a representative example of a general object detection task.

Users were instructed to locate the empty chair among multiple distractors in an environment bounded by cardboard partitions, as shown in Fig. 7. The task combined frontal obstacle/free-space discrimination with lateral target object identification. The left and right motor vibrations denoted the direction of an empty chair. For example, a user could start with a random walk until receiving a signal from either the left or right motor. The user then turned towards the side of that motor until it stopped vibrating. Turning too far in a given direction would activate the opposite-side motor. In this manner the user walked forward while adjusting his or her walking direction using the left and right signals. The front motor signaled the proximity of the chair or of another obstacle.

We conducted with 4 users 20 chair-finding trials, of which 18 were successful. The average time to find the target chair was 23.8 seconds. In each trial there was one empty as the target, and 0 to 3 non-chair distractors. The supplemental video demonstrates several chair-finding trials.

Furthermore, we conducted an extensive systematic test with one blind user over 53 trials in multiple sessions. The user performed the chair-finding task in the environment shown in Fig. 7, with one target chair and two distractors. At each trial, we randomly varied the method used to find the target (cane only, system only, or cane and system together). We also varied the position of the target, both absolute and also relative to the distractors, at each trial. Thus, performance differences between conditions should not be due to learning an absolute or relative target position, or practice effects over time. We measured the time to completion and contacts, see Table III.

Overall times to completion were comparable across conditions, but the user incurred significantly fewer non-target contacts and accidental collisions when using the system, either by itself or together with a cane, compared to using the cane only. Comparing the use of a cane only with the use of the system only suggests that even with relatively little practice, the system is better at facilitating navigation to a target among distractors while reducing contacts.



(a) Maze 1: Sighted User (b) Maze 1: White Cane (c) Maze 1: System Only (d) Maze 2: System (e) Maze 3: System

Fig. 6: Maze environments overlaid with sample walking trajectories, recorded by a motion capture system. (a) Maze 1 with 11 turns, the purple trajectories of two sighted users are shown as reference. (b) Maze 1 with the trajectories of users B1, B2, and B4 using a white cane. (c) Maze 1 with the trajectories of B1 to B5 using the proposed system. (d) Maze 2 with 11 turns, showing the trajectories of B1 to B5 using the proposed system. (e) Maze 3 with 8 turns, showing the trajectories of B1 to B4 using the proposed system.

Fig. 7: Object localization and discrimination task. The target was an empty chair, with non-target distractors consisting of an occupied chair or a recycling bin.

TABLE III: User performance in the chair-finding task. Non-target contacts are cane or body contacts with objects other than the target. Accidential collisions are inadvertent contacts, a subset of non-target contacts. The asterisks indicate that average non-target contacts for the cane-only condition were significantly greater than for either of the two conditions using the system (rank sum tests, p<.01).

Task	**Cane**	**System**	**Cane+System**
Total Trials	18	18	17
Avg. Time (s)	21.1	30.3	21.8
Avg. Non-Target Contacts	1.6*	0.3*	0.3*
Avg. Accidential Collisions	0.2	0.2	0.06

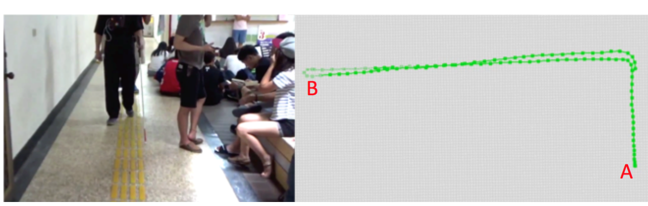


Fig. 8: Path-following experiment in a crowded environment. Left: User following tactile paving with cane or cane plus system. Right: User trajectory from a to B and back for a total length of 90 m.

D. *Navigation through Realistic Environment*

To test the system's performance in a real-world environment, we asked three blind users to follow a path down the center of a long hallway at Tamkang University, Taiwan. The path was in the center of a crowded hallway and consisted of tactile paving tiles (Fig. 8). In addition to other pedestrians, a stationary experimenter occluded the path at three locations. Thus, navigating the path required navigation in the presence of both static and dynamic obstacles. Due to local restrictions on human subject testing in real-world environments, we were not allowed to test the proposed system without a cane. We instructed users to avoid any contact with obstacles along the path, including with their canes, and measured collisions as the number of cane or body contacts. In a practice round with cane and cane plus system, users contacted nearly every static obstacle they encountered, but drastically reduced cane contacts in a third round of cane plus system (Table IV). The increased success in avoiding contacts indicates the efficacy of the system. A $\sim$12% increase in walking time is due to less familiarity and training with the system.

We further tested the real-world performance of the system at the MIT laboratory. We asked one experienced user to navigate the length of several indoor hallways without a cane, thus using only the system to determine the presence and direction of obstacles and navigable space. The user was supervised by the experimenters throughout the trials. The initial test lasted about 2 min.; the second test involved several hallway turns and lasted about 5 min. The user was able to avoid walls, corners, shelves, a knee-height bench, and a free-hanging sign with only one error in each of the two trials (a shoulder contact and one misjudgment of a doorway).

TABLE IV: User performance in tracing a tiled path. Users utilize the system to avoid cane contacts with obstacles.

User #	B6	B7	B8
Cane Only			
No. of Trials	2	2	2
Average Time (s)	65	71	62
Collision / Obstacles	1 (6/6)	1 (6/6)	1 (4/4)
Cane + Sys: Practice			
No. of Trials	2	2	2
Average Time (s)	65	68	62
Contacts	1 (6/6)	1 (6/6)	0.75 (3/4)
Cane + Sys: Testing			
No. of Trials	4	4	2
Average Time (s)	73	80	65
Contacts	**0.17** (2/12)	**0.17** (2/12)	**0** (0/4)

V. CONCLUSION

We presented a real-time wearable system, which includes a camera, an embedded computer and a belt with embedded vibration motors that provides vibration feedback to signal obstacles to its users. Using depth information from a camera, the system distinguishes walkable free space from obstacles and can identify a few important types of objects such as the location of a chair. These descriptions of the surroundings are communicated to the person wearing the device and translated into safe navigation directions. We conducted user studies with blind users who completed tasks requiring shorelining; maze traversal without collisions or the assistance of a cane; localizing an unoccupied chair; and path tracing while avoiding obstacles within a crowded environment. For every task, the use of the system decreased the number of collisions compared to the use of a cane only.

Surveys administered on cane use indicated that some important tasks, like finding a target object and avoiding collisions in crowded areas, were difficult to perform with a cane alone, revealing a functional gap that the proposed system can fill. In post-testing questionnaires, users reported rapid acclimation, readily interpretable signals, and moderate comfort with the haptic belt interface.

Haptic devices provide a high frame rate and low-latency feedback, which is desirable for navigation tasks, including walking through a maze or finding an empty chair. Braille displays offer richer high-level feedback in a discrete way, but with longer reaction times due to sweeping of the fingers on the braille cells. Audio feedback was deemed undesirable given the low refresh rate and long latency, as well as potential obstruction of other sounds, a major source of environmental cues for blind people.

Future enhancements include outdoor capability using a stereo vision camera, a larger library of useful object recognition, and collaborative work [28] utilizing a low-power energy-efficient implementation of the surface normal estimation and dynamic frame skipping.

ACKNOWLEDGMENTS

This work was inspired by the vision of the late Professor Seth Teller. We are grateful to him for his leadership in this important field. We are also thankful for the help by Jordan Allspaw, Jih-Shi Chen, Yi-Wei Huang, Nicole Jimenez, Ni-Ching Lin, Paul Parravano, and Grace Teo. The research was supported by the National Science Foundation (grant IIS1226883), the Andrea Bocelli Foundation, the MIT Vice President for Research and MOST, Taiwan (grant 105-2218-E-009-015).

REFERENCES

[1] World Health Organization. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs282/en/

[2] R. Golledge, J. Marston, and C. Costanzo, "Attitudes of visually impaired persons towards the use of public transportation," *J. Vis. Impairment Blindness*, vol. 90, pp. 446–459, 1997.

[3] N. Kanwal, E. Bostanci, K. Currie, and A. F. Clark, "A navigation system for the visually impaired: A fusion of vision and depth sensor," *Applied Bionics and Biomechanics*, vol. 2015, 2015.

[4] R. Velázquez, "Wearable assistive devices for the blind," in *Wearable and autonomous biomedical devices and systems for smart environment*. Springer, 2010, vol. 75 LNEE, pp. 331–349.

[5] R. Manduchi, *Mobile vision as assistive technology for the blind: An experimental study*. Springer, 2012, vol. 7383.

[6] J. A. Hesch and S. I. Roumeliotis, "Design and analysis of a portable indoor localization aid for the visually impaired," *International Journal of Robotics Research*, vol. 29, no. 11, pp. 1400–1415, Sep. 2010.

[7] A. Cosgun, E. A. Sisbot, and H. I. Christensen, "Guidance for human navigation using a vibro-tactile belt interface and robot-like motion planning," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE, 2014, pp. 6350–6355.

[8] Y. H. Lee and G. Medioni, "Wearable rgbd indoor navigation system for the blind," in *Computer Vision-ECCV 2014 Workshops*. Springer, 2014, pp. 493–508.

[9] A. Wachaja, P. Agarwal, M. Zink, M. Reyes Adame, K. Möller, and W. Burgard, "Navigating blind people with a smart walker," in *Proc. of the IEEE Int. Conf. on Intelligent Robots and Systems (IROS)*, Hamburg, Germany, 2015.

[10] A. Culhane, J. Hurdus, D. Hong, and P. D'Angio, "Repurposing of unmanned ground vehicle perception technologies to enable blind drivers," in *Association for Unmanned Vehicle Systems International (AUVSI) Unmanned System Magazine, 2011*. IEEE, 2011.

[11] Y. H. Lee and G. Medioni, "A rgb-d camera based navigation for the visually impaired," in *RSS 2011 RGBD: Advanced Reasoning with Depth Camera Workshop*, 2011.

[12] A. S. Huang, A. Bachrach, P. Henry, M. Krainin, D. Maturana, D. Fox, and N. Roy, "Visual odometry and mapping for autonomous flight using an rgb-d camera," in *Int. Symposium on Robotics Research (ISRR)*, Flagstaff, Arizona, USA, Aug. 2011.

[13] S. Koenig and M. Likhachev, "Fast replanning for navigation in unknown terrain," *Robotics, IEEE Transactions on*, vol. 21, no. 3, pp. 354–363, 2005.

[14] J. Glover, S. Thrun, and J. T. Matthews, "Learning user models of mobility-related activities through instrumented walking aids," in *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, vol. 4. IEEE, 2004, pp. 3306–3312.

[15] V. Filipe, F. Fernandes, H. Fernandes, A. Sousa, H. Paredes, and J. Barroso, "Blind navigation support system based on microsoft kinect," *Procedia Computer Science*, vol. 14, pp. 94–101, 2012.

[16] Y. Tian, "Rgb-d sensor-based computer vision assistive technology for visually impaired persons," in *Computer Vision and Machine Learning with RGB-D Sensors*. Springer, 2014, pp. 173–194.

[17] F. Ribeiro, D. Florêncio, P. A. Chou, and Z. Zhang, "Auditory augmented reality: Object sonification for the visually impaired," in *Multimedia Signal Processing (MMSP), 2012 IEEE 14th International Workshop on*. IEEE, 2012, pp. 319–324.

[18] R. Manduchi and J. Coughlan, "(computer) vision without sight," *Commun. ACM*, vol. 55, no. 1, pp. 96–104, Jan. 2012. [Online]. Available: http://doi.acm.org/10.1145/2063176.2063200

[19] D. Dakopoulos and N. G. Bourbakis, "Wearable obstacle avoidance electronic travel aids for blind: A survey," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 40, no. 1, pp. 25–35, 2010.

[20] H. Badino, U. Franke, and D. Pfeiffer, "The stixel world-a compact medium level representation of the 3d-world," in *Pattern Recognition*. Springer, 2009, pp. 51–60.

[21] D. Pfeiffer, F. Erbs, and U. Franke, "Pixels, stixels, and objects," in *Computer Vision–ECCV 2012. Workshops and Demonstrations*. Springer, 2012, pp. 1–10.

[22] T. Scharwächter, M. Enzweiler, U. Franke, and S. Roth, "Stixmantics: A medium-level model for real-time semantic scene understanding," in *Computer Vision–ECCV 2014*. Springer, 2014, pp. 533–548.

[23] M.-Y. Liu, S. Lin, S. Ramalingam, and O. Tuzel, "Layered interpretation of street view images," in *Proceedings of Robotics: Science and Systems*, Rome, Italy, July 2015.

[24] J. M. Coughlan and A. L. Yuille, "Manhattan world: Orientation and outlier detection by bayesian inference," *Neural Computation*, vol. 15, no. 5, pp. 1063–1088, 2003.

[25] S. Song and J. Xiao, "Sliding shapes for 3d object detection in depth images," in *Computer Vision–ECCV 2014*. Springer, 2014, pp. 634–651.

[26] A. S. Huang, E. Olson, and D. C. Moore, "LCM: Lightweight communications and marshalling," in *Intelligent robots and systems (IROS), 2010 IEEE/RSJ international conference on*. IEEE, 2010, pp. 4057–4062.

[27] R. C. Arkin, *Behavior-based robotics*. MIT press, 1998.

[28] D. Jeon, N. Ickes, P. Raina, H.-C. Wang, D. Rus, and A. Chandrakasan, "A 0.6V, 8mW 3D vision processor for a navigation device for the visually impaired," in *International Solid-State Circuits Conference (ISSCC)*. IEEE, 2016.