A Wearable Mobility Aid for the Visually Impaired based on embedded 3D Vision and Deep Learning

Matteo Poggi
University of Bologna,
Department of Computer Science and Engineering (DISI)
Viale Risorgimento, 2 40136 Bologna, Italy
matteo.poggi8@unibo.it

Stefano Mattoccia
University of Bologna,
Department of Computer Science and Engineering (DISI)
Viale Risorgimento, 2 40136 Bologna, Italy
stefano.mattoccia@unibo.it

***Abstract*—In this paper we propose an effective and wearable mobility aid for people suffering of visual impairments purely based on 3D computer vision and machine learning techniques. By wearing our device the users can perceive, guided by audio messages and tactile feedback, crucial information concerned with the surrounding environment and hence avoid obstacles along the path. Our proposal can work in synergy with the white cane and allows for very effective and real-time obstacle detection on an embedded computer, by processing the point-cloud provided by a custom RGBD sensor, based on passive stereo vision. Moreover, our system, leveraging on deep-learning techniques, enables to semantically categorize the detected obstacles in order to increase the awareness of the explored environment. It can optionally work in synergy with a smartphone, wirelessly connected to the the proposed mobility aid, exploiting its audio capability and standard GPS-based navigation tools such as Google Maps. The overall system can operate in real-time for hours using a small battery, making it suitable for everyday life. Experimental results confirmed that our proposal has excellent obstacle detection performance and has a promising semantic categorization capability.**

Fig. 1. Overview of the proposed mobility aid. The figure highlights main features such as object detection and categorization, haptic interface and smartphone/tablet integration for GPS navigation and voice synthesis.

I. Introduction

Autonomous mobility can be a serious problem for people suffering of visual impairments and, even when walking on a well-known route, there are hazards along their path. While some are stationary and thus their position can be learned, many others are not (people, cars, etc.) and hence can't be predicted beforehand. The white cane enables to detect nearby obstacles, but it requires physical contact, so it cant't detect farther objects. Moreover, it can't detect floating objects, such as bars or branches. Trained dogs can help in these circumstances but, as reported in [1], not without limitations due to the training cost and the relative short life of a dog.

In recent years, many devices exploiting different technologies such as GPS, sonar, vision and others described later have been proposed. However, most of them are cumbersome and not be suited for everyday activities due to the short battery life or other specific constraints related to the adopted technology. On the other hand, the rapid technology progress, mainly driven by the mobile/embedded market, enables the deployment of very powerful computing devices characterized by limited energy requirements. This fact has lead many researchers, with different degrees of effectiveness, to propose systems aimed at improving everyday life to people suffering of disabilities. One of the most critical activity is concerned with autonomous mobility: a navigation device should effectively detect any potential hazard and promptly inform the user. Moreover, an effective mobility aid should be designed with a deep understanding of the user's requirements. In [1] is reported a detailed study, based on the feedback gathered from 30 adults, providing useful insights for the design of mobility aid suited for indoor and outdoor navigation and [2], [1] pointed out the importance to select a representative set of users for testing new systems.

Our system relies on the dense and accurate 3D data provided in real-time by a custom RGBD camera worn on glass frames for blind users and with a different arrangements for other kind of visual impairments. Depth data and 2D images provided by such sensor are processed, by an embedded CPU board, to accurately detect potential obstacles and to categorize them by means of *deep-learning* techniques.The user perceives the surrounding environment sensed by the 3D camera by means of an haptic interface and audio messages that summarize the outcome of scene interpretation. The overall system, depicted in Figure 2, is extremely small and lightweight, about 250 g including a small battery enabling 3 hours of navigation,

978-1-5090-0679-8/16/$31.00 ©2016 IEEE

making it truly *wearable*. The evaluation reported in this paper shows that our proposal has excellent performance in terms of obstacle detection reliability and promising categorization capabilities.

II. Related work

Existing mobility aids can be broadly classified in two, not mutually exclusive, categories:

- Electronic Travel Support (ETS): systems aimed at enabling autonomous navigation typically by means of dynamic obstacle detection. They can be *Non-vision based*, exploiting ultrasound sensors [3], [4], [5], [6] or laser scanner coupled with Inertial Measurements Unit (IMU) [7], GPS signals [8], [9] and/or RFID [10], or *Vision based*, relying on stereo [11], [12], [13] or active sensors like Kinect [14], [15], [16].
- Self Localization Support (SLS): systems aimed at enabling self localization according to different technologies (e.g., "Blindsquare"[1]).

While factors such as weight/size, battery life and responsiveness are crucial for both categories, for ETSs sensing capability, algorithms for obstacle detection/categorization and intuitive feedback to the user clearly play a major role. People suffering of visual impairments often improve their capability to perceive the surrounding environment through other senses and in particular relying on touch and hearing as noted in [17] and [18]. For this reason, most ETS systems provide feedback according to these two senses by means of *haptic interfaces* and/or audio signals.

- Haptic-interface. The sense of touch, frequently not fully exploited by normally sighted people, is on the other hand essential for the visually impaired and for this reason it has been widely adopted in ETSs [3], [19], [14], [13], [20], [15], [16], [21], [22], [12], [5], [11], [4].
- Audio. The sense of hearing is also crucial for people suffering of visual disabilities and for this reason it has been exploited following two main approaches: audio information, by means of messages describing the environment to the user [23], [14], [13], [24], [8], [25], [7], and *sonification*, encoding in audio signals information concerned with the sensed area [26], [27], [28].

Our proposal, outlined in Figure 1, extends a very preliminary prototype [29] enabling dynamic obstacle detection and categorization, GPS localization and navigation capability combining features of ETS and SLS in a compact and lightweight setup. It aims at overcoming limitations of previously proposed mobility aids taking advantage of state-of-the-art embedded vision technology. Moreover, compared to other solutions based on active 3D vision technology (e.g., those based on the Kinect sensor) our proposal is suited not only for indoor but also for outdoor environments. Finally, it is worth to point out that we do not aim at replacing the white cane but rather at using it in synergy with our proposal in order to enrich understanding of the explored environment.

[1] www.blindsquare.com

Fig. 2. On the left, hardware components of the proposed mobility aid: custom RGBD camera ($\approx$ 90 g with lenses and holders), embedded computing platform ($\approx$ 50 g), glove with micro-motors for tactile feedback (few grams), the pocket battery ($\approx$ 70 g), bone-conductive headset ($\approx$ 45 g) and smartphone. In purple, the optional components. On the right, the overall system worn by a user.

III. Proposed mobility aid

In this section, we describe the main components of our wearable mobility aid shown in Figure 2.

A. 3D sensing and computing platform

3D sensing is carried out by a custom RGBD sensor [30] based on stereo vision technology. It provides dense and accurate depth map processing synchronized stereo images at more than 30 fps (up to 640 $\times$ 480 resolution) according to state-of-the-art stereo vision algorithms implemented into a low cost FPGA (Spartan 6 model 75 in the reported setup). Specifically, we have mapped into the FPGA a complete stereo vision pipeline including a custom and modified version of the SGM algorithm [31]. The output of the RGBD sensor (reference rectified image and disparity map) is sent, via USB at about 20 fps, to the embedded computer Odroid U3, in charge of obstacle detection/categorization. Given the 3D map and the left image provided by the camera, the Odroid performs all the vision processing and sends feedback to the user. It is worth to observe that without mapping 3D sensing algorithms into the FPGA, the overall frame rate enabled by the Odroid would be much lower.

B. Feedback interface

In order to enable a safe and effective autonomous navigation, a well designed feedback strategy is required to perceive the explored environment. For this purpose we devised a hybrid interface, exploiting tactile and audio stimuli. Following the guidelines provided in [1], our solution was aimed at obtaining an intuitive, yet effective, user interface.

According to the suggestions gathered from visually impaired users during the development of the first prototype, we designed a vibro-tactile glove, to be worn on a single hand, that does not prevent to use the same hand for other tasks (e.g, for the white cane or a mobile phone). Three micro-motors, driven by the GPIO of the Odroid, are placed on different fingers (inch, middle and pinky), each one related to a different Volume of Interest (VOI), as depicted in Figure 3, analyzed by the obstacle detection module. Compared to other solutions (e.g., [29]), this approach, suggested by visually impaired, increases sensitivity to vibrations with respect to arms and back regions. Moreover, this strategy is also less

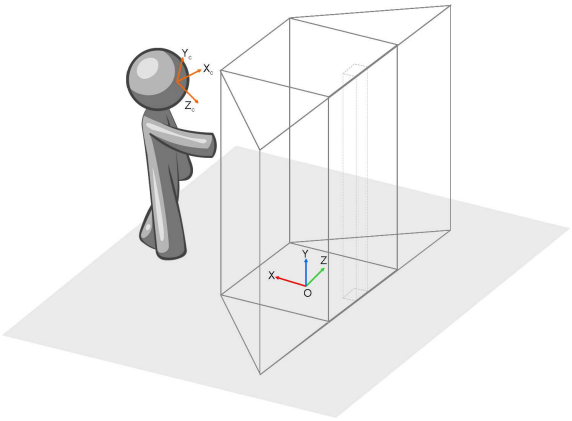


Fig. 3. The three VOIs in front of the user sensed by our system.

cumbersome. In the current setup, in order to have a more intuitive strategy, information concerned with the distance to the closest obstacle is provided by means of audio messages without exploiting depth modulation (e.g., vibration frequency provided by micro-motors inversely proportional to depth).

Audio messages are synthesized by the Odroid and sent to a bone-conductive headset in order to not isolate the person from the environment. Alternatively, audio messages can be wirelessly sent by the Odroid to a client App running on a smartphone (with the iOS operating system) and synthesized by its audio subsystem. The system allows a flexible configuration of the frequency of messages (e.g., warning continuously the user or when the distance between user and obstacle significantly changes). When using the client App, the feedback can jointly work with other navigation systems, fading route indications when an obstacle appears. To improve user's experience, it also provide several configuration settings by means of assistive touch technology, which aids many people with visual impairments to use smartphone in everyday tasks [1].

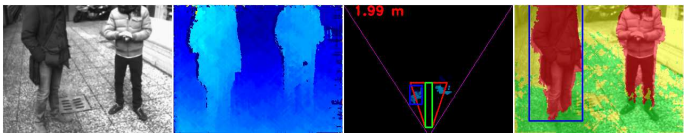
IV. Processing pipeline

In this section, we describe the computer vision pipeline for obstacle detection and categorization implemented on the embedded device Odroid U3. The output is forwarded to the user for a prompt reaction to obstacles.

A. Obstacle detection

Our system, starting from the dense disparity map provided by the RGBD sensor, computes on the embedded CPU the point-cloud according to (1) mapping each point with a valid disparity value to the corresponding 3D point of coordinates (X_c, Y_c, Z_c) w.r.t. the camera reference system by knowing the baseline of the stereo camera b, the focal length f, the optical center (u_0, v_0) and the coordinate (u, v) of the point at disparity d.

$$Z_c = \frac{bf}{d} \qquad X_c = \frac{Z_c(u - u_0)}{f} \qquad Y_c = \frac{Z_c(v - v_0)}{f} \tag{1}$$



a) b) c) d)

Fig. 4. Obstacle detection pipeline. a) left image acquired by the 3D sensor b) disparity map provided by the depth sensor c) bird-view re-projection, highlighting the 3 analyzed VOIs, green when no obstacles are present inside, red otherwise d) Final segmentation w.r.t. the left image, encoding in green the ground plane, in red the detected obstacles, in yellow the rest of the scene. The blue bounding boxes in c) and d) highlight the closest obstacle detected

From the point-cloud, a robust RANSAC framework [32] allows us to obtain a reliable estimation of the ground plane equation. This information enables to discriminate between ground plane (the safe region on which the user can walk onto) and any other object not laying on this surface. Once such plane has been detected, the point-cloud is reprojected on it as a bird-view map. For this purpose, we determine the intersection between the plane equation, in the camera reference system, and the optical axis of the camera. On such point, we build a new reference as depicted in Figure 3.

Nevertheless, the RANSAC method detects the ground plane as the largest planar surface in the sensed point-cloud; it fails when this assumption is violated, for example in presence of a large obstacle or a planar surface different from the ground (e.g., a wall). Also noisy depth measurements affects RANSAC algorithm. In order to deal these issues, we apply a Kalman filter [33]. It improves the overall stability of the raw plane equation computed from noisy point-clouds, allows to filter out outliers and leads, consequently, to a more accurate bird-view map. Moreover, its predictive model can also provide plane estimation when the previous module fails (e.g., in presence of large occlusions). Despite this positive fact, in this latter case the predicted plane equation gets less reliable increasing the amount of consecutive missing measurements. Of course, an IMU, not deployed in our current pure vision-based system, might improve effectiveness in plane estimation under similar circumstances as well as the overall plane estimation phase by fusing visual measurements with inertial data.

Once obtained the reference system, the point-cloud is processed to obtain two bird-view maps: an Occupancy Map (OM), encoding the volumes occupied by obstacles, and a Digital Elevation Map (DEM), showing their maximum height. To build them, the observed environment is split into bins laying on the detected plane. The number of 3D points from the cloud assigned to a single bin represents the occupancy, while the Y coordinate (according to reference system in Figure 3) of the highest point inside a bin represents its elevation. We process these maps by applying threshold values on both occupancy and height, in order to further filter out noisy measurements. In particular, the threshold value o_{occ} applied to the occupancy is obtained dynamically according to the Otsu's method [34]. In the current setup we do not take

advantage of the height map. However, we plan to use it to detect suspended hazards higher than the user itself and, thus, not hindering the way. The remaining elements on the map, representing possible obstacles, are filtered by a hysteresis with thresholds t_{low} and t_{high}; looking at the last N frames, an element is considered obstacle when it reaches t_{high} detections and until it falls below t_{low} again. Finally, we look for the obstacle closest to the user into the three VOIs, depicted in Figure 3. At the end of this process, as reported in Figure 4 column c), we obtain an image showing the ground plane (superimposed in green), the closest obstacle (in red) and the background objects (in yellow). Observing Figure 4, it is worth to note that, how suggested by visually impaired users, the three VOIs do not cover the whole field of view (about 60 degrees). In fact, all of them agreed that only a small portion of the scene in front of the user needs to be analyzed, corresponding to the center volume shown in Figure 4. A wider zone would make difficult to detect some important elements in the scene like open doors or narrow passages.

B. Obstacle categorization with deep learning

Machine learning techniques have been widely adopted in many practical applications and deep-learning is one of the most effective techniques for scene understanding. A deep neural network is a multilayer architecture with layers connected by non-linear transformations. In computer vision, CNNs are deep neural networks made of several layers, called *convolutional layers*, that extract features from the images by applying several normalization and filtering operations, and a final classifier, typically, a *Multi Layer Perceptron* (MLP). Compared to other machine learning techniques, such as Bag of Visual Words [35], that rely on an explicit feature extraction phase, a CNN allows for a higher level of abstraction deploying adaptive convolutional layers. LeCun et al. [36] reported how such multistage architectures yield to significant improvements compared to a single layer architecture.

In our semantic labeling module, we adopt a LeNet architecture [37], in particular a 2-layers plus a 2-levels MLP network, as shown in Figure 5. The Bank Filter and Feature Pooling blocks works with 16 and 256 5×5 kernels, while a spatial convolution is applied, randomly linking nodes in the network with a fan-in of 1 for the convolutional layers and 4 for the hidden layer of the MLP. Finally, the MLP works on 128 5x5 kernels. Training has been carried out deploying the Negative-Likelihood as loss function, minimized by means of back-propagation performed using BFGS [38]. The framework used to train the CNN and integrated for real-time semantic labeling in the embedded CPU board is Torch 7 [39].

Our strategy, focusing on the recognition of the closest obstacle, is driven by practical and performance issues. While it is important for the user to be aware of the presence of possible hazards on the scene, it would be difficult to perceive the nature of all of them intuitively and in a small amount of time, so we decided to focus on the recognition of the closest (and, probably, the one the user will meet first) to not overwhelm the user with too many details. Moreover, to properly categorize each element on the scene, more complex networks could be deployed (e.g., R-CNN [40]), able to deal with the task, but unsuitable for an embedded device such as the Odroid U3, while LeNet is able to process data in a reasonable amount of time on such device as we'll report in the next section.

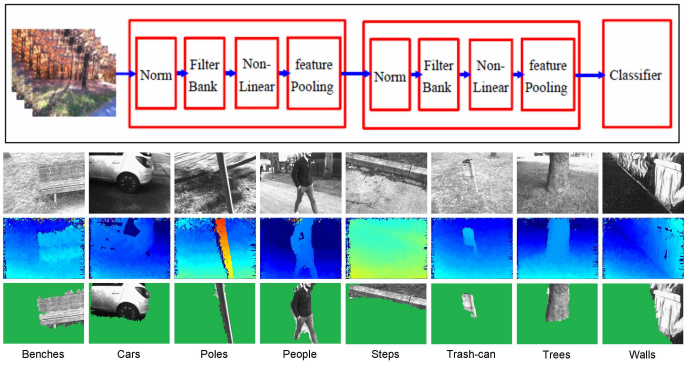


Fig. 5. Overview of the obstacle categorization module. On top, LeNet [37] model architecture. On bottom, an example of each class (one class per column) from the training dataset, containing 1000+ instances per class, showing reference image, disparity map computed by the RGBD sensor and automatically segmented obstacle.

V. Experimental results

In this section we provide experimental results concerned with several aspects of our wearable mobility aid. The camera was configured with monochrome imaging sensors, a resolution of 320×240, a baseline of 6.1 cm, focal length of 3.8 mm, enabling a 20+ fps for the Odroid embedded platform. The DEM and OM maps have been built according to 2×2 cm size bins. The sensed area for obstacle detection was set from 0.5 m up to 3 m. In this range the camera setup allows depth resolution of 0.17 cm for nearest point and 6.15 cm for the farthest ones. We applied hysteresis thresholds of 1 and 3 on the last 5 frames, drastically reducing misdetection/false alarms with a negligible detection delay at 20+ fps. Our evaluation was carried out in natural and man-made environments. We trained our CNN to recognize eight categories of obstacles, shown in Figure 5, providing 1000+ samples for each class. We'd like to observe that a larger set of categories and a larger training set for each category could be deployed by the network without any additional overhead during real-time processing.

A. Time analysis and battery life

Obstacle detection on the Odroid U3 takes about 30 ms, a value compatible with the frame rate of the depth sensor. The categorization process is performed by a different thread, in charge to recognize an obstacle when it is detected and to send the outcome to the client application (by means of an *ad-hoc* wifi network). Due to its higher execution time (about 140 ms on the Odroid), it is carried out on a subset of the total number of frames, in particular 1 out of 5. However, it

#Sequence	#TP	#TN	#FP	#FN	Precision	Recall
1	230	3354	71	2	76.41%	99.57%
2	1557	4677	3	101	99.81%	93.91%
3	742	2348	31	75	95.99%	90.82%
4	5824	638	15	46	99.74%	99.22%
5	2142	8278	82	82	96.31%	96.31%
6	1552	3797	0	72	100.00%	95.57%
7	498	1513	0	134	100.00%	78.79%
8	676	1319	0	32	100.00%	97.62%
9	1428	1812	107	171	93.02%	89.30%
Overall	14649	27736	309	715	97.93%	95.34%

TABLE I

AVERAGE RESULTS CONCERNED WITH DETECTION ON 9 SEQUENCES (43409 FRAMES) ACQUIRED IN URBAN AND NATURAL ENVIRONMENTS.

#Sequence	Environment	#TP	#Recognized	%Recognized
1	A	230	167	72.60%
2	C	1557	1339	86.00%
3	B	742	442	59.57%
4	B	5824	4298	73.80%
5	A	2142	1824	85.15%
6	C	1552	978	63.01%
7	B	498	384	77.10%
8	B	676	514	76.04%
9	B	1428	581	40.69%
Overall	-	14649	10527	71.86%

TABLE II

AVERAGE RESULTS CONCERNED WITH CATEGORIZATION ON 9 SEQUENCES (43409 FRAMES), ACQUIRED IN NATURAL (A), URBAN (B) AND MIXED (C) ENVIRONMENTS.

is worth to note that the user is unlikely to face two different obstacles in such a short amount of time (i.e., less than 200 ms). The overall system, including RGBD sensor and haptic interface, has a power consumption of about 5 Watt (4.95 V and 1.05 A). Using the small battery (68 g, 3000 mA/h) shown in Figure 2, the overall system runs for about 3 hours, up to 10 hours with a slightly weighter (231 g) 10000 mA/h battery.

B. Obstacle detection and categorization

We also carried out an evaluation of the proposed mobility aid on 9 different sequences (43409 frames), acquired in urban and natural environments, in order to measure the reliability in terms of obstacle detection and scene interpretation capability. Some of these sequences are available on Youtube[2].

Tables I reports experimental results concerned with obstacle detection on each of the 9 sequences. Moreover, the table also summarizes, in the bottom row, the average results gathered on all the 9 sequences. From left to right, the columns contain the sequence (total for the bottom row), true positives, true negatives, false positives, false negatives, precision and recall. Precision is the ratio between true positives and true positives plus false positives. Recall is the ratio between true positive and true positive plus false negatives. Observing the bottom row of Table I we can notice that, on average, our system achieve high precision and recall indexes (respectively, 97.93% and 95.34%).

In table II we report the outcome of obstacle categorization for each of thee 9 sequences and, at the bottom, the average results on all the sequences. The columns report the number of the sequence, the true positive related to detection, the amount of correct recognitions and its percentage (withe respect to the whole amount of detected obstacles). On average, we can notice that the correctness (last column) is nearly 72%. The best result (86%) has been obtained on sequence #2, natural environment, while the worst result has been obtained with sequence #9, mostly concerned with an urban environment. Sequences #6, #8 and #9 are available on Youtube.

C. Evaluation with blind people

We also carried out extensive tests with the help of blind users during different stages of the development phase. A first prototype [29] was tested by a single user, who mainly provided a critical suggestion regarding the feedback strategies (previously made of bracelets put on both arms and the back, now improved with a tactile glove). He also suggested to include in our system GPS navigation capabilities. An improved and more compact version of our mobility aid was tested by about 15 people during a workshop concerning IT applied to disabilities (Handimatica 2014[3]), with very positive feedbacks and additional suggestions. Finally, 5 more blind users tested the prototype described in this paper (without object categorization enabled) with the support of Institute of Blind people (Istituto dei Ciechi) Francesco Cavazza[4] in Bologna.

According to what stated in [1] and [2], the overall testing phase included more than 20 users, with different degrees of visual impairments. Some of them accustomed to the white cane while some others not. This latter fact in particular led to different behaviors during the evaluation: while some of them were rapidly confident with the haptic feedback, others, accustomed to the white cane, took more time to trusts in the feedback provided by our system. However, once used to our system both of them have been able to safely explore the surrounding environment. Another important factor is the frequency of the audio messages (some users feel a continuous voice to be overwhelming, while others are glad to be constantly noticed of the presence of obstacles, even when they stand still), suggesting us to allow the user to setup this strategy according to the smartphone app. As future work we plan to include in the proposed mobility aid cross-walk recognition capability [41].

VI. CONCLUSIONS

In this paper we have described a wearable mobility aid for people suffering of visual impairments. The system relies on a custom RGBD camera, providing dense depth measurements, and an embedded CPU board where we have mapped a vision based obstacle detection pipeline. Our system also provides semantic categorization of detected obstacles by means of a Convolutional Neural Network. Our system is extremely small, lightweight and enables to obtain real-time accurate details concerning the explored environment by means of an

[2]https://www.youtube.com/channel/UChkayQwiHJuf3nqMikhxAlw

[3]www.handimatica.com

[4]www.cavazza.it

haptic interface and audio messages. The experimental results reported show that our system has excellent detection performance (close to 98% of detection rate) and very promising object categorization capability (close to 72% of correctness). In particular, this latter fact could be further improved, by extending the training dataset, without affecting the overall execution time.

REFERENCES

[1] M. A. Williams, A. Hurst, and S. K. Kane, ""pray before you step out": describing personal and situational blind navigation behaviors." in *ASSETS*. ACM, 2013, p. 28.

[2] A. Sears and V. Hanson, "Representing users in accessibility research," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11. New York, NY, USA: ACM, 2011, pp. 2235–2238.

[3] C. Shah, M. Bouzit, M. Youssef, and L. Vasquez, "Evaluation of ru-netra - tactile feedback navigation system for the visually impaired," in *Virtual Rehabilitation, 2006 International Workshop on*, 2006, pp. 72–77.

[4] K. Ito, M. Okamoto, J. Akita, T. Ono, I. Gyobu, T. Takagi, T. Hoshi, and Y. Mishima, "Cyarm: an alternative aid device for blind persons." in *CHI 2005*, 2005.

[5] S. Cardin, D. Thalmann, and F. Vexo, "Wearable system for mobility improvement of visually impaired people," in *In Visual Computer journal*, 2006.

[6] J.-H. Lee, E. Choi, S. Lim, and B.-S. Shin, "Wearable computer system reflecting spatial context," in *Semantic Computing and Applications, 2008. IWSCA '08. IEEE International Workshop on*, July 2008, pp. 153–159.

[7] J. M. Loomis, R. G. Golledge, and R. L. Klatzky, "Navigation system for the blind - auditory display modes and guidance," *Presence*, vol. 7, no. 2, pp. 193–203, 1998.

[8] L. Ran, S. Helal, and S. Moore, "Drishti: An integrated indoor/outdoor blind navigation system and service," in *in Proc. of the Second IEEE Annual Conference on Pervasive Computing and Communications*, 2004.

[9] A. S. Helal, S. E. Moore, and B. Ramachandran, "Drishti: An Integrated Navigation System for Visually Impaired and Disabled," in *ISWC '01: Proceedings of the 5th IEEE International Symposium on Wearable Computers*, 2001.

[10] B. U. Ceipidor, C. M. Medaglia, F. Rizzo, and A. Serbanati, "RadioVirgilio/Sesamonet:an RFID-based Navigation system for visually impaired," in *Mobile Guide '06 (http://mobileguide06.di.unito.it/programma.html)*.

[11] P. Arcara, L. Di Stefano, S. Mattoccia, C. Melchiorri, and G. Vassura, "Perception of depth information by means of a wire-actuated haptic interface," in *In IEEE Int. Conf. on Robotics and Automation*, 2000.

[12] D. Dakopoulos, S. K. Boddhu, and N. G. Bourbakis, "A 2d vibration array as an assistive device for visually impaired." in *BIBE*, 2007.

[13] H. Fernandes, P. Costa, V. Filipe, L. Hadjileontiadis, and J. F. Barroso, "Stereo vision in blind navigation assistance," in *2010 World Automation Congress, WAC 2010*, 2010.

[14] D. Ni, L. Wang, Y. Ding, J. Zhang, A. Song, and J. Wu, "The design and implementation of a walking assistant system with vibrotactile indication and voice prompt for the visually impaired." in *ROBIO*, 2013.

[15] Kinecthesia, "Kinecthesia 2.0," http://www.kinecthesia.com/.

[16] M. Zöllner, S. Huber, H.-C. Jetter, and H. Reiterer, "Navi: A proof-of-concept of a mobile navigational aid for visually impaired based on the microsoft kinect," in *Proceedings of the 13th IFIP TC 13 International Conference on Human-computer Interaction - Volume Part IV*, ser. INTERACT'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 584–587.

[17] R. Velázquez, "Wearable Assistive Devices for the Blind," *Wearable and Autonomous Biomedical Devices and Systems for Smart Envi- ronment: Issues and Characterization*, 2010.

[18] K. A. Kaczmarek, J. G. Webster, P. Bach-y Rita, and W. J. Tompkins, "Electrotactile and vibrotactile displays for sensory substitution systems," *IEEE Transactions on Biomedical Engineering*, 1991.

[19] S. Meers and K. Ward, "A vision system for providing 3d perception of the environment via transcutaneous electro-neural stimulation." in *IV*, 2004.

[20] J. Zelek, R. Audette, J. Balthazaar, and C. Dunk, "A stereo-vision system for the visually impaired," University of Waterloo, Tech. Rep., 2000.

[21] G. Flores, S. Kurniawan, R. Manduchi, E. Martinson, L. Morales, and E. Sisbot, "Vibrotactile guidance for wayfinding of blind walkers," *Haptics, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.

[22] L. A. Johnson and C. M. Higgins, "A navigation aid for the blind using tactile-visual sensory substitution." *Conf Proc IEEE Eng Med Biol Soc*, 2006.

[23] N. G. Bourbakis and D. Kavraki, "An intelligent assistant for navigation of visually impaired people," in *Proceedings of the 2Nd IEEE International Symposium on Bioinformatics and Bioengineering*, 2001.

[24] B. Huang and N. Liu, "Mobile navigation guide for the visually disabled," *Journal of the Transportation Research Board*, no. 1885, pp. 28–34, 2004.

[25] A. Hub, J. Diepstraten, and T. Ertl, "Augmented indoor modeling for navigation support for the blind," in *Conference Proceedings: CPSN'05 - The International Conference on Computers for People with Special Needs, Las Vegas*, 2005.

[26] A. Rodríguez, L. Bergasa, P. Alcantarilla, J. Yebes, and A. Cela, "Obstacle avoidance system for assisting visually impaired people," in *IEEE Intelligent Vehicles Symposium, Proceedings of Workshop Perception in Robotics*, 2012.

[27] P. R. Sanz, B. R. Mezcua, J. M. S. Pena, B. N. Walker, S. J. Tipton, D. J. White II, C. Sershon, Y. B. Choi, F. Yu, J. Zhang *et al.*, "1 evaluation of the sonification protocol of an artificial vision system for the visually impaired," *Evaluation*, 2014.

[28] G. Balakrishnan, G. Sainarayanan, R. Nagarajan, and S. Yaacob, "Wearable real-time stereo vision for the visually impaired."

[29] S. Mattoccia and P. Macri', "3d glasses to improve autonomous mobility of people visually impaired." in *Second Workshop on Assistive Computer Vision and Robotics (ACVR2014), ECCV Workshop*. Springer, 2014.

[30] S. Mattoccia and M. Poggi, "A passive rgbd sensor for accurate and real-time depth sensing self-contained into an fpga," in *Proceedings of the 9th International Conference on Distributed Smart Cameras*, ser. ICDSC '15. New York, NY, USA: ACM, 2015, pp. 146–151.

[31] H. Hirschmüller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, 2008.

[32] S. Choi, T. Kim, and W. Yu, "Performance evaluation of ransac family." in *BMVC*, 2009.

[33] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.

[34] N. Otsu, "A Threshold Selection Method from Gray-level Histograms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[35] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *In Workshop on Statistical Learning in Computer Vision, ECCV*, 2004, pp. 1–22.

[36] K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun, "What is the best multi-stage architecture for object recognition?" in *Proc. International Conference on Computer Vision (ICCV'09)*. IEEE, 2009.

[37] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Intelligent Signal Processing*, S. Haykin and B. Kosko, Eds. IEEE Press, 2001, pp. 306–351.

[38] Q. V. Le, J. Ngiam, A. Coates, A. Lahiri, B. Prochnow, and A. Y. Ng, "On optimization methods for deep learning." in *ICML*, L. Getoor and T. Scheffer, Eds. Omnipress, 2011, pp. 265–272.

[39] R. Collobert, K. Kavukcuoglu, and C. Farabet, "Torch7: A matlab-like environment for machine learning," in *BigLearn, NIPS Workshop*, 2011.

[40] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation." *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 142–158, 2016.

[41] M. Poggi, L. Nanni, and S. Mattoccia, "Crosswalk recognition through point-cloud processing and deep-learning suited to a wearable mobility aid for the visually impaired," in *New Trends in Image Analysis and Processing - ICIAP 2015 Workshops - ICIAP 2015 International Workshops: BioFor, CTMR, RHEUMA, ISCA, MADiMa, SBMI, and QoEM, Genoa, Italy, September 7-8, 2015, Proceedings*, 2015, pp. 282–289.