3D glasses as mobility aid for visually impaired people

Stefano Mattoccia and Paolo Macrı'

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

Abstract. This paper proposes an effective and wearable mobility aid aimed at improving the quality of life of people suffering of visual disabilities by enabling autonomous and safe navigation in unknown environments. Our system relies on dense and accurate depth maps, provided in real-time by a compact a stereo vision system mapped into an FPGA, in order to detect obstacles in front of the user and to provide accordingly vibration feedbacks as well as audio information by means of a bone-conductive speakers. Compared to most approaches with similar purposes, even in the current prototype arrangement deployed for testing, our system is extremely compact, lightweight and energy efficient thus enabling hours of safe and autonomous navigation with standard batteries avoiding the need to carry cumbersome devices. Moreover, by conceiving the 3D sensing device as a replacement of standard glasses typically worn by visually impaired people and by using intuitive feedbacks provided by means of lightweight actuators, our system provides an ergonomic and comfortable user interface with a fast learning curve for its effective deployment. This fact has been extensively verified on the field by means of an experimental evaluation, in indoor as well as in outdoor environments, with different users simulating visual impairment including a blind person.

Keywords: wearable, visually impaired, 3D, stereo vision, obstacle detection

1 Introduction

For people suffering of visual impairments even common everyday tasks, such as freely walking on a sidewalk to reach a well-known destination, can represent a serious problem due to the potential dangerous situations encountered along the path. In the context figured out, hazards can be represented by other pedestrians, animals or objects (boxes, traffic signals, cars, waste containers, etc). Some of these hazards can be *learned* by the user day by day, however many obstacle are not stationary and hence must be detected dynamically. Remaining in the specific context figured out, other dangerous hazards can be represented by moving cars, motorbikes, bikes at cross-roads. Although for the former hazards the white cane can be effective (but quite intrusive requiring physical contacts with the ”sensed” people/objects), in the latter cases it is almost useless. In fact, even when using a white cane, the visually impaired is not aware, excluding the audio perception of the surrounding, if vehicles/objects are approaching him/her, at what speed and from which directions. These facts, according to our discussions with blind people, create a sense of fear that is constantly perceived by the user. Specific signals for visually impaired available in some urban environments, such as road markers, can provide additional cues for people suffering of visually impairments. However, these signals can’t help in dynamic context such as that previously highlighted. Appropriately trained guided dogs can help in such contexts but they also have some well-know limitations (e.g. the short life of a dog, costs and efforts for training, etc).

The social relevance of the the problem and the large amount of people suffering of visual impairments has lead to the development of interesting mobility aids by means of various technologies such as GPS, sonar, vision sensors that will be thoroughly reviewed in the next sections. Among these technologies, vision based approaches have the potential to overcome all other technologies but so far these systems haven’t found practical deployment. The reason can be found in two key factors. The first is the actual effectiveness provided by the mobility aid in typical contexts required by people suffering of visual impairments. The other key factor is the limited usability; existing systems are in most cases cumbersome and not suited for everyday deployment for many hours as would be required for an effective practical deployment.

Our proposal aims at overcoming these limitations by using an effective methodology for obstacle detection based on robust and dense 3D data that, thanks to a 3D camera with on board FPGA processing capable of delivering very accurate depth maps according to state-of-the-art stereo vision algorithms, can be implemented on embedded devices with minimal energy and computational requirements. For these reasons, the overall system proposed in this paper is truly wearable and lightweight. Moreover, we propose in this paper a very intuitive and effective user interface that enables to perceive the surrounding environments without reducing the already limited perception capability of the visual impaired.

2 Related work

The dramatic technology progress, mainly driven by the mobile/embedded market, nowadays 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 mobility aids for people suffering of visual disability. These systems can be broadly categorized in two, not mutually exclusive, categories:

- Electronic Travel Support (ETS): systems aimed at enabling autonomous navigation typically by means of obstacle detection methodologies. Sometimes these systems also enable path planning.

- Self Localization Support (SLS): systems aimed at providing to the user the capability to localize himself according to different technologies. An example, available as a mobile phone application, is "Blindsquare"[1] that provides to the user, by means of audio messages, navigation and localization information according to the position retrieved by the GPS receiver of a mobile phone.

We'd like also to point out that exist other interesting aids for visually impaired that do not strictly fall into one of the two highlighted categories. Among these we'd like to remember the ORCAM system[2] that, by means of machine learning algorithms applied to a small camera linked to the glasses, enables to identify learned objects or to infer the status of traffic lights at cross roads. Another interesting system that does not strictly fall into one of the two previous categories is the "Lend an eye" device[3] that enables to stream to the mobile phone of a remote human assistant what is seen by a camera worn by the visually impaired in order to receive suggestions.

Our system, in its current development stage, being aimed at enabling autonomous navigation by means of obstacle detection methodologies falls in the category of ETS systems. Despite this fact, it could be deployed in conjunction with SLS systems or other systems reviewed so far.

ETS systems rely on sensors, based on different technologies, in order to sense the surrounding environment and, by inferring relevant features, to provide a feedback to the user. Therefore three factors are crucial in ETS systems: effective sensing capability, robust algorithms for obstacle detection (or path planning) and intuitive feedback to the user. Three further crucial factors of an ETS system, essentials for a practical deployment, are weight, size and fast responsiveness. Excluding these latter underlying factors, we can further classify ETS according to two notable features: the sensing technology adopted and the feedback perceived by the user.

2.1 Sensing

For ETS system the sensing device used to perceive the environment is crucial to obtain robust and accurate suggestions with a *safe* amount of time before any potential collision. In the literature we can identify two broad classes of sensing technologies used for this purpose: non vision based and vision based devices.

- Non-Vision Based
  Systems based on this kind of sensing devices mostly rely on ultrasound or laser scanners technologies, often coupled with Inertial Measurements Unit (IMU) [1], GPS signals [2,3] and/or RFID [4]. These latter technologies are often used in conjunction with vision based sensing devices. Ultrasound technology provides a *localized* measure of the distance, in the form of an

---

[1] www.blindsquare.com

[2] www.orcam.com

[3] www.lendaneye.com

electrical signal, between the sensor and the area where it is pointed. Positive aspects of this technology are the limited cost and the small size. On the other hand, ultrasound sensors are directional, thus covering only a small area of the scene and can't provide an image of the sensed area. Laser scanners (e.g. [5]) enable sensing in larger areas but they are typically cumbersome and, as for ultrasound sensors, do not provide images of the sensed area. In [6], [7] and [8] are described ETS systems based on ultrasound sensors for directional depth measurements. The same technology coupled with optical sensors in order to cover a larger area was proposed in [9]. Other ETS systems, such as [5] and [9], rely on GPS coupled with optical sensors to detect obstacles.

- Vision Based

  Vision, and in particular vision devices for depth perception, have been widely used for ETS systems. In [10–16] the sensing technology used consists in stereo vision. The same technology was adopted by [17]; however, in this cases, the reference image of the stereo camera is also used to detect specific markers put on the road in order to define a pre-configured path. It is worth observing that most of these stereo vision systems rely on the conventional fixed window algorithm [18] for stereo matching that, as well-known, is not very accurate near depth discontinuities and provides unreliable results in poorly textured regions. Moreover, these stereo vision systems often have a very low frame rate. Despite the widespread deployment of stereo vision technology for ETS systems, other sensing technologies have been used by other researchers; [5] uses a laser scanner coupled with two monocular camera while [9] relies on a Kinect and ultrasound sensors. Of course, due to the structured pattern projected by the Kinect, the latter system is suited only for indoor environments.

  Compared to these systems, mainly based on heavy and cumbersome devices, the 3D sensing device proposed in this paper has several advantage: it is extremely lightweight, has minimal energy requirements and provides very accurate depth maps according to a state-of-the-art processing pipeline at high frame rate (about 18 fps in the configuration reported in this paper). Moreover, compared to most previous approaches, the computing platform for obstacle detection, thank to the FPGA implementation of the stereo matching algorithm, can be a compact embedded device as in the current prototype or a smartphone/tablet.

2.2 Feedback

People suffering of visual disability often improve their capability to perceive through other senses. In [19] and [20] was highlighted that visually impaired users mostly rely on touch and hearing to compensate for their visual loss and for this reason most ETS systems provide feedback according to these two senses.

- Haptic

  The sense of touch, frequently not fully exploited by normally sighted people, is on the other hand essential for visually impaired. For this reason this sense has been widely used to provide feedback in ETS systems. In [6] a hand held cylinder containing moving rings provides a tactile feedback of the sensed area. Two gloves with *transcutaneous electrical nerve stimulation* (TENS) on each finger were proposed for depth perception in [21]. The sensed area is divided in ten regions and for each region a finger receives a stimulation proportional to the depth sensed in that area. Vibrotactile devices (vibration actuators), placed on sensitive parts of the human body such as hands or arms are widely used to provide haptic feedback. In [9] an array of vibrotactile devices is used to identify regions without obstacles while [17] relies on this stimulation to provide modification to the detected trajectory. A glove with a vibrotactile device for each finger was proposed in [11] in order to provide a feedback for five not overlapping regions of the sensed area. A similar approach, enhanced by a vibration modulated according to the perceived depth in each area, can be found in [22]. Array of vibration actuators with modulation proportional to the distance are reported in [16] and [8]. However, it is worth observing that [19] and [20] reported that increasing the number of vibration actuators can reduce the overall effectiveness of the feedback device. Another feedback approach consists in constraining a finger of the visually impaired by means of actuated wires as reported in [10] and [7].

- Audio

  The sense of hearing is also crucial for people suffering of visual disabilities and for this reason it has been exploited by means of two main approaches: audio information and *sonification*. In the first case the information consists in audio messages that can help the user to better understanding the surrounding environment according to a localization system such as a GPS (e.g. "you are at this location") or the perception module (e.g. "there is an obstacle in front of you"). Examples of systems based on this approach are [5], [9],[17], [23], [2], [14] and [1]. The other strategy, referred to as sonification, consists in encoding information concerned with the sensed area in audio signals. For instance by changing the frequency of the audio signal according to the sensed distance. Examples of systems based on this strategy are [15], [13] and [12].

  We conclude this section observing that this latter strategy requires a significant amount of training and, similarly to the former approach based on standard audio information, isolates the user from the environment. Moreover, the audio signal should be used only when it is strictly requested by the user (e.g. "Where am I?") or triggered by potentially dangerous situations (e.g. obstacles). For these reasons, in our system we deploy a bone conductive headset to provide standard audio messages triggered by specific events (e.g. an obstacle in front of the user).

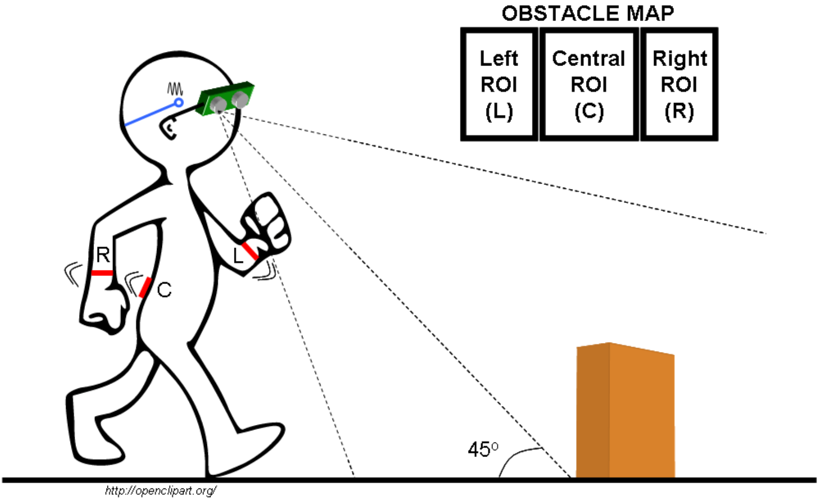

Fig. 1. Overview of the proposed system, shown in details in Figure 2, with the embedded 3D camera mounted on conventional glasses frames as shown in Figure 3. In red the three vibrotactile actuators and in blue the bone conductive headset for audio messages. The sensed area is divided in three regions, each one corresponding to a vibrotactile actuator.

3 Overview of the proposed ETS system

Our proposal aims at enabling people suffering of visual impairments to move autonomously in unknown environments in presence of dynamic objects/people by wearing *glasses* with 3D sensing capability and an effective feedback approach as depicted in Figure 1. Our system, reported in details in Figure 2, consists of a compact 3D sensor mounted on conventional glasses frames, worn as a pair of conventional glasses and tilted of about 45 degrees downward looking as shown in Figure 3, an embedded PC that detects obstacles in the area sensed by the 3D camera. The user perceives the feedback according to three vibrotactile actuators depicted in red in Figure 1 (the small armband/bracelet worn around each arm and a further armband put in another location of the human body such as the back of the neck) and a bone conductive headset according to a strategy that will be discussed in the remainder.

Observing Figure 2 we can notice the 3D camera (worn as in Figure 3), the bone conductive headset, two of the armband containing the vibrotactile actuators (the other vibrotactile device in this configuration is mounted on the back of the belt and hence not visible in the figure) and a belt containing the processing units and a battery. Concerning the processing units: the obstacle detection algorithm has been implemented on the Odroid U3 [24] platform (on the left of figure, enclosed with a plastic case) while the Arduino Due board [25], connected by means of a serial communication to the Odroid, merely controls the vibrotactile actuators. The battery, at the right, provides energy to the entire system

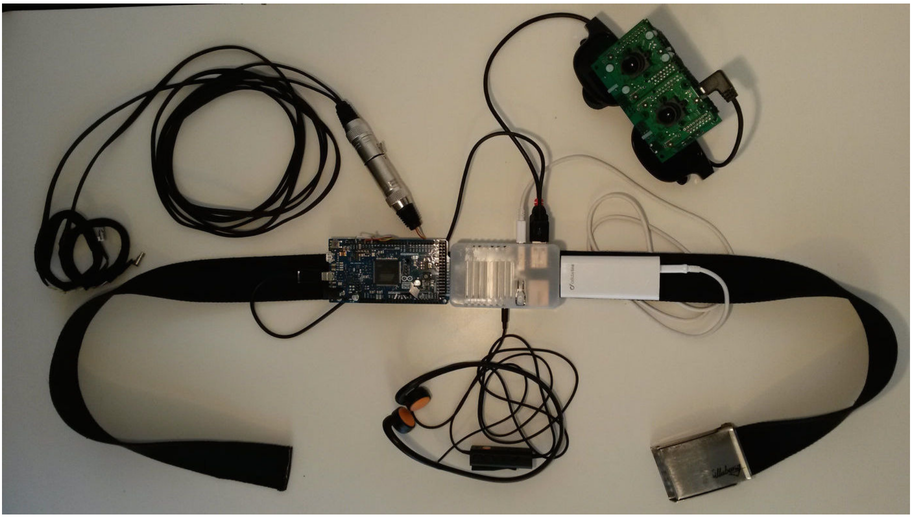

Fig. 2. Components of the proposed ETS system: 3D camera with FPGA on board processing, Odroid U3 computing platform, Arduino Due, three vibrotactile actuators, bone conductive headset and battery. The overall weight of all the components is about 300 g (3D camera + glasses + lenses + holders account for about 110 g).

for hours. The weight of all the electronic devices shown in Figure 2 including the camera, the headset, vibrotactile actuators and battery is about 300 g. The processing module shown in the figure can be worn as a conventional belt, as depicted in Figure 5, and this fact coupled with the embedded 3D camera makes our system truly wearable. Finally we observe that, although in the currently development stage the prototype is composed of multiple distinct modules plus a battery, a smartphone or a tablet coupled with the 3D sensor could be used in place of all the processing devices (i.e. the embedded system and the Arduino board) in order to further increase compactness and usability of the proposed ETS system.

4 Hardware

In this section we describe the hardware components of our current prototype focusing our attention on the 3D sensing device at the core of our proposal.

4.1 3D camera

The key component of our system is an embedded stereo vision camera with on board FPGA processing shown in Figure 3 that allows us to obtain accurate and robust depth maps of the scene in front of the user at almost 20 fps without introducing any computational overhead to the processing modules implemented on the Odroid CPU board [24].

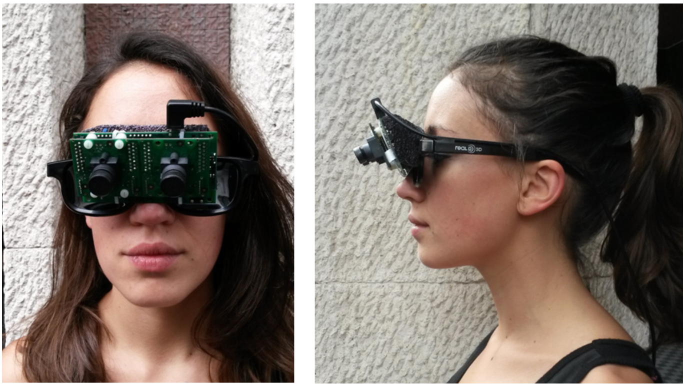

Fig. 3. Embedded 3D camera mounted on conventional glasses. The camera is self powered through the USB cable.

The embedded 3D camera, designed by our research group [26, 27], can be configured with monochrome or color sensors at a maximum resolution of 752 × 480 and 60 fps. In the experiments reported in this paper the camera is configured with a resolution of 320× 240 and a frame rate of about 20 Hz, sufficient to handle fast moving object in sensed area. As will be shown in the experimental results section, the camera provides very accurate and dense depth maps according to a state-of-the-art stereo vision processing pipeline fully implemented into the FPGA that, in this specific case, is a Spartan 6 model 75. The processing pipeline consists of the following three main components: pre-processing based on the Census transform [28], rectification of the stereo pair and matching according to the Hamming distance. In particular, stereo matching algorithm mapped into the FPGA is a memory efficient implementation of the SGM algorithm [29] and the processing pipeline also includes an effective outliers detection module and a $\frac{1}{8}$ subpixel depth interpolation module. The overall processing pipeline, thank to a stream processing methodology [26], relies only on the internal memory of the FPGA. The reduced power consumption of the embedded 3D camera (about 2 Watt at 640 × 480 image resolution and processing stereo pairs at 40+ fps), enables us to provide power supply by means of the same USB connector used for the data and connected to the Odroid platform. The weight of the current overall camera shown in Figure 3 is about 90 g, including the M12 lenses and holders, and for this reason it does not result uncomfortable to the user wearing the 3D glasses. The camera is configured with a baseline of about 6 cm that, with the lenses adopted in our experimental validation and the hardware configuration mapped into the FPGA, enables to infer depth from 1 m to 5 m.

4.2 Embedded processing

The data streamed by the 3D camera are sent, via USB, to the Odroid U3 [24] processing platform in order to detect the ground plane and thus potential obstacles in the sensed area. This device, one completed this task, performs a reasoning on the outcome of the obstacle detection module and accordingly sends audio messages, by means of its audio interface, to the speakers and commands, by means of a serial link, to the Arduino Due board to activate the vibrotactile actuators.

The Odroid U3 contains an ARM SoC, model Exynos 4412, Cortex A9 architecture with four cores clocked at 1.7 Ghz, 1 MB level 2 cache, 2 GB of LP-DDR2 880 Mhz RAM, a Mali-400 graphic processor at 440 Mhz developed by ARM, an USB controller and an Ethernet controller. The input/output ports provided by the Odroid board are: 3 USB, 1 Micro HDMI, 1 Ethernet. The Odroid has a size of 83 x 48 mm and weights 48 g including the heat sink. The operating system installed is Linux, specifically the distribution used for our tests was Xubuntu 13.10. The energy requirement of this platform is below 10 W (typically around 5 W), making this device an ideal candidate for our compact and lightweight ETS system.

The Arduino Due [25] board is connected, through a serial interface mapped on the USB interface, to the Odroid U3 in order to receive command to activate the vibrotactile actuators according to the outcome of the obstacle detection module. However, it is worth observing that the Arduino Due board wouldn't be strictly required being available I/O ports on the Odroid platform. Nevertheless, it has minimal weight, size and power consumption and for these reasons it was adopted to speed-up the development phase of our prototype.

5 Obstacle detection

Once completed the description of the sensing device and of the processing modules, let's take a look to the algorithm for obstacle detection implemented into the Odroid U3. Our approach consists in the first and crucial stage in determining, for each frame, the ground plane from the disparity maps provided by the 3D camera under the hypothesis that a significant portion of the ground plane is observed in the sensed area. To accomplish this task there are two possible alternatives: processing the point-cloud associated to the disparity map or detecting the ground plane directly in the disparity domain. Although the former solution is completely invariant to camera tilting, after an appropriate experimental evaluation we decided to compute the ground plane directly in the disparity domain. The reason for this choice is twofold: efficiency and specific application domain constraints. Concerning the first fact, working in the disparity domain is less demanding in terms of computational requirements than working in the point-cloud domain. The second reason is concerned with the specific application tackled in our project: in fact in the envisioned system the user, in normal conditions, can tilt the camera on the left or on the right side of few degrees and this fact can be easily managed by a roll-detection algorithm implemented in software and/or by means of an IMU (currently not used in our system). Finally, we observe that tilting the camera by moving the head up and down would not create any problem to the plane detection algorithm working in the disparity domain.

Our algorithm, starting from the disparity map provided by the stereo camera, detects the ground plane applying a robust RANSAC-based approach [30] in the *v-disparity* domain [31]. In order to obtain a robust detection of the ground plane in the v-disparity histogram, we have implemented and tested variants of the RANSAC approach, recently reviewed an evaluated in [32], and aimed at improving efficiency and robustness to outliers. Our final framework, used for the experimental results reported in this paper, includes the following modifications to the original RANSAC framework:

- A degeneracy check in the sampling phase that enables to avoid degenerate model hypotheses such as, in our specific case, collinear vertical points. This strategy quickly allows us to discard a model hypothesis in the first stage of the RANSAC approach without any further computationally demanding verification stage.
- An R-RANSAC $\mathrm{T}_{d,d}$ test switch, as proposed in [33], that can be optionally triggered in the sampling stage, enabling a further pre-verification stage for fast rejection of bad hypotheses. This strategy enables to verify the model considering only a small subset of randomly selected points. If these contain outliers the model is quickly discarded with a minimal amount of computation being the verification process performed on a small subset of points. Otherwise, the model is validated according to the remaining points. Although, this strategy requires more hypotheses compared to a standard RANSAC approach, this fast pre-verification phase enables in most cases a significant runtime reduction according to our evaluation.
- An effective cost function as proposed in [34]. Compared to the original RANSAC approach that assigns to all inliers the same cost, this method weights each inlier according on how the point fits with the model. In our experimental evaluation, this strategy effectively enabled to improve plane detection accuracy as well as to deal with noisy disparity measurements.
- An optional adaptive termination criteria that enables to determine the maximum number of iterations at every cycle [35] in order to minimize the runtime.

Our software processing module implemented on the Odroid, handles small lateral tilting of the camera, determining this parameter and rotate the image accordingly before the ground detection approach previously outlined in order to avoid ground plane mis-detection. The roll-detection algorithm, similarly to the plane detection algorithm, works in the v-disparity histogram domain and relies on the enhanced RANSAC framework previously outlined. Specifically, this approach determines the *mode* of the disparity in regions of the disparity map where the ground plane is more likely to be found (e.g. those areas of the disparity map that were identified as belonging to the ground plane in the previous frame) or more simply in prefixed areas in the disparity map combined by means of a voting scheme. Once determined the most frequent disparity in these regions (i.e. the mode), we build the v-disparity space considering only those points with disparity values similar to the detected mode and apply a robust linear regression based on the RANSAC framework. This allows us to determine the tilting angle and thus to compensate for lateral tilting before the ground detection algorithm. It is worth to note that for small lateral tilting, as frequently occurs in the considered application domain, the roll-detection wouldn't be required at all being the v-disparity approach invariant to small misalignments between the camera and the ground plane. Moreover, we point out that our approach does not rely, in the current stage of development, on IMUs or a Kalman filter that could further improve the robustness of our ground plane detection approach.

Once detected the ground plane we set as obstacle those points that do not lie on this surface according to a prefixed tolerance threshold and to the resulting *obstacle map* we provide feedback to the user.

6 User feedback

By analyzing the lower portion of disparity map, the subsystem implemented on the embedded CPU board provides vibrotactile and audio messages to the user. In order to design a comfortable and effective system we decided to minimize the feedbacks according to a strategy outlined in the reminder. Moreover, we observe that the proposed system let hands and arms free; this fact is extremely important because enables exploration by touch and/or enables the deployment of the conventional white cane (e.g. for balancing).

6.1 Tactile feedback

To accomplish an effective tactile perception, we divide the lower portion of the disparity map in three regions as depicted in Figure 1 and link each region to a vibrotactile actuator (the two associated to L and R ROIs are visible in Figure 2, the one associated to the central ROI is positioned on the back of the belt). The relationship between the actuators and the regions of the obstacle map is as follows:

- obstacle in ROI-L: left vibrotactile motor enabled
- obstacle in ROI-C: central vibrotactile motor enabled
- obstacle in ROI-R: right vibrotactile motor enabled

For each of the three regions we trigger an alarm only if there are a sufficient number of points marked as obstacle. An example of the described approach on real-images is shown in Figure 5; green means no obstacle and red means obstacle detected in that specific region. This setup allows the user to obtain an intuitive perception of the obstacle detected in the field-of-view of the camera. Moreover, the user can easily explore the surrounding area by simpling rotating the head in order to detect obstacle in the area pointed by the glasses These operations, for a constrained area around the user, are typically carried out by means of a white cane.

It is worth observing that we deliberately decided to use only three vibrotactile actuators to design a simple and effective interface that requires only few minutes for training. Moreover, we decided to avoid modulation of the tactile feedback (e.g. proportional to the distance from the average obstacle detected in one of the sensed areas shown in Figure 1) since our evaluation clearly highlighted that this strategy requires a longer training period and, more importantly, the modulated vibration is difficult to perceive and thus less effective compared to the simpler, yet effective, solution proposed in this paper.

6.2 Audio feedback

When an obstacle is detected in the central ROI shown in Figures 1 and 5, an audio message containing the average distance to the object is provided to the user. In order to not provide excessive and counterproductive feedbacks, the audio message is sent again only if the user/obstacle moves of at least 1 m with respect to the position of the previous obstacle. The audio feedbacks are provided according to an headset based on the bone conduction technology. This fact, as previously pointed out, is very important in order to not acoustically isolate the visually impaired from the surrounding environment. For the same reasons outlined in the previous section, concerned with usability factors, we deliberately decided to avoid sonification in our system.

7 Experimental validation

We have extensively tested, in indoor and outdoor scenarios, the proposed ETS system shown in Figure 1 with about ten users (three of them are shown in Figure 4) including a person visually impaired that has suggested specific feedbacks during the development stage and proposed some effective modifications to our initial prototype. For normally sighted users we simulated visual impairments by covering their eyes. Experimental results concerned with three normally sighted users and a visually impaired person are available at this link[4]. It is worth noting that experimental results reported in the videos are concerned with a preliminary prototype of the proposed ETS system. Compared to the current prototype shown in Figures 2, 3 and in the rightmost image of Figure 4, the main differences are basically in the larger camera case and battery. Figure 4 shows at the left and at the center two users testing the preliminary prototype.

During our evaluation all the users agreed on the fact that the current prototype is truly *wearable* and that the proposed user interface is extremely intuitive and effective. In fact, they were able to effectively deploy the system after only few minutes it was worn. Only the visually impaired person required a slightly longer period to be fully accustomed to the system because in his daily living he

---

[4] `www.vision.deis.unibo.it/smatt`

Fig. 4. Experimental evaluation with three users. The rightmost image is concerned with latest prototype shown in Figures 2 and 3 while the left and center images are concerned with the initial prototype of the system.

heavily relies on the white cane for perception and balance. For these reasons, he initially didn't trust on the additional feedbacks provided by our system. Despite this fact, after a training period of about 30 minutes he was ready to use the proposed system and, mainly for balancing, his white cane. This latter fact is not a problem being, in our proposal, arms and hands free to move without constraints. After this short training period and an extensive evaluation in an urban scenario his feedback on our ETS system was extremely positive. He also highlighted that a notable feature that currently is not available in our system is a GPS-based audio navigation support that would provide a direction during obstacle detection. Nevertheless, it is worth pointing out that during the development stage described in this paper we focused our attention on the more challenging obstacle detection task because the highlighted GPS feature could be eventually added to our system by using a standard navigator (such as Google Maps or similar), available nowadays even in cheapest smart phones, or by using a GPS module attached or connected via Bluetooth to the Odroid board.

In Figure 5 are depicted three screenshot taken in an outdoor urban scenario. The upper portion of the figure contains the reference images acquired by the stereo camera and the lower portion the obstacle maps (in white the detected ground plane and in black the detected obstacle points). Superimposed to the obstacle maps are shown, for each image reported, the three sensed regions with green encoding *no obstacle* and red *obstacle detected*. In this latter case superimposed to the image we can read the average distance from the camera to the detected obstacle. From the figure we can notice that the disparity maps provided by the 3D camera are quite accurate; in fact, the obstacle detection algorithm, processing these disparity maps accurately detects obstacles and ground plane in the real application scenarios reported in the figure. Despite these positive aspects, during our experimental evaluation we have detected some failures; these

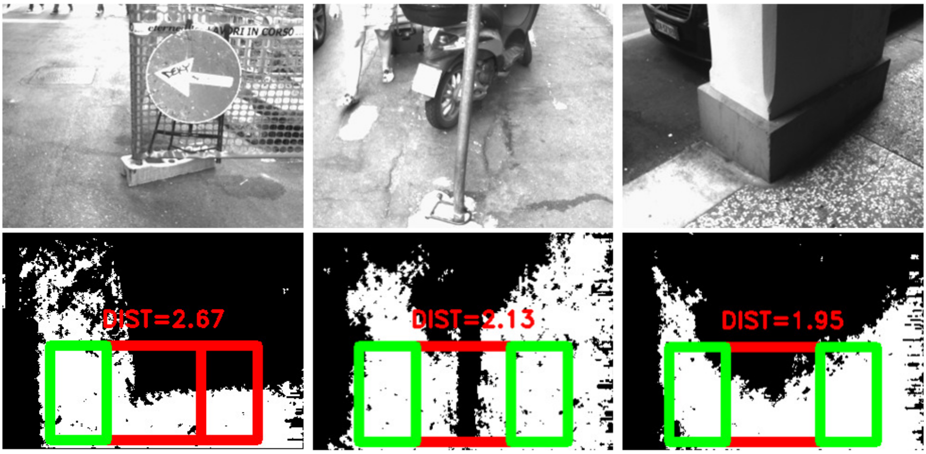

Fig. 5. Outcome of the obstacle detection algorithm in an urban scenario processing the 3D data provided by the camera worn by an user during the experimental evaluation.

facts mainly occur in indoor environments when the sensed area is completely uniform or made of transparent objects. Nevertheless, in these circumstances the resulting disparity map tends to be very noisy and thus, in most cases, these areas are correctly identified as obstacles.

Finally, we observe that compared to most ETS systems proposed in literature our prototype is extremely compact and lightweight. Moreover, thank to reliable depth map and the high frame rate (about 18 fps) our system is effective even when the sensed scene contains fast moving objects/people.

8 Conclusions and future work

In this paper we have proposed an effective and wearable system aimed at improving the autonomous mobility of people suffering of visual impairments. Our proposal relies for depth sensing on an embedded stereo camera mounted on a pair of glasses that enables, by means of a robust obstacle detection algorithm implemented into an embedded device, to provide effective and intuitive feedbacks to the user according to vibrotactile and audio actuators. Thank to the FPGA-based 3D camera and the embedded devices used the overall system, even in the current form, is extremely compact, lightweight and with a standard battery can work for hours.

Future work is aimed at porting the obstacle detection algorithm from the embedded Odroid U3 to mobile phones or tablet in order to implement the whole software system into a single device as well as to take advantage of the GPS receiver available in most of these devices. Moreover, we are going to apply machine learning techniques for object categorization that would allow to provide further audio information concerning the kind of obstacle sensed by our system.

References

1. Loomis, J.M., Golledge, R.G., Klatzky, R.L.: Navigation system for the blind - auditory display modes and guidance. Presence **7**(2) (1998) 193–203
2. Ran, L., Helal, S., Moore, S.: 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)
3. Helal, A.S., Moore, S.E., Ramachandran, B.: Drishti: An Integrated Navigation System for Visually Impaired and Disabled. In: ISWC '01: Proceedings of the 5th IEEE International Symposium on Wearable Computers. (2001)
4. Ceipidor, B.U., Medaglia, C.M., Rizzo, F., Serbanati, A.: RadioVirgilio/Sesamonet:an RFID-based Navigation system for visually impaired. In: Mobile Guide '06 (http://mobileguide06.di.unito.it/programma.html)
5. Bourbakis, N.G., Kavraki, D.: An intelligent assistant for navigation of visually impaired people. In: Proceedings of the 2Nd IEEE International Symposium on Bioinformatics and Bioengineering. (2001)
6. Shah, C., Bouzit, M., Youssef, M., Vasquez, L.: Evaluation of ru-netra - tactile feedback navigation system for the visually impaired. In: Virtual Rehabilitation, 2006 International Workshop on. (2006) 72–77
7. Ito, K., Okamoto, M., Akita, J., Ono, T., Gyobu, I., Takagi, T., Hoshi, T., Mishima, Y.: Cyarm: an alternative aid device for blind persons. In: CHI 2005. (2005)
8. Cardin, S., Thalmann, D., Vexo, F.: Wearable system for mobility improvement of visually impaired people. In: In Visual Computer journal. (2006)
9. Ni, D., Wang, L., Ding, Y., Zhang, J., Song, A., Wu, J.: The design and implementation of a walking assistant system with vibrotactile indication and voice prompt for the visually impaired. In: ROBIO. (2013)
10. Arcara, P., Di Stefano, L., Mattoccia, S., Melchiorri, C., Vassura, G.: Perception of depth information by means of a wire-actuated haptic interface. In: In IEEE Int. Conf. on Robotics and Automation. (2000)
11. Zelek, J., Audette, R., Balthazaar, J., Dunk, C.: A stereo-vision system for the visually impaired. Technical report, University of Waterloo (2000)
12. Balakrishnan, G., Sainarayanan, G., Nagarajan, R., Yaacob, S.: Wearable real-time stereo vision for the visually impaired
13. Sanz, P.R., Mezcua, B.R., Pena, J.M.S., Walker, B.N., Tipton, S.J., White II, D.J., Sershon, C., Choi, Y.B., Yu, F., Zhang, J., et al.: 1 evaluation of the sonification protocol of an artificial vision system for the visually impaired. Evaluation (2014)
14. Hub, A., Diepstraten, J., Ertl, T.: 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)
15. Rodríguez, A., Bergasa, L., Alcantarilla, P., Yebes, J., Cela, A.: Obstacle avoidance system for assisting visually impaired people. In: IEEE Intelligent Vehicles Symposium, Proceedings of Workshop Perception in Robotics. (2012)
16. Dakopoulos, D., Boddhu, S.K., Bourbakis, N.G.: A 2d vibration array as an assistive device for visually impaired. In: BIBE. (2007)
17. Fernandes, H., Costa, P., Filipe, V., Hadjileontiadis, L., Barroso, J.F.: Stereo vision in blind navigation assistance. In: 2010 World Automation Congress, WAC 2010. (2010)
18. Szeliski, R.: Computer Vision: Algorithms and Applications. 1st edn. Springer-Verlag New York, Inc., New York, NY, USA (2010)

19. Velázquez, R.: Wearable Assistive Devices for the Blind. Wearable and Autonomous Biomedical Devices and Systems for Smart Envi- ronment: Issues and Characterization (2010)
20. Kaczmarek, K.A., Webster, J.G., Bach-y Rita, P., Tompkins, W.J.: Electrotactile and vibrotactile displays for sensory substitution systems. IEEE Transactions on Biomedical Engineering (1991)
21. Meers, S., Ward, K.: A vision system for providing 3d perception of the environment via transcutaneous electro-neural stimulation. In: IV. (2004)
22. Johnson, L.A., Higgins, C.M.: A navigation aid for the blind using tactile-visual sensory substitution. Conf Proc IEEE Eng Med Biol Soc (2006)
23. Huang, B., Liu, N.: Mobile navigation guide for the visually disabled. Journal of the Transportation Research Board (1885) (2004) 28–34
24. Hard-Kernel: Odroid u3. `http://hardkernel.com/main/main.php`
25. Arduino: Arduino due. `http://www.arduino.cc`
26. Mattoccia, S.: Stereo vision algorithms for fpgas. In: The 9th IEEE Embedded Vision Workshop, CVPR 2013 Workshop. (2013)
27. Mattoccia, S., Marchio, I., Casadio, M.: A compact 3d camera suited for mobile and embedded vision applications. In: The Fourth IEEE Workshop on Mobile Vision, CVPR 2014 Workshop. (2014)
28. Zabih, R., Woodfill, J.: Non-parametric local transforms for computing visual correspondence. In: Proceedings of the Third European Conference-Volume II on Computer Vision - Volume II. ECCV ’94 (1994) 151–158
29. Hirschmüller, H.: Stereo processing by semiglobal matching and mutual information. IEEE Trans. Pattern Anal. Mach. Intell. **30**(2) (2008) 328–341
30. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM **24**(6) (1981) 381–395
31. Labayrade, R., Aubert, D.: In-vehicle obstacles detection and characterization by stereovision. In: in Proceedings the 1st International Workshop on In-Vehicle Cognitive Computer Vision Systems. (2003) 13–19
32. Raguram, R., Chum, O., Pollefeys, M., Matas, J., Frahm, J.M.: Usac: A universal framework for random sample consensus. IEEE Trans. Pattern Anal. Mach. Intell. **35**(8) (2013) 2022–2038
33. Matas, J., Chum, O.: Randomized ransac with td, d test. Image Vision Comput. **22**(10) (2004) 837–842
34. Torr, P.H.S., Zisserman, A.: Mlesac: A new robust estimator with application to estimating image geometry. Computer Vision and Image Understanding **78** (2000) 2000
35. Choi, S., Kim, T., Yu, W.: Performance evaluation of ransac family. In: BMVC. (2009)