Sound of Vision – 3D scene reconstruction from stereo vision in an electronic travel aid for the visually impaired

M. Owczarek, P. Skulimowski, P. Strumillo

Institute of Electronics, Lodz University of Technology, Poland,
{mateusz.owczarek, piotr.skulimowski, pawel.strumillo}@p.lodz.pl

Abstract. The paper presents the preliminary results for the parametrization of 3D scene for sonification purposes in an electronic travel aid (ETA) system being built within the European Union's H2020 Sound of Vision project. The ETA is based on the concept of sensory substitution, in which visual information is transformed into either acoustic or haptic stimuli. In this communication we concentrate on vision-to-audio conversion i.e. employing stereovision for reconstruction of 3D scenes and building a spatial model of the environment for sonification. Two prerequisite approaches for the sonification are proposed. One involves the direct sonification of the so-called "U-disparity" representation of the depth map of the environment, while the other relies on the processing of the depth map to extract obstacles present in the environment and presenting them to the user as auditory icons reflecting specific size and location of the sonified object.

Keywords: electronic travel aid, stereo vision, UV-disparity, ground plane detection, obstacle detection

1 Introduction

Numerous solutions for navigation of the blind and visually impaired were proposed over the past few years. Such solutions can be divided into two groups. The first group are assistive applications aimed at supporting the visually impaired in solving specific, usually isolated issues. Very often they use off-the-shelf solutions such as mobile phones or tablets with such examples as text-to-speech applications, image magnifiers, audio description devices, banknote recognizers etc. [1, 2, 3]. The second group, referred to as Electronic Travel Aids (ETAs), are systems which aim at enhancing orientation and mobility (O&M) of the blind users. Research work on O&M devices for the visually impaired dates back to the turn of 19th and 20th centuries. First ETA device is attributed to Kazimierz Noiszewski, Polish optician who built Electroftam. This was a device that used light-sensitive selenium cells to convert light energy into sound and tactile stimulations. Another notable attempt was made

by Bach-y-Rita in the 1970s who built a system converting images recorded by a video-camera into a pattern of haptic stimulations generated by a matrix of vibrating actuators positioned at the back of the user [4]. These devices, however, were bulky and consumed too much power to be practical in use.

In spite of nearly 100 year's effort since first attempts of Noiszewski, none of the solutions has found a wide-spread acceptance among the blind community [5]. That is why research and development studies in this area are being carried out extensively and involve auditory display and/or tactile interfaces [6, 7].

Contemporary, solutions for O&M and ETA for the visually impaired can be subdivided into a number of solutions:

- Simple obstacle detectors – small hand-held devices that sample distance to objects of the environment.
- Navigation and telenavigation systems – employing either GPS (navigation) or establishing wireless connection with a remote guide (telenavigation) to help the blind user in local and global wayfinding tasks in the environment.
- Environment imagers – more complex vision based systems that convert images (or their regions) of the environment into auditory or haptic representations.

Examples of simple obstacle detectors are Ultracane, Laser Cane, Teletact or Miniguide [8].

Examples of the GPS navigation systems for the blind are the American Trekker Breeze by Human Ware and Polish Nawigator by Migraf. Telenavigation systems on the other hand, are still at a prototype technology readiness level and were mainly developed by Brunel University, UK [9] and Lodz University of Technology [10].

Examples of environment imager systems are: the Sonicguide (and its newer version KASPA), the VOICe [11] and similar SVETA [12], Espacio Acustico Virtual [13], and the Sonic Pathfinder [14]. The Sonic Pathfinder was the first environment imager system that processed the sensed data about the environment to limit the amount of information for acoustic presentation. This is an important advance in the approach to ETAs operation since most of the visually impaired users complain about overabundance of auditory information that is used to convey spatial features of the environment. Similar approach was followed in the Naviton project [15]. A pair of camera in a stereovision setting and algorithms for reconstructing 3D structure and building a simplified scene model of the environment were employed.

In this paper we present the preliminary results for the parametrization of 3D scene for sonification purposes – one of the concepts within the Sound of Vision project [6]. The idea behind the project is to create a wearable ETA for assisting the visually impaired individuals by rendering the image of the environment through the auditory display. The device captures the 3D image of the environment, processes it, and presents the relevant information about the environment to the user.

2 Image acquisition and 3D reconstruction

In order to filter out relevant information that is communicated to the blind user the images of the surrounding environment need to be segmented into disjoint regions that are further classified as ground plane regions and regions representing other objects. So detected objects are described by a set of geometric parameters. Thus, a specific 3D model of the environment is built for the sonification purposes. Such an approach is novel in comparison to the environmental imager devices outlined in the earlier chapter.

The task of three dimensional scene reconstruction can be considered as a problem of determination of points' coordinates of the objects within the scene. Among the existing reconstruction methods stereovision was chosen due to its simplicity, effectiveness in natural light conditions (passive method), and small size and weight – particularly important in the case of the blind users. What is more, the use of video cameras allows to implement additional, auxiliary image processing algorithms, such as text recognition [3], objects semantic analysis, etc.

The schematic diagram of the proposed method for 3D scene modelling and parametrization is depicted in Fig. 1. It consists of the three key processing steps:

1. Estimation of the ground plane location based on disparity and depth maps.
2. Obstacle detection after background and ground plane removal.
3. Parametric description for each of the detected obstacle.

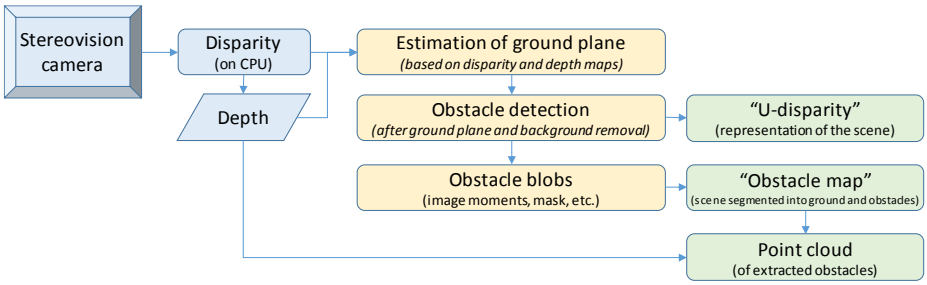


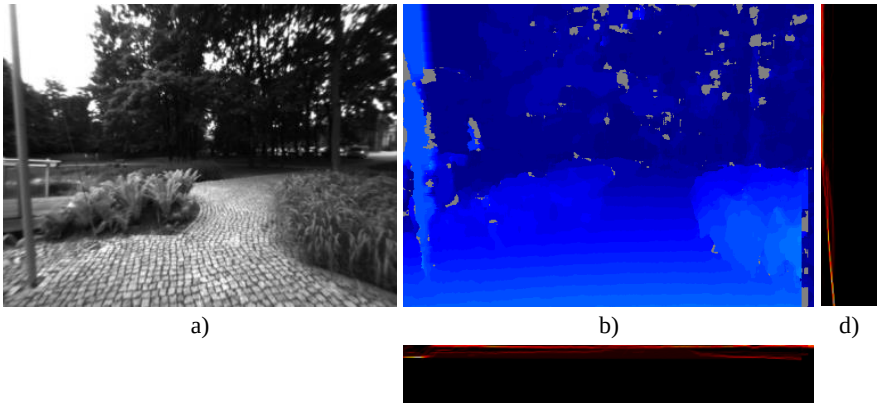
Fig. 1. Schematic diagram of the proposed method for 3D scene reconstruction, and ground plane and obstacle detection

For the sake of simplicity, let us assume that images from the stereo vision camera (Fig. 2a) are devoid of geometric distortions and rectified, which is a prerequisite for the calculation of the disparity map. The disparity map in turn is used to convert pixel coordinates from the captured images into world point coordinates in the camera coordinate system, from which depth of the obstacles can be computed [5, 17].

a) b) c)

Fig. 2. Prototype wearable image acquisition device: **a)** Duo MLX stereo vision camera build into a custom-made "glasses" and **b)** user wearing them during early tests of the device, while **c)** processing unit is held in a backpack

Our broad literature search [2, 7, 16, 17, 18 and 19] shows that one of the most effective methods for ground plane and obstacle detection, provided that the stereo vision camera is used, are solutions based on the so-called "UV-disparity" representation of the disparity map. The key feature in such an approach is that pixels representing the ground plane are well reproduced in the "V-disparity" map (represented by a distinctive line segment – see Fig. 3d) while obstacles are clearly visible in the "U-disparity" map.



a) b) d)

c)

Fig. 3. Example test scene: **a)** (reference) image as seen by the right camera, **b)** disparity image, **c)** "U-disparity" and **d)** "V-disparity" representations of the disparity map

The line segment in the "V-disparity" representation, which corresponds to the ground plane area, is found using the Hough transform. All the pixels in the disparity map for which the disparity value fits the detected line equation are removed from the disparity map. Those pixels are also taken into consideration while calculating the ground plane equation. In the considered application, there is no need to present the obstacles that are very distant from the observer (more than 8 m). The same applies to the obstacles located high enough above the system user's head.

3 Obstacle detection and sonification

Further analysis is performed based on the disparity map with all the pixels corresponding to the ground plane, background and those for which distance from the ground plane is greater than the pre-selected value removed. Example of such disparity image is depicted in Fig. 5b. It may be noticed, that it contains only the elements of the environment that may endanger the blind pedestrian. Having this information as input for the next steps, we proposed two approaches, depending on the preferred way of presenting the information about the environment to the user.

3.1 Direct sonification of the "U-disparity"

One of the ideas is a direct sonification of the "U-disparity" representation, calculated for the disparity map after pre-processing steps. Note that in the "U-disparity" map the obstacles are represented by distinctive line segments (see Fig. 4). What is more, based on the y-coordinate (vertical coordinate) in the "U-disparity" map it is possible to estimate the distance to the face of the nearest obstacle. It is worth mentioning, that such an approach may lead to a simple sonification scheme such as left-to-right scene scanning of the environment.

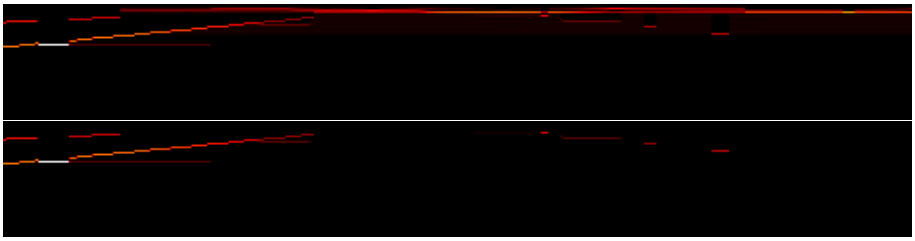


Fig. 4. "U-disparity" representation of the reference disparity image (top) and after applying the preprocessing steps described to this point (bottom)

3.2 The concept of the simplified sound icons

The second approach consists of binarization of the earlier processed disparity map (Fig. 5b). This approach was implemented by the contour finding method based on [20]. The aim of this next step is to find the minimum-area bounding rotated rectangles for the detected obstacles (Fig. 5c). Regions of the area below a certain threshold are not taken into account. Results of obstacles detection for two example 3D scenes are shown in Fig. 5d.

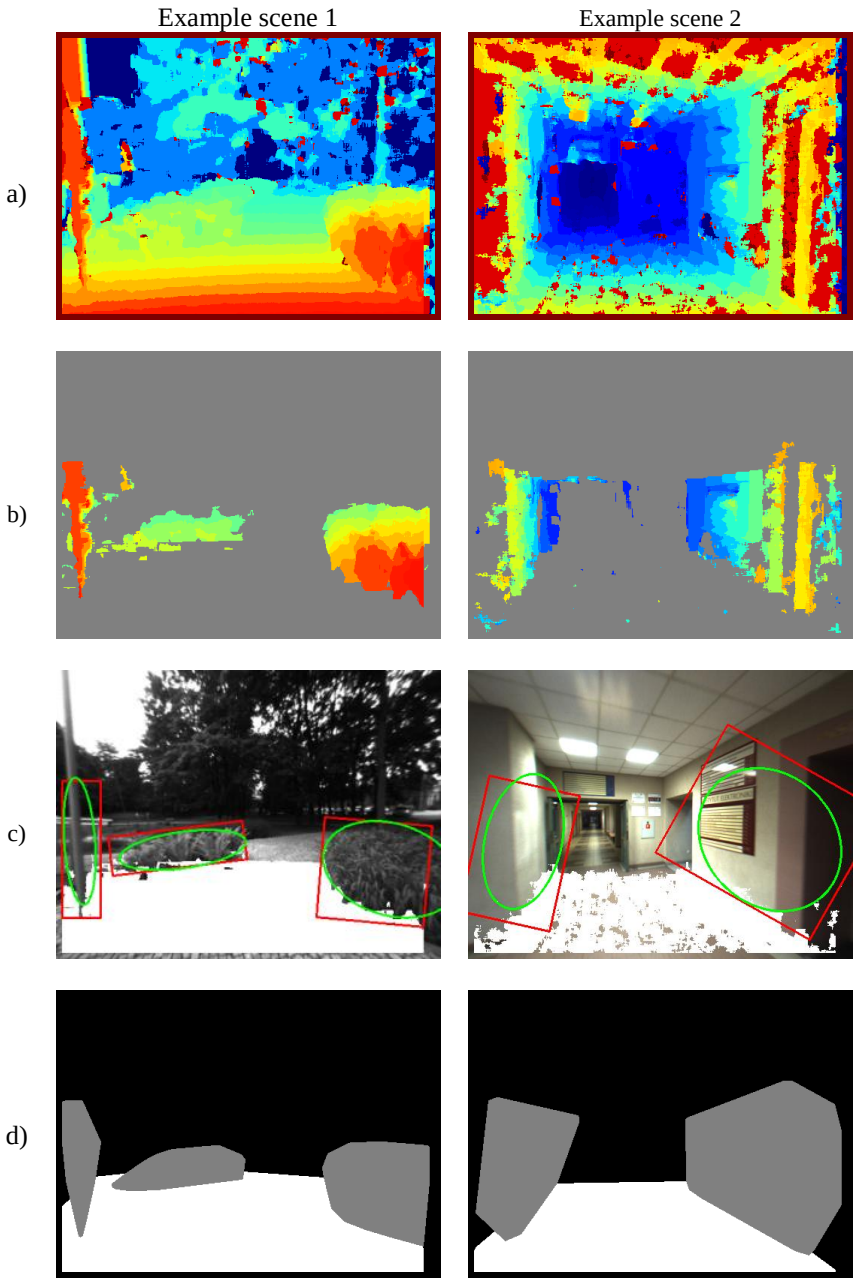


Fig. 5. Results of obstacles detection from stereovision images: **a)** raw disparity map, **b)** result of the processing of the disparity map, **c)** minimum-area bounding rotated rectangles outlining the detected obstacles, **d)** result segmentation map

Based on the extracted ROIs (Regions of Interest) a set of obstacles' parameters can be calculated. For the sonification purposes we proposed a set of parameters for the detected obstacles that were calculated based on the obstacles regions in color images and the computed disparity (and depth) maps. These parameters are grouped into the following categories: color-related features (e.g. dominant color), depth-related features, disparity-related features, two dimensional features (e.g. area, centroid or image moments, etc.) and three dimensional features (e.g. bounding rectangular cuboid, volume, etc.). Selected parameters will serve as the data source for various scene sonification schemes [21].

4 Conclusions

In this study we have undertaken the problem of 3D scene modelling and parametrization from stereo vision. To address this issue, a simple and effective method of detecting obstacles and the ground plane has been developed. It works in real-time, confidently in the outdoor environment and is satisfactory for indoor scenes (we encourage the reader to view our supplementary material, e.g. video sequences, available at http://icchp2016.naviton.pl/). Obstacles can be detected even if the quality of the disparity map is low, provided that the ground plane occupies significant part of the image and is clearly visible. The weakness of this solution is that only significantly large obstacles can be detected for now, as it is very hard to detect small and potentially dangerous obstacles using the stereo vision camera with a short baseline.

Acknowledgment. This work received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 643636 "Sound of Vision".

References

1. Hersh, M., Johnson, M.: Assistive technology for visually impaired and blind people, Springer, London (2008)
2. Ivanchenko, V., Coughlan J., Huiying, S.: Detecting and locating crosswalks using a camera phone. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshop, CVPRW '08. on, pp.1–8, 23–28 June 2008
3. Huiying, S., Coughlan, J.: Reading LCD/LED displays with a camera cell phone. IEEE Conference on Computer Vision and Pattern Recognition Workshop, CVPRW '06, pp.119–119, 17–22 June 2006
4. Bach-y-Rita P.: Brain mechanisms in sensory substitution. Academic Press (1972)
5. Strumiłło, P.: Electronic personal navigation systems for the blind and visually impaired (in Polish), Faculty of Electrical, Electronic, Computer and Control Engineering, Lodz University of Technology, Lodz (2012)
6. Sound of Vision: Natural sense of vision through acoustics and haptics, available online: http://www.soundofvision.net/ [accessed on March 25th, 2016]

7. Flores, G., Kurniawan, S., Manduchi, R., Martinson, E., Morales, L.M., Sisbot, E.A.: Vibrotactile guidance for wayfinding of blind walkers. IEEE Transactions on Haptics, vol. 8, no. 3, pp. 306–317, July–Sept. 2015
8. Dakopoulos, D., Bourbakis, N.G.: 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)
9. Garaj, V., Jirawimut, R., Ptasinski, P., Cecelja, F., Balachandran, W.: A system for remote sighted guidance of visually impaired pedestrians. British Journal of Visual Impairment, vol. 21, pp.55–63 (2003)
10. Barański, P., Strumiłło, P.: Emphatic trials of a teleassistance system for the visually impaired. Journal of Medical Imaging and Health Informatics, vol. 5, no. 8, pp. 1640–1651 (2015)
11. vOICe, available online: www.seeingwithsound.com [accessed on March 25th, 2016]
12. Balakrishnan, G., Sainarayanan, G., Nagarajan, R., Sazali, Y.: A stereo image processing system for visually impaired. International Journal of Signal Processing, vol. 2, no. 3, pp. 136–145 (2006)
13. González–Mora, J., Rodríguez–Hernández, A., Rodríguez–Ramos, L., Díaz–Saco, L., Sosa N.: Development of a new space perception system for blind people, based on the creation of a virtual acoustic space. Engineering Applications of Bio-Inspired Artificial Neural Networks, pp. 321–330, Springer Berlin/Heidelberg (1999)
14. Heyes, D., The Sonic Pathfinder: a new electronic travel aid, Journal of Visual Impairment and Blindness, vol. 77, pp. 200–202, 1984
15. Bujacz, M., Skulimowski, P., Strumillo, P.: Naviton – a prototype mobility aid for auditory presentation of 3D scenes, Journal of Audio Engineering Society, vol. 60, no. 9, 696–708 (2012)
16. Chunlei, Y., Cherfaoui, V., Bonnifait, P.: Evidential occupancy grid mapping with stereovision. IEEE Intelligent Vehicles Symposium (IV), pp. 712–717, June 28 2015–July 1 2015, doi: 10.1109/IVS.2015.7225768
17. Zhencheng, H., Uchimura, K.: U-V-disparity: an efficient algorithm for stereovision based scene analysis. IEEE Intelligent Vehicles Symposium, pp. 48–54, 6–8 June 2005, doi: 10.1109/IVS.2005.1505076
18. Labayrade, R., Aubert, D., Tarel, J.-P.: Real time obstacle detection in stereovision on non flat road geometry through "V-disparity" representation. IEEE Intelligent Vehicle Symposium, vol.2, pp. 646–651, 17–21 June 2002, doi: 10.1109/IVS.2002.1188024
19. Zhencheng, H., Lamosa, F., Uchimura, K.: A complete U-V-disparity study for stereovision based 3D driving environment analysis. Fifth International Conference on 3-D Digital Imaging and Modeling, pp. 204–211, 13-16 June 2005, doi: 10.1109/3DIM.2005.6
20. Suzuki, S. and Abe, K.: Topological structural analysis of digitized binary images by border following. Computer Vision, Graphics, and Image Processing, vol. 30, no. 1, pp. 32–46, 1985
21. Bujacz, M., Kropidlowski, K., Rotnicki, M., Witek, P., Ivanica, G., Moldoveanu, A., Saitis, C., Spagnol, S., Johannesson, O.I., Unnborsson, R.: Sound of Vision - spatial audio output and sonification approaches In: Computers Helping People with Special Needs of the series Lecture Notes in Computer Science 2016 (in press)