Development of a Versatile Assistive System for the Visually Impaired Based on Sensor Fusion

Nicolae Botezatu, Simona Caraiman
„Gheorghe Asachi” Technical University of Iasi
Romania

Dariusz Rzeszotarski, Pawel Strumillo
Lodz University of Technology
Poland

Abstract — **In this paper we describe the 3D acquisition component integrated in the Sound of Vision (SoV) system. SoV is a computer vision based sensory substitution device (SSD) for the visually impaired. Its main objective is to provide the users with a 3D representation of the environment around them, conveyed by means of the hearing and tactile senses. One of the biggest challenges for the SoV system is to ensure pervasiveness, i.e., to be usable in any indoor or outdoor environments and in any illumination conditions. To this end, the proposed 3D acquisition system was designed based on a fusion of data from different types of sensors. We present the hardware and software solution for developing this acquisition system and provide insight on its exploitation in various computational scenarios of the SoV system. In the first indoor trials with the system, carried out in modeled indoor environments, the visually impaired volunteers were capable of performing simple navigation tasks and avoiding cardboard box obstacles.**

Keywords—sensory substitution; data fusion; structured light; stereo vision; inertial measurement device

I. INTRODUCTION AND RELATED WORK

Visual impairment or vision loss is a severe condition that seriously affects the life of the individuals suffering from it. The ability to see, gives people access to the world around, from details of facial expressions to color, shape, size of the objects, and to the hazards from the environment that could endanger their lives. Blind persons face challenges performing everyday activities that sighted take for granted, like reading or walking. One of the most important problems for blind people when moving in open or closed environments is the lack of external references, creating a distortion of absolute directions, of the position of their heads and bodies related to streets or furniture.

The development of aids for helping the visually impaired to perceive the environment, to orientate and navigate has been the subject of many research works in the past two decades. The reported efforts to support the rehabilitation of visually impaired have been directed towards the development of electronic travel aids (ETAs) and sensory substitution devices (SSDs). An ETA is a form of assistive technology with the purpose of enhancing mobility for the blind user [1, 2]. Sensory substitution devices are designed to convey visual information to the visually impaired by substituting visual information into one of their intact senses [3-9]. These devices employ non-invasive human–machine interfaces, which, in the case of the blind, transform visual information into auditory or tactile representations using a predetermined transformation algorithm.

Although other environment sensing techniques, like sonar or radar, have shown promising results, computer vision methods have more potential for providing an appropriate representation of the environment in real-world settings, which are noisy and difficult to interpret. Creating such a representation implies acquiring information and filtering it in order to provide the user with information that is not confusing and does not sensory overload the user [10-14]. Moreover, as a general trend, higher quality image sensing devices are becoming cheaper, smaller and more widely available.

Analyzing the state of development for these assistive systems from the perspective of the end-user, we find that a plethora of works have been reported in the literature [7-9, 15-31]. However, there are still some important steps to be taken before large communities of visually impaired users embrace this technology. The reasons for not having such consumer grade systems largely available for the end-user are related to many factors, such as form factor, lack of efficient training program or general limitations of visual rehabilitation.

In this paper we describe the 3D acquisition component of the Sound of Vision (SoV) system [32]. The SoV system is a non-invasive, wearable sensory substitution device that assists visually impaired people by creating and conveying an auditory and tactile (haptic) representation of the surrounding environment. This representation is created based on computer vision techniques, updated and conveyed to the blind users continuously and in real time. The objective of the SoV system is to aid both the perception and the navigation of visually impaired users in unknown environments.

The main challenging requirements for the SoV system are (i) to provide users with real-time feedback regarding the structure of the environment, (ii) to work in both indoor and outdoor environments, (iii) irrespective of the illumination conditions and, (iv) to be wearable. These general requirements translate into technical requirements for the computer vision component of the SoV device. They specifically have an important impact on the design of the 3D acquisition system.

Many other computer vision based assistive systems for the blind have tackled the problem of environment sensing and understanding. However, very few of them consider the pervasiveness aspect [29, 30, 31] and work either in indoor or outdoor environments. This limitation mainly comes from the integrated 3D sensors. The infrared-based sensors, e.g., Kinect, do not cope with bright illumination from the sun. Stereo sensors provide unreliable depth estimations in the presence of poor artificial lighting or uniformly colored/textured surfaces specific to indoor environments. The system described by Kurata et al. [30] obtains positioning data from several sensing sources such as GPS, Wi-Fi, PDR (Pedestrian Dead Reckoning), image-based registration, and active RFID (if the infrastructure is in place), and integrates them based on each uncertainty. Road-network data is also employed for map matching. A local navigation aid for impaired users is included through the use of a laser range finder (LRF) to detect obstacles in the path. An obstacle-map is rendered on a tactile display, also used for Braille output. The authors also report the positioning error and coverage of each sensor and method on an indoor and outdoor trial course. The VeDi system [31] provides another showcase for indoor and outdoor navigation by integrating vision-based with pedestrian-localization systems. The authors report a custom designed system that demonstrates how partially sighted people could be aided by the technology in performing an ordinary activity, like going to a mall and moving inside it to find a specific product. Computer vision techniques for detection, recognition, and pose estimation of specific objects or features in the scene are combined with a hardware-sensor pedometer. This enables the system to derive an estimate of user location with sub-meter accuracy. The user is guided through the outdoor section of the route (from a bus stop to the sliding doors representing the entrance of the mall) using a PDR system. Navigation in the indoor environment is performed using a Visual Navigator that searches for specific visual beacons (signs or environmental features). It notifies the user as soon as a target is detected in the scene, it communicates the direction to follow in order to reach it, and the action needed to pass from one target to the next. Moreover, the navigator warns the user in the case that dangerous situations are detected (i.e. a wet floor sign). A Structure from Motion algorithm is used to predict the movement of the user with respect to a starting point. The drifts introduced by the algorithm are compensated by the PDR.

Sensor fusion has also been exploited in the Navig project [28]. It uses GPS, two IMUs (Inertial Measurement Unit), one for body heading and a second one for head orientation, an adapted GIS and a stereo vision module. The vision module serves two functions: object localization and user positioning. Both functions rely on the estimation of the distance between a target and the user. The global navigation task is supported by the user positioning mode. In this mode, the system looks for geolocated landmarks tagged in the GIS. During the journey, according to the estimated location of the user, models corresponding to nearby targets are automatically loaded (e.g. signs, facades, logos, etc.). When detected, these visual landmarks (called visual reference points) are not rendered to the user but are used to refine the current GPS position. This function can provide a position estimate relying exclusively on embedded vision. It can then be fused with GPS data to improve positioning; it can also be used in situations where the GPS positioning is faulty or absent.

While all these systems employ a form of data fusion from different sources, they only address the navigation problem and do not focus on the sensory substitution approach. The SoV system tackles the pervasiveness requirement by integrating both an Infra Red (IR) depth sensor and a stereo vision system, together with an IMU device for recovering the head orientation. The main goal is to provide depth information in any environment (indoor or outdoor) and in any illumination conditions.

In the reminder of this paper, we present the hardware and software solution devised for the acquisition component of the SoV system. The presentation modules of the system, however, are not covered in this work. For example see [8] where a more detailed discussion is provided on "auditory display" methods of 3D scene images. Finally, we provide a general insight on the SoV system exploitation possibilities for ensuring its overall intended functionality.

II. System design and Implementation

The underpinning idea of the designed system is to build a wearable assistive device aiding the visually impaired in independent mobility and travel both in indoor and outdoor environments. This is a challenging task since different space scales and illumination conditions characterize theses environments. The novelty of the designed assistive system, not discussed in detail in this work, is represented by the employment of non-visual presentation techniques of the environment that combine audio and haptic modalities. Such a solution increases the capacity of the sensory channels aiming at substituting the lost vision. However, the key part of the system is the acquisition sub-system of the 3D processing module responsible with the identification and selection of the objects of 3D space for presentation.

This acquisition sub-system was designed with three objectives in mind: (i) to provide consistent input for the processing algorithms used by the SoV system in multiple usage scenarios (e.g. indoor/outdoor, different lighting conditions), (ii) to use affordable acquisition devices and (iii) to have a flexible structure.

The result is a system with a modular design, with off-the-shelf components and fusion techniques of our design to preprocess raw input data. The system is powered by an i7 Intel processor notebook with most of the critical image processing procedures ported to a Graphical Processing Unit (GPU) platform.

A. Hardware

The acquisition devices are placed onto a rigid structure, which can be easily attached to various headgear designs (Fig. 1).

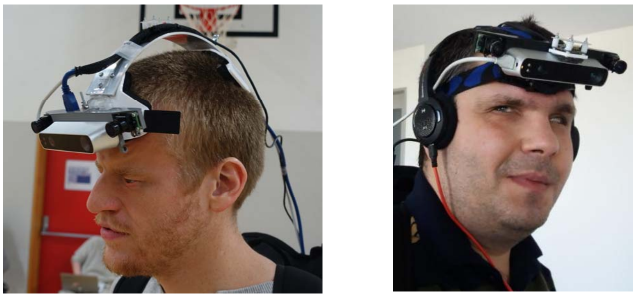

Fig. 1. Acquisition devices support attached to two different headgear designs. (Left) Rigid acrylic structure resembling VR headset implementations. (Right) Lightweight headgear with elastic strap bands.

TABLE I. MAIN CHARACTERISTICS OF THE ACQUISITION DEVICES

Device	Parameter	Value
LI-OV580 Stereo Camera	Max. resolution	4416 x 1242 @ 15fps
	Sensor type	OmniVision 4M CMOS
	Sensitivity	1900 mV/lux
	Format	1/3”
	Pixel size	2 x 2 µm
	FOV	109°/88°/72° (D/H/V)
Structure Sensor PS1080	Max. resolution	640 x 480 @ 30fps
	Operating range	0.4 - 3.5 meters
	Precision	0.5 mm @ 0.4 m 30 mm @ 3 m
	FOV	58°/45° (H/V)
LPMS-CURS2 IMU	Max. data rate	400 Hz
	Output data format	Raw data / Euler angle / Quaternion
	Resolution	< 0.01°
	Accuracy	< 2° dynamic < 0.5° static

The acquisition devices are described below. Table I also presents their main characteristics. All the devices are connected to the SoV central processing unit via a USB 3.0 hub.

- A stereo RGB camera (SC) - LI-OV580 from Leopard Imaging - used for outdoor image capture. The two cameras are mounted on separate Printed Circuit Boards (PCBs) and are connected by wire to the central unit. The main advantage of this design is that the baseline can be configured specifically for the application.
- A Depth-of-Field (DoF) camera (SS) - Structure Sensor PS1080 from Occipital - used for indoor or low light image capture. It has an integrated battery that can extend its usage period by 3 to 4 hours.
- An Inertial Measurement Unit (IMU) - LPMS-CURS2 from Lp-Research - is used for tracking the head/body movement. It integrates three sensors (accelerometer, gyroscope, magnetometer) and a fusion engine that processes sensor data and outputs heading data by means of a Kalman filter.

B. Operating modes

The 3D Acquisition system has four operating modes - Stereo camera input, Structure sensor input, Stereo-Structure dual input and recording playback (the first three are pictured in Fig. 2). The first three modes are designed for the real-time use of the SoV system, whereas the playback one is implemented for offline use in order to test, tune, and customize the processing algorithms used further down the pipeline.

From a functional software perspective, the 3D Acquisition system is based on four distinct modules - one input module for each acquisition device (i.e. stereo camera, structure sensor, IMU device) and one main module for data synchronization, aggregation and preprocessing.

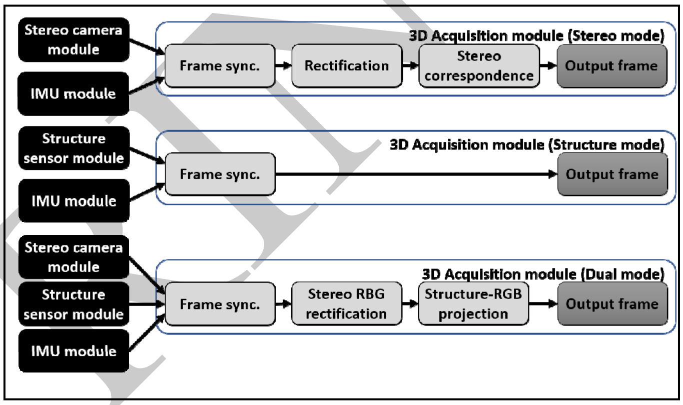

Fig. 2. Overview of the operating modes of the 3D Acquisition module

1) Stereo mode

The Acquisition module captures Stereo frames, synchronizes them with the IMU data, rectifies the left and right images and then applies a stereo correspondence algorithm (Elas or SGBM) in order to compute the disparity map. The main improvement compared to the previous SoV prototype consists in acquiring and rectifying stereo image pairs of a larger resolution, i.e., 1280 x 720. In the previous prototype the 640 x 480 resolution of the sensor was used. However, by experimental testing, it was found that with this resolution, the actual field of view of the camera differed from the technical specification provided by the manufacturer.

2) Structure mode

The Acquisition module captures Structure frames (depth frames) and synchronizes them with the IMU data. No further preprocessing is performed in the acquisition module.

3) Dual mode

The Acquisition module captures Stereo frames, synchronizes them with the IMU data, rectifies the left and right images and then optionally runs a mapping procedure between RGB and depth frames or disparity and depth frames. If the second mapping is active, a stereo correspondence algorithm (Elas or SGBM) is used to compute the disparity map.

C. Calibration procedure

In order to function properly, processing algorithms that make use of stereo images require the physical characteristics of the cameras used. Due to the manufacturing process of the headgear (i.e. 3D printing accuracy and glitches) and due to the mounting process (i.e. manual positioning and fastening, parts tensioning), each assembled headgear is a unique item and the cameras must be calibrated individually (a special re-calibration procedure to be run by the user was also developed). Moreover, for the dual acquisition mode (i.e. projection of the depth output of the Structure sensor onto the RGB data from the Stereo camera) both devices must be calibrated together.

The calibration process makes use of a chessboard calibration pattern that must be placed at about 2 meters in front of the cameras. About 50 images are taken (Fig. 3) with the pattern in different orientations in order to compute the geometric distortions introduced by the cameras as well as the extrinsic parameters describing the transform between each pair of cameras (left + right, left + IR, right + IR), where by IR we denote the infra-red sensor of the Structure Sensor device. The processing of the images is done with a custom developed calibration tool.

Some difficulties arise from the fact that the images from the Structure sensor (i.e. the device is equipped with an IR camera) must be taken beside the images from the stereo camera: the projected IR pattern from the Structure sensor is not suited for the chessboard pattern (i.e. the contour of the squares loses sharpness) and as a result the IR pattern generator must be masked and an alternate IR light source must be used (e.g., an incandescent lamp).

The output of the calibration process consists in a set of intrinsic, extrinsic and distortion parameters between the left-right RGB cameras, left RGB-IR cameras, and IR-right RGB cameras.

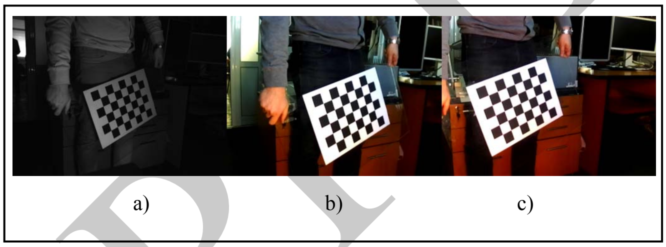

Fig. 3. Sample calibration images – a) IR camera output from the Structure sensor; b) left and c) right RGB Stereo camera output.

D. Synchronization

The data from the input devices (Stereo camera, Structure sensor, IMU) is captured on separate threads of the application, in order to maintain an acquisition rate close to the rates provided by the hardware. Thus, the need to synchronise the input streams arises.

The synchronization achieved is a “loose” one. This means that we take into account the fact that the acquisition threads present switching jitter due to the load of the system and the OSs task switching mechanics (Fig. 4). Tests show that on a Windows 7 system with modest resources the jitter with both negative/positive values does not exceed 25% of the capture period at 30fps and cancels itself after 10-15 frames.

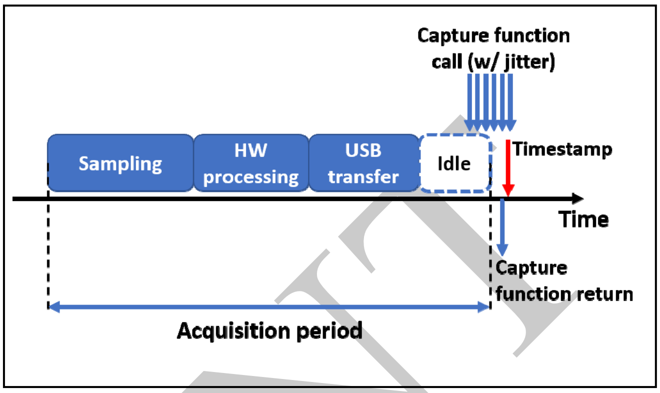

Fig. 4. Generic acquisition timeframe

All captured data is timestamped (based on one of the system's steady clocks) and then a matching process is run based on the acquisition mode used:

- *Stereo mode* - The IMU acquisition rate (100 fps) is higher than the one from the Stereo camera (15 fps), so for every camera frame, all IMU samples that were captured in the 1/15 s period that ended when the Stereo timestamp was set are matched to the video frame.
- *Structure mode* - Similar to the Stereo mode with different acquisition rate from the Structure sensor (30 fps).
- *Dual mode* - In this mode, the time reference for the process is represented by the Stereo frame timestamp (i.e. it has the smallest sample rate) and for each Stereo frame, two Structure frames are matched. Afterwards, the IMU samples are matched as described above.

E. Depth map remapping

The fusion between the output of the two imaging devices is necessary for the myriad of processing algorithms employed by the SoV system. The module implements two types of remapping: depth onto RGB and depth onto disparity.

The depth map from the Structure Sensor is reprojected onto the left rectified image from stereo camera using the following steps:

1. Using intrinsic parameters of the Structure Sensor, a 3D point cloud map is obtained from the depth map;

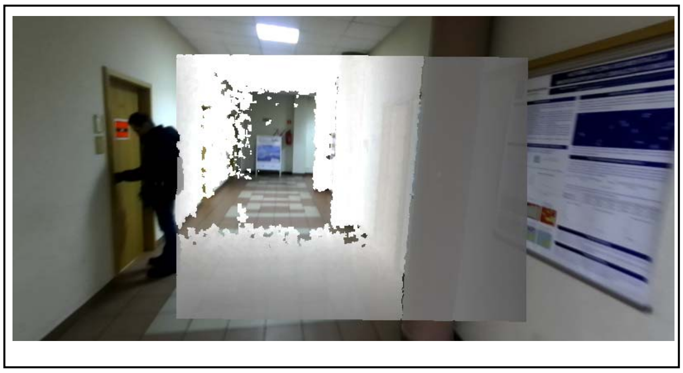

Fig. 5. Reprojection of the depth map from the Structure Sensor onto left rectified image

2. For each element from the 3D points cloud map a respective 3D coordinate in the left rectified camera reference frame is calculated;
3. Each 3D coordinate obtained in the previous step is projected onto the rectified image using intrinsic parameters of the left rectified camera image;
4. The depth map reprojected onto the left rectified image is a two-dimensional array of the same size as the size of the left rectified image;
5. The image coordinates obtained in the previous step are rounded to integer points to the element of two-dimensional array to which a respective depth value should be assigned. If more than one depth value is assigned to the same element of the array the smallest value is chosen.

Figure 5 shows the depth image reprojected onto the left rectified image from the Stereo camera and illustrates the difference between the FoV of the two input devices.

For the depth onto disparity remapping, the reprojection of the depth map onto the left rectified RGB image is aligned with the computed disparity map. Therefore, the two maps can be fused together in the following way:

- Recalculate the disparity values from stereo into depth values;
- For each element in the depth map check if it is valid (has nonzero depth value). For valid elements, substitute an element in the disparity map from the Stereo camera with the respective element from the reprojected depth map from the Structure Sensor. Valid elements from the Structure Sensor depth map are regarded as superior over the respective elements from the Stereo camera depth map.

As a result of this procedure we get a precise depth map for the Structure Sensor FoV and less accurate for the larger FoV of the stereovision system. The former enables high precision tracking of the ground plane and the latter is required for detection of obstacles for a wider FoV.

III. Applications

A. General usage

The 3D Processing module of the SoV project exploits different combinations of sensor data to maximize the system usability in different situations and still provide environmental information in conditions atypical to normal CMOS sensors. Table II summarizes the supported operation scenarios.

TABLE II. USAGE SCENARIOS FOR DIFFERENT ENVIRONMENTS AND ILLUMINATION

Env.	Lighting conditions	Inputs	Processing
Indoor	Normal light	Depth map (SS) Left color map (SC)	Planar-surface detection and combination into objects; Door, texts, sign detection (color & depth maps fusion)
	Low light / complete dakness	Depth map (SS)	Planar-surface detection and combination into objects;
Outdoor	Normal light	Left/right color maps (SC)	Object detection (disparity map histogram segmentation); Camera movement estimation for object tracking (Stereo pairings)
	Low light / complete darkness	Depth map (SS)	Object detection (depth map histogram segmentation);

One important topic when it comes to binding the output of the acquisition system to the processing pipeline is that of the extents of the encodable space around the user. The encodable space is defined by the hardware parameters of the cameras that make up the 3D acquisition system. Specifically, the encodable space is in the form of a pyramidal frustum defined by the field of view and the depth range. The two types of cameras embedded in the system have different values for these parameters resulting in two definitions of the encodable space.

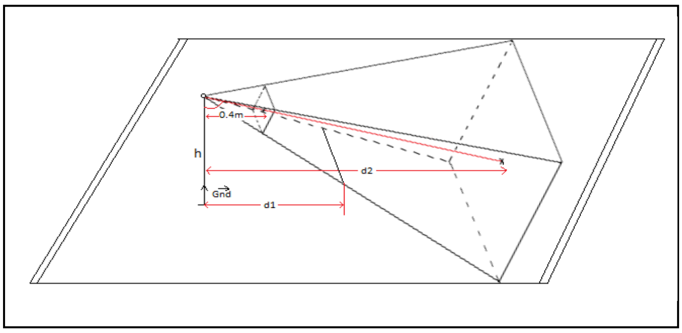

Fig. 6. Encodable space in front of the user depends on the FOV and depth range of the sensor. d1 represents the distance to the closest point on the ground that can be measured by the sensor. d2 represents the maximum depth of points that can be measured

The user height and the orientation of the cameras with respect to the ground determine the minimum and maximum distances at which the system can report the presence of head-height and ground level obstacles (Fig. 6). The tilt of the cameras' support can be adjusted to optimize these parameters for each user height.

The depth of objects closer than 40 cm to the cameras cannot be retrieved neither by the structure, nor by the stereo sensor. Thus, these objects will not be identified by the 3D processing module. However, integrating an inter-frame consistency mechanism, based, for example, on camera motion estimation, allows a global 3D model to be build. With such a model, objects further away are "memorized" by the system and reported to the user even when they become closer than 40 cm to the cameras. If such an approach is not available, the task of keeping track ("memorizing") of the object is up to the user. This is the case of the indoor, structure sensor based, operating mode, where camera motion estimation cannot be reliably exploited.

The depth resolution refers to the accuracy with which a system can estimate changes in the depth of a surface. If estimated with a stereo sensor, it is proportional to the square of the depth and is inversely proportional to the focal length and baseline. Thus, the depth estimation errors increase quadratically with the distance from the sensor. The range of depth refers to the minimum and maximum distances of objects that can be measured by the stereo vision system for a given maximum disparity. While the stereo vision sensor can provide depth estimations at distances larger than 10 m, they are very much affected by errors (at 10m the depth error is approx. 25 cm, while at 20 m the depth error is approx. 104 cm).

Table III summarizes the actual values for the field of view and the ranges of distances to obstacles on ground level and at head level for the two acquisition modalities. The cameras are tilted individually for each user such that head height obstacles (including 0.3 m above and below the camera level) are observed by the sensors at the minimum distance (0.4 m). Thus, the minimum distance to the closest observable object on the ground only depends on the user height.

TABLE III. ENCODABLE SPACE PARAMETERS. THE DISTANCE TO THE CLOSEST OBSERVABLE POINT CORRESPONDS TO A USER HEIGHT OF 1.7M

Parameter	Structure sensor	Stereo camera
Field of view - horizontal	58 deg.	88 deg.
Field of view - vertical	45 deg.	72 deg.
Min distance to observable ground level object (d1)	1.70 m	1.04 m
Min distance to observable head level object	0.4 m	0.4 m
Max distance to observable ground/head level object (d2)	5 m	10 m

B. Ground plane detection

A very important step in the detection of objects in the environment is represented by the correct estimation of the ground surface. One approach is to segment the 3D global model into objects after excluding the regions corresponding to the ground. Moreover, the ground information can be exploited for computing the best free space in front of the user and for the detection of holes on the pathway.

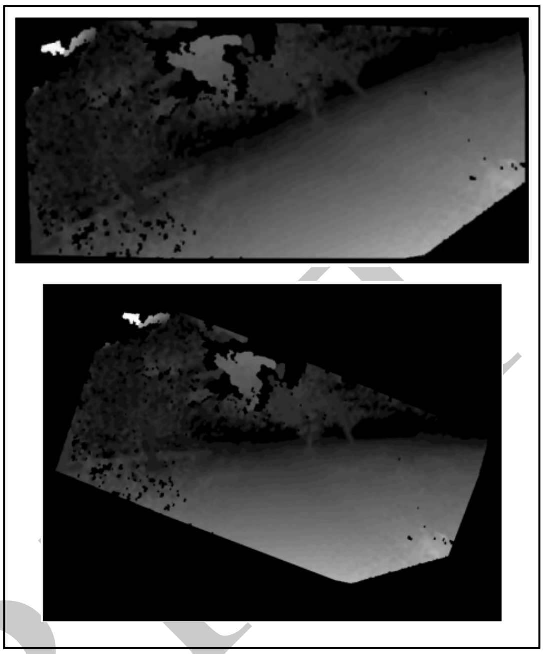

Fig. 7. Example of stereo-IMU data fusion for accurate ground plane detection. Disparity maps are rotated based on the IMU data.

The main approach for ground detection in the stereo processing pipeline is to compute a vDisparity histogram in which the pixels corresponding to the ground surface are expected to form a line [33]. This assumption does not hold in the presence of camera roll. To overcome this limitation, we recover the orientation of the camera based on the synchronized data from the IMU device mounted on the SOV headgear. We rotate the image along the z axis around the optical center (cx, cy) of the camera using cubic interpolation. This rotation ensures that the ground surface pixels correspond to a straight line in the vDisparity image (Fig. 7).

C. Doors detection

The problem of door detection is very important in systems aiding visually impaired users in independent travel in urban spaces. Detecting doors using a white cane can be problematic because they may be confused with corners, niches in the corridors or with pillars.

The detection mechanism makes use: i) of the ground plane detection algorithm presented in the previous subsection (i.e. with the synchronized data from the Stereo camera and the IMU device) and ii) of line detection and door model fitting algorithms that are based on the remapping procedure of the depth map onto the left RGB image [34] (Fig. 8).

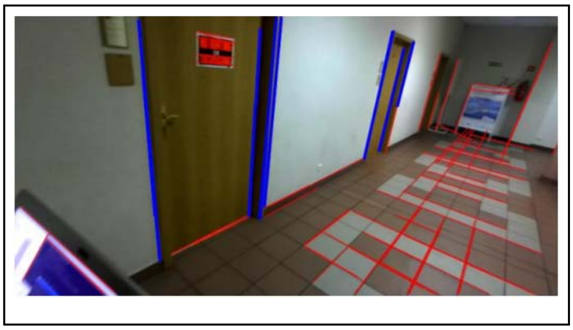

Fig. 8. Door detection example. The detection is performed in indoor scenes based on the fusion of depth and color data coming from the structured light sensor and the stereo vision system, respectively.

IV. Conclusions

This paper presents a novel approach to the subject of building a flexible 3D acquisition system with off-the-shelf components. The proposed system relies on input data fusion from a stereo RGB camera, a structured light sensor and an IMU device in order to deliver consistent data to image processing algorithms in conditions atypical to normal CMOS sensors. This will ensure pervasiveness of the sensory substitution system in which the proposed acquisition system is integrated. The tests of the system are planned at different illumination and indoor/outdoor environments. The wearability aspect of the system is tackled by mounting the acquisition components on a rigid structure easily attachable to various headgear designs. The height of the user is an important factor that influences the extents of the encodable space around the user. More specifically, it influences the minimum and maximum distance at which ground level and head level obstacles can be sensed by the system. The tilt of the headgear can be easily customized in order to optimize these distances for each user.

Further improvement of the design for the acquisition system will consider incorporating latest releases of more lightweight structured light based sensors. The total weight of the acquisition component currently does not exceed 200 g, which is less than the weight of state of the art VR headgears. However, usability studies carried out with participation of the visually impaired volunteers show that prolonged use of the headgear (more than 1 hour) can cause discomfort or fatigue to the users.

Acknowledgment

The work presented in this paper has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 643636 "Sound of Vision".

References

[1] M. Hersh and M. A. Johnson, Assistive Technology for Visually Impaired and Blind People, 1st ed. Springer Publishing Company, Incorporated, 2008.

[2] D. Dakopoulos and N. G. Bourbakis, "Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 1, pp. 25–35, Jan 2010

[3] M. Auvray, S. Hanneton, and J. ORegan, "Learning to perceive with a visuo-auditory substitution system: Localization and object recognition with the voice," Perception, vol. 36, no. 3, p. 416430, 2007.

[4] G. Bologna, B. Deville, T. Pun, and M. Vinckenbosch, "Transforming 3d coloured pixels into musical instrument notes for vision substitution applications," EURASIP International Journal of Image and Video Processing, vol. 2007, no. 2, pp. 1–15, 2007

[5] S. Levy-Tzedek, D. Rimer, and A. Amedi, "Color improves 'visual' acuity via sound," Frontiers in Neuroscience, vol. 8, no. 358, 2014.

[6] S. Abbouda, S. Hanassya, S. Levy-Tzedek, S. Maidenbaum, and A. Amedi, "Eyemusic: Introducing a visual colorful experience for the blind using auditory sensory substitution," Restorative Neurology and Neuroscience, vol. 32, no. 2, 2014.

[7] L. Dunai, G. Peri-Fajarnes, E. Lluna, and B. Defez, "Sensory navigation device for blind people," THE JOURNAL OF NAVIGATION, vol. 66, p. 349 362, 2013.

[8] M. Bujacz, "Representing 3d scenes through spatial audio in an electronic travel aid for the blind," 2010. PhD Thesis, Technical University of Lodz.

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

[10] R. Jafri, S. Ali, H. Arabnia, and S. Fatima, "Computer vision-based object recognition for the visually impaired in an indoors environment: a survey," Visual Computing, vol. 30, pp. 1197 – 1222, 2014.

[11] J. Terven, J. Salas, and B. Raducanu, "New opportunities for computer vision-based assistive technology systems for the visually impaired," Computer, vol. 47, pp. 52 – 58, 2014.

[12] T. Pun, P. Roth, G. Bologna, K. Moustakas, , and D. Tzovoras, "Image and video processing for visually handicapped people," EURASIP Journal on Image and Video Processing, pp. 1 – 12, 2007.

[13] S. Maidenbaum, S. Abboud, and A. Amedi, "Image and video processing for visually handicapped people," Neuroschience and Biobehavioral Reviews, vol. 41, pp. 3 – 15, 2014.

[14] R. Manduchi and J. Coughlan, "(computer) vision without sight," Commun. ACM, vol. 55, pp. 96–104, Jan. 2012.

[15] J. M. Saez Martinez and F. Escolano Ruiz, "Stereo-based Aerial Obstacle Detection for the Visually Impaired," in Workshop on Computer Vision Applications for the Visually Impaired, (Marseille, France), James Coughlan and Roberto Manduchi, Oct. 2008.

[16] A. Rodriguez, J. J. Yebes, P. F. Alcantarilla, L. M. Bergasa, J. Almazan, and A. Cela, "Assisting the visually impaired: Obstacle detection and warning system by acoustic feedback," Sensors, vol. 12, no. 12, pp. 17476–17496, 2012.

[17] S. Mattoccia and P. Macri, "3d glasses as mobility aid for visually impaired people," in Proc. of the ECCV2014 Workshop, 2014.

[18] G. Balakrishnan, G. Sainarayanan, R. Nagarajan, and S. Yaacob, "A stereo image processing system for visually impaired," International Journal of Signal Processing, vol. 2, no. 3, p. 136, 2008.

[19] E. Peng, P. Peursum, L. Li, and S. Venkatesh, "A smartphone-based obstacle sensor for the visually impaired," in Ubiquitous Intelligence and Computing (Z. Yu, R. Liscano, G. Chen, D. Zhang, and X. Zhou, eds.), vol. 6406 of Lecture Notes in Computer Science, pp. 590–604, Springer Berlin Heidelberg, 2010.

[20] J. Jose, M. Farrajota, J. M. Rodrigues, and J. H. du Buf, "The smartvision local navigation aid for blind and visually impaired persons," JDCTA: International Journal of Digital Content Technology and its Applications, vol. 5, no. 5, pp. 362 – 375, 2011.

[21] L. Chen, B.-L. Guo, and W. Sun, "Obstacle detection system for visually impaired people based on stereo vision," in Genetic and Evolutionary Computing (ICGEC), 2010 Fourth International Conference on, pp. 723–726, Dec 2010.

[22] J. Saez, F. Escolano, and M. Lozano, “Aerial obstacle detection with 3d mobile devices,” IEEE J Biomed Health Inform, vol. 19, pp. 74 – 80, 2015.

[23] R. Tapu, B. Mocanu, A. Bursuc, and T. Zaharia, “A smartphone-based obstacle detection and classification system for assisting visually impaired people,” in Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on, pp. 444–451, 2013.

[24] P. Costa, H. Fernandez, P. Martins, J. Barroso, and L. Hadjileontiadis, “Obstacle detection using stereo imaging to assist the navigation of visually impaired people,” Procedia Computer Science, vol. 12, pp. 83 – 93, 2012.

[25] V. Filipe and et. al., “Blind navigation support system based on microsoft kinect,” Procedia Computer Science, vol. 14, no. 0, pp. 94 – 101, 2012.

[26] S. Wang, H. Pan, C. Zhang, and Y. Tian, “Rgb-d image-based detection of stairs, pedestrian crosswalks and traffic signs,” Journal of Visual Communication and Image Representation, vol. 25, no. 2, pp. 263–272, 2014.

[27] Y. H. Lee, T.-S. Leung, and G. Medioni, “Real-time staircase detection from a wearable stereo system,” in 21st International Conference on Pattern Recognition (ICPR 2012), pp. 3770–3773, 2012.

[28] S. Kammoun, G. Parseihian, O. Gutierrez, A. Brilhault, A. Serpa, M. Raynal, B. Oriola, M.-M. Mac, M. Auvray, M. Denis, S. Thorpe, P. Truillet, B. Katz, and C. Jouffrais, “Navigation and space perception assistance for the visually impaired: The NAVIG project,” IRBM, vol. 33, no. 2, pp. 182 – 189, 2012.

[29] L. Ran, S. Helal, and S. Moore, “Drishti: An integrated indoor/outdoor blind navigation system and service,” in IEEE International Conference on Pervasive Computing and Communications, pp. 23–32, 2004.

[30] T. Kurata, M. Kourogi, T. Ishikawa, Y. Kameda, K. Aoki, and J. Ishikawa, “Indoor-outdoor navigation system for visually-impaired pedestrians: Preliminary evaluation of position measurement and obstacle display,” in Wearable Computers (ISWC), 2011 15th Annual International Symposium on, pp. 123–124, June 2011.

[31] P. Chippendale, V. Tomaselli, V. D’Alto, G. Urlini, and C. Modena, “Personal shopping assistance and navigator system for visually impaired people,” in Proc. of the CVPR2014 Workshop”, 2014.

[32] Sound of Vision (SOV) - Natural sense of vision through acoustics and haptics, Horizon 2020 No 643636, http://www.soundofvision.net/

[33] P. Herghelegiu, A. Burlacu and S. Caraiman, "Robust ground plane detection and tracking in stereo sequences using camera orientation," *2016 20th International Conference on System Theory, Control and Computing (ICSTCC)*, Sinaia, 2016, pp. 514-519.

[34] P. Skulimowski, M. Owczarek, P. Strumillo, Door Detection in Images of 3D Scenes in an Electronic Travel Aid for the Blind, Submitted: 10th International Symposium on Image and Signal Processing and Analysis (ISPA 2017)