Door Detection in Images of 3D Scenes in an Electronic Travel Aid for the Blind

Piotr Skulimowski, Mateusz Owczarek, Pawel Strumillo
Institute of Electronics, Lodz University of Technology, Poland
Email: {piotr.skulimowski, mateusz.owczarek, pawel.strumillo}@p.lodz.pl

***Abstract*—In this paper we propose a fast method for detecting doors in images of 3D scenes. First, the equation estimating the orientation and location of the ground surface is computed. This information is used in further processing steps of the algorithm. Then, the edge image is calculated (using the Canny edge detector) and line segments justifying predefined conditions are searched for by applying the Probabilistic Hough Transform method. Pairs of parallel line segments perpendicular to the ground surface located at a distance range 80-110 cm are identified. The detection performance has been also enhanced by detecting door handles. The proposed method was successfully verified on the recorded indoor RGB-D video sequences acquired by a vision based Electronic Travel Aid (ETA) for the blind. The achieved door detection performance for the tested sequences is at a level of 63% for sensitivity and 84% for positive predictivity values.**

I. Introduction

The number of visually impaired people living on Earth approaches 300 million. Interestingly, the use of mobile phones and to a large extend personal computers is quite ubiquitous among people with sight disability. This is because, numerous applications such as: speech synthesizers, optical magnifiers, text and color readers help the visually impaired in efficient use of computing and communicating technologies. On the other hand, there have been also numerous research attempts to develop systems that would successfully aid the visually impaired in independent mobility and travel [1]–[3]. Here, in spite of dynamical advancements in Information and Communication Technologies (ICT) practical introduction of Electronic Travel Aids (ETAs) has been slow. The attempts to develop efficient personal ETA technologies range from simple devices integrated with white canes like optical or ultrasound scanners (making the white canes smart) to more complex devices converting 3D scene geometry into audio or haptic modalities non-visually displaying the environment for a blind user. ETAs can also be subdivided into two categories that are either dedicated to micro-navigation tasks (like obstacle avoidance or shore-line following) or macro-navigation tasks such large-space spatial orientation and positioning [4]. Examples of micro-navigation devices are *K-Sonar Cane*, *UltraCane* or *Mini-guide* [2], whereas examples of systems for macro-navigation are *Trekker-breeze* (GPS-based navigation aid for the blind pedestrian) or prototype/academic systems employing radio or optical beacons embedded into the environment infrastructure [5].

Recently, due to availability of powerful libraries for image processing and analysis (e.g. OpenCV) more and more approaches attempt to recognize special, characteristic objects in a scene that can serve as an important navigation information for the blind persons. In particular, the problem of door detection is very significant in systems aiding the visually impaired users in the space perception and mobility [1], [6]–[8]. Detecting doors using just a white cane is not a trivial task, even if the cane is "smart" [2], [9], since ultrasonic or laser-based [10] sensors are no use if the door is closed. Another problem that blind people are facing with is the location of the door's handle and the direction or type of door opening (e.g. automatic sliding, folding, lateral, swing-door and so on).

In this paper we propose a fast method for detecting doors in 3D scenes. The method is one of the functional modules of an ETA device developed within the "Sound of Vision" project [3]. The device is equipped with a stereovision camera and the Structure Sensor [11] (structured-light 3D scanner). The Structure Sensor allows to obtain precise depth information in the indoor environment even in poor lighting conditions, stereovision camera is used as a source of color images and as a depth information in the outdoor environment. The aim of the project is to detect obstacles, identify their locations and their geometric parameters and present them using auditory or haptic display methods. The assumption is generally not to recognize type of objects but rather to interpret them as a generic object category. However, some of the scene objects like: ground surface and doors have been assigned a special interpretation. Ground detection is important, since this is the scene regions devoid of obstacles. Whereas doors detected in a scene can be important navigation cues in open spaces like in-door corridors or can be helpful in out-door environment in locating entrances to buildings. The aim of the door detection procedure in the "Sound of Vision" device is to locate both the door wing and its handle in the scene.

This kind of information will be sonified using one of the models presented in [3], [12], [13] or a message will be rendered using a text-to-speech (TTS) system to inform the user about the presence of doors in the processed scene.

The rest of this paper is organized as follows: in Section II we review door detection algorithms and discuss advantages and disadvantages of different image processing approaches to this problem. The proposed algorithm is explained in Section III. Results verifying the performance and robustness of the proposed algorithm are presented and commented in

978-1-5090-4011-7/17/$31.00 © 2017 IEEE

Section IV. Finally, Section V concludes the paper with a summary of the presented work and outlines the foreseen application of the algorithm in an ETA for the visually impaired.

II. Related work

The findings of our literature search on applications of door detection show that the problem is mainly addressed in studies concerning systems guiding autonomous vehicles [14] and robots [10], [15]–[17]. In such target applications the problem of door detection is simplified in comparison to the applications dedicated for the visually impaired. First of all, the camera in robot systems usually remains at a constant height above the floor level. Robots and automotive vehicles move smoothly on the floor what allows to estimate the equation of the floor plane reliably [15]. Secondly, the camera is specially mounted so that it allows to visualize the ground plane and ground obstacles. This is why many proposed methods use information from the bottom parts of the door, like door gap or kick area which are visible in the images captured during movement [15].

Shalaby et al. [7] propose an algorithm based on the geometric properties of the quadrilaterals. The method is proposed in a solution dedicated for the blind users and is based on image domain without referring to depth maps. Because they use weak assumptions like polygon model fitting that is based on the distance between corners in image domain and polygon size threshold compared to the whole image area, the proposed method allows to detect doors in front of the user from a predefined distance only.

The method proposed by Zhao et al. [8] relies on the information from built-in sensors in modern smartphones. It measures the change of magnetic field when the blind user is passing through the door.

Kaleci et al. [10] use laser range data to detect open doors only.

Yuan et al. [14] propose a door detection method dedicated for the autonomous vehicles. It is designed for an indoor environment and uses depth information only. The authors assumed that the camera is always perpendicular to the plane defined by the door and the door dimensions are known *a priori*. They use information about random error of depth measurement (which increases with an increasing distance to the sensor) which results in a large area of missing values in depth map corresponding to the opened doors.

Chen et al. [15] detect doors using the camera mounted on a mobile robot. The proposed solution uses the AdaBoost [18] method working on a collection of features characteristic for doors like: color difference, texture change between doors and walls, a kick plate or door gap between the bottom door edge and the floor plane.

Zhou et al. [16] use Kinect sensor that is mounted nearly vertically to the ground. Their method is based on detecting vertical straight lines and compute depth difference between the left and right side lines.

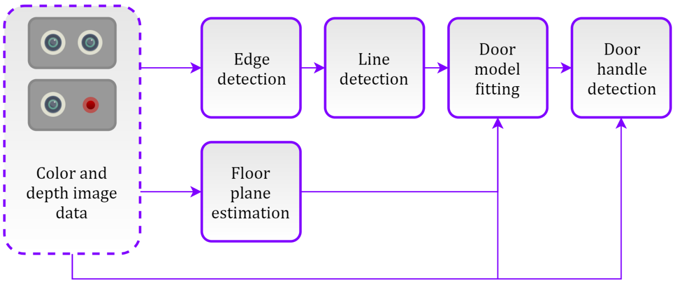


Fig. 1. Consecutive steps of the proposed door detection method

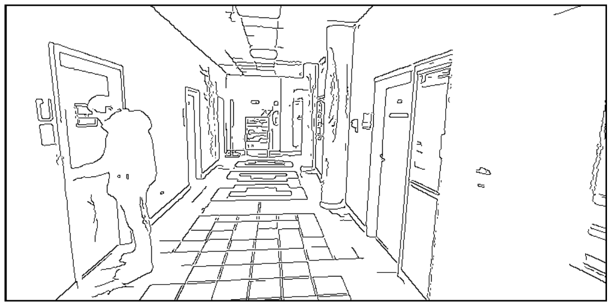


Fig. 2. Example result of Canny edge detection algorithm [20] applied to a color image of a corridor with doors

Meyer et al. [17] propose detection of a single-wing door based on normal vectors. Because handles are usually made of reflective materials that distort the depth map, this characteristic regions of the map are used to detect a door handle.

Finally, Sekkal et al. [19] use a model of a corridor in their door detection algorithm. The authors detect the ground plane and walls by identifying the vanishing point. The door search algorithm is executed for walls areas and is based on detecting rectangular shapes in 3D domain that match a predefined door model.

III. A method for detecting the door

A general scheme for detecting doors in images of 3D scenes is shown in Fig. 1. In the first step of the algorithm the color image is converted into a grey scale image. In the next step the image is filtered by the result of Canny edge detection algorithm [20] (Fig. 2). The points, for which depth values were not calculated or distance from the camera is larger than a predefined threshold (5 meters) are removed. This maximum distance value is a result of the assumed range of the Sound-of-Vision device and the maximum range of the Structure Sensor camera. All other parameters of the proposed algorithm were chosen experimentally.

In the further step of the algorithm the Probabilistic Hough Transform [21], [22] is applied to the edge image in order to detect the line segments. Depth values of these lines allow to choose line segments for further processing which can belong to doors. Then, each line segment is compared with each other and the (shortest) distance between them is calculated (in 3D space). The distance is defined as the minimum distance between any two points of these line segments. If the distance

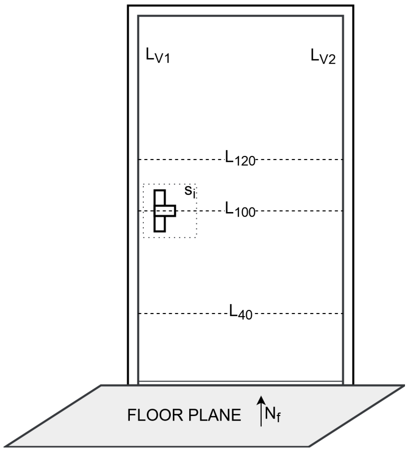


Fig. 3. A door model adopted in the proposed method

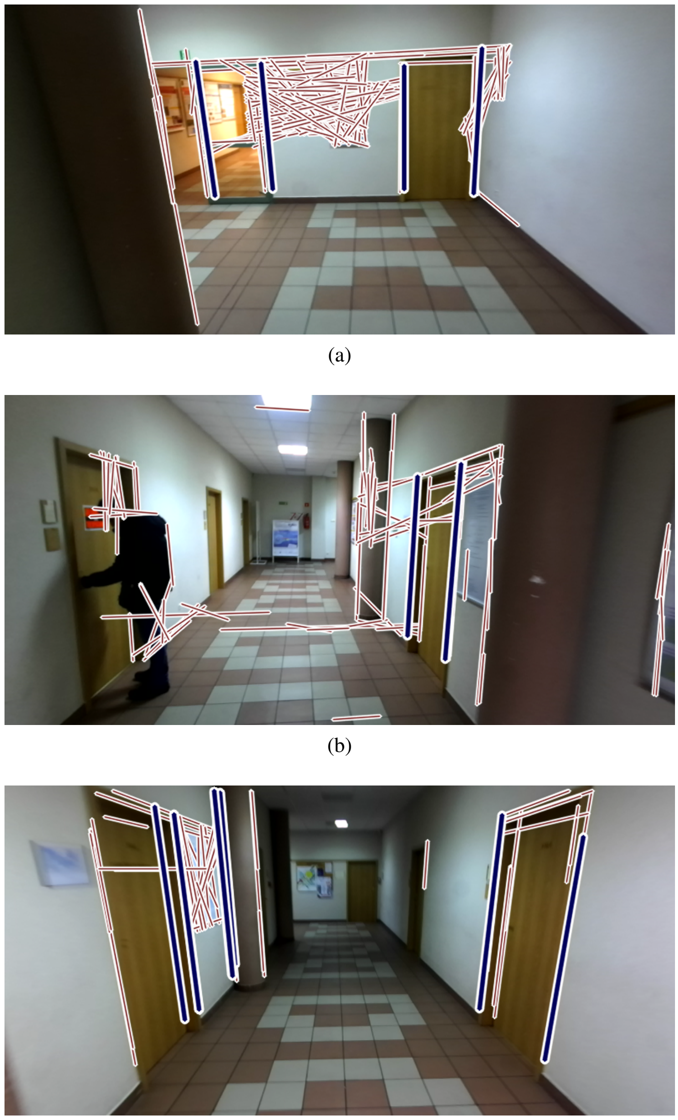
is within the predefined threshold (80–110 cm) these line segments are identified as belonging to doors. In order to eliminate false positive results the following filters are also applied:

— the angle between the floor plane's normal vector $\vec{N}_f$ and one of the door edges L_{V1}, L_{V2} must be less than 20 deg,
— the angle between door vertical lines' vectors L_{V1} and L_{V2} must be less than 20 deg,
— the distance from the floor plane of one of the ends of each vertical line segment L_{V1}, L_{V2} must be less than 30 cm and greater than 180 cm for another end,

The proposed door model is shown in Fig. 3.

These filters operate provided the parameters of the plane estimating the detected ground plane are given. For the test purposes of the proposed method a set of points from the bottom part of the image was chosen which fulfill the following criteria: Z values of these points come from the Structure Sensor, for each horizontal line of the image we choose points for which distance from the camera is the largest (they should belong to the ground plane). Ground plane equation is calculated with the assumption, that the summed squared distance between points and plane candidate is minimized. Location of the horizontal lines is not of much use since they are often invisible in the edge images or they are out of field of view of the camera. In order to eliminate false positive results (like one can notice in the images in Fig. 4) it was decided to detect and localize door handles. At first, a line segment L_{100} in the image domain connecting points from the left and right vertical line segments of doors which are located 1 meter above the floor plane is found.

Then, 20 points equidistant from this line segment are detected, for each point i sum of pixels s_i for a square neighborhood in the edge image is calculated. If for the first or last 3 points (the handle can be on the left or right side in the door wing) s_i reaches maximum value $s_{max} = max(s_i)$ and this value is larger than 1.3 of a minimum value for the all s_i values, the region of interest corresponding to this point is identified as a candidate door handle. Again, to eliminate



(a)

(b)

(c)

Fig. 4. Example result of door model fitting: vertical black lines correspond to door's edge candidates (a–c)

false positive results the same procedure is applied to the line segments located 40 cm and 120 cm above the ground plane (line segments L_{40} and L_{120}). For these line segments s_i values should be less than 0.7 of s_{max}. This condition allows to eliminate false positive results which can appear e.g. on wall posters between doors. The results of door handles' detection are shown in Fig 5.

The proposed method requires that almost the entire surface of the door is visible in the image and depth values should be calculated for this area. To solve this problem it was proposed to use points representing door handles as the key points that can be tracked through the video sequence. A pyramidal implementation [23] of the Lucas-Kanade feature tracker [24] was employed for this task. If the consecutive frames captured by the camera are not blurry, it is possible to track door handles even if a previously detected door is only partially visible in

(a)

(b)

(c)

Fig. 5. Examples results of door edges (represented as vertical black lines) and door handle (represented as a red horizontal line with a black dot on one of the ends) detection: two different kinds of doors are properly detected (a), door on the left is occluded by a person, therefore is left undetected; doors located at a distance further than 5 meters are not taken into account (b), a false-positive result, where a poster is detected as a door (c)

the image. The use of ego-motion parameters [25] to predict the location of door handles might also improve detection performance of this algorithm step.

The depth information comes from the fusion of Structure Sensor depth data and the disparity map from a stereovision camera (recalculated into depth appropriately), using the method proposed in [26]. The processing speed on Intel® Core™ i7-6700HQ 2.6 GHz machine slightly depends on the number of lines found by the Hough Transform and estimates at about 15 fps including the Canny image processing and a simple method for ground plane estimation based on a method proposed in [27] and with the use of points from the Structure Sensor depth data. It is worth noting that it is possible to use any different method of edge detection and ground plane equation estimation in the proposed algorithm.

IV. Results

The proposed method for door detection was verified on test image sequences which consist of 50 randomly selected frames from 850 frames for which ground truth annotations of door location were prepared. We annotated doors which are fully visible and are located at a distance of no larger than 5 meters. A detected door is defined as a true positive if the Jaccard similarity coefficient (JC) between its bounding rectangle and the bounding rectangle of the corresponding annotated (ground truth) door is larger than or equal to 50%. A false positive result means that the detected door has no corresponding annotated door in the ground truth set. Only one-to-one matches are supported (no splits and merges). A confusion matrix for door detection results results is shown in Table I.

TABLE I
Confusion matrix for door detection results

True positive $(TP) = 46$	False negative $(FN) = 27$	Sensitivity $(TPR) = TP/(TP + FN)$ $= 46/(46 + 27)$ $= 63\%$
False positive $(FP) = 9$	True negative $(TN) = 3$	Specificity $(SPC) = TN/(FP + TN)$ $= 3/(9 + 3)$ $= 25\%$
Positive predictive value $(PPV) = TP/(TP + FP)$ $= 46/(46 + 9)$ $= 84\%$	Negative predictive value $(NPV) = TN/(FN + TN)$ $= 3/(27 + 3)$ $= 10\%$	

False positive rate $(FP/(FP+TN))$ is 75%, false negative rate $(FN/(TP+FN))$ is 37%, accuracy $((TP+TN)/(TP+FP+FN+TN))$ is 58%. Fig. 6 shows the number of annotated and detected doors in each frame from the test sequence.

The proposed method was also verified qualitatively on the sequence recorded at a different location. Example frames which show door detection results are presented in Fig. 7.

Detailed analysis of the results shows that the main source of errors is the quality of images recorded by a camera in motion. Some images are blurry which results in a poor line detection and in inability in calculation of depth values. Moreover, the proposed door detection method has been improved by detecting door handles also. Frequently, however, door handles are difficult to detect because they are poorly visible. This in mainly due to shadows of door frames cast onto the handles. Qualitative analysis of the results for the test sequence shows that all doors were detected at least several times. Further extension of the method will focus on tracking the detected doors.

V. Summary

In this paper we propose an efficient algorithm for door detection in images of 3D scene acquired by a geometrically

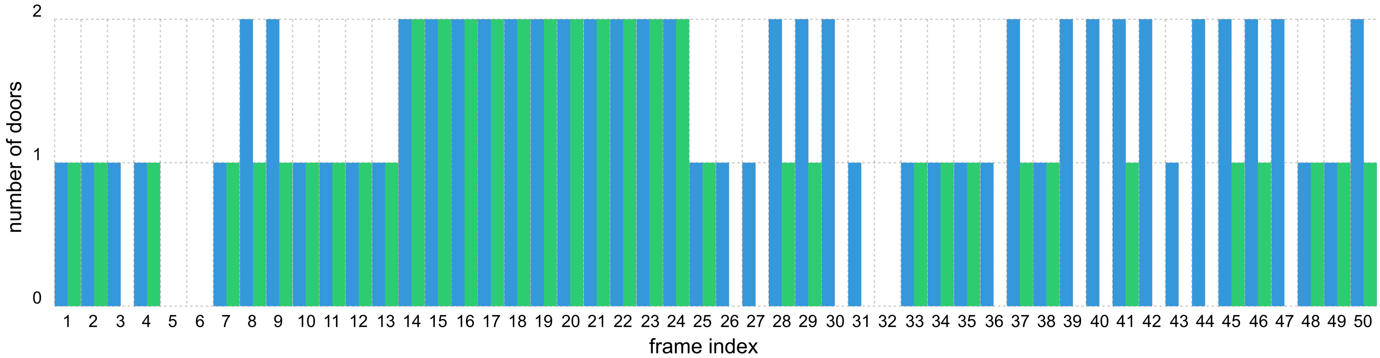
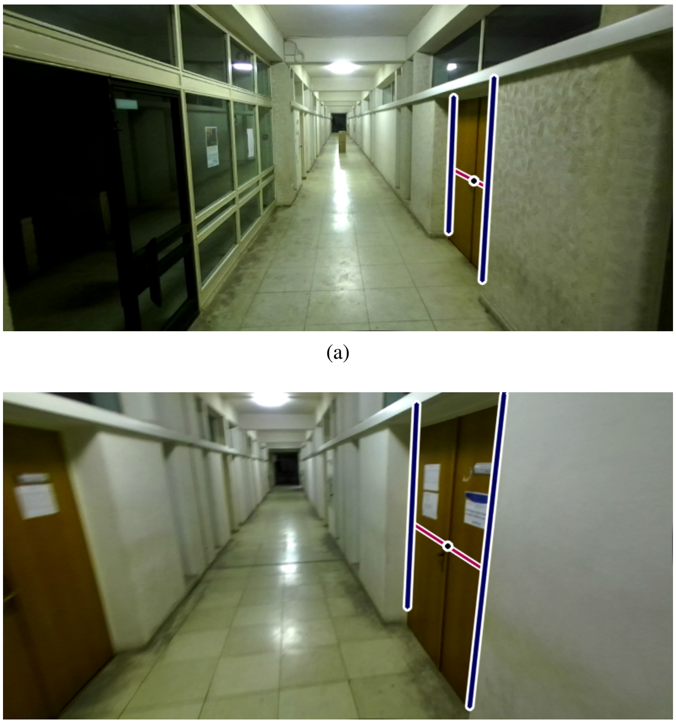


Fig. 6. Annotated (TP+FN – blue color) and detected (TP – green color) doors in each frame



(a)

(b)

Fig. 7. Examples of door detection results.

calibrated system comprising stereovision rig and a structured light camera.

The proposed method allows to detect doors in the predefined width and height ranges. It is based on the depth map of the scene and its efficiency strongly depends on the estimation accuracy of the ground plane equation. The main difficulties encountered during verification of the methods are the following: doors can be made of different materials (for example wood and glass), can be of different dimensions than the assumed ones, can be located behind stairs, e.g. in indoor environments. Also, narrow field of view of the cameras causes that doors that are partially visible cannot be correctly detected.

High performance of the method for door detection was verified on real-world indoor image sequences acquired by a blind individual using our system. Whereas the sensitivity of door detections is at a level of 64% the positive predictivity values rich 84%, i.e. once the system informs the user that a door is detected for 84 out of 100 cases this is in fact a door.

Acknowledgment

This project was supported by the European Union's Horizon 2020 Research and Innovation Programme 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] P. Strumillo, M. Bujacz, P. Baranski, P. Skulimowski, P. Korbel, M. Owczarek, K. Tomalczyk, A. Moldoveanu, and R. Unnthorsson, *Different approaches to aiding blind persons in mobility and navigation in the "Naviton" and "Sound of Vision" projects*, 1st ed. Springer International Publishing, 2017.

[4] P. Baranski and P. Strumillo, "Enhancing Positioning Accuracy in Urban Terrain by Fusing Data from a GPS Receiver, Inertial Sensors, Stereo-Camera and Digital Maps for Pedestrian Navigation," *Sensors*, vol. 12, no. 6, pp. 6764–6801, 2012.

[5] P. Wawrzyniak and P. Korbel, "Wireless indoor positioning system for the visually impaired," in *2013 Federated Conference on Computer Science and Information Systems*, Sept 2013, pp. 907–910.

[6] M. Moreno, S. Shahrabadi, J. Jos, J. du Buf, and J. Rodrigues, "Realtime Local Navigation for the Blind: Detection of Lateral Doors and Sound Interface," *Procedia Computer Science*, vol. 14, pp. 74–82, 2012.

[7] M. M. Shalaby, M. A. M. Salem, A. Khamis, and F. Melgani, "Geometric model for vision-based door detection," in *2014 9th International Conference on Computer Engineering Systems (ICCES)*, December 2014, pp. 41–46.

[8] Y. Zhao, C. Qian, L. Gong, Z. Li, and Y. Liu, "LMDD: Light-Weight Magnetic-Based Door Detection with Your Smartphone," in *2015 44th International Conference on Parallel Processing*, Sept 2015, pp. 919–928.

[9] V. Singh, R. Paul, D. Mehra, A. Gupta, V. Sharma, S. Jain, C. Agarwal, A. Garg, S. Gujral, M. Balakrishnan, K. Paul, P. Rao, and D. Manocha, "'Smart' Cane for the Visually Impaired: Design and Controlled Field Testing of an Affordable Obstacle Detection System," in *Proc. 12th Int. Conf. on Mobility and Transport for Elderly and Disabled Persons (TRANSED)*, June 2010.

[10] B. Kaleci, C. M. Senler, H. Dutagaci, and O. Parlaktuna, "Rule-Based Door Detection Using Laser Range Data in Indoor Environments," in *2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI)*, November 2015, pp. 510–517.

[11] "3D scanning, augmented reality, and more for mobile devices," 2017. [Online]. Available: https://structure.io/

[12] M. Bujacz, K. Kropidlowski, G. Ivanica, A. Moldoveanu, C. Saitis, A. Csapo, G. Wersenyi, S. Spagnol, O. I. Johannesson, R. Unnthorsson, M. Rotnicki, and P. Witek, *Sound of Vision - Spatial Audio Output and Sonification Approaches*. Cham: Springer International Publishing, 2016, pp. 202–209.

[13] S. Spagnol, C. Saitis, M. Bujacz, O. I. Johannesson, K. Kalimeri, A. Moldoveanu, A. Kristjansson, and R. Unnthorsson, "Model-based obstacle sonification for the navigation of visually impaired persons," in *19th International Conference on Digital Audio Effects (DAFx-16)*, Brno, 2016, pp. 309–316.

[14] T. H. Yuan, F. H. Hashim, W. M. D. W. Zaki, and A. B. Huddin, "An automated 3D scanning algorithm using depth cameras for door detection," in *2015 International Electronics Symposium (IES)*, September 2015, pp. 58–61.

[15] Z. Chen and S. T. Birchfield, "Visual detection of lintel-occluded doors from a single image," in *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, June 2008, pp. 1–8.

[16] Y. Zhou, G. Jiang, G. Xu, X. Wu, and L. Krundel, "Kinect depth image based door detection for autonomous indoor navigation," in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, August 2014, pp. 147–152.

[17] S. Meyer, Z. Borgsen, M. Schpfer, L. Ziegler, and S. Wachsmuth, "Automated Door Detection with a 3D-Sensor," in *2014 Canadian Conference on Computer and Robot Vision*, May 2014, pp. 276–282.

[18] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[19] R. Sekkal, F. Pasteau, M. Babel, B. Brun, and I. Leplumey, "Simple monocular door detection and tracking," in *2013 IEEE International Conference on Image Processing*, September 2013, pp. 3929–3933.

[20] J. Canny, "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, November 1986.

[21] R. Stephens, "Probabilistic approach to the Hough transform," *Image and Vision Computing*, vol. 9, no. 1, pp. 66–71, 1991.

[22] H. Kalviainen, P. Hirvonen, L. Xu, and E. Oja, "Probabilistic and non-probabilistic Hough transforms: overview and comparisons," *Image and Vision Computing*, vol. 13, no. 4, pp. 239–252, 1995.

[23] J.-y. Bouguet, "Pyramidal implementation of the Lucas Kanade feature tracker," *Intel Corporation, Microprocessor Research Labs*, 2000.

[24] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *In IJCAI81*, 1981, pp. 674–679.

[25] P. Skulimowski and P. Strumillo, "Verification of visual odometry algorithms with an OpenGL-based software tool," *Journal of Electronic Imaging*, vol. 24, no. 3, p. 033003, 2015.

[26] D. Rzeszotarski, "Fusion of RGB image and depth maps from stereovision and Structure Sensor," 2017, internal report.

[27] E. Ernerfeldt, "Fitting a plane to many points in 3D," 2015. [Online]. Available: http://www.ilikebigbits.com/blog/2015/3/2/plane-from-points