

# Ground plane detection in 3D scenes for an arbitrary camera roll rotation through “V-disparity” representation

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

**Abstract**—In this paper we propose a fast method for detecting the ground plane in 3D scenes for an arbitrary roll angle rotation of a stereo vision camera. The method is based on the analysis of the disparity map and its “V-disparity” representation. First, the roll angle of the camera is identified from the disparity map. Then, the image is rotated to a zero-roll angle position and the ground plane is detected from the V-disparity map. The proposed method was successfully verified on a simulated 3D scene image sequences as well as on the recorded outdoor stereo video sequences. The foreseen application of the method is the sensory substitution assistive device aiding the visually impaired in the space perception and mobility.

## I. INTRODUCTION

THE TASK of ground plane detection in images of 3D scenes is an important step in many computer vision algorithms [1], [2], [3], [4], [5], [6], [7], [8]. Segmentation of the ground plane region and estimation of its spatial orientation allows for detecting free space that is devoid of obstacles in the imaged 3D scenes. Such knowledge is of high importance for depth sensing stereo vision based techniques that are applied e.g. in an automotive industry and systems for guiding autonomous robots [1], [2], [3], [9], [10]. Stereo vision camera modules in such systems are mounted in rigs, which limit camera movement versus the world coordinate system to just a single degree of freedom (1 DoF), that is left and right turn (the yaw angle). Such a constraint simplifies image analysis techniques of the scene that are based on the depth maps computed from the stereo matching algorithms [11].

There are, however, mobile applications of the computer vision systems (e.g. in humanoid robots, or electronic travel aids (ETAs) for the visually impaired and blind [8], [12], for which this work is intended to) in which camera movements are not restricted and need to be defined by 6 DoF ego-motion parameters [8], [13]. That is, three parameters defining 3D translational motion vector  $T = [U \ V \ W]$  and three parameters  $\omega = [\alpha \ \beta \ \gamma]$  defining angular motion of the camera. These rotation angles are known as *pitch*, *yaw* and *roll*, respectively (Fig. 1). In such systems the value of roll angle changes during the movement. Moreover, if a camera is attached to the user’s body, a constant error value may be added if the camera is not positioned properly or undergoes slight position changes

This project was supported by the European Union’s Horizon 2020 Research and Innovation Programme under grant agreement No 643636 “Sound of Vision.”

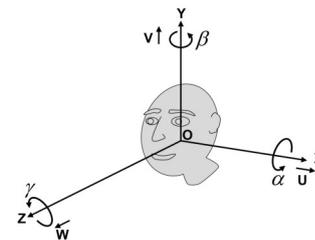


Fig. 1. Parameters defining the 3D translational and rotational motion vectors

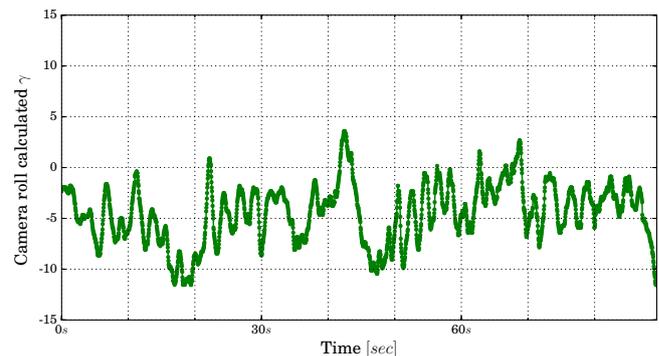


Fig. 2. An example illustrating how the camera roll angle changes during the walk in an open-space outdoor environment with a limited number of obstacles. The camera was mounted on a helmet and the roll angle was estimated using the method described in the article

during the user’s movement. Fig. 2 shows how the roll angle of the camera varies during the walk.

In this paper we propose a fast method for detecting the ground plane in 3D scenes for an arbitrary roll rotation of a stereo vision camera. The method is based on the analysis of the disparity map and its histograms termed “V-disparity” representation [14] (Fig. 3). The disparity map is the horizontal displacement  $d = x_l - x_r$  of a position at which the scene object is projected onto the left and right image of the stereo vision camera. Note that the larger the disparity the smaller is the depth of the scene point in relation to the position of the stereo vision camera [15]. An example disparity map is shown in Fig. 4a, in which the disparity value is coded by a greyscale level. The larger the disparity the brighter is the pixel in the disparity image [14].

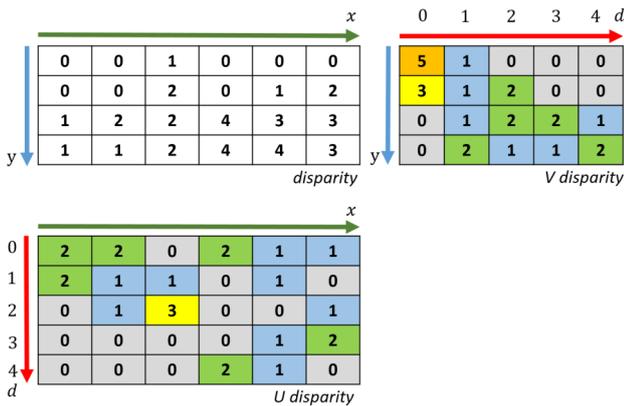


Fig. 3. A visual explanation of how the “UV-disparity” representation of the disparity is calculated

The so-called “V-disparity” representation is built by computing histograms of consecutive rows of the disparity map and presenting them as a monochrome image. Similarly, the “U-disparity” representation contains histograms of consecutive columns of the disparity map. The scheme for calculating those representations is explained in Fig. 3. The number of columns of the U-disparity map equals the number of columns of the disparity map and number of rows of the V-disparity map equals the number of rows of the disparity map. The remaining dimensions (rows of the U-disparity and columns of the V-disparity, respectively) are the histogram bins defined by disparity values  $d$ . The UV-disparity maps can be directly built for the disparity maps calculated with pixel-accuracy only. It is worth noting, that the maximum value of the U-disparity image is the number of columns in the disparity map, and the maximum value for the V-disparity image is the number of rows of the disparity map.

The rest of this paper is organized as follows: in Section II we review the ground plane detection algorithms and discuss advantages and disadvantages of different image processing approaches to this problem. The proposed algorithm for estimating camera roll angle is explained in Section III. Results verifying the performance and robustness of the proposed algorithm are presented and commented in 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 electronic travel aid for the visually impaired.

## II. RELATED WORK

For arbitrary pitch and yaw angles and zero roll rotation of the camera, horizontal line segments of constant depth in a 3D scene are represented by line segments aligned along rows of the disparity map (for a calibrated and rectified stereo vision camera [16]). However, for non-zero roll camera rotations (Fig. 4a) these horizontal lines are no longer aligned along disparity map rows. Thus, detection of the ground plane based on the V-disparity map (note that the V-disparity is computed as a collection of histograms of consecutive rows of the dispar-

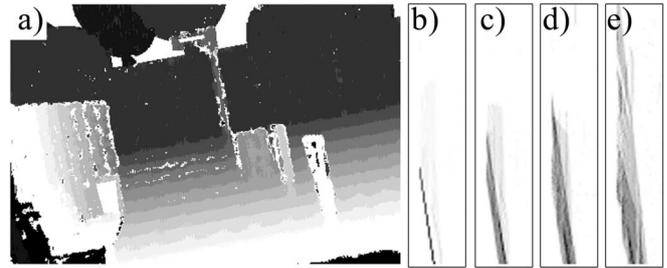


Fig. 4. Test scene imaged by a camera rotated by a roll angle  $\gamma = -10^\circ$ : disparity map calculated by using the *Block Matching* (BM) technique [13] for the scene from Fig. 5 (a), V-disparity maps computed for the camera rotated by  $\gamma = 0^\circ$  (b),  $\gamma = -10^\circ$  (c),  $\gamma = -20^\circ$  (d) and  $\gamma = -45^\circ$  (e), respectively



Fig. 5. Results of the ground plane detection based on the V-disparity and Hough Transform in two almost identical artificial scenes rotated by a roll angle  $\gamma = -10^\circ$ . The scenes differ just by a presence of a small bench in one of the scenes. Bright regions represent the detected ground plane (note the poor result, especially for the scene shown in the right hand image)

ity map) becomes a difficult task. This is because the ground plane in the V-disparity domain is no longer represented by a single line segment but by a rather “fuzzy” region for which its angular orientation is difficult to identify (see Fig. 4c–e obtained for the increasing camera roll rotations).

It can be noticed, that results of ground plane detection with the use of the *Hough Transform* (HT) on the V-disparity map are very sensitive even to minor changes in the content of the scene. Note an example of two almost identical scenes shown in Fig. 5. The two scenes differ just by a presence of a small bench in the scene shown on the left. This seemingly minor change has yielded significantly different ground plane detection results (indicated by white regions). In order to improve the plane detection precision, prior to application of the HT technique, the scene image should be rotated by an adequate angle to compensate for the roll angle of the camera (Fig. 6, step 3).

The findings of our literature search on applications of the V-disparity representation for ground plane detection show that the problem of a non-zero camera roll angle is addressed or noticed in very few studies, e.g. [1], [3], [9], [17] among others.

Cong et al. [1] propose a method for detecting ground surface based on the maximum local energy in the V-disparity map. This approach seems to work even if the ground is not a flat surface. However, the problem of a non-zero roll angle was not directly addressed.



Fig. 6. Consecutive steps of the proposed V-disparity based ground plane detection method (including step 3<sup>rd</sup>, which is specific to our method)

Wu et al. [3] introduce a special method for mounting the camera that allows to ignore a non-zero roll angle. On the other hand the ground plane parameters are calculated by using the V-disparity map after removing large image regions representing obstacles identified in the U-disparity domain. An initial road profile is calculated without using the Hough Transform. Instead, the assumption that the maximum intensity in each row of the V-disparity map corresponds to the road lanes is used.

Lin et al. [9] use a RANSAC-based plane fitting algorithm to find the plane equation. The method allows to calculate the road lane of the same depth, which need not to be parallel to the horizontal axis of the disparity image. The authors have noticed the problem of non-zero roll angle for images of sloping roads but they assumed that the proposed plane fitting algorithm is performing well for small values of the roll angles.

Finally, Labayrade and Aubert [17] propose an estimation of the roll, pitch and yaw camera angles. A combined iterative and linear regression methods were applied to the projections of the plane to the V-disparity map to estimate the roll and pitch angle. The Authors noted, however, that this method can be computationally expensive. The value of yaw angle was estimated indirectly by determining the vanishing point.

### III. A METHOD FOR ESTIMATING CAMERA ROLL ANGLE

A general scheme for detecting the ground plane in images of 3D scenes is shown in Fig. 6. The region corresponding to the ground plane is detected in the disparity map through its V-disparity representation. Namely, a plane equation that best fits the surface of the ground is computed on the basis of the line identified in the V-disparity map, e.g. by applying the classical *Hough Transform* [3], [4], [6], [12], [14].

Our method for camera roll angle estimation is based on the observation, that for zero-roll angles any line segment, that is taken from the ground plane and is coplanar with the line  $O_L O_R$  connecting optical centers of the stereo vision cameras, is projected onto the same  $y$ -coordinates in the stereo vision images and in the corresponding disparity map. Note that any point from such a line assumes the same depth. However, for non-zero camera roll angles these ground plane lines are no longer coplanar with the  $O_L O_R$  line. Consequently, these scene line segments (of equal depth) are projected onto the disparity map at an angle that is equal to the camera roll angle. In order to identify the camera roll angle a method is proposed in which the disparity map is cross-sectioned by a series of lines  $l$  at varying angles. For each angular position of line  $l_i, i \in \mathbb{N}$  disparity map values at points  $P_1$  and  $P_2$  equidistant to  $l_i$  are collected (see Fig. 7).

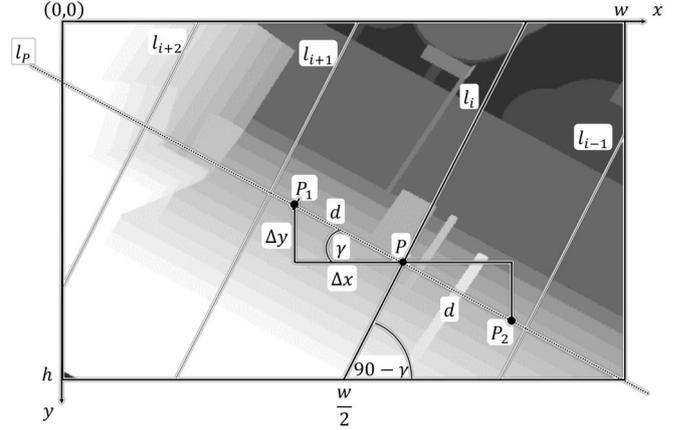


Fig. 7. An example ground truth depth image for camera roll angle  $\gamma = 30^\circ$  with superimposed line  $l_i$  and points  $P_1, P_2$  ( $w$  is the number of pixels in a single row of the disparity map and  $d = 0.1w$  was selected in computations)

Note, that only for lines  $l$  that are vertical to  $l_p$  the disparity map points  $P_1$  and  $P_2$  take similar values.

The slope intercept form of line  $l_i$  in the image coordinate system is  $y = Ax + B_1$  with  $A = -\tan(90^\circ - \gamma)$  and  $B_1 = h + \frac{w}{2} \tan(90^\circ - \gamma)$ . Likewise, line  $l_p$  such that  $l_p \perp l_i$  is given by  $y = -\frac{x}{A} + B_2$ . If  $|PP_1| = |PP_2| = d$ , then:

$$(\Delta x, \Delta y) = \left( \frac{d}{\sqrt{1 + (\tan \gamma)^2}}, |\Delta x \tan \gamma| \right) \quad (1)$$

Because  $|\tan \gamma| = \left| \frac{1}{A} \right|$ , coordinates of points  $P_1$  and  $P_2$  are:

$$\begin{aligned} P_1 & \left( \lfloor x - \Delta x \rfloor, \left\lfloor y + \frac{\Delta x}{A} \right\rfloor \right) \\ P_2 & \left( \lfloor x + \Delta x \rfloor, \left\lfloor y - \frac{\Delta x}{A} \right\rfloor \right) \end{aligned} \quad (2)$$

where  $\lfloor x \rfloor$  denotes the floor function of  $x$ . Let us assume, that the total number of points  $P = (x, y)$  is  $Q$ , i.e. it is equal to the number of analyzed point pairs.

The proposed method assumes that: the dominant part of the disparity image is occupied by a ground plane and if the image is rotated by  $\gamma$ -degrees, the disparity values at points  $P_1$  and  $P_2$  shall remain the same for a series of lines  $l$  for a given  $\gamma$ . To estimate the roll rotation angle of the camera, the disparity map is dissected by lines  $l$  at different  $\gamma$  angles ( $\gamma \in [\gamma_{min}, \gamma_{max}]$ ), with a predefined step of  $\Delta\gamma = 0.5^\circ$ . For each  $\gamma$  value the parameter  $q(\gamma)$  is calculated:

$$q(\gamma) = \frac{N_E}{N_A} \quad (3)$$

where  $N_E$  is the number of pairs  $(P_1, P_2)$  for which the disparity values are the same and  $N_A$  is the total number of all analyzed pairs.

Finally, the camera roll angle is such a  $\gamma_r$  value for which (3) reaches the maximum (see Fig. 9). The algorithm of the proposed method is shown in Fig. 8.

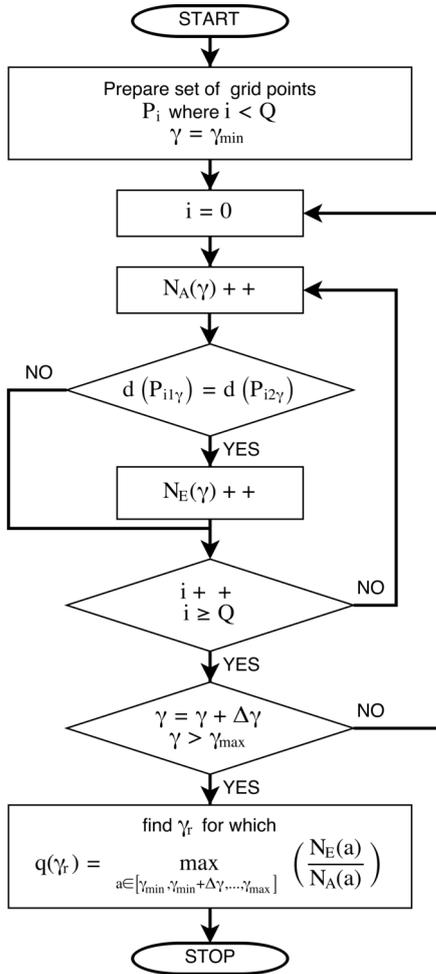


Fig. 8. The block diagram of the proposed method for roll angle estimation of a stereo vision camera

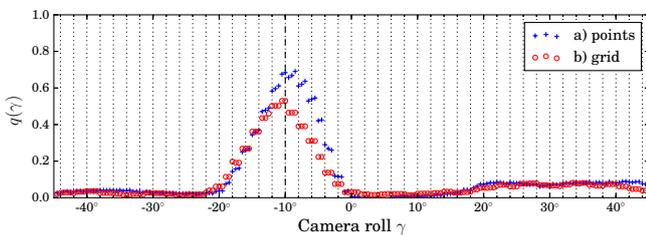


Fig. 9. Plots of  $q(\gamma)$  values for different angles of line  $l_i$  dissecting the disparity map obtained from the two proposed methods (based either on point or grid resolution). Note pronounced maxima for  $\gamma = -10^\circ$  obtained from both methods. These are correct estimations of camera roll angle (see Fig. 4a)



Fig. 10. Example results of the ground plane detection for the scene from Fig. 4: Ground truth region ( $M_{GT}$ ) from the SESGen [13] software (a), Region detected using only the classical HT-based approach (b), Region detected using the proposed method (c)

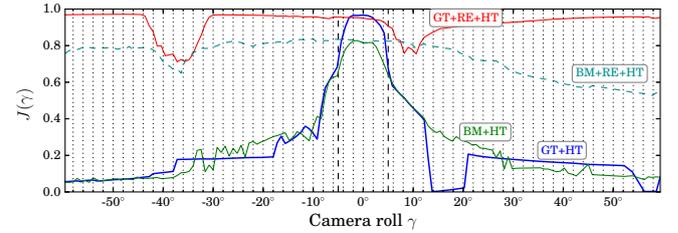


Fig. 11. Ground plane detection accuracy measured by means of the JC obtained for different camera roll rotations and variants of the ground plane estimation algorithms

GT: ground truth disparity map

BM: disparity map obtained using the *Block Matching* technique

RE: camera roll angle estimation (Fig. 6, step 3)

HT: plane detection using the Hough Transform (Fig. 6, steps 4–5)

#### IV. RESULTS

The proposed method for ground plane detection was verified on test image sequences rendered by our SESGen software [13]. The sequence consists of 600 images, for which the roll rotation of the camera ranges from  $-60^\circ$  to  $+60^\circ$  with a step of  $0.2^\circ$ . For each rendered image the SESGen computes the ground truth segmentation map and the ground truth disparity map with a pixel and subpixel accuracy correspondingly.

To measure the accuracy of the plane detection results we used the *Jaccard similarity coefficient* (JC):

$$J(M_D, M_{GT}) = \frac{\text{area}(M_D \cap M_{GT})}{\text{area}(M_D \cup M_{GT})} \quad (4)$$

where  $M_D \cap M_{GT}$  denotes the intersection of the detected and “ground truth” ground plane regions, and  $M_D \cup M_{GT}$  is their union. Those regions are represented by bright regions shown in Fig. 10. The JC is calculated for both the ground truth disparity maps and for the disparity maps calculated using the *Block Matching* (BM) technique [13]. Results are shown in Fig. 11. Note, that the ground plane detection algorithms with no camera roll angle correction tend to fail for roll rotations of more than  $\pm 5^\circ$  for which a significant drop of the JC occurs.

In order to reduce the computational complexity, both the number of steps and the number of lines dissecting the disparity map can be adjusted appropriately. Additionally, in most cases just the bottom part of the disparity image shall be taken into account. We also tested a modification of our method in which  $P_1$  and  $P_2$  are nodes of a grid (with grid size up to 30 pixels). Such a modification slightly decreases the roll angle estimation accuracy (see Fig. 9), but significantly

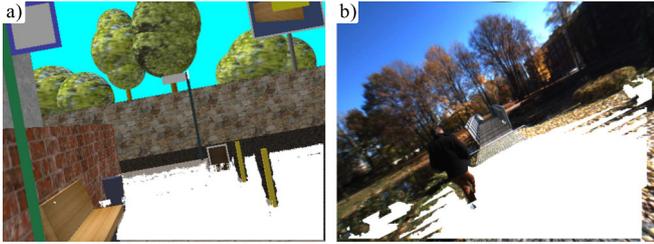


Fig. 12. Results of the ground plane estimation using the proposed method



Fig. 13. Point pairs from Fig. 12b for which the disparity values are the same for the camera roll angle is such a  $\gamma$  value for which (3) reaches the maximum

reduces the computational complexity which is essential in mobile and wearable platforms. Fig. 12 shows example results of the ground plane estimation methods for the pre-recorded and artificial sequences. Then, Fig. 13 shows point pairs from the scene shown in Fig. 12b for which the disparity values are equal to the camera roll angle for such a  $\gamma$  value for which (3) reaches the maximum. Please note, that these points can be successfully used in the plane fitting algorithm.

An average calculation time of the proposed algorithm (Fig. 6 step 3) for the test images is 0.7 ms on an Intel Core i7-4770 3.4 GHz processor. The computational complexity is estimated as  $O(n^2)$ . The obtained Root-Mean-Squared Error (RMSE) for the SESGen sequences equals  $RMSE = 0.466^\circ$ .

The proposed method was also verified on a set of disparity images captured in an indoor environment along with the readouts from a digital inclinometer permanently attached to the stereo vision camera. Images were captured using the ZED Stereo Camera ( $1920 \times 1080$  image resolution,  $110^\circ$  field of view and 120 mm baseline [18]). Camera roll estimation results are shown in Fig. 14. The obtained Root-Mean-Squared Error value for this sequence is  $RMSE = 1.76^\circ$

We encourage the reader to view our material supplementary to this paper [19] (e.g. result video sequences, images in higher resolution, etc.) available at <http://uv-disparity.naviton.pl/>.

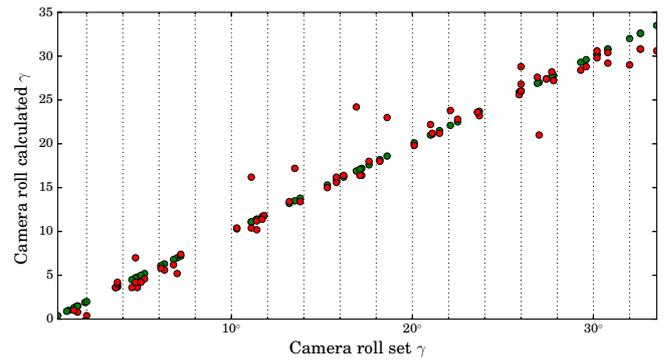


Fig. 14. Results of the roll angle estimation for the recorded indoor sequence. Images were captured using the ZED stereo vision camera. The disparity map was computed using the API provided by the camera manufacturer [18]. Green dots denote roll angle read from the digital inclinometer, red dots denote roll angle values for the corresponding readouts from the inclinometer and calculated using the proposed method

## V. CONCLUSIONS

In this paper we propose a reliable algorithm for ground plane detection in 3D scene images from the disparity maps and their V-disparity representation under large roll angle values. From our literature survey we note that the problem of non-zero roll angle in ground plane detection tasks has been noticed in just few earlier studies [1], [3], [9], [17]. Moreover, only the authors of the latter work undertook the problem of roll angle estimation. They, however, did not provide any time performance of their iterative algorithm.

The strong advantage of the algorithm we propose is its computing efficiency and capability of estimating camera roll rotations for large angles (tested from  $-60^\circ$  to  $+60^\circ$ ) with the  $RMSE < 0.5^\circ$ . Such rotations can occur for 6DoF motions of the camera, e.g. in cameras mounted on drones, robots or 3D scene analysis systems aiding the visually impaired. The identified roll angle allows to rotate the disparity image to zero-roll angle. The so corrected disparity map is then used for detecting the ground plane through the corresponding V-disparity map. The reliably detected ground plane region is a basis for successful performance of further 3D scene analysis algorithms. Finally, we have shown high robustness of our ground plane detection algorithm on simulated 3D scene image sequences and real-world outdoor image sequences.

## 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] Y. Cong, J.J. Peng, J. Sun, L.L. Zhu and Y.D. Tang, "V-disparity based UGV obstacle detection in rough outdoor terrain," *Acta Automatica Sinica*, vol. 36 (5), 2010, pp. 667–673, [http://dx.doi.org/10.1016/S1874-1029\(09\)60029-X](http://dx.doi.org/10.1016/S1874-1029(09)60029-X)
- [2] Y. Li and Y. Ruichek, "Occupancy grid mapping in urban environments from a moving on-board stereo vision system," *Sensors*, vol. 14, 2014, pp. 10454–10478, [http://dx.doi.org/10.1016/S1874-1029\(09\)60029-X](http://dx.doi.org/10.1016/S1874-1029(09)60029-X)

- [3] M. Wu, S.K. Lam and T. Srikanthan, "Nonparametric Technique Based High-Speed Road Surface Detection," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 16 (2), 2015, pp. 874–884, <http://dx.doi.org/10.1109/TITS.2014.2345413>
- [4] C. Yu, V. Cherfaoui and P. Bonnifait, "Evidential occupancy grid mapping with stereo vision," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, June 2015, pp. 712–717, <http://dx.doi.org/10.1109/IVS.2015.7225768>
- [5] D. Yiruo, W. Wenjia and K. Yukihiro, "Complex ground plane detection based on V-disparity map in off-road environment," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, June 2013, pp. 1137–1142, <http://dx.doi.org/10.1109/IVS.2013.6629619>
- [6] A. Iloie, I. Giosan and S. Nedeveschi, "UV disparity based obstacle detection and pedestrian classification in urban traffic scenarios," in *Proceedings of the IEEE Int Intelligent Computer Communication and Processing (ICCP) Conference*, September 2014, pp. 119–125, <http://dx.doi.org/10.1109/ICCP.2014.6936963>
- [7] X. Zhu, H. Lu, X. Yang, Y. Li and H. Zhang, "Stereo vision based traversable region detection for mobile robots using u-v-disparity," in *Proc. 32nd Chinese Control Conference (CCC)*, July 2013, pp. 5785–5790.
- [8] T. S. Leung and G. Medioni, "Visual Navigation Aid for the Blind in Dynamic Environments," in *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Columbus, OH, 2014, pp. 579–586, <http://dx.doi.org/10.1109/CVPRW.2014.89>
- [9] Y. Lin, F. Guo and S. Li, "Road Obstacle Detection in Stereo Vision Based on UV-disparity," *Journal of Information & Computational Science*, vol. 11 (4), 2014, pp. 1137–1144, <http://dx.doi.org/10.12733/jics20103012>
- [10] Z. Hu, F. Lamosa and K. Uchimura, "A complete U-V-disparity study for stereovision based 3D driving environment analysis," in *Fifth International Conference on 3-D Digital Imaging and Modeling (3DIM'05)*, 2005, pp. 204–211, <http://dx.doi.org/10.1109/3DIM.2005.6>
- [11] J. Suhr, H. Kang and H. Jung, "Dense stereo-based critical area detection for active pedestrian protection system," *Electronic Letters*, vol. 48 (19), 2012, pp. 1199–1201, <http://dx.doi.org/10.1049/el.2012.1176>
- [12] M. Owczarek, P. Skulimowski and P. Strumillo, "Sound of Vision – 3D Scene Reconstruction from Stereo Vision in an Electronic Travel Aid for the Visually Impaired," in: *Computers Helping People with Special Needs*, ICCHP 2016, Lecture Notes in Computer Science, vol. 9759, pp. 35–42, 2016, [http://dx.doi.org/10.1007/978-3-319-41267-2\\_6](http://dx.doi.org/10.1007/978-3-319-41267-2_6)
- [13] P. Skulimowski and P. Strumillo, "Verification of visual odometry algorithms with an OpenGL-based software tool," *Journal of Electronic Imaging*, vol. 24 (3), 2015, pp. 033003, <http://dx.doi.org/10.1117/1.JEI.24.3.033003>
- [14] R. Labayrade, D. Aubert and J.P. Tarel, "Real Time Obstacle Detection in stereo vision on Non Flat Road Geometry Through "V-disparity" Representation," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, 2002, pp. 646–651, <http://dx.doi.org/10.1109/IVS.2002.1188024>
- [15] M.Z. Brown, D. Burschka, and G.D. Hager, "Advances in computational stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25 (8), August 2003, pp. 993–1008, <http://dx.doi.org/10.1109/TPAMI.2003.1217603>
- [16] D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*, Pearson Education, Inc., 2003.
- [17] R. Labayrade and D. Aubert, "A single framework for vehicle roll, pitch, yaw estimation and obstacles detection by stereo vision," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, 2003, pp. 31–36, <http://dx.doi.org/10.1109/IVS.2003.1212878>
- [18] "Developer Center - ZED," 2017, visited on 2017-04-28. [Online]. Available: <https://www.stereolabs.com/developers/>.
- [19] P. Skulimowski, "UV-disparity analysis: Ground plane estimation results for simulated and real outdoor scenes," 2017, visited on 2017-04-28. [Online]. Available: <http://uv-disparity.naviton.pl/>.