Obstacle Detection in Stereo Sequences using Multiple Representations of the Disparity Map

Adrian Burlacu*, Simona Bostaca*, Ioana Hector*, Paul Herghelegiu*, Gabriel Ivanica†,
Alin Moldoveanu† and Simona Caraiman*
*Faculty of Automatic Control and Computer Engineering,
"Gheorghe Asachi" Technical University of Iasi, Romania,
Email: aburlacu@ac.tuiasi.ro
†Faculty of Automatic Control and Computer Science
University Politehnica of Bucharest, Romania

Abstract—**Object detection represents an important task in various application fields, such as automotive or assistive technologies. In this context stereo cameras are among the most common devices used for acquiring images and information from the environment. Efficient processing of the disparity maps computed from stereo images represents a crucial step in obstacle detection algorithms. In this paper we provide an application-independent framework for obstacle detection based on disparity processing. In order to identify regions in the 3D environment that can be classified as obstacles we employ multiple representations of the disparity map: V-disparity, U-disparity, θ-disparity. We provide a comprehensive overview of those representations and their use in obstacle detection algorithms. We evaluate the proposed framework in the context of automotive and visually impaired assistive applications, using data from both real and virtual environments.**

Keywords: *disparity map, u-disparity, v-disparity, θ-disparity, obstacles detection*

I. INTRODUCTION

Obstacle detection has been an active research area, with results widely used in many fields such as automotive or assistive systems. The most important goal is to accurately locate obstacles in front of the camera and measure the distance to the obstacles. Existing solutions employ either active or passive sensors. In the category of passive sensors, stereo vision-based systems can be used to directly compute distance evaluating the disparities between the two images. An important focus on disparity processing has been reported in the automotive and mobile robotics research areas.

In [1] the authors use the properties of the V-disparity image to develop a solution for the obstacle detection problem for autonomous driving cars. The longitudinal profile of the road is estimated and the objects located above the road surface are then extracted as potential obstacles. The same authors present in [2] a system that can be divided into two main stages. The first one deals with on-board road obstacles detection (the focus is put on obstacle areas and free road surface extraction), whereas the second one focuses on obstacle characterization (e.g., car/truck discrimination). In [3], Chen et al. segment the stereo disparity map by employing a depth slicing technique and then accurately marking the object boundaries

using a region growing method to improve on-road obstacle segmentation. Another region growing technique for vehicle detection is suggested by Kormann et al. [4]. The obstacles are detected using UV-disparity maps, and splines are used for the road model. Lee et al. [5] perform vehicle detection using a road feature and disparity histogram. In [6], they present another stereo-vision based obstacle detection approach using UV-disparity maps and bird-eye view mapping.

Another research subject present in many works in the past two decades is the development of systems dedicated to help the visually impaired to perceive the environment, for orientation and navigation purposes. The reported efforts to support the rehabilitation of visually impaired have been directed towards the development of electronic travel aids and sensory substitution devices. The development of many of these devices exploits stereo vision systems and processing methods for obstacle detection.

Using disparity maps instead of intensity images significantly simplifies the obstacle detection process. However, calculating a dense disparity map requires a significant amount of computation which is burdensome for the overall system. Moreover, such assistive devices also pose constraints regarding wearability and real time operation. With fast growing computing systems, dense disparity matching can be achieved in real-time. Once the disparity map is available, the potential obstacles in the scene can be identified using different techniques.

Saez et al. [7] propose a stereo-based system that relies on ground plane information to detect aerial obstacles in front of a blind user. The proposed method doesn't explicitly extract the ground plane, i.e., using model fitting techniques. However, it maintains the floor of the local map parallel to the horizontal plane by exploiting an information based method, more precisely an entropy minimization schema. The work of Bujacz [8], consists in an iterative algorithm that detects planes based on the acquired depth map. The spatial points, which remain after subtraction of the detected planes or form groups too small to qualify as surfaces, are grouped with their spatial neighbors and marked as obstacles. Rodriguez et al. [9] also designed an obstacle detection system based on 3D ground detection. Ground plane detection is achieved using a model

fitting technique, i.e., RANSAC. This is a global approach, in contrast with Bujacz's local approach which exploits the similarity between neighboring patches in order to group them into larger surfaces. Mattoccia et al. [10] developed a mobility aid based on stereo vision which detects obstacles in front of the user. Ground plane detection is achieved using a modified RANSAC approach in the v-disparity domain. The obstacles are considered to be the points that do not lie on this surface according to a prefixed tolerance threshold.

In this paper we present an obstacle detection algorithm, which makes use of multiple representations of the disparity maps such as: U-disparity [11], V-disparity [1], and the recently defined θ - disparity representation [12]. Combining the main advantages of all representations, the proposed algorithm is able to accurately identify regions in stereo sequences that can be classified as obstacles. To the best of our knowledge, the framework presented in this paper is the first to employ θ - disparity processing for complex applications.

II. DISPARITY MAP REPRESENTATIONS

A. Disparity Map

Any reliable stereo vision system requires to go through several phases before it can be practically exploited. A very important step is the stereo calibration phase. Once the intrinsic and extrinsic parameters are recovered, the left and right image are acquired and denoised. Next, the rectification is performed to ensure distortion removal and stereo images alignment. By this means, the correspondences can be restricted to the same line in both images and thus the computation burden in the stereo matching step can be reduced.

The disparity map refers to the displacement of the relative features or pixels between two views. Disparity maps are essential for various applications like 3D reconstruction, image based rendering, or robotic navigation. The most basic tool needed for finding corresponding points in the stereo pair is a matching cost function that measures image similarity. Most widely used are matching cost functions that compare image intensities by their absolute or squared differences [13], [14].

Recently new computational approaches were proposed. Among them, the ELAS (Efficient Large-Scale Stereo Matching) [15] algorithm proved to have very good results regarding the density of the generated disparity map. This algorithm uses a Bayesian approach and performs quite well even in relatively low-textured images. First, the disparities of a sparse set of support points are computed using relatively strict constraints. The image coordinates of the support points are then used to create a 2D mesh via Delaunay triangulation. Finally, a prior is computed to reduce the matching ambiguities. In particular, this prior is formed by calculating a piecewise linear function induced by the support point disparities and the triangulated mesh.

In the following subsections different representations of the disparity map will be presented. For this purpose consider the disparity map to be denoted by $D(x, y)$, where (x, y) represents the position of a pixel in the image.
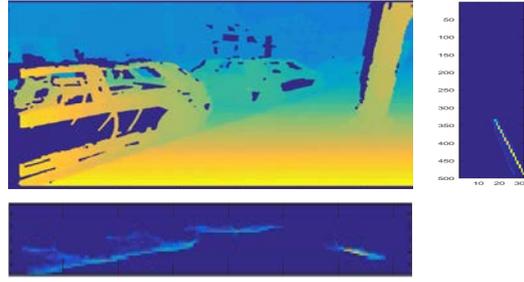
Fig. 1. Disparity image and its row-wise and column-wise histograms

B. V-Disparity

The V-Disparity image [1] provides a good interpretation of the geometrical content in a scene and has a high accuracy when it comes to detecting the ground plane. This type of disparity image can be understood as the disparity histogram of each line in the disparity map. One of the most important features of the V-Disparity map is the fact that major planar surfaces in the scene have corresponding line representations in the V-Disparity image. Vertical surfaces are mapped into vertical line segments in the V-disparity image, while the ground plane corresponds to a slanted line segment.

The V-Disparity map is built based on a disparity map generated from a stereo image pair (Fig. 1). Consider F, a function attached to the input disparity image such that $F_v(D(x, y)) = D_v(x, d)$, where $D_v(x, d)$ is the V-Disparity map. The function F sums up all the points with the same disparity value that appear on every given row of the image.

In Algorithm 1 we present the steps that allow the computation of the V-Disparity map.

Algorithm 1: V-Disparity Computation

Input: Disparity map $D(x, y)$
Output: V-Disparity Map $D_v(x, d)$

1 **for** *each ith column in D* **do**
2   **for** *each jth line in D* **do**
3   **if** $D(i, j) > 0$ **then**
4          $D_v(j, D(i, j)) + +$

C. U-Disparity

The U-Disparity map [11] has the same building concept as the V-Disparity map. The main and most important difference is that the U-Disparity map is a column-wise representation of the disparity values (Fig. 1). It provides information regarding the obstacles found in a scene, by marking them with multiple horizontal lines.

In order to build a U-Disparity map consider a function F_u, linked to the disparity map, such that $F_u(D(x, y)) = D_u(d, y)$, where $D_u(d, y)$ is the desired U-Disparity map. The $D_u(d, y)$ space sums up all the pixels in the initial disparity map $D(x, y)$ that have the same disparity value and are found
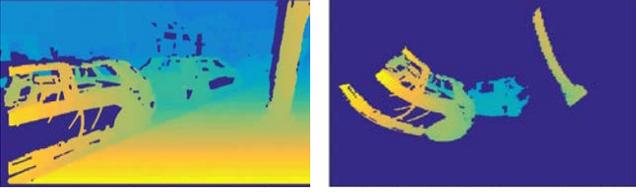
Fig. 2. left - Disparity image; right - Polar representation of the disparity map

across the same column y. The steps required to compute the U-Disparity map are presented in Algorithm 2 .

Algorithm 2: U-Disparity Computation

Input: Disparity map $D(x, y)$
Output: U-Disparity Map $D_v(x, d)$

1 **for** *each ith column in D* **do**
2   **for** *each jth line in D* **do**
3   **if** $D(i, j) > 0$ **then**
4              $D_v(D(i, j), i) + +$

D. θ-Disparity

θ-disparity [12] is a recent approach for representing the 3D information of a scene. The main idea behind it is to obtain a radial representation of the significant objects in a set, with respect to a point of interest, based on the disparity map.

The first step in building the θ transform is to apply a polar transform over the disparity map, where the pole of the transform is the middle pixel of the bottom line of the disparity map. This corresponds to the endpoint of the end effector in a robotics application or of the head-worn camera position in the context of assistive systems . Assuming that $D(x, y)$ has x_{max} lines and y_{max} columns, the coordinates of the pole will be $(x_{max}, \frac{y_{max}}{2})$. An example of polar transformed disparity map $P(\rho, \theta)$ is illustrated in Fig.2, where

$$\rho \in \left\{ 0, ..., \sqrt{x_{max}^2 + \left(\frac{y_{max}}{2}\right)^2} \right\} \quad (1)$$

and $\theta \in \{0, ..., 180\}$

The new map displays the set of disparity values that are laying along the direction angle θ in the original disparity map, $D(x, y)$ relative to a point of interest, the pole of the transform. In order to achieve the desired θ disparity map a column wise histogram will be computed. Moreover, a weighting factor $\sin(\theta)$, can be applied to each element in order to emphasise the nearby obstacles and to avoid the noticeable degeneration in the polar transformed disparity map $P(\rho, \theta)$. As we get closer to the extreme angle values (0 and 180), the values of the pixels corresponding to those columns tend to have the same value.

The θ-disparity map has only positive integer values and each point indicates the number of pixels from the initial disparity map, $D(x, y)$, that lie across a certain direction

---

Algorithm 3: θ-Disparity Computation

Input: Disparity map $D(x, y)$
Output: θ-Disparity Map $D_\theta(\theta, d)$

1 Initialization $D_\theta \leftarrow 0$
2 Compute $P(\rho, \theta)$ polar transform of $D(x, y)$ around $(x_{max}, \frac{y_{max}}{2})$
3 **for** *each angle of P: $\theta = 0$ to 180* **do**
4   **for** *each disparity level $d = D_{min}$ to D_{max}* **do**
5   **for** *each row of P: $\rho = 1$ to ρ_{max}* **do**
6   **if** $P(\rho, \theta) = d$ **then**
7                  $D_\theta(\theta, d) + +$
8          $D_\theta(\theta, d) = D_\theta(\theta, d) \sin(\theta)$

---

and have a disparity value d. The information held in the θ-disparity map can be interpreted in many different ways, depending on the goal of the application.

The steps required to obtain a theta disparity map are presented in Algorithm 3.

III. OBSTACLE DETECTION ALGORITHM

In this section we describe the use of the disparity representations in the previous section and detail the proposed obstacle detection algorithm. This algorithm is structured in a series of steps as depicted in Fig. 3.
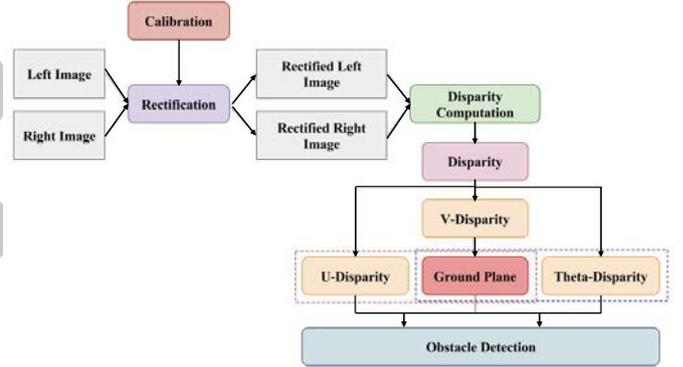
Fig. 3. The steps performed in the obstacles detection framework

The algorithm input is the left and right raw images acquired by the stereo vision system. Using calibration parameters, a rectification procedure is applied on the input images, the result being the rectified left and right images. Once the rectified images are available the disparity map can be computed. This task is performed using the ELAS algorithm described in section II-A.

The dense disparity map is further exploited using the three representations: discussed in the previous section. Each of these representations has its own contribution to the obstacle detection algorithm as detailed in the following.

A. Ground Plane Detection

Usually, regions corresponding to the ground plane match a line in the V-Disparity image. In cases where the ground

plane occupies a significant region in the image, it corresponds to the most dominant line in the V-disparity map. To detect such lines in the V-Disparity map we employ an adapted version of the Hough Transform [16]. The standard Hough transform detects all the straight lines that match certain restrictions in an image. To speed up the algorithm, we use a simple thresholding algorithm on the V-Disparity image. This operation removes pixels with a low gray value that correspond to poorly represented disparities in the original image. This operation is based on the assumption that the ground plane should not have less than a user defined number of pixels located on each line.

One shortcoming associated with the standard V-disparity based ground plane detection method is that it cannot be applied on images where the ground plane is not horizontal. For non-zero roll camera rotations, estimation of the ground plane based on the V-disparity using the Hough Transform becomes a difficult task. This is due to the fact that the ground plane in the V-disparity domain is no longer represented by a single line. For larger roll angles the disparity map should be rotated before calculating the V disparity map.

To detect the roll rotation angle of the camera we can make use of stereo vision motion estimation procedures. *Libviso2* [17] is a fast algorithm for computing the 6 DOF motion of a moving mono/stereo camera. For stereo sequences, *libviso2*, uses a procedure to extract "circular" feature matches and project feature points from the previous frame into 3D via triangulation, using the calibration parameters of the stereo camera rig. Assuming squared pixels and zero skew, instead of minimizing the residuals in Euclidean space, the *Libviso2* framework makes use of the intrinsic parameters of the stereo camera to minimize the residuals in the image space, where the noise level is similar for all components of the measurement vector, thus recovering the motion in a 6 parameters representation.

Once the camera motion is recovered, the roll-pitch-yaw parameterization can be easily computed and the V-disparity image can be rotated accordingly, thus ensuring robust ground plane detection results.

B. Obstacle Detection using U and θ-disparity maps

The effectiveness of the U-disparity representation is revealed in automotive applications where the detection of the road boundaries is of interest. In the U-disparity map, obstacles are usually represented as horizontal lines. This represents the main technique for obstacles identification using U-disparity map. However, when an obstacle is passing near the camera side, both the front and side faces of the obstacle are observed. The obstacle is then represented by a polyline in the V-disparity image: a horizontal part for frontage and a connected oblique part for its side face. This observation allows us to define an approach in which the U-disparity image can be easily segmented, after the ground plane is removed from the disparity map. The obstacles are represented in the image by connected-regions. Noisy pixels are removed from the V-disparity map by means of morphological operations (erosion and dilation). Each connected-region resulting after these pre-processing steps indicates a potential obstacle.

The segmentation of the θ-disparity image employs similar steps as in the processing of the U-disparity maps: the ground plane pixels are removed from the image, followed by thresholding, morphological operations and finally, detection of connected components. The strong points of the θ representation of the disparity result from the following observations: it makes no assumption about the sensor-environment geometry and it preserves the direction and angular distribution of obstacles and obstacle-free regions in a scene, while it can be estimated in a very efficient way.

The robustness of the obstacle detection algorithm is increased by performing various set operations on the obstacles pixels detected with the two methods. Depending on the actual obstacle detection application, the final segmentation results can be obtained using the intersection or the union of the results produced by the two methods. For example, in scene understanding for human assistive applications it is usually important to also produce a description of the obstacles, regarding size, shape and distance to the user. In such cases, the intersection of the two results can avoid under-segmentation and lead to better object separation and tracking in consecutive frames. Moreover, a description of the free navigable space in the environment can be efficiently extracted based on the θ-disparity segmentation.

IV. EXPERIMENTAL RESULTS

In this section we describe the experiments performed for evaluating the proposed obstacle detection framework. The evaluation is performed using data acquired from both real and virtual environments. To simulate an automotive application context for our framework, we used the KITTI Vision Benchmark Suite [18]. Images and ground truth information are generated from a virtual environment in order to test new applications for visually impaired assistive devices. The results obtained with the proposed obstacle detection framework are also made available online[1][2].

A. Results Obtained in a Real Environment Setup

The KITTI benchmark dataset was captured using a station wagon dedicated to mobile robotics and autonomous driving research. The vehicle was equipped with two color and two grayscale PointGrey Flea2 video cameras (10Hz, resolution: 1392×512 pixels, opening: $90^o \times 35^o$). Camera setup is chosen such that we obtain a baseline of roughly 54 cm between the same type of cameras and that the distance between color and grayscale cameras is minimized (6cm). The scenarios are diverse, capturing real-world traffic situations and range from freeways over rural areas to inner-city scenes with many static and dynamic objects. The data from the KITTI benchmark is calibrated, synchronized and timestamped, including both rectified and raw image sequences.

---

[1]https://www.youtube.com/watch?v=JWneKqtPFZs
[2]https://www.youtube.com/watch?v=Wc3CeFBf3FQ
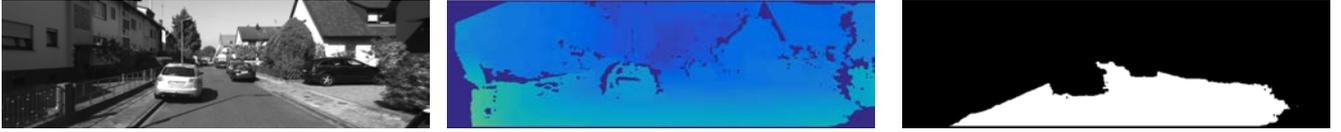
Fig. 4. Experiments with the KITTI dataset: Left image; Disparity map; Ground plane mask
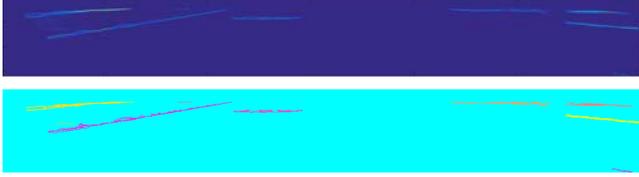
Fig. 5. U-disparity and U-disparity labeled for an image in the KITTI sequence

Fig. 7. Obstacles detection results for the KITTI dataset: overlaid segmentation results obtained in both θ and U-disparity images

Fig. 6. Experiments with the KITTI dataset: Polar representation of the disparity map after ground plane removal, θ-disparity before and after labelling
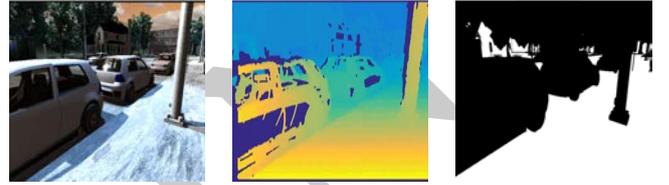
Fig. 8. Virtual environment: Left image; Disparity map; Ground plane mask

The disparity maps were computed using the ELAS algorithm. As the resolution of the disparity values decreases exponentially with the distance to the camera, the disparity images were thresholded to exclude pixels with disparity values less than 25. A ground plane mask was computed using the approach described in section III-A. An example image in the sequence, along with its disparity map and extracted ground plane is presented in Fig. 4.

The U-disparity map is computed for the disparity image after excluding the ground plane pixels. Then, morphological opening, with a square structural element of size 1, is applied to the U-disparity image after thresholding with a value of 0.05. This results in the labelled regions presented in Fig. 5.

The θ-disparity segmentation is similarly obtained: the θ-disparity image is thresholded with a value of 0.01, and morphologically opened with the same structural element (Fig. 6).

The results of both segmentations for the same image can be comparatively assessed in Fig. 7 as they are both overlaid on the initial intensity image. The obstacles pixels segmented in the θ-disparity map are marked with blue, the U-disparity results are marked with red, while their intersection is marked in green.

B. Results Obtained with Synthetic Data in Virtual Environments

A series of 3D virtual environments were designed to generate benchmark testing stereo sequences for human assistive devices. Since virtual scenes can provide ground truth information, testing using such synthetic data can offer valuable information about the efficiency of the algorithms and acknowledge worst case scenarios. Moreover, different environment scenarios can be tested without the need to physically find these locations or recreate some special situations in real-life environments.

The scenes were designed to mimic common outdoor locations. The development of the 3D scenes was done using the Unity game engine. Ground truth segmentation information was obtained by assigning a label and unique ID to each object in the scene. Thus, for the virtual environment testing scenarios the ground plane as well as camera orientation are straightforward to compute. Although accurate disparity information for each pixel is also available, the disparity maps were generated following the same procedure as for the real environments dataset.

Both U and θ-disparity segmentations were performed in the same manner (Figures 8, 9 and IV-B), using the same pre-processing parameters, except for the disparity threshold which was set to 10, as better disparity resolution was obtained in
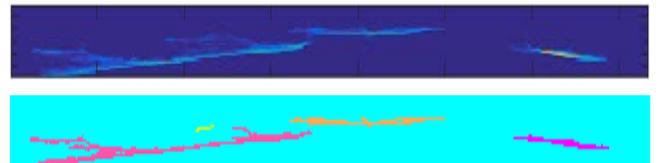
Fig. 9. U-disparity and U-disparity labeled for an image in the virtual environment sequence
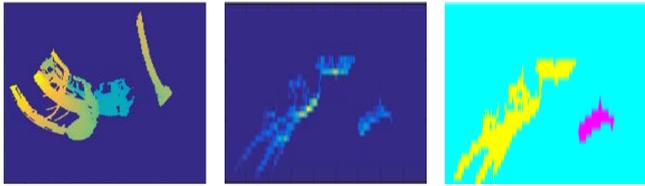
Fig. 10. Virtual environment: Polar representation of the disparity map after ground plane removal, θ-disparity before and after labelling

Fig. 11. Obstacles detection results for the virtual environment dataset: overlaid segmentation results obtained in both θ and U-disparity images

the virtual environment. The combined results of the θ and U-disparity can be observed in Fig. 11.

V. Conclusions

In this paper we proposed a framework for obstacle detection in stereo sequences that makes use of multiple representations of the disparity maps, namely the U-disparity, V-disparity and θ-disparity histograms. The framework is application independent and suited for real-time processing requirements. This is a consequence of approaching the obstacle detection problem in the 2D space instead of processing the 3D point cloud associated with each frame in the sequence. The V-disparity maps are processed for ground plane extraction, while the regions in the image corresponding to obstacles are detected in the U and θ representations together. The results extracted from the θ-disparity image efficiently complement the U-disparity segmentation, as the union or intersection of the two outputs can provide more accurate descriptions of the obstacles, depending on the application. Moreover, the segmentation in the θ-disparity map reveals the radial disposition of the obstacles with respect to the camera/user as well as the free navigable directions in a straightforward way.

The verification of the proposed approach with datasets specific to various applications (automotive, assistive devices) revealed that the algorithm can provide reliable segmentation of the obstacle occupied regions in the environment. In the following we would like to assess this approach in the development of obstacles identification frameworks that would also employ obstacles characterisation, either by means of classification or features descriptions alone.

Acknowledgment

References

[1] R. Labayrade, D. Aubert, and J. P. Tarel, "Real time obstacle detection in stereovision on non flat road geometry through "v-disparity" representation," in *Intelligent Vehicle Symposium, 2002. IEEE*, vol. 2, June 2002, pp. 646–651 vol.2.

[2] R. Labayrade and D. Aubert, "In-vehicle obstacles detection and characterization by stereovision," in *Proc. of the 1st Intl. Workshop on In-Vehicle Cognitive Computer Vision Systems*, 2003, pp. 13–19.

[3] L.-C. Chen, X.-L. Nguyen, and C.-W. Liang, "Object segmentation method using depth slicing and region growing algorithms," in *Intl. Conf. on 3D Systems and Applications*, 2010, pp. 87–90.

[4] B. Kormann, A. Neve, G. Klinker, and W. Stechele, "Stereo vision based vehicle detection," in *VISAPP*, 2010, pp. 431–438.

[5] C. Lee, Y. Lim, S. Kwon, and J. Lee, "Stereo vision based vehicle detection using a road feature and disparity histogram," *Optical Engineering*, vol. 50, no. 2, pp. 1–10, 2011.

[6] C. Lee, "Stereo vision-based obstacle detection using dense disparity map," in *Intl. Conf. on Graphic and Image Processing. International Society for Optics and Photonics*, 2011.

[7] 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, Oct 2008. [Online]. Available: https://hal.inria.fr/inria-00325455

[8] M. Bujacz, "Representing 3d scenes through spatial audio in an electronic travel aid for the blind," 2010, phD Thesis, Technical University of Lodz. [Online]. Available: http://www.bujaczm.os1.kappa.pl/Bujacz Michal - PhD Thesis.pdf

[9] 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. 17 476–17 496, 2012.

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

[11] Z. Hu, F. Lamosa, and K. Uchimura, "A complete u-v-disparity study for stereovision based 3d driving environment analysis," in *5th Intl. Conf. on 3D Digital Imaging and Modeling (3DIM 2005), 13-16 June 2005, Ottawa, Ontario, Canada*, 2005, pp. 204–211.

[12] L. Nalpantidis, D. Kragic, I. Kostavelis, and A. Gasteratos, "Theta-disparity: An efficient representation of the 3d scene structure," in *Intelligent Autonomous Systems 13 - Proc.of the 13th Intl. Conf. IAS-13, Padova, Italy, July 15-18, 2014*, ser. Advances in Intelligent Systems and Computing, vol. 302. Springer, 2014, pp. 795–806.

[13] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 328–341, Feb 2008.

[14] R. A. Hamzah, R. A. Rahim, and Z. M. Noh, "Sum of absolute differences algorithm in stereo correspondence problem for stereo matching in computer vision application," in *Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on*, vol. 1, July 2010, pp. 652–657.

[15] A. Geiger, M. Roser, and R. Urtasun, "Efficient large-scale stereo matching," in *Asian Conference on Computer Vision (ACCV)*, 2010.

[16] P. V. C. Hough, "Machine analysis of bubble chamber pictures," in *Proceedings, 2nd International Conference on High-Energy Accelerators and Instrumentation, HEACC 1959*, vol. C590914, 1959, pp. 554–558.

[17] A. Geiger, J. Ziegler, and C. Stiller, "Stereoscan: Dense 3d reconstruction in real-time," in *Intelligent Vehicles Symposium (IV)*, 2011.

[18] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, June 2012, pp. 3354–3361.