Ground Geometry Assessment in Complex Stereo Vision Based Applications

Adrian Burlacu, Andrei Baciu, Vasile Ion Manta and Simona Caraiman
"Gheorghe Asachi" Technical University of Iasi, Romania,
Faculty of Automatic Control and Computer Engineering,
Email: aburlacu@ac.tuiasi.ro

***Abstract*—Accurate ground area detection is one of the most important tasks in various stereo vision based applications, such as autonomous driving or assistive technologies for visually impaired. Correct assertion of the ground geometry improves obstacle detection algorithms by eliminating false positive locations in image. In this paper we provide an application oriented evaluation on the correlation of the ground geometry with the quality of the disparity map. The disparity map is processed in its V-map representations and two methods for 3D ground area points identification are discussed. Next, different types of surfaces are fitted and evaluated using real data from automotive and visually impaired assistive applications.**

Keywords: *ground geometry, disparity map, evaluation, automotive, visually impaired*

I. Introduction

Free space and obstacle detection represent two of the most active research areas that involve active or passive sensors. Both are undisputed related with an accurate ground area identification. Stereo vision systems are passive sensors that can be used to evaluate data and measure distances in front of the camera. Multiple areas, such as robotics, automotive and assistive systems, benefit from research results involving stereo vision cameras.

The key factor of distance computation is a correct evaluation of the disparities between the left and right image. Processing algorithms for the disparity map were initially reported in applications involving automotive and mobile robotics. In [1] an algorithm that analyses the V-disparity map is used to solve the obstacles detection problem for autonomous driving applications. In this work the goal is to identify free road space and obstacles. Wedel, et.al., [2], tackle the non-planarity of the road surface in real autonomous driving applications. They reveal an algorithm that uses flexible B-spline curves for piecewise planar or quadratic ground detection. The algorithm was evaluated mostly on highways were the free space in front of the vehicle is large and the B-spline curves can be accurately recovered.

The standard assessment for ground area is that it can be modeled as a plane. This assumption is true in applications that can generate dense disparity maps with low noise [3]. Otherwise, in conditions of non-uniform illuminations, sun glare, shadows, etc., the plane fitting involves multiple post filtering steps. These problems can be solved by modeling the ground using high order surfaces.

In the past two decades an important amount of work was reported on developing systems that can improve the quality of life for visually impaired. Visual impairment, or vision loss is a severe condition that seriously affects the life of the individuals suffering from it. Blind persons face challenges doing everyday things we take for granted, like reading or walking. One of the main aspects that underlie spatial navigation is the allocentric sense, the awareness of one's body relative to the environment. For sighted people, the brain relies mostly on information from the eyes to accomplish the navigation tasks. One of the most important problems for blind people when moving in open or closed environments is precisely the lack of external references, creating a distortion of absolute directions, of the position of their heads and bodies related to free space and obstacles.

Regarding ground area identification for visually impaired applications most techniques assume a plane model. This plane model can be recover using a local approach, such as the one presented in [4], or a global one, as the one depicted in [5]. The local approach assumes the ground to be a plane that can be revered by exploring neighboring patches. The global one makes use of RANSAC for ground plane equation identification. Multiple variations of the RANSAC approach were reported later, each trying to improve the classic approach [6], [7].

In this paper we present a ground geometry assessment scheme using new disparity map analysis methods. Application wise, we evaluate different ground geometries which can further increase the robustness of obstacles and best free space detection. The targeted applications involve autonomous driving (KITTI database) and visually impaired assistive devices (Sound of Vision Project database). To the authors knowledge, this is the first attempt at evaluating outdoor ground plane geometry for visually impaired assistive solutions.

In the section II the main properties of the disparity map are discussed. The ground geometry assesment scheme is presented in section III, while section IV illustrates the experimental results. In the end of the paper we present conclusions and future work.

II. Stereo Vision

This section presents the main properties of the disparity map and the properties of one of its representation: the V-disparity map.

Algorithm 1: V-Disparity Map Computation

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

```
for each i^th column in d do
    for each j^th line in d do
        if d(i, j) > 0 then
            d_v(j, d(i, j)) + +
```

A. Disparity Map

Stereo vision system require a set of intermediary steps before the practical exploitation can start. These steps are needed for the recovery of intrinsic and extrinsic parameters, which represent the calibration phase of any visual acquisition device. Once the calibration phase is finalized, the acquisition can start. Next, rectification ensures distortion removal and stereo alignment.

For multiple types of applications and especially for autonomous driving and blind assistive technologies, the correspondence problem between the left and right image is of key importance. Solving this problem leads to the computation of the disparity maps [8]. In order to easily find the solution, the correspondence problem can be simplified to a one-dimensional search along the epipolar lines. Thus, building a cost function that measures image similarity and following an optimization procedure, the disparity map can be recovered [9], [10].

Among the methods proposed for computing the disparity, the Efficient Large-Scale Stereo Matching (ELAS) algorithm [11] proved very good results for various types of stereo acquisition systems. Using a Bayesian approach, this algorithm reported good results even for images with relatively low texture. The main idea of the algorithm is to use a triangulation on a set of robustly matched support points to reduce the matching ambiguities of the remaining points [11]. This allows an efficient analysis of the disparity map, yielding accurate dense reconstruction without the need for global optimization.

B. V-Disparity

The geometrical content in 3D reconstructed scene can be interpreted by using the V-Disparity map [1] which embeds information that can be used to recover the ground area and vertical surfaces. Building the V-disparity map requires the computation of disparity histogram for each line in the disparity map. This construction leads to two important features in the V-Disparity map, which are that major planar surfaces have corresponding line representations with predefined slopes and vertical surfaces are connected with vertical line segments.

The V-Disparity map can be computed using the following steps. Let $d(x, y)$ denote the disparity map for a stereo pair, where (x, y) is a position in the image. Consider h, a function that sums up all the positions with the same disparity value on every row of the image $h(d(x, y)) = d_v(x, d)$, where $d_v(x, d)$ denotes the V-Disparity map.

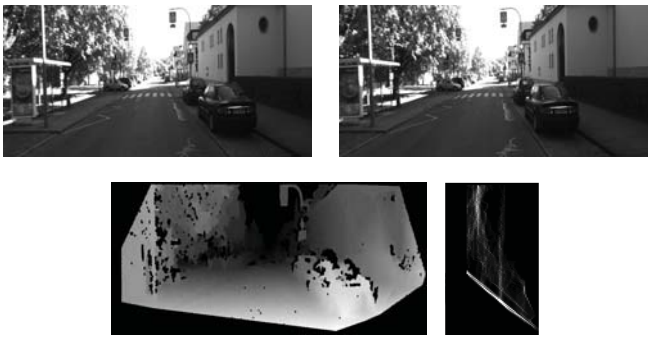

Fig. 1: First row KITTI - left and right frame, Second row: disparity map and v-disparity

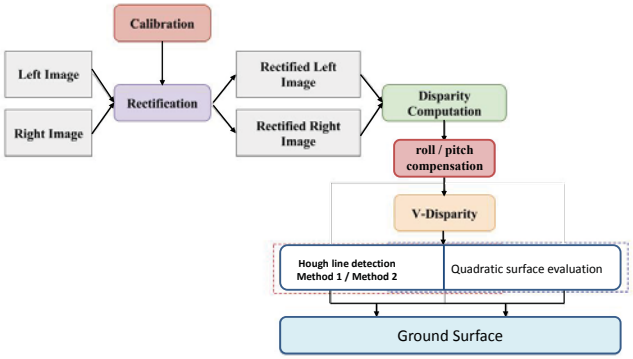

Fig. 2: Ground surface recovery scheme

The implementation of the V-disparity computation steps can be summed up in Algorithm 1. An example of V-disparity map is depicted in fig.1. The left and right frames are part of the KITTI automotive database.

III. Ground Identification and Geometry Assessment Algorithm

In this section, we detail some of the problems that emerge when dealing with ground recovery in complex stereo vision guidance applications. When the ground occupies a significant region in the image it usually corresponds to a dominant line in the V-disparity map and is associated with a planar surface. This is the reason for most of the reported works being on ground plane detection. An important role in establishing the accuracy of the ground plane detection is the quality of the disparity map, which is a key factor that will be further detailed in the experimental results subsection.

As a solution to these problems, a ground area detection scheme is proposed (fig. 2). The first step is to rectify the left and right images using the stereo system calibration parameters. The disparity map is computed using the rectified images. For the proposed scheme, the disparity map is recovered by

employing the ELAS algorithm (section II-A). The dense disparity map is further analyzed and the following ground identification algorithm is obtained.

An important implementation issue rises when the ground is not horizontal. This configuration does not have a direct correspondence in the V-disparity map, which makes the estimation of the ground using the Hough Transform a difficult task. High values of roll angle imply a counter rotation of the disparity map before generating the V disparity representation.

The roll angle of the camera can be recovered by using an extra sensor such as an Inertial Measurement Unit (IMU) [12] or using visual odometry (VO) [13]. Visual odometry is a technique for estimating the motion of a moving vehicle using video input from its on-board cameras. Once the camera motion is available, the values for the roll-pitch-yaw angles are recovered. Using the roll angle, the V-disparity image is rotated to be horizontally aligned, which ensures a robust ground area detection. Once the roll bias is corrected, the detection of the ground area lines in the V-Disparity map can be done by employing the Hough Transform [14].

In order to identify the line in the V-disparity map that is the true correspondent to the ground area we first need to have a value for the tilt angle (pitch) of the stereo vision camera. For automotive applications the value of the pitch has a very small bias from the one obtained in the extrinsic calibration stage. For the visually impaired applications the estimation of the tilt angle tends to be more complex because of the higher number of degrees of freedom for the camera motion.

The initial value of the pitch can be obtained using [2]

$$\tan\theta = \frac{h}{bf}d + \frac{1}{f}(c_y - v_y), \tag{1}$$

where h is the ground related height of the stereo system, b is baseline, f is the focal length, c_y is the s the y-coordinate of the principal point in the image and the pair (v_y, d) is a point in the V-disparity map. A robust estimate of the tilt angle using a set of V-disparity points is obtained by analyzing the histogram of the $\tan\theta$ values calculated in (1). The tilt angle is found as the maximum in this histogram. In addition, as a quality measure, the variance and the number of V-disparity points supporting the found tilt angle are used.

For the resulted pitch, the perfect slope of the line from V-disparity that corresponds to the best ground area candidate is given by [15]:

$$g = \frac{h}{b\cos\theta}. \tag{2}$$

Using the value of the perfect slope, the Hough transform can be used to recover all the straight lines that match a certain confidence interval in relation to g.

To speed up the line detection algorithm and still maintain high robustness, one of the following two methods can be considered:

1) extract the line with the closest slope to g only from the lower half of the V-disparity map;
2) extract the line with the closest slope to g only from the lower quarter of the V-disparity map. Add points from V-disparity that verify the line equation;

The previous methods are employed under the assumption that the ground surface corresponds to a line that contains not have less than a user defined number of pixels located on each line.

Once the line is recovered from V-disparity the 3D fitting procedure can start. For the applications targeted by this work, we have chosen four types of surfaces, each with a different number of parameters denoted by (S $\#parameters$):

$$(S3)z = ax + by + c$$

$$(S4)z = ax + by + cxy + d$$

$$(S5)z = ax^2 + by^2 + cx + dy + e$$

$$(S8)z = ax + by + cxy + dx^2y + ey^2x + fx^2 + gy^2 + h$$

Each of these surfaces will be evaluated in comparison with ground truth annotated stereo frames. The experimental results are presented in the next section.

IV. Experimental results

This section presents the experimental results obtained by our evaluation. Two types of applications were considered: autonomous driving and visually impaired people assistance. Regarding experimental data, for autonomous driving applications we used the KITTI Vision Benchmark Suite [16] (fig. 3 up). For the visually impaired assistive device (fig. 3 down) application we used the Sound of Vision Project (SOV) database (https://www.soundofvision.net).

Fig. 3: up - KITTI acquisition setup; down - SOV acquisition setup

A. Experimental data

1) The KITTI benchmark - Autonomous driving: The KITTI benchmark dataset [16] was design for evaluating autonomous driving research algorithms. The autonomous platform was equiped with high resolution stereo PointGrey Flea2 camera systems. The design of the stereo system lead to a baseline of roughly $b = 0.54[m]$ between the cameras. The sequences were captured by driving on highways or around a mid-size city. Both static and dynamic objects appear in sequences. Manual labeled objects in 3D point clouds provide accurate 3D bounding boxes for multiple object classes including the ground area.

2) The SOV benchmark - Visually impaired people assistive devices: The goal of the Sound of Vision project is to design, implement and validate an original non-invasive hardware and software system to assist visually impaired people by creating and conveying an auditory representation of the surrounding environment. Regarding stereo vision acquisition, a stereo RGB camera - LI-OV580 from Leopard Imaging - is used for outdoor image capture. The two cameras are mounted on separate 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. 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. Regarding resolution, the system acquires and rectifies stereo image pairs of a larger resolution, i.e., 1280 x 720. In order to function properly, processing algorithms that make use of stereo images require the physical characteristics of the cameras. After calibration, the reported intrisic parameters were: focal length $f = 374.742$, projection point coordinates $c_x = 403.958, c_y = 194.223$ and baseline $b = 0.148221[m]$.

B. Experimental Results

In this section we reveal the evaluation results of the proposed ground surface recovery scheme. We target two types of applications: autonomous driving and visually impaired assistive awareness. Each of these applications is challenging, with a higher complexity degree for the assisting devices for visually impaired.

The standard assessment for ground area is that it can modeled as a plane. This assumption is true in applications that can generate dense disparity maps with low noise. Otherwise, in conditions of non-uniform illuminations, sun glare, shadows, etc., the plane fitting involves multiple post filtering steps (e.g. tunning of the height above ground threshold that decides whether a 3D point belongs to the ground or not).

The evaluation process consist in the following steps:

(i) Use the disparity map to compute the V-disparity representation
(ii) The two methods described for line recovery in section III are employed and two corresponding 3D point clouds are computed
(iii) For each of the two point clouds resulted in the previous step, fit the four surfaces (S3, S4, S5, S8) presented in section III.
(iv) Use ground truth labeled images of the tested frames to compute a fitting percentage. This will reveal the accuracy of the line recovery methods and surface model fitting.

A Matlab implementation of the proposed ground surface fitting scheme was used for experimental results. In fig. 4 and fig. 5 two results of the proposed scheme are illustrated. In both cases we can observe that recovering the ground equivalent line from the V-disparity map using method 2 gives better results than method 1 for any of the considered surface models. Also, in comparison with classic planar fitting, adding xy factors and squares of the individual x and y coordinates in the surface equation leads to more accurate results.

These conclusions are sustained by the numerical results that are depicted in Tables I, II and III. In these table we have the fitting percentages for three sequences: a sequence with 30 frames from KITTI, a second sequence with 25 frames from KITTI and a sequence with 35 frames from SOV.

	S3	S4	S5	S8
Method 1	70.03%	70.22%	61.2%	75.13%
Method 2	71.3%	71.48%	62.24%	76.35%

TABLE I: Kitti sequence 1, 30 frames

	S3	S4	S5	S8
Method 1	81.55%	81.71%	74.69%	80.79%
Method 2	82.18%	82.15%	75.33%	82.45%

TABLE II: Kitti sequence 2, 25 frames

	S3	S4	S5	S8
Method 1	71.86%	71.83%	69.35%	70.19%
Method 2	86.5%	88.12%	83.12%	76.74%

TABLE III: SOV sequence, 35 frames

For the KITTI sequences the highest percentage is obtained by the surface model S8, where we have the highest number of parameters. The percentage difference between method 1 and method 2 of v-disparity line detection is lower than 3%. This is a result of large baseline of the stereo acquisition system. In comparison, for the SOV sequence the percentage difference between method 1 and method 2 of v-disparity line detection approximately 15%, which underlines the complexity induced by the shorter baseline and higher motion degrees of freedom. The combination between method 2 and surface S4 gives the highest fitting percentage $(88, 12\%)$.

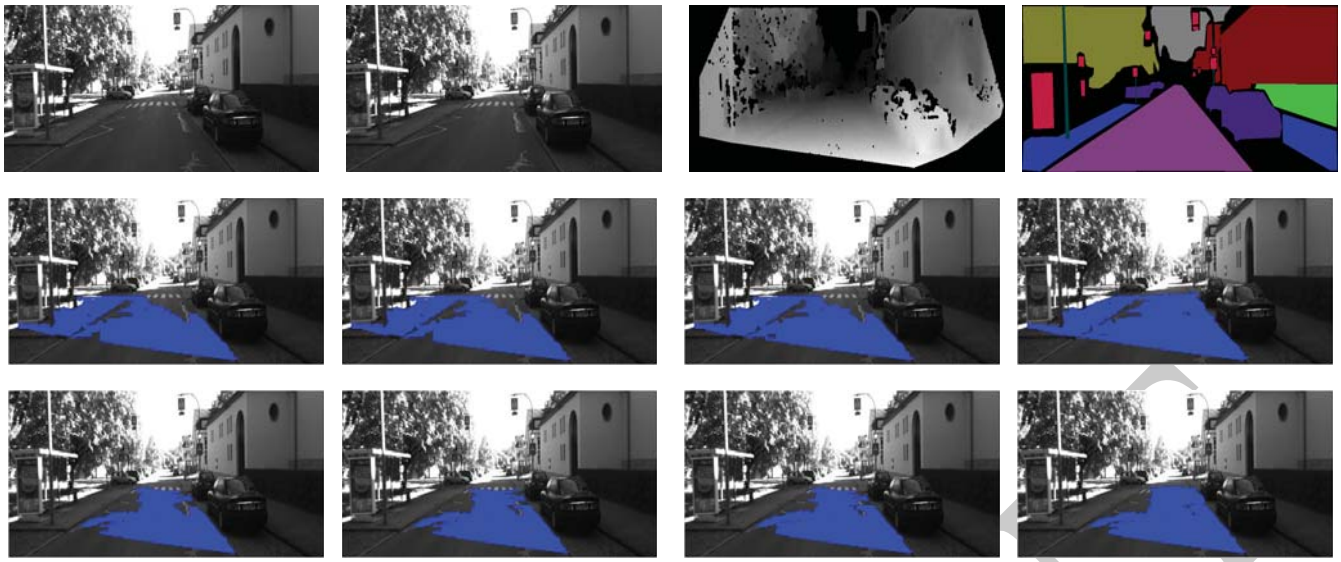

Fig. 4: First row: KITTI - left and right frame, disparity map and ground truth labels
Second row: ground area for each surface from 1 to 4 obtained using method 1 3D points
Third row: ground area for each surface from 1 to 4 obtained using method 2 3D points

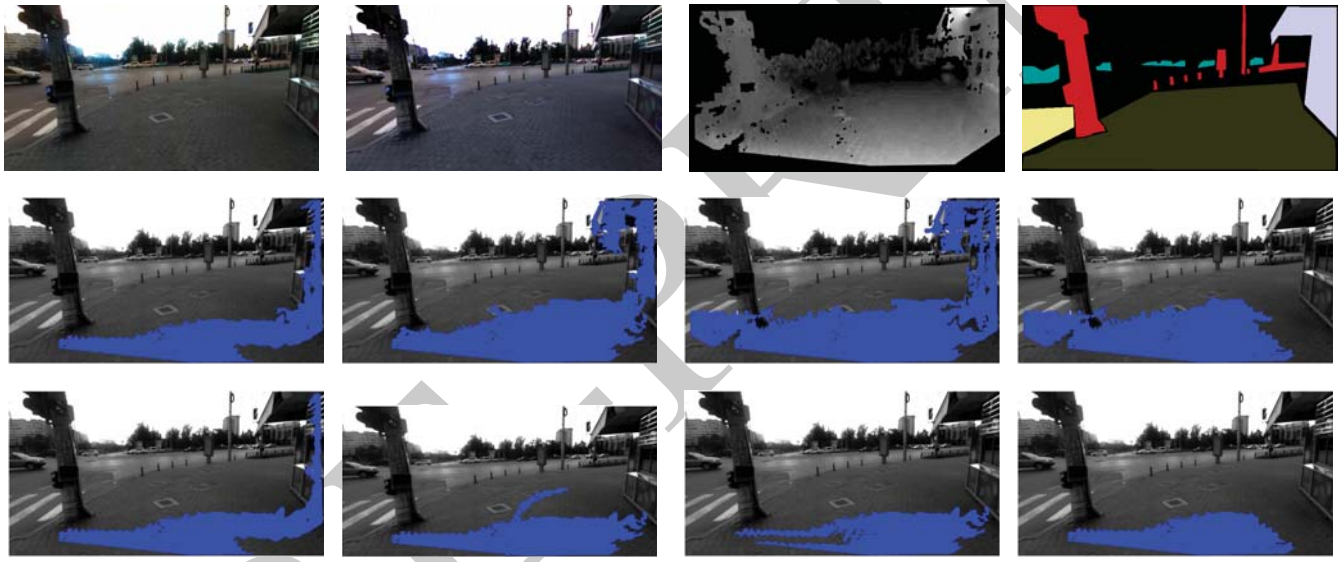

Fig. 5: First row: SOV - left and right frame, disparity map and ground truth labels
Second row: ground area for each surface from 1 to 4 obtained using method 1 3D points
Third row: ground area for each surface from 1 to 4 obtained using method 2 3D points

V. Conclusions

In this paper we proposed a ground geometry assessment scheme for stereo vision based applications. This research was focused on the evaluation of different types of surface models that can accurately describe the ground area in complex applications. Two types of applications areas were chosen for testing benchmarks: stereo based autonomous driving and stereo based visually impaired assistive devices. For autonomous driving the KITTI database was considered, while for visually impaired assistive device the choice was the database of the Sound of Vision project.

The experimental results showed that a key factor is the quality of the disparity map and the length of the baseline of the acquisition system. The accuracy for KITTI dataset can be increased if an equation model with high number of parameters is used $z = ax + by + cxy + dx^2y + ey^2x + fx^2 + gy^2 + h$. For the SOV dataset the best results were obtained by the $z = ax + by + cxy + d$ model. These results lead to the conclusion that in non-ideal acquisition conditions the addition to the standard ground plane equation of xy factor and squaring factors for x

and y coordinates could increase the accuracy of the ground detection.

Future work will be dedicated to real time obstacle detection in complex environments with the visually impaired assistive devices.

ACKNOWLEDGMENT

This project has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement No 643636 Sound of Vision.

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] A. Wedel, H. Badino, C. Rabe, H. Loose, U. Franke, and D. Cremers, "B-spline modeling of road surfaces with an application to free-space estimation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 4, pp. 572–583, 2009.

[3] N. Bernini, M. Bertozzi, L. Castangia, M. Patander, and M. Sabbatelli, "Real-time obstacle detection using stereo vision for autonomous ground vehicles: A survey," in *IEEE 17th International Conference on Intelligent Transportation Systems (ITSC*, 2014, pp. 1–6.

[4] 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

[5] 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.

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

[7] K. Yang, K. Wang, W. Hu, and J. Bai, "Expanding the detection of traversable area with realsense for the visually impaired," *Sensors*, vol. 10, no. 4, pp. 1–20, 2016.

[8] A. Burlacu, A. Bostaca, I. Hector, P. Herghelegiu, G. Ivanica, A. Moldoveanu, and S. Caraiman, "Obstacle detection in stereo sequences using multiple representations of the disparity map," in *Int. Conf. on System Theory, Control and Computing, October 2016*, 2016, pp. 1–6.

[9] 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.

[10] 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.

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

[12] J. Lin and D. Kulic, "Human pose recovery using wireless inertial measurement units," *Physiological Measurement*, vol. 33, no. 12, pp. 2099–2115, 2012.

[13] Y. Jiang, H. Chen, G. Xiong, and D. Scaramuzza, "Icp stereo visual odometry for wheeled vehicles based on a 1dof motion prior," in *IEEE International Conference on Robotics and Automation, May 2014*, 2014, pp. 1–6.

[14] 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.

[15] J. Zhao, J. Katupitiya, and J. Ward, "Global correlation based ground plane estimation using v-disparity image," in *IEEE International Conference on Robotics and Automation, pages = 529–534, year = 2007,*.

[16] 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.