Turkish Journal of Electrical Engineering and Computer Sciences

Volume 32 | Number 3 Article 8

5-20-2024

Stereo-image-based ground-line prediction and obstacle detection

EMRE GÜNGÖR

AHMET ÖZMEN

Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2024) 32: 465 – 482

doi:10.55730/1300-0632.4081

Stereo-image-based ground-line prediction and obstacle detection

Emre GÜNGÖR[1], **Ahmet ÖZMEN**[2*]
[1]Department of Computer Engineering, Kütahya Health Sciences University, Kütahya, Turkiye
[2]Department of Software Engineering, Sakarya University, Sakarya, Turkiye

Received: 08.02.2023 • **Accepted/Published Online:** 27.04.2024 • **Final Version:** 20.05.2024

Abstract: In recent years, vision systems have become essential in the development of advanced driver assistance systems or autonomous vehicles. Although deep learning methods have been the center of focus in recent years to develop fast and reliable obstacle detection solutions, they face difficulties in complex and unknown environments where objects of varying types and shapes are present. In this study, a novel non-AI approach is presented for finding the ground-line and detecting the obstacles in roads using v-disparity data. The main motivation behind the study is that the ground-line estimation errors cause greater deviations at the output. Hence, a novel ground plane is defined as a region in the v-disparity map by using random variables to minimize these errors. In this new approach, weighted least squares regression, outlier detection, and camera height approximation were utilized for determining the ground region with higher accuracy. KITTY-2 dataset was chosen to conduct validation and evaluation experiments of the proposed approach. The experiment results were presented in GitHub, and the performance comparison shows that the proposed approach provides at least 20% improvement over Hough transform, which is a widely used non-AI algorithm. The results were also compared with a recently published article data and the best outcome was obtained among them for the recall metric.

Key words: Stereo image processing, obstacle detection, ground-line detection, v-disparity

1. Introduction

In recent years, image processing studies have attracted more attention for driver perception applications, which are fundamental to advanced driver assistance systems. In addition to obstacle detection or autonomous vehicle control systems, stereo image processing has also been used in various fields, and many scientific researches have been conducted to solve problems in image recognition, video game and movie industry, etc. It is anticipated that the need for autonomous vehicles in the fields of service and asset distribution will increase in the near future. In this study, a ground-line prediction method based on stereo vision and v-disparity, especially for autonomous vehicles, is presented. The ground-line describes a clean path on the ground that is free from obstacles and has more general meaning, whereas the road-line is mostly used for discovering the road (or trajectory) on the image for traffic problems [1]. While this work is more focused on ground-line detection, the solutions found can also be used for road-line detection using region of interest features.

Ground-line estimation and obstacle detection rely on real-time 3D sensor systems. Radar, LiDAR, or stereo vision systems are widely used sensor systems for detecting objects in the scene. Radars, in particular, offer robust solutions in varying environmental conditions like rain, dust, or sunlight. Operating at very high frequencies (76-81 GHz), they face challenges in obtaining sufficient output power and maintaining linearity.

*Correspondence: ozmen@sakarya.edu.tr

[2]. Another alternative is light detection and ranging (LIDAR), which has been used to detect and classify both static and dynamic obstacles using voxel-based representation [3]. While LIDAR systems can measure the distances more precisely, commercially available LIDAR systems tend to be rather expensive and require computationally intensive hardware modules [4]. Stereo-vision systems, on the other hand, offer economical and long lasting solutions. However, challenges such as varying lighting conditions and the need for rapid conversion from 2D images to 3D data space persist. Nonetheless, the output of such systemshelps us solve two major driving system's assistance problems: 1) Obstacle detection, and 2) Road-line estimation.

Several studies have explored identifying obstacles or ground-lines using image features, such as corner detection on stereo images [5].For instance, a real-time obstacle detection approach utilized stereo correspondence with matching scores, with the disparity space utilized to identify optimal object boundaries [6]. In another study, ego-motion was employed to identify optimized obstacle boundary points [7], while inertia measurements were combined with image segmentation from stereo cameras on boats to detect obstacles at sea [8].

In general, the calculation of disparity is a primary step for ground-line estimation or obstacle detection, which affects overall system performance. Therefore, new solutions in disparity calculations that improve computational complexity or outcome accuracy are important contributions in the field [9]. In another study, Yuan and his team developed a non-AI method that utilizes both U and V disparity maps to detect dynamically appearing objects in autonomous driving scenarios [10]. By focusing only on moving objects in the area (narrowing the processed area), they shortened the processing time while increasing obstacle detection accuracy. In this study, the proposed approach was evaluated using the KITTI Vision Benchmark Suite. Aside from v-disparity and u-disparity, θ-disparity has been employed for multiple representations in the literature [11, 12]. Moreover, non-flat ground geometry estimations are calculated using cubic B-spline curves, as road profiles do not always change linearly. Such non-linear road profile estimations are modeled using cubic B-spline curve [13].

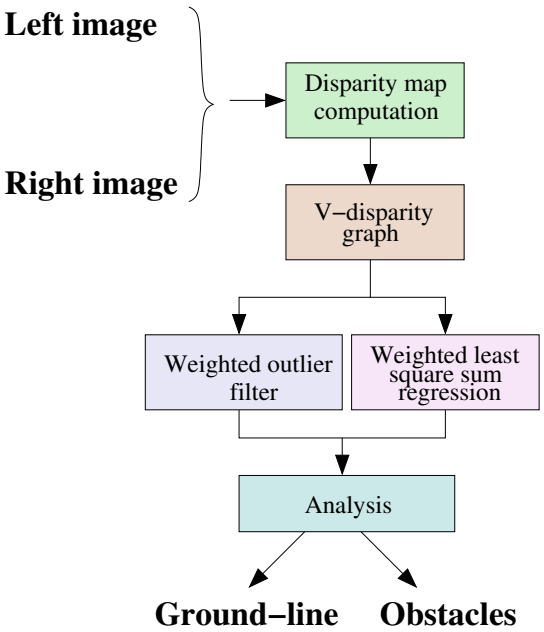


Figure 1. The block diagram of the proposed method.

The Hough transform is a widely used non-AI-based approach for detecting obstacles and estimating ground-lines in stereo images [11, 13–16]. However, it can yield erroneous results under certain conditions, such as when the standard road structure is absent, such as in the presence of potholes, roads passing through forests, or bridges on the road. Therefore, this research aims to develop an improved non-AI algorithm that is less susceptible to road defects or environmental factors without increasing computational requirements.

Recently, deep learning methods have shown promising results in obstacle (object) detection or ground-line prediction. However, they have some limitations: 1) Excessive computing requirements lead to latency issues. Even light-weight deep learning model for stereo video processing can cause unacceptable delays in real -time applications when sufficient hardware, such as a GPU, does not exist [17, 18], 2) State-of the art deep learning models may lack robustness [19, 20]. When input sets or parameters change in AI-based algorithms, it often requires careful consideration and potentially significant adjustments. Reevaluation, retraining, and fine-tuning may be necessary to ensure that the algorithm continues to perform effectively in the new context. To address these shortcomings, a voxel-based approach is proposed to confine the computation within a region of interest (ROI) and align the resolution of the disparity estimation with the occupancy grid [21]. In a recent review study in the field of image processing-based obstacle detection in railways, it was noted that AI methods may be inadequate or yield problematic results, and in some cases, cannot be used at all. An example provided in the article illustrates that if the training data for detecting a tree fallen on the road is based on synthetic data and does not adequately resemble real-world scenarios, the success rate of detection significantly decreases [23]. Similar findings were also reported in another review study, highlighting that despite significant advancements, deep learning techniques encounter challenges in complex and unfamiliar environments with objects of various types and shapes present. [24].

In this study, a novel v-disparity graph-based method that does not rely on AI has been proposed. The new algorithm utilizes weighted least squares (WLS) regression and outlier filtering. The block diagram of the proposed method is depicted in Figure 1. Experiments conducted during the study revealed that the new approach increased the success rate by 20% compared to the Hough transform algorithm. Moreover, both qualitative and quantitative outcomes demonstrated that the new approach is more robust and less prone to errors caused by obstacle distribution or erroneous data in the input image.

The remaining parts of the paper are outlined as follows: In Section 2, a summary of the related literature and the environmental setup for the study is presented. In Section 3, the algorithm, data structures, and computation flow are presented in details. Section 4 presents experimental results with image database, comparison metrics, comments and discussions about the results are presented. Section 5 concludes the paper.

2. Related work

Calculating the v-disparity map is the first step for ground-line estimation in non-AI methods, then the obstacles on or around the roadway can be located. Disparity map is fundamentally created using pixel differences between projected points in the left and right images, as shown in Figure 2a. The ground-line can be extracted from v-disparity graph; however, this graph includes more points representing either errors, outliers, or more distant objects. Figure 2b displays an v-disparity graph derived from a sample of stereo road images in the KITTI dataset.

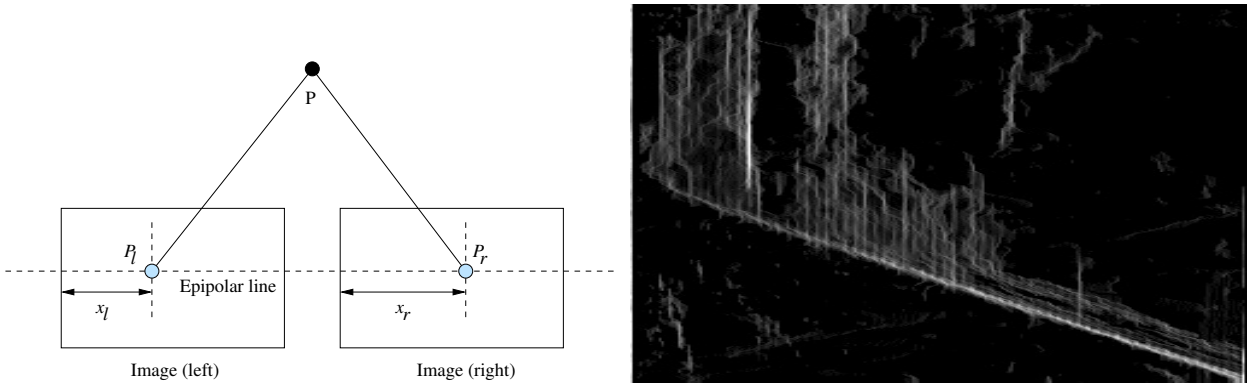


Figure 2. a) A target scene at point P and its associated left and right images. b) Example v-disparity map. The x-axis represents the disparity values or the depth differences between corresponding points in the left and right images. The y-axis represents the vertical position of points in the scene.

2.1. Disparity map, u-disparity and v-disparity

Creating 3D information from 2D images necessitates a stereo camera setup equipped for this purpose. The generation of a disparity map or depth map involves computing pixel disparities between two images captured by left and right cameras. Various methods can be employed to calculate pixel differences. In this study, the semiglobal block matching algorithm is utilized to produce the disparity map for distance data [22]. In the real world, a single point is projected onto a 2D stereo environment where pixel offsets occur relative to its distance from the cameras. Equation 1 illustrates the fundamental calculation structure for determining the disparity value based on the pixel location in the left and right images.

$$d = |P_l - P_r| \tag{1}$$

Since the left and right camera images are in the same plane, the disparity d corresponds to the difference between their horizontal coordinates $(x_l - x_r)$ in calibrated camera systems. To determine the disparity values in block matching algorithm, the associated pixels in the scene must be identified. The block matching algorithm finds pixel similarities using block search. Equation 2 defines $\vec{D}$ as a matrix of pixel differences between stereo images denoted as $I(x, y)$. The algorithm produces block vector lists, which encapsulate differences between associated pixels. A search in these lists aims to identify the minimums for each pixel using Equation 3. The resulting output of the algorithm is an image known as the disparity map. Subsequently, to detect gradual changes in the disparity map and locate obstacles, their derivatives are computed in the next step. The ground plane also shows gradual changes in the disparity map from horizon to camera location, which can be used to extract the ground plane.

$$\vec{D} = \sum_{(i,j)\in I(x,y)}^{N} |P_l(i,j) - P_r(i-d,j)| \tag{2}$$

$$\vec{d} = \min_{d\in[d_{min},d_{max}]} (\vec{D}) \tag{3}$$

In stereo vision, u-disparity and v-disparity maps are used to extract depth information from a stereo image pair. These maps provide an alternative representation of the depth map that can be useful in certain applications. The U-disparity map represents the discrepancy values along the horizontal axis of the image, while the v-disparity map does so along the vertical axis. Each pixel in the v-disparity map corresponds to a specific vertical position in the stereo image pair and contains the corresponding disparity value [25, 26]. Analyzing the v-disparity map helps in understanding the depth distribution along the vertical direction. V-disparity values are calculated using Equation 4. The primary distinction between u-disparity and v-disparity lies in the direction they represent. Both maps offer valuable insights into depth distribution in stereo images and can be utilized in various computer vision applications, such as obstacle detection, scene understanding, and 3D reconstruction.

$$v = \frac{[v_0 sin(\theta) + \alpha cos(\theta)] + [[v_0 cos(\theta) - \alpha sin(\theta)]}{(Y + h)sin(\theta) + Zcos(\theta)} \tag{4}$$

Figure 2b illustrates an example v-disparity map. The x-axis illustrates the disparity values or the depth differences between corresponding points in the left and right images. Each disparity level on the x-axis corresponds to a specific range of disparities, with smaller values on the left and larger values on the right. On the other hand, the y-axis represents the vertical position of points in the scene, where the top of the v-disparity map corresponds to the top of the image, and the bottom corresponds to the bottom of the image. Each row on the y-axis corresponds to a specific vertical position in the scene. The y-axis helps us understand how the depth or disparity information varies with height in the scene. Consequently, it reveals the heights of objects and their relative positions in the environment. The inverse diagonal line in this figure represents the ground-line. This study focuses on developing a systematic approach for estimating this ground-line with higher precision. Disparity values below the diagonal line, depicted in gray, indicate disparity errors, while those above the diagonal suggest a higher likelihood of obstacles existing around the ground-line. Although the slope of the road exhibits similar behavior, obstacles are generally observed near perpendicular angle relative to the horizontal axis of the v-disparity.

Figures 3a–3c show how obstacles are located using the estimated ground-line. In Figure 3a, the ground-line is observed as a straight line at a certain angle. Obstacles tend to be represented perpendicularly in the same graph. After the ground-line estimation, the v-disparity graph is divided into two segments by the ground-line diagonal: objects above the line are classified as obstacles, while the rest is regarded as part of the ground-line. This transformation is depicted in Figures 3b and 3c.

2.2. Hough transform

Hough transform method is commonly employed in numerous non-AI based studies for ground-line estimation [11, 13, 15, 16]. This technique can be defined as a representation of all line segments in the v-disparity graph using Equation 5, where r denotes the length of a normal from the origin to this line, and θ represents the orientation of r with respect to the x-axis.

$$r = xsin(\theta) + ycos(\theta) \tag{5}$$

$$y = -x\frac{cos(\theta)}{sin(\theta)} + \frac{r}{sin(\theta)} \tag{6}$$

So, all lines in the v-disparity graph become known after using Equations 5 and 6 where $\theta \in [0, 180]$ and $r \in \boldsymbol{R}$ [27].

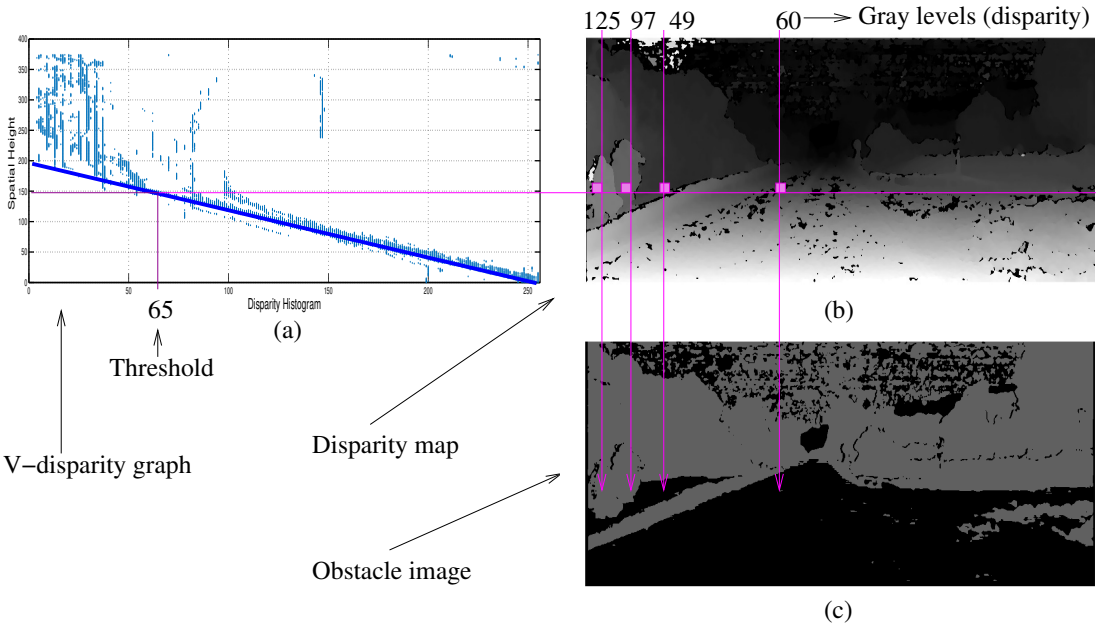


Figure 3. a) V-disparity map, b) depth map with gray color levels, c) ground-plane (black) and obstacles (dark gray). An obstacle in the image is located by making calculations on the depth map image.

In our test results, it is seen that using Hough transform could yield lower success rate or erroneous results under different conditions where standard structure of the road is not present. The situations where Hough transform fails according to the studies and our experimental results are described as follows:

- The Hough transform might struggle to accurately identify the ground-line in the v-disparity graph, especially when distant obstacles affect the distribution. During the experiments, it has been observed that Hough transform may fail on unstructured roads because it is harder to find a suitable threshold on unstructured roads [26]. While bounding the interval search of plane parameters in the Hough space to specific intervals is one approach, it does not resolve the issue in all cases since a poor distribution can lead to an erroneous estimate close to the correct one.
- In the images where some pixels corresponding to the trees on the horizon line accumulate to form a vertical alignment in the v-disparity image. The frequency and accumulation of vertical similarities could result in erroneous outcomes, particularly in the near and far end regions of the estimated line.
- The Hough transform has the tendency to interpret the natural ground-line as multiple smaller line segments based on selected regions in the stereo image. This segmentation into multiple line segments poses a challenge for selecting the best possible fit, often leading to a higher error rate in predictions.

Ground-line estimations using Hough transform represent a logical and generally a sufficient solution for well-structured roads. Therefore, our study becomes an important alternative in non-AI approaches, particularly when the Hough transform performs poorly in environmental conditions such as vehicles turning or advancing in urban areas or forested regions, where the distribution of obstacles at near or far ends exceeds the expected amount.

3. Methodology

In stereo vision systems, the initial step typically involves computing the disparity map from the stereo image, followed by obtaining either the u-disparity or v-disparity graph, depending on which is needed for ground-line estimation. The block diagram in Figure 1 illustrates the proposed method for ground-line prediction and obstacle detection in stereo vision systems.

The ground-line generally lies in a region that is visible as an inverse diagonal shape in the v-disparity graph. One concrete method to obtain the ground-line involves just taking the y-axis value of the minimum visible pixels in v-disparity graph. However, the graph usually contains erroneous data and outliers, as observed in the lower left corner in Figure 2b. In the v-disparity graph, the x-axis values are normalized integers ranging from 0 to 255, while the y-axis values represent the image height. The light-gray stains indicate that the frequencies at certain rows are capped at 255, the maximum value. Since disparity errors generally fall into the numerical zero value category, the first column of the v-disparity graph is omitted in equations to reduce their inclusion.

In this study, the depths to objects on the v-disparity graph are represented as random variables. Subsequently, the ground-plane is defined as an area formed by these random variables. Both treating the disparity as a random variable and defining the ground as an area are novel approaches in the literature. The weighted outlier and weighted least squares regression methods are then employed to determine this area.

3.1. Ground-line as a region and random variable

In disparity maps, the minimum value in a row ideally indicates the ground disparity in the absence of obstacles or errors. Leveraging this assumption, we can infer that observations of ground disparity in the data should coincide with minimum values. Moreover, considering the relative distance from the camera, the disparity values of the ground exhibit gradual changes from higher values to lower (zero) values, provided there are no obstacles causing discontinuities.

Such gradual change forms a solid line in v-disparity graph. However, real-world data may deviate from this ideal scenario, making perfect gradual changes unattainable. Therefore, if we take distribution of disparity data as a random variable X in each observation, and assume the intensity of distribution lies in the region of ground line in a concentrated form, this ground region could be represented to reduce errors in both ground plane and obstacle detection. Equation 7 is used to define the ground region. The concentration of random points on v-disparity converges to a linear region in a mean-squared sense. These values fluctuate due to noise and other factors such as sidewalks, pavements, or road cavities. Equation 7 holds with the assumption that there should always be a ground plane present in the stereo images.

$$\lim_{\delta t \to 0} E[\frac{X(t-\delta t)-X(t)}{\delta t} - X'(t)^2] = 0$$
$$X'(t) = \frac{X(t)}{\delta t} \tag{7}$$

Since the samples stem from observation of the real world, the inner product of random variables is used. This inner product could be represented as $< X, Y >= E\{XY\}$ then this should obey $E^2\{XY\} \leq E\{X^2\}E\{Y^2\}$. By choosing $X = Xn(\omega) - X(\omega)$ and $Y = 1$, we obtain inequality $E^2\{(X_n(\omega) - X(\omega)\} \leq E\{(X_n(\omega) - X(\omega))^2\}$. So, by taking $n \to \infty$ we reach:

$$\lim_{n\to\infty} E^2\{(X_n(\omega) - X(\omega)\} = \lim_{n\to\infty} E\{(X_n(\omega) - X(\omega))^2\} = 0 \tag{8}$$

In the light of Equation 8, if we take infinite stereo road image frames and extract v-disparity, convergences of the ground data could easily be seen on the superposition of probabilities (or observations). This situation also persist in even single v-disparity graph, where the ground region in that case converges to a definite line (see Figure 4). Consequently, according to wide sense stationary (WSS) random process, two conditions are satisfied: 1) The ground plane exhibits a stationary expectation and this property can be seen in the v-disparity graph. 2) Another criterion is the relation between the height of the camera and disparity values.

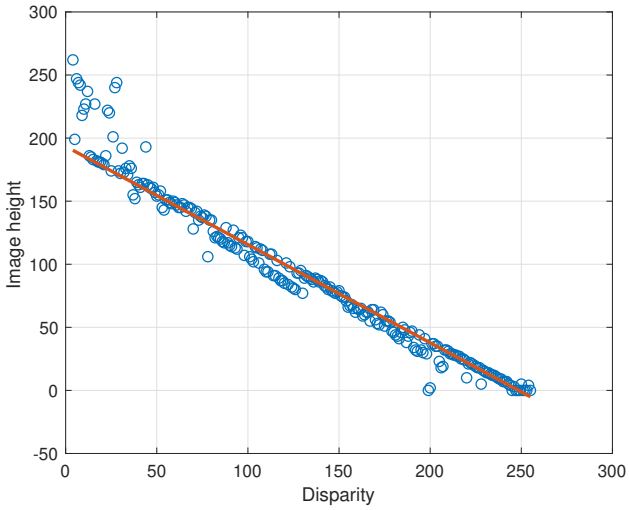


Figure 4. V-disparity of minimum values and weighted least square (WLS) sum regression result (red line).

Since the v-disparity ground region is a random variable, t in $X(t)$ signifies the distance from the camera. Consequently, each random variable that represents the road becomes a region in the v-disparity graph. Therefore, derivatives and distribution of random data are used as tools to define the region of ground plane in v-disparity. However, noise and outliers may also be in the data, and they must be filtered out.

3.2. The ground line and region estimation

The minimum values in each column are selected after excluding nondominant low-frequency outliers, based on the assumptions that the ground is a region and lies at the minimum points in each row. Subsequently, the ground line or region can be readily calculated by thresholding the intensities of the image as a $[MxN]$ matrix, where M and N are ranges of height and disparity values, respectively. Thus, low-frequency disparity data with a high probability of error is eliminated. After eliminating the possible erroneous data, the minimum values are determined for each disparity value based on the defined disparity resolution.

According to the differentiability of the random variables specified in Equation 7, two operations are applied to define a better region and estimate the ground-line with minimum error. The first operation involves applying weighted least square (WLS) regression. According to our definitions, the ground region is convergence, and as the frame number approaches infinity, it becomes a ground-line. Therefore, to minimize the errors in

estimation, WLS regression is applied to find a vector $\beta \in \mathbb{R}$ $(n < m)$, $y \approx W\beta$ in an Euclidean sense. In this way, $||y - W\beta||_2$ will be minimum. When errors are included as random variables, the measurements become $y \approx W\beta + \epsilon$, where β and ϵ are both random variables. To find the best fit (where $||y - W\beta||_2$ is minimum) and to reduce the error variable introduced by the ϵ term, two different statistical methods are used: 1) WLS, as mentioned before, and 2) an outlier filter to limit the erroneous sample inclusion.

The reason why WLS is used in obtaining the ground-line can be explained as follows: when analyzing the ground region data, it becomes apparent that the closer road region contains more ground data, represented with higher frequencies, whereas the faraway road region nearing the horizon is represented with fewer pixels. Hence, the WLS approach is employed to approximate better and reduce the error. Equation 9 shows the fundamental procedure to calculate WLS.

$$WLS(\beta, \vec{\omega}) = \sum_{i=1}^{n} \omega_i (y_i - \vec{x_i}.\beta)^2 \tag{9}$$

The outlier filter is a well-known method to reduce noise effects in finding the region if the data distribution is concentrated. Instead of a simple outlier region, a ground area region (GAR) value is calculated where the values lie within the range $[Q_1 - (1.5)IQR, Q_3 + (1.5)IQR]$. The GAR value is defined at a distance by the inter-quartile range (IQR) from quartiles to reduce the errors. The ground region is also bounded by this GAR value (see Figure 5a). The median and the first quartile calculations are shown in Equation 10, where L represents the lower level of the median class, i.e. the class containing the middle observation in the distribution, and p.c.f. denotes the predicting the cumulative frequency to the median class. Here, i signifies the class-interval of the median class.

$$\begin{aligned} Median &= L + \frac{N/2 - p.c.f.}{f} \times i \\ Q_1 &= L_{Q1} + \frac{N/4 - C}{f_{Q1}} \times h \\ IQR &= Q_3 - Q_1 \\ IQR_cam &= \pm(1).IQR \end{aligned} \tag{10}$$

In Equation 10, L_{Q1} is the lower limit of the first quartile class, h is its width, f_{Q1} is its frequency and C is the cumulative frequency of classes preceding the first quartile class.

In this study, a ground-plane is defined first in the v-disparity graph as a region, and then the ground-line is estimated within this region to track the changes to provide a better obstacle detection. In contrast, the Hough transform considers only a single ground-line without any regional considerations. The noise and the other factors are also considered during the calculations in this study. Therefore, to improve ground line estimation, the interquartile range is used instead of just the outliers. Figure 5b depicts the GAR graph obtained by using derivatives of minimum height function and its values located based on WLS lines, which creates a tolerance region.

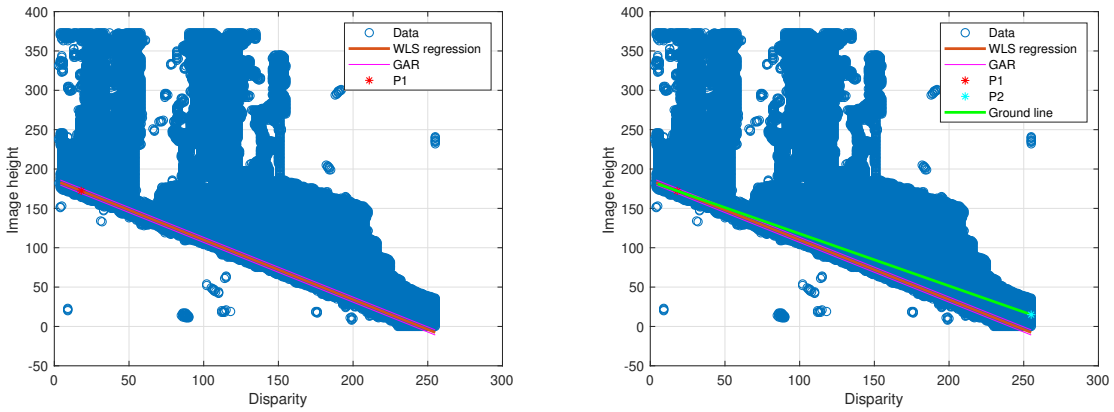


Figure 5. Definition of the random ground area region (GAR) and WLS with tolerance region. a) Derivatives of the minimum height function and GAR values, b) WLS line in the GAR region.

3.3. Camera height on obstacle detection

The stereo camera setup is typically positioned above the car, parallel to the road surface. Any changes in camera height and/or angle will result in corresponding alterations in the v-disparity values. Therefore, this relationship must be integrated into the estimation formula. The v-disparity graph and the camera's height and angle share a relationship that must be considered during estimations. There are two possible approaches to achieve this: 1) incorporating a constant value into the formula to account for the camera height, or 2) automatically estimating the camera height using the ground and obstacles as references (adaptive system). Equation 11 can be employed for an adaptive system. In this equation, statistical outliers are automatically computed by extracting the IQR value from the camera, defined by the maximum v-disparity values at each column of the v-disparity matrix (Figure 6).

$$\begin{aligned} IQR_{cam} &= Outlier_w(threshold_{maxvals}(V_{disp})) \\ P_{cam} &= P_{gnd} + IQR_{cam} \end{aligned} \tag{11}$$

If the error rate in the disparity map is high, it's crucial to limit the slope change of the ground line. This limitation arises due to the potential for incorrect weighting in formulas, caused by the high frequencies present in low-disparity data near the horizon. Consequently, the camera estimations are constrained by the height defined in Equation 12. The parameters in the equation are used as random variable $X(i)$, where $X(i) = (D_i, H_i)$. Here the point at the horizon is defined as $X_{horizon} = (D_0, H_0)$, while the camera and ground variables are represented as $X_{cam} = (D_{cam}, H_{cam})$ and $X_{gnd} = (D_{cam}, H_{gnd})$, respectively.

$$\begin{aligned} H_{lowTh} &= slope_{gnd} * (D_i - D_{max}) + (H_{gnd} - 2 * IQR_{cam}) \\ H_{highTh} &= slope_{gnd} * (D_i - D_{max}) + (H_{gnd} + 4 * IQR_{cam}) \end{aligned} \tag{12}$$

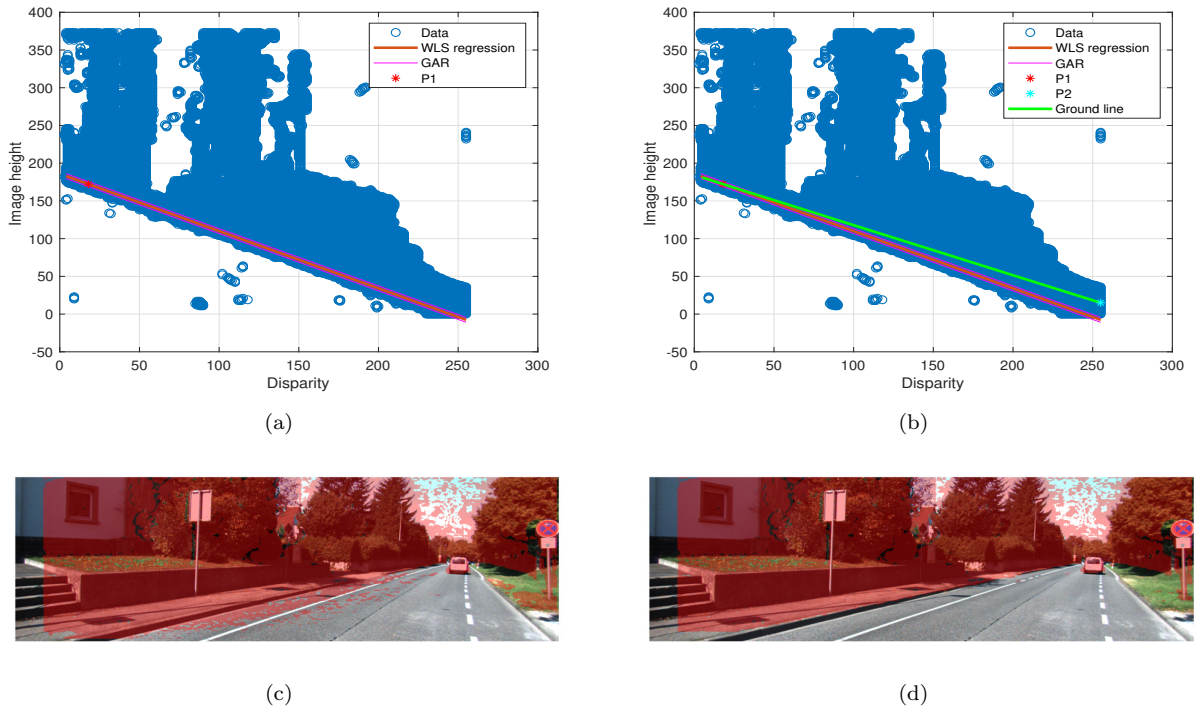


Figure 6. Effect of the camera height on ground line estimation. a) estimation using only WLS, the point P1 shows the value of $X_{Horizon}$ variable for this specific example, b) camera height consideration, the point P2 shows the value of X_{Cam} variable, c) the resulting image without camera height consideration, d) the resulting image with camera height correction.

3.4. Performance evaluation of the proposed algorithm

Pixel-based performance metrics are utilized in the experiments to assess the proposed algorithm and compare its outcomes with the Hough transform. The evaluation requires a calibrated stereo dataset to conduct performance tests and compare approaches with other methods. KITTI Vision Benchmark Suite was found suitable in the literature, and the images were used for testing [28, 29]. Researchers often analyze the performance of their proposed algorithms using a single dataset. In this regard, the KITTI dataset offers a sufficient number of samples to thoroughly test every aspect of the proposed method, including its robustness [30–32]. Ground truth instance segmentation data, based on pixel-level semantics, are obtained from the KITTI Benchmark Suite for conducting comparisons.

Accuracy, precision, and recall are widely adopted metrics for evaluating the performance of segmentation or object detection. These metrics typically quantify the percentage of pixels in the image that were correctly classified. Pixel accuracy is commonly used both for each class individually and globally across all classes. In per-class pixel accuracy, a true positive (TP) signifies a pixel correctly predicted to belong to the given class, while a true negative (TN) represents a pixel correctly identified as not belonging to the given class. A false positive (FP) occurs when a predicted pixel lacks an associated ground truth pixel, and a false negative (FN) arises when the ground truth lacks an associated predicted pixel. The formulas for accuracy, precision, and recall are given in Equation 13.

$$\begin{aligned} Accuracy &= (TP+TN)/(TP+TN+FP+FN) \\ Precision &= (TP)/(TP+FP) \\ Recall &= (TP)/(TP+FN) \end{aligned} \tag{13}$$

An alternative pixel-based performance metric, Intersection of Union (IoU), is also used to evaluate the accuracy of segmentation and object detection algorithms. It provides a measure of how effectively the model delineates object boundaries, thus enhancing segmentation accuracy. Consequently, IoU is utilized to evaluate the quality of segmented regions by quantifying the overlapping area between the predicted region and the ground truth. The formula used in experiments is given in Equation 14.

$$IoU = (Ground\_truth \cap Predicted)/(Ground\_truth \cup Predicted) \tag{14}$$

4. Results and discussion

The complexity of obtaining a v-disparity map from stereo image pairs is typically considered to be linear with respect to the pixel count in the images. This means that the computational effort required to compute the v-disparity map scales linearly with the number of pixels in the left and right images. While the complexity is linear with respect to the pixel count, some optimizations and algorithms can reduce the computational load in practice for real-time applications as we did in this research. For example, the disparity maps are calculated using semiglobal block matching algorithm using 9-pixel blocks [22]. However, it should be noted that the use of 9-pixel blocks may sometimes result in detail lost, such as the blurring of high-frequency textured roads, which may appear as stone-paved roads.

An important contribution of this work is the inclusion of camera height considerations. A case study result is presented in Figure 6b, illustrating v-disparity data, ground region, and line approximation concerning camera height estimation. The system's overall success rate has improved with the inclusion of adaptive tilting property. Moreover, limiting slope changes enhances the system's robustness against errors in the calculation of the disparity map (Figures 6c and 9d).

Another fact is that the ground-line typically appears as an inverse diagonal shape in the v-disparity data. As a specific method, the ground-line can be derived by extracting the y-axis value corresponding to the minimum visible pixels in the v-disparity graph. However, the graph often contains erroneous data, as evident in the lower left corner of Figure 2b, which affects ground-line estimations. An additional significant contribution of this study the proposal of a robust method for determining the ground-line. The impact of this contribution can be observed in Figure 6d.

In the images, there are more pixels of the road plane in the parts closer to the camera, and the pixel count gradually decreases in the distant parts (Figure 7). It is observed that the ground-line estimations near camera become more resistant to noise because there are more pixels representing the road. However, in the horizon estimations, sensitivity to error increases due to the reduction in pixel count, and even one pixel in distant regions may significantly affect the approximation. This finding prompted us to use weighted formulas in both approaches, outlier filtering, and weighted regression, during the study to achieve a more accurate determination of the road plane.

Figure 8 illustrates another example where the road does not extend straight ahead or where environmental obstacles, such as sideways trees, are present. In this instance, the Hough transform incorrectly detects the ground-line (Figures 8a and 8c). However, the proposed method eliminates these errors and reveals a more accurate ground-plane and obstacle identification, as shown in Figures 8b and 8d. In addition to

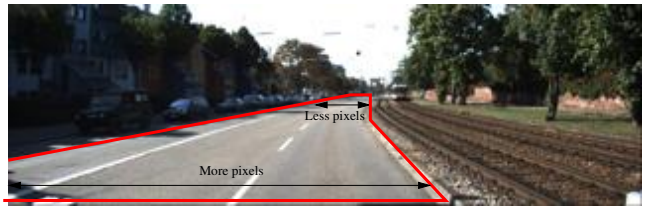
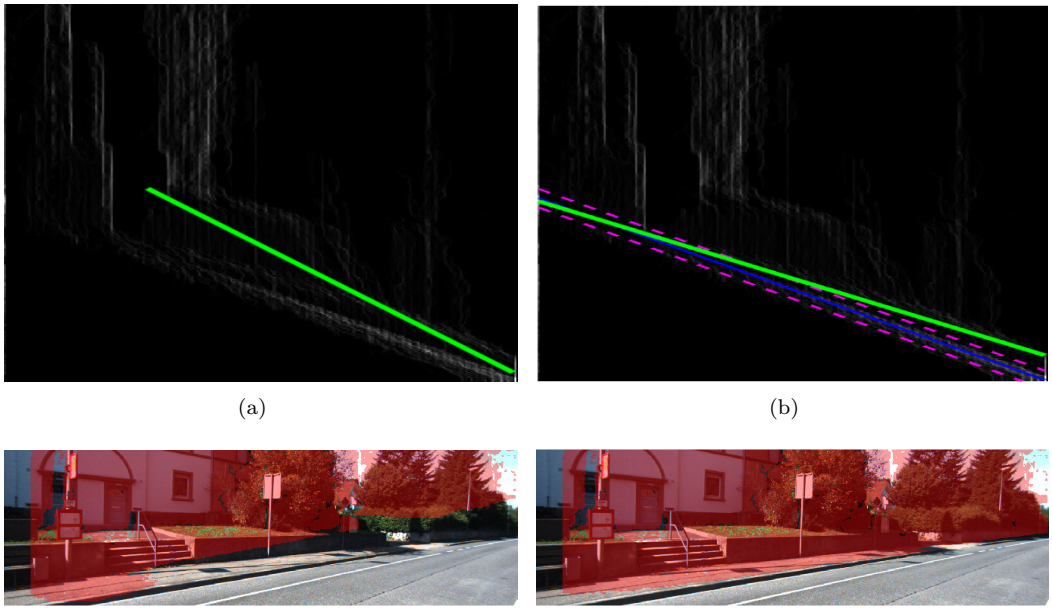


Figure 7. The road shown in the picture is represented by more pixels nearby but fewer pixels farther away.



(a) (b)

(c) (d)

Figure 8. Ground-line output: a) found by the Hough transform, b) found by the proposed method, Obstacle detection results: c) using the Hough transform, d) using the proposed method.

the qualitative comparison, quantitative comparative results are given in Table 1 between the Hough transform and proposed approach using the same image samples from KITTI-2 Benchmark Suite. The results show that at least 20% improvement is achieved using pixel-based metrics such as accuracy, precision, recall, and IoU. The detailed list of all the experiments can be accessed at GitHub repository at the following link: *https://github.com/mregungor/obstacle_detection*.

During the experiments, execution times of both algorithms were measured five times for each example and the average was taken as representative since the variation between samples was not high (see Table 1). It was seen that the Hough transform was on average 2 times faster than the developed algorithm. This is because our proposed algorithm spends more time to improve the accuracy.

Table 1. Comparative results between the Hough transform and proposed method using all samples in KITTI dataset. Four pixel-based metrics are used in the experiments: accuracy, precision, recall, and IoU. The last column of the table shows the average elapsed time of each experiment in seconds. The upper part of the table shows 15 arbitrary samples and the remaining part shows the average performance of all 200 samples (for details please visit *https://github.com/mregungor/obstacle_detection*).

Performance of the proposed algorithm for arbitrary selected samples							
Exp.no	**Img.Id**	**Method**	**Accuracy(%)**	**Precision(%)**	**Recall(%)**	**IoU(%)**	**E.time(s)**
1	0	Hough t.	70.4	45.8	58.1	34.5	**0.22564**
		Proposed M.	**89.5**	**72.5**	**98.0**	**71.4**	0.50434
2	6	Hough t.	44.1	15.0	99.2	15.0	**0.21668**
		Proposed M.	**75.6**	**28.8**	**98.7**	**28.7**	0.57644
3	13	Hough t.	63.7	45.5	53.2	32.5	**0.23118**
		Proposed m.	**85.1**	**70.8**	**93.1**	**67.3**	0.23260
4	27	Hough t.	64.8	47.6	73.2	40.5	**0.21472**
		Proposed m.	**83.6**	**77.3**	**70.7**	**58.5**	0.39622
5	42	Hough t.	51.4	24.6	**72.2**	**22.5**	0.20594
		Proposed m.	75.9	**36.2**	30.9	20.0	**0.18368**
6	60	Hough t.	19.8	10.0	53.6	9.2	**0.21680**
		Proposed m.	**75.3**	**38.0**	**99.1**	**37.8**	0.44452
7	83	Hough t.	72.5	47.6	52.6	33.3	**0.21294**
		Proposed m.	**93.9**	**83.1**	**96.2**	**80.5**	0.41938
8	104	Hough t.	37.2	33.4	85.2	31.5	**0.21238**
		Proposed m.	**48.6**	**39.8**	**100.0**	**39.8**	0.31380
9	121	Hough t.	57.9	25.4	85.0	24.3	**0.20962**
		Proposed m.	**78.2**	**42.1**	**99.3**	**42.0**	0.29312
10	133	Hough t.	49.5	20.3	22.6	12.0	0.21384
		Proposed m.	**95.7**	**99.1**	**86.5**	**85.9**	**0.19724**
11	148	Hough t.	67.0	51.5	78.6	45.2	**0.21136**
		Proposed m.	**87.7**	**75.9**	**94.3**	**72.6**	0.23610
12	165	Hough t.	70.8	28.8	**99.9**	28.8	**0.22768**
		Proposed m.	**81.1**	**38.4**	99.4	**38.4**	1.01898
13	177	Hough t.	66.3	39.0	60.0	31.0	**0.22340**
		Proposed m.	**90.0**	**74.4**	**91.6**	**69.6**	0.66998
14	189	Hough t.	74.5	54.8	84.3	49.8	**0.21512**
		Proposed m.	**95.1**	**88.3**	**96.4**	**85.5**	0.39096
15	199	Hough t.	65.7	37.5	100.0	37.5	**0.21762**
		Proposed m.	**85.6**	**58.9**	**99.2**	**58.6**	0.65478
Average performance of all (200) samples							
-	All	Hough t.	63.4	42.8	69.5	34.7	**0.21699**
		Proposed m.	**85.9**	**67.6**	**94.6**	**64.7**	0.435476

The results of our study were compared in a separate table with a very recent study in the literature that proposes a different non-AI method [10]. In the referenced publication, only 100 selected images were used from the KITTI-2 dataset, however we used all images in our experiments from the same dataset. The experimental results of the reference paper were presented only with precision, recall and F1-score, so we dropped the accuracy and added the F1-score in this table (see Table 2).

Table 2. Comparative results between several non-AI methods published in [10] and our proposed method. Since it was not given in the referenced paper, the accuracy metric is not included in the table.

Method	**Precision (%)**	**Recall (%)**	**F1-score (%)**
Zhou's [33]	61.0	76.7	68.0
Lin's [34]	45.2	88.3	59.3
Zhang's [35]	**85.5**	61.3	71.4
Yuan's [10]	78.3	85.3	**81.6**
Ours	67.6	**94.6**	78.9

Although the computation power consumed was not specifically measured during the experiments, the proposed algorithm require less computation power comparing AI based algorithms. This conclusion can also be drawn from the complexity of the algorithm. Different AI-based algorithms have varying computational complexities, especially some deep learning methods are intrinsically more dependent on computing power than other techniques because these models have more parameters and require more data to train. In order to achieve the desired performance goal with AI-based approaches, it is necessary to add GPU board(s) to the computer. Some examples given in the references use deep learning approaches to achieve the same goal running on GPU boards, but the proposed approach does not need extra hardware for performance related issues [18, 21].

Figure 9 shows an image of example mobile robot developed for use in the logistics industry [1]. This robotic system can go to given targets autonomously, and on the way to the target, it can detect the road plane and the obstacles that suddenly appear in front of it in real time. The system generally creates the road plane from images obtained from the stereo camera system. The mobile system is also equipped with LIDAR sensors and can determine distances to surrounding objects. Battery consumption of these systems, which are in continuous operation, is important. The industry target of the work presented in this paper is such applications.

5. Conclusions

This study introduces a novel approach for identifying the ground-line and detecting obstacles in road images using v-disparity data. The proposed method's complexity scales linearly with the input pixel count, and it does not demand high-performance hardware like a powerful GPU. Conversely, AI approaches may offer superior accuracy but necessitate extensive training and encounter challenges in intricate, unfamiliar environments with diverse objects of varying types and shapes.

In this novel approach, WLS regression, outlier filtering, and camera height adjustment are applied to v-disparity data to enhance detection within the region of interest. The accuracy of ground-line detection is boosted by incorporating weight values derived from depth data, contingent on image proximity. Additionally, the ground-line is depicted using random variables, which ultimately form a ground region, reducing errors in ground and obstacle detection. Outliers, including data irrelevant to ground-line detection and erroneous block matching outputs, are mitigated through statistical methods, leading to a significant enhancement in

[1]Bottobo Robotics (2024). Bottobo Mobile Robot [online]. Website `https://bottobo.com` [accessed 19 February 2024].

overall performance. Experimental studies using stereo images from the KITTI-2 benchmark suite validate the effectiveness of the proposed method. Comparative results demonstrate a minimum 20% improvement over the Hough transform, also a non-AI algorithm. Furthermore, comparison with numerical results from a recent article highlights the proposed method's superior performance, particularly in the recall metric.

(a) (b)

Figure 9. Example stereo vision application used in logistics. a) The mobile robot uses a non-AI approach to process real-time stereo images coming from the cameras to find the road line and detect obstacles. b) Side view while it is traveling autonomously.

To further mitigate errors, future iterations of the algorithms developed in this study could incorporate additional data processing blocks. Integrating new hardware sensors, such as inertia and tilt sensors, into the system has the potential to significantly enhance output accuracy by leveraging data from these sensors.

References

[1] Fan R, Dahnoun N. Real-time stereo vision-based lane detection system. Measurement Science and Technology. 2018; 29 (7): 074005. http://doi.org/10.1088/1361-6501/aac163

[2] Alaba SY, Ball JE. A survey on deep-learning-based lidar 3D object detection for autonomous driving. Sensors, 2022; 22 (24): 9577. http://doi.org/10.3390/s22249577

[3] Asvadi A, Premebida C, Peixoto P, Nunes U. 3D lidar-based static and moving obstacle detection in driving environments: an approach based on voxels and multi-region ground planes. Robotics and Autonomous Systems, 2016; vol. 83, pp. 299-311. http://doi.org/10.1016/j.robot.2016.06.007

[4] Catapang AN, Ramos M. Obstacle detection using a 2D lidar system for an autonomous vehicle. In: 2016 6th IEEE International Conference on Control System; Computing and Engineering (ICCSCE); 2016. pp. http://doi.org/10.1109/ICCSCE.2016.7893614

[5] Huh K, Park J, Hwang J, Hong D. A stereo vision-based obstacle detection system in vehicles. Optics and Lasers in Engineering, 2008; vol. 46, pp. 168-178. http://doi.org/10.1016/j.optlaseng.2007.08.002

[6] Kubota S, Nakano T, Okamoto Y. A global optimization algorithm for real-time on-board stereo obstacle detection systems. In: 2007 IEEE Intelligent Vehicles Symposium; 2007. pp. 7-12. http://doi.org/10.1109/IVS.2007.4290083

[7] Singh D. Stereo visual odometry with stixel map based obstacle detection for autonomous navigation. In: Proceedings of the Advances in Robotics 2019 (AIR 2019), Association for Computing Machinery, 2020. pp. 1-5. http://doi.org/10.1145/3352593.3352622

[8] Bovcon B, Mandeljc R, Pers J, Kristan M. Stereo obstacle detection for unmanned surface vehicles by imu-assisted semantic segmentation. Robotics and Autonomous Systems, 2018; vol. 104, pp. 1-13. http://doi.org/10.1016/j.robot.2018.02.017

[9] Kok KY, Rajendran P. A review on stereo vision algorithm: challenges and solutions. ECTI Transactions on Computer and Information Technology (ECTI-CIT), 2019; 13 (2) 112-128. http://doi.org/10.37936/ecti-cit.2019132.194324

[10] Yuan J, Jiang T, He X, Wu S, Liu J et al. Dynamic obstacle detection method based on U–V disparity and residual optical flow for autonomous driving. Scientific Reports, 2023; vol. 13, pp. 7630-7639. http://doi.org/10.1038/s41598-023-34777-6

[11] Wang B, Fremont V, Florez RS. Multiple obstacle detection and tracking using stereo vision: application and analysis. In: 13th International Conference on Control Automation Robotics & Vision (ICARCV 2014); 2014. pp. 1074-1079. http://doi.org/10.1109/ICARCV.2014.7064455

[12] Burlacu A, Bostaca S, Hector I, Herghelegiu P, Ivanica G, Moldoveanul A, Caraiman S. Obstacle detection in stereo sequences using multiple representations of the disparity map. In: 20th International Conference on System Theory, Control and Computing (ICSTCC 2016); 2016. pp. 854-859. http://doi.org/10.1109/ICSTCC.2016.7790775

[13] Suhr J, Jung H. Dense stereo-based robust vertical road profile estimation using Hough transform and dynamic programming. IEEE Transactions on Intelligent Transportation Systems 2015; vol. 16, pp. 1528-1536. http://doi.org/10.1109/TITS.2014.2369002

[14] Khalid Z, Mohamed EA, Abdenbi M. Stereo vision-based road obstacles detection. In: 2013 8th International Conference on Intelligent Systems: Theories and Applications (SITA) 2013; 1-6. http://doi.org/10.1109/SITA.2013.6560817

[15] Leng J, Liu Y, Du D, Zhang T, Quan P. Robust obstacle detection and recognition for driver assistance systems. IEEE Transactions on Intelligent Transportation Systems 2020; 21 (4) 1560-1571. http://doi.org/10.1109/TITS.2019.2909275

[16] Sabe K, Fukuchi M, Gutmann S, Ohashi T, Kawamoto K et al. Obstacle avoidance and path planning for humanoid robots using stereo vision. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA '04); 2004. vol. 1, pp. 592-597. http://doi.org/10.1109/ROBOT.2004.1307213

[17] Shamsafar F, Woerz S, Rahim R, Zell A. MobileStereoNet: Towards lightweight deep networks for stereo matching, 2021. arXiv CoRR. vol. abs/2108.09770. http://doi.org/10.48550/arXiv.2108.09770

[18] Hongsheng X, Tianyu C, Qipei Z, Jixiang L, Zhihong Y. A deep learning and depth image based obstacle detection and distance measurement method for substation patrol robot. IOP Conference Series: Earth and Environmental Science; 2020. vol. 582. http://doi.org/10.1088/1755-1315/582/1/012002

[19] Wang BH, Diaz-Ruiz C, Banfi J, Campbell M. Detecting and mapping trees in unstructured environments with a stereo camera and pseudo-lidar, 2021. arXiv CoRR. vol. abs/2103.15967. http://doi.org/10.48550/arXiv.2103.15967

[20] Eppenberger T, Cesari G, Dymczyk M, Siegwart R, Dube R. Leveraging stereo-camera data for real-time dynamic obstacle detection and tracking, 2020. arXiv CoRR. vol. abs/2007.10743. pp. 10528-10535. http://doi.org/10.48550/arXiv.2007.10743

[21] Li H, Li Z, Akmandor NU, Jiang H, Wang Y et al. Stereovoxelnet: Real-time obstacle detection based on occupancy voxels from a stereo camera using deep neural networks, 2023. arXiv CoRR. vol. abs/2209.08459. http://doi.org/10.48550/arXiv.2209.08459

[22] Hirschmüller H. Accurate and efficient stereo processing by semi-global matching and mutual information. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005; vol. 2, pp. 807-814. http://doi.org/10.1109/CVPR.2005.56

[23] Ristić-Durrant D, Franke M, Michels K. A review of vision-based on-board obstacle detection and distance estimation in railways. Sensors (Basel). 2021;21 (10):3452. http://doi.org/10.3390/s21103452

[24] Badrloo S, Varshosaz M, Pirasteh S, Li J. Image-based obstacle detection methods for the safe navigation of unmanned vehicles: a review. Remote Sensing. 2022; 14 (15):3824. http://doi.org/10.3390/rs14153824

[25] Labayrade R, Aubert D. In-vehicle obstacles detection and characterization by stereo vision. In: Proceedings of the 1st International Workshop on In-Vehicle Cognitive Computer Vision Systems, 2003;1-7.

[26] Soquet N, Aubert D, Hautiere N. Road segmentation supervised by an extended v-disparity algorithm for autonomous navigation. In: 2007 IEEE Intelligent Vehicles Symposium, 2007; 160-165. http://doi.org/10.1109/IVS.2007.4290108

[27] Cantoni V, Mattia E. Hough Transform. New York, NY, USA: Springer, 2013; 917-918. http://doi.org/10.1007/978-1-4419-9863-7_1310

[28] Geiger A, Lenz P, Stiller C, Urtasun R. Vision meets robotics: the kitti dataset. The International Journal of Robotics Research 2013; 32: 1231-1237. http://doi.org/10.1177/0278364913491297

[29] Menze M, Geiger A. Object scene flow for autonomous vehicles. In: Conference on Computer Vision and Pattern Recognition (CVPR) 2015; 1-10. http:/doi.org/10.1109/CVPR.2015.7298925

[30] Jin X, Yang H, He X, Liu G, Yan Z et al. Robust LiDAR-based vehicle detection for on-road autonomous driving. Remote Sensing, 2023; 15 (12):3160. http://doi.org/10.3390/rs15123160

[31] Wu J, Huang S, Yang Y, Zhang B. Evaluation of 3D LiDAR SLAM algorithms based on the KITTI dataset. The Journal of Supercomputing, 2023. 79 (14). http://doi.org/10.1007/s11227-023-05267-3

[32] Burnett K, Yoon DJ, Wu Y, Li AZ, Zhang H et al. Boreas: a multi-season autonomous driving dataset, 2023. vol. abs/2203.10168. http://doi.org/10.48550/arXiv.2203.10168

[33] Zhou D, Wang L, Cai X, Liu Y. Detection of moving targets with a moving camera. IEEE International Conference on Robotics and Biomimetics (ROBIO), Guilin, China, 2009; 677-681. http://doi.org/10.1109/ROBIO.2009.5420591

[34] Lin SF, Huang SH. Moving object detection from a moving stereo camera via depth information and visual odometry. In: Proceedings of the IEEE ICASI, Chiba, Tokyo, Japan 2018; 437-440. http://doi.org/10.1109/ICASI.2018.8394278

[35] Zhang J, Henein M, Mahony R, Ila V. Robust ego and object 6-DoF motion estimation and tracking. in 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2020. http://doi.org/10.1109/IROS45743.2020.9341552