



HAL
open science

Vehicles Detection in Stereo Vision Based on Disparity Map Segmentation and Objects Classification

Djamila Dekkiche, Bastien Vincke, Alain Mériqot

► **To cite this version:**

Djamila Dekkiche, Bastien Vincke, Alain Mériqot. Vehicles Detection in Stereo Vision Based on Disparity Map Segmentation and Objects Classification. Symposium on Visual Computing, Jan 2016, Las vegas, United States. hal-01697760

HAL Id: hal-01697760

<https://hal.science/hal-01697760>

Submitted on 31 Jan 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Vehicles Detection in Stereo Vision Based on Disparity Map Segmentation and Objects Classification

Djamila Dekkiche^{1,2}, Bastien Vincke² and Alain Mériçot²

¹ IRT SystemX Palaiseau, France

² Université Paris Sud, Institut d'Electronique Fondamentale, Orsay, France

Abstract. This paper presents a coarse to fine approach of on-road vehicles detection and distance estimation based on the disparity map segmentation supervised by stereo vision. Scene segmentation is first performed relying on the robustness of the UV-disparity maps to generate free space and obstacles space. This last is investigated for on-road vehicles detection. The detection process starts with off-road objects subtraction based on the connected component labeling algorithm which is also used for on-road segments extraction instead of the traditional hough transform for more robust, precise and fast detection. Objects classification is then applied to the on-road segments by using some cues describing the geometry of vehicles like width, height. However, these latter have been measured not in meter but rather in pixels in function of the disparity. The whole approach is presented and the experimental results of evaluation are shown.

1 Introduction

Stereo vision systems have recently emerged in the domain of robotics and autonomous cars. These systems provide the 3D perception of the environment which is employed in Advanced Driver Assistance Systems (ADAS) to support a variety of functions including obstacles detection [1], lane departure warning [2] and collision warning systems [3]. While the depth measurement precision of stereo vision systems is not as high as with active sensors such as RADAR and LIDAR, the stereo camera can compete with these active technologies due to its low cost in one hand and the amount of traffic scene information it provides in the other hand.

The litterature describes several works on stereo vision based vehicles detection. The majority of the proposed approaches rely on a depth map obtained usually from a disparity map through stereo matching. The study of the state of the art shows two major axes in this field. The first one includes motion based approaches [4]. The idea is to perform features tracking in the monocular image plane of one of the stereo cameras and 3-D localization in the disparity and depth maps. The second axis deals with the transformation of the disparity map in a more representative and compact form including occupancy grid [5]

and ground surface modeling [6]. These different transformations aim mainly to facilitate scene segmentation and reduce computation time.

The V-disparity [6] is a popular approach for ground plane estimation and road scene analysis [7]. It is a compact representation of the disparity map in a new space more robust and more representative for obstacles detection. This transformation models the road surface as a slanted line and vertical obstacles as vertical lines. Using curve fitting techniques such as Hough transform [8], lines can be detected. The same principle has been used to generate the U-disparity map which has been used for free space estimation [9] as well as for obstacles detection [7].

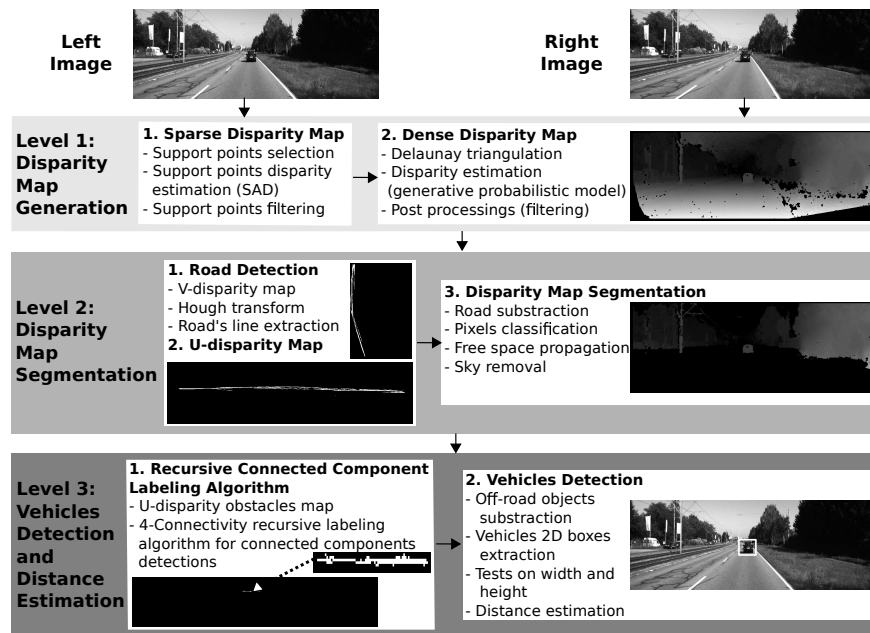


Fig. 1: Algorithm's Functional Diagram

In this paper, we propose a coarse to fine approach for vehicles detection and distance estimation based on stereo vision. The approach can be divided into three processing levels as shown in Fig. 1 (for better lisibility of the different images in this article, readers are encouraged to view them on a computer). The first level deals with the generation of a dense disparity map, the approach proposed in [10] has been used for this purpose. The second level performs scene segmentation relying on the V-disparity map and a set of post processings. Finally, vehicles detection is processed in the third level based on the U-disparity map generated from the obstacles disparity map and a set of post processings.

One of the common issues in on-road obstacles detection are the off-road

information such as high walls, buildings and trees along the road which may affect the precision of the detection and increase the false alarms. The important contribution of this paper is the approach proposed to remove the off-road objects based on the connected segments labeling algorithm. While this approach has been used in some articles [11] mainly for on-road obstacles detection, we propose to use it also for off-road objects subtraction. The second contribution of the paper is the proposed approach for obstacles classification. We rely on a set of tests based on the geometry of the obstacles like the width and height which are not measured in meters but in pixels as functions of disparity. The idea is based on the fact that close objects are projected with many pixels and far objects with few pixels. This solution reduces the errors of disparity quantification and allows more accurate detection.

In the remainder of this paper, the next section describes the complete on-road vehicles detection algorithm. Section 3 presents the experimental results. Finally, the last section sums up the contributions and concludes with future works.

2 Vehicles Detection Algorithm

The proposed algorithm performs vehicles detection as well as distance estimation based on stereo vision. The disparity map is first processed for scene segmentation to get a confident map of obstacles that simplifies the process of vehicles detection. Then, on-road vehicles detection is performed starting from removing off-road objects belonging to the static obstacles space like trees and sidewalks panels. For this purpose, a recursive connected labeling algorithm is applied to the U-disparity map. Finally, a set of post processings is applied for selecting vehicles and identifying their corresponding distances.

2.1 Level 1: Disparity Map Generation

In this paper, we used the Efficient Large scale Stereo matching (ELAS) algorithm [10]. It is a Bayesian approach for dense stereo matching (Fig. 2(b)). It starts by selecting a set of supporting points and identifying their corresponding disparity. These support points are then used to build a 2D mesh via Delaunay triangulation which is processed to find the disparity of the remaining pixels based on a generative probabilistic model.

2.2 Level 2 : Scene Segmentation

The dense disparity map is segmented into two distinctive spaces; free space and obstacles space (Fig. 2). The former one includes road, sidewalks and sky, the later one covers static obstacles (trees, building, panels) and dynamic ones (moving vehicles and pedestrians). The obstacles map can be viewed as a confident map since the probability of detecting vehicles is higher in this map compared to the complete dense disparity map.

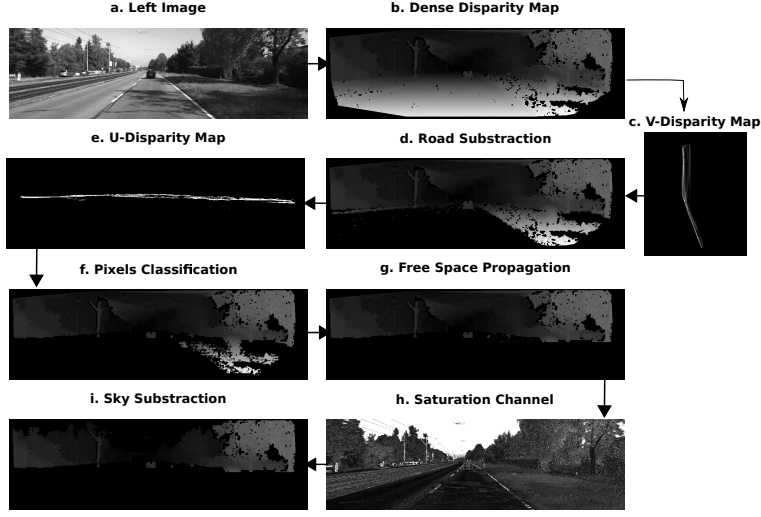


Fig. 2: Scene Segmentation Approach

Road Detection The detection of the road profile is based on the V-disparity space [6] where the x axis plots the disparity d and the y axis plots the image row number. The intensity value of a pixel $p_{vdisp}(d, v)$ on this map is identified According to Eq. (1).

$$p_{vdisp}(d, v) = \sum_{u=0}^{cols} \lambda, \quad \lambda = \begin{cases} 1 & \text{if } p_{disp}(u, v) = d . \\ 0 & \text{otherwise} . \end{cases} \quad (1)$$

The ground plane is projected as a slanted line under the hypothesis that the road's surface is plane and horizontal. The hough transform [8] has been used to detect this line and to identify its equation $ad+b$. Then, for each pixel $p_{disp}(u, v)$ in the disparity map, we calculate its distance $dist$ (Eq. (2)) with respect to the road's line which should be less than a threshold ϵ (few pixels) to classify this pixel as as road's pixel. Through this process, road's pixels are substracted from the disparity map. Fig. 2(b, c, d) shows results of road segmentation (d) based on the dense disparity map (b) and the V-disparity map (c).

$$p_{disp}(u, v) = \begin{cases} 0 & \text{if } dist < \epsilon \\ p_{disp}(u, v) & \text{otherwise} \end{cases}, \quad dist = \frac{|au - vb|}{\sqrt{1 + a^2}} . \quad (2)$$

Pixels Classification The classification phase aims to recover the non-road free space pixels by using the U-disparity map. The intensity of a pixel $p_{udisp}(u, d)$ in the U-disparity map is determined according to Eq. (3).

$$p_{udisp}(u, d) = \sum_{v=0}^{rows} \lambda, \quad \lambda = \begin{cases} 1 & \text{if } p_{disp}(u, v) = d . \\ 0 & \text{otherwise} . \end{cases} \quad (3)$$

If the intensity of a pixel on the U-disparity map is higher than a certain threshold τ , this means that in a certain column u of the disparity map, there are too many pixels with the same distance to the camera and these points belong to potential obstacles. Based on this observation, pixels with high intensity are kept and the others are set to 0 (Eq. (4)). The threshold τ refers to the height of obstacles measured in pixels, in our case, it has been set to 40 pixels. Fig. 2(d, e, f) shows results of pixels classification (f) based on the disparity map after road subtraction (d) and by using the U-disparity map (e).

$$p_{disp}(u, v) = \begin{cases} p_{disp}(u, v) & \text{if } p_{udisp}(u, d) > \tau, d = p_{disp}(u, v) . \\ 0 & \text{otherwise .} \end{cases} \quad (4)$$

Propagation of Free Space Pixels classification has been performed to get two sets of pixels candidates belonging to free space and obstacles space. The neighbour of these initial seed pixels is then examined to determine the class of the non-classified pixels. The idea is to count the contribution of classified free space and obstacles space pixels on the neighbour of each pixel. An accumulator *accum* has been used to count this contribution. If the neighbour's pixel belongs to free space, *accum* is decremented, otherwise it is incremented (Eq. (5)). Fig. 2(g) shows the results of free space propagation.

$$d(u, v) = \begin{cases} d(u, v) & \text{(free space pixel) if } accum < 0 . \\ 0 & \text{(obstacles space pixel) otherwise .} \end{cases} \quad (5)$$

Sky Substraction While many free space pixels have been recovered through free space propagation, wrong classification may happen concerning the sky's pixels classified as obstacles pixels (Fig. 2(g)). The reason is that, the propagation task relies on the analysis of the surrounding pixels, since the sky's pixels are on the top part of the image far from the free space and close to obstacles space, the contribution of obstacles pixels is higher. To remove sky from disparity map, the saturation (S) channel (Fig. 2(h)) of the HSL color space is used and applied as a mask on the last segmented disparity map by using Eq. (6).

$$d(u, v) = \begin{cases} 0 & \text{if } s(u, v) \text{ is } (\text{black}(0) \vee \text{white}(255)) \wedge (v < b - \epsilon) . \\ d(u, v) & \text{otherwise .} \end{cases} \quad (6)$$

The saturation channel has been used because the sky's pixel $s(u, v)$ on this channel is either white or black. It is white in case the sky is blue or grey on the RGB color space and it is black when the sky is white. We may then apply the S channel as a mask to each pixel $d(u, v)$. However, to avoid removing pixels belonging to potential vehicles, the mask is applied only for the part above the horizon line b identified previously (see section. 2.2) with a tolerance range ϵ (30 pixels). The result is shown on Fig. 2(i).

2.3 Level 3: Vehicles Detection Approach

Vehicles detection task relies on the U-disparity map which is less noisy compared to the the V-disparity map. This choice is based on the fact that, in the U-disparity map, obstacles in general are represented by separate horizontal lines even side to side vehicles which is not the case on the V-disparity map where obstacles at the same distance are overlaid and represented by the same line. However, the crucial point with the U-disparity map is how to remove the off-road features. To cope with this issue, connected segments labeling algorithm has been used. Finally, on-road vehicles are detected and recognised based on their geometry. The whole process can be divided into two phases: an off-road features substraction phase and an on-road vehicles detection phase.

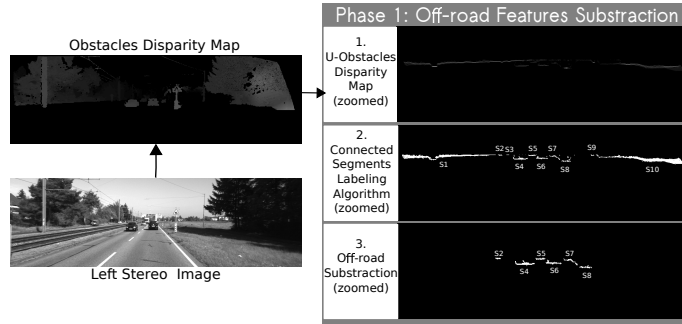


Fig. 3: Vehicles Detection Algorithm: Phase 1 Off-road Features Substraction

Fig. 3 shows an example of the complete off-road substraction phase. First, the U-disparity map is generated from the obstacles disparity map. Then, connected segments algorithm is applied based on the 4-connected neighborhood approach (Fig. 5) as follows:

1. Scan the U-disparity map from left to right and from top to bottom
2. The first non zero pixel is taken as the seed point
3. check for the horizontal (left and right) and the vertical (top and bottom) neighbours
4. Join each non-zero neighbour to the seed point
5. Apply recursively the 4-connected neighborhood for each new joined pixel (steps 3 and 4)

Once the whole U-disparity map is processed, a list of segments is recovered and each one is treated according to its width. To take into consideration the fact that an obstacle looks smaller when it is far and bigger when it is close, the width is measured in function of the disparity. Fig. 4 illustrates the principle. Let the pixel $P_i(u, v)$ be a projection on the image plane of a point $P_w(X, Y, Z)$

in the real world plane. To recover X based on the image coordinate system we rely on Eq. (7). C_u is the projection of the x coordinate of the camera's optical center in the image plane, Z is the distance and f is the focal distance.

$$X = (u - C_u) \times Z/f . \quad (7)$$

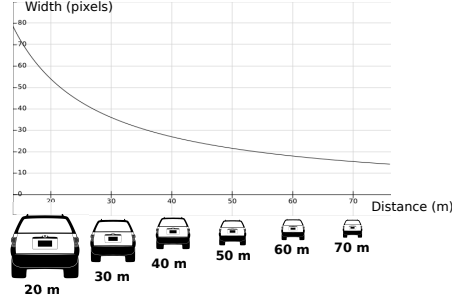


Fig. 4: Vehicle's Width Variation in Pixels with Distance for a 1.5 m Vehicle's Width

Let suppose u_{min} and u_{max} are the minimum and maximum vertical limits of a vehicle in the image plane. By using Eq. (7), we can determine the width W in meter as shown in Eq.(8). Fig. 4 shows the variation of the width in pixels ($u_{max} - u_{min}$) with distance for a fixed vehicle's width of 1.5 m.

$$\begin{cases} X_{min} = (u_{min} - C_u) \times Z/f \\ X_{max} = (u_{max} - C_u) \times Z/f \end{cases} \Rightarrow W = (X_{max} - X_{min}) = (u_{max} - u_{min}) \times Z/f. \quad (8)$$

To remove off-road segments represented by long segments, we have fixed the width interval from 1.5m to 3m and the detection distance from 30m to 70m. Based on this, we have determined the variation interval of the width in pixel to take into consideration. Fig. 3 shows the result of applying this to subtract the off-road features.

Phase 2 : On-road Vehicles Detection

Hypothesis Generation To detect on-road vehicles, the list of on-road segments generated is investigated. To distinguish between on-road vehicles and other on-road objects we rely on the geometry features of vehicles. From the previous phase, the vertical position of each segment on the image plane (u_{min}, u_{max}) is recovered (Fig. 5) and hence the width in meter is deduced according to Eq. (8). Also, the disparity range is determined (Fig. 5). Then, for each on-road segment, in the disparity map region limited by u_{min} and u_{max} , we recover

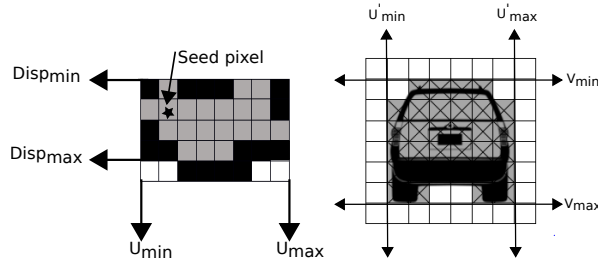


Fig. 5: 4-Connected Neighborhood Approach

Fig. 6: Vertical and Horizontal Vehicle's Position Identification

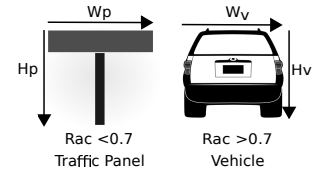


Fig. 7: 2D Box Area/External Contour Ratio Principle

all pixels having disparity in the disparity range $[disp_{min}, disp_{max}]$ (Fig. 5). The horizontal position is then determined (v_{min}, v_{max}) and the height is found. Also we refine the vertical position (U'_{min}, U'_{max}) as shown in Fig. 6. Fig. 8 shows these different steps at the hypothesis generation level.

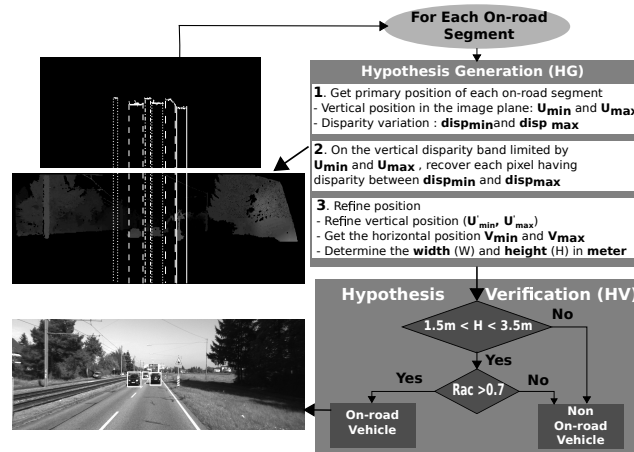


Fig. 8: Vehicles Detection Algorithm: Phase 2 On-Road Vehicles Detection

Hypothesis Verification To select vehicles among other on-road obstacles, two tests are applied (Fig. 8). First, the height in pixels is checked (Eq. 8). Vehicles have height between 1.5m and 3.5m, any obstacle with height outside this range is rejected. If the test is verified, the ratio between the 2D box area and the external contour is computed. The external contour is limited by the pixels which have been recovered (Fig. 6). The reason of using this ratio as a metric

is that, for vehicles, this ration is supposed to be high which is not the case for panels as shown in Fig. 7.

3 Experimental results

The algorithm has been evaluated on KITTI datasets. For each dataset, a list of tracklets is available as the ground truth representing the different objects available on each frame. Each tracklet is presented in velodyne coordinates as a 3D box with its corresponding width, height and length.

The algorithm is implemented on a standard PC with an Intel CPU (i5) of 1.8 GHz. The operating system is Ubuntu 14.04. The disparity map as mentioned previously is generated from LibELAS library based on the approach explained on section 2.1.

3.1 Experimental Design

Before using the tracklets for evaluation, there are some points to take into consideration. The first one deals with the obstacles type; since our algorithm deals only with vehicles, we need first to filter all non-vehicles tracklets. The second point concerns the off-road objects which are not taken into consideration in our algorithm like parked vehicles. Finally, the detection distance range of our algorithm is set up from $30m$ to $70m$, hence, we have to take into consideration only tracklets having distance within this range. To cope with these issues the position and motion history data of the tracklets have been used, for instance, to remove non-vehicles objects, the object's type has been employed. For evaluation, we determine the rate of correct detections or missed ones and false alarms. For better evaluation, these criteria are checked manually when the data provided by the tracklets is not sufficient. To test, we need first to perform a 2D projection of the 3D boxes representing the tracklets. Then, we determine the percentage of intersection between the 2D box of each tracklet taken as a ground truth and the 2D box generated by our algorithm. A detection is then validated once the percentage of intersection is higher than 70%. For evaluation we have selected 3 datasets; Dataset 1: High way, Dataset 2: Urban road and Dataset 3: Rural road .

3.2 On-Road Vehicles Detection Results

Table 1 shows the evaluation results of the algorithm whitout the aid of a tracking module on the 3 different datasets. Two hundreds frames have been selected for each dataset. Dataset 1 contains 270 on-road detections, 241 of them have been well detected. In dataset 2 there exists 236 detections, 211 have been well detected, while the rest have been missed. Dataset 3 contains 128 on-road vehicles associated with tracklets, 211 have been correctly detected.

For evaluation, the three criteria previously presented have been determined for

Table 1: On-road Vehicles Detection Evaluation

Dataset	Correct detections	Missed detections	False alarms
Dataset 1	89.26%	10.74%	8.88%
Dataset 2	89.40%	10.59%	2.96%
Dataset 3	92.19%	7.81%	11.71%

each dataset. During the evaluation, we noticed the absence of redundant detections due to the use of connected component algorithm to recover the segments on the U-disparity map. This solution also increases the accuracy of distance estimation. The results show that high successful detection rate can be achieved as shown in the top left image in Fig. 9 and the estimated distance reaches 70m. Also the classification task works well since different types of vehicles have been detected like cars and trucks, the bottom left image in Fig. 9 illustrates this situation.



Fig. 9: Some Detections Results Obtained from KITTI Datasets

Although the algorithm gives sufficient results, it has still some deficiencies. The first point concerns the use of the connected component algorithm to remove off-road features, although this techniques works well it fails when we deal with on-road vehicles close to high off-road features, in this case, we will loose the vehicles, also this approach is highly sensitive to the stereo matching errors since it is related to the U-disparity map generated from the projection of the disparity map which may increase the false alarms and the rate of missed detections as shown in the top right image of Fig. 9. The false alarms are generally sidewalks panels and road markers as illustrated on the top and bottom right images in Fig. 9. They are usually generated because of the stereo matching errors that affect directly the scene segmentation task and hence the on-road vehicles detection phase. This can be solved by merging the proposed algorithm with a lane marking detection module. We have compared the performances of our algorithm with some contributions cited in Table 2 in terms of the Farthest Detection Distance (FDD), the Correct Detections Rate (CDR) and the False Detections FD .

Table 2: Comparison of our Algorithm with other on-road Vehicles Detection Algorithms

Article	FDD	CDR	FD	Detection Type
Sun et al. [12]	32×32 image region	98.5%	2%	Rear
Alonso et al. [13]		92.63%	3.63%	Rear and front
Bergmiller et al. [14]		83.12%	16.7%	Rear
Southall et al. [15]	40m	99%	1.7%	Single lane rear
Chang et Cho [16]	32×32 image region	99%	12%	Rear
Kowsari et al. [17]	120m	98.6%	13%	Multi view
Our algorithm	70m	90.28%	7.85%	Multi view

4 Conclusion

In this paper, a stereo vision based scene segmentation and on-road vehicles detection algorithm has been proposed. While a variety of vehicles detection algorithms exist in the literature, the proposed algorithm provides improvements to cope with some crucial issues. The first one concerns the off-road features subtraction. It is an important point to deal with when on-road objects are targeted for detection. We proposed to use the connected segments labeling algorithm which has been also used to recover the on-road segments instead of the traditional hough transform for better and fast obstacles detection. For the second improvement, width and height of objects are not measured in meter but rather in pixels in function of the disparity. This solution increases the detection precision and distance estimation. Also, some cues have been presented for objects classification like the ratio between the 2D box area and the external contour of the object. The algorithm has been evaluated on the KITTI vision dataset and the experimental results show that it can detect the most on-road vehicles and determine their distance up to 70m.

For futur works and to improve the current algorithm, a tracking component will be added to increase the rate of correct detections while decreasing the rate of false and missed detections. Also, since ADAS applications are real time, we intend to integrate this algorithm in a dedicated hardware architecture.

5 ACKNOWLEDGMENT

This research work has been carried out in the framework of the Technological Research Institute SystemX, and therefore granted with public funds within the scope of the French Program "Investissements d'Avenir".

References

1. Yu, Q., Araújo, H., Wang, H.: A stereovision method for obstacle detection and tracking in non-flat urban environments. *Autonomous Robots* **19** (2005) 141–157

2. Jiang, Y., Gao, F., Xu, G.: Computer vision-based multiple-lane detection on straight road and in a curve. In: International Conference on Image Analysis and Signal Processing (IASP), IEEE (2010) 114–117
3. Khan, W., Klette, R.: Stereo accuracy for collision avoidance for varying collision trajectories. In: Intelligent Vehicles Symposium (IV), IEEE (2013) 1259–1264
4. Rabe, C., Franke, U., Gehrig, S.: Fast detection of moving objects in complex scenarios. In: Intelligent Vehicles Symposium, 2007, IEEE (2007) 398–403
5. Perrollaz, M., Yoder, J.D., Negre, A., Spalanzani, A., Laugier, C.: A Visibility-Based Approach for Occupancy Grid Computation in Disparity Space. IEEE Transactions on Intelligent Transportation Systems **13** (2012) 1383–1393
6. Labayrade, R., Aubert, D., Tarel, J.P.: Real time obstacle detection in stereovision on non flat road geometry through "v-disparity" representation. In: Intelligent Vehicle Symposium. Volume 2., IEEE (2002) 646–651
7. Gao, Y., Ai, X., Rarity, J., Dahnoun, N.: Obstacle detection with 3d camera using UV-Disparity. In: 7th International Workshop on Systems, Signal Processing and their Applications (WOSSPA), IEEE (2011) 239–242
8. Duda, R.O., Hart, P.E.: Use of the Hough transformation to detect lines and curves in pictures. Communications of the ACM **15** (1972) 11–15
9. Perrollaz, M., Spalanzani, A., Aubert, D.: Probabilistic representation of the uncertainty of stereo-vision and application to obstacle detection. In: Intelligent Vehicles Symposium (IV). (2010) 313–318
10. Geiger, A., Roser, M., Urtasun, R.: Efficient large-scale stereo matching. In: Computer Vision–ACCV. Springer (2011) 25–38
11. Wang, B., Rodriguez Florez, S.A., Frémont, V.: Multiple Obstacle Detection and Tracking using Stereo Vision: Application and Analysis. In: 13th International Conference on Control, Automation, Robotics & Vision (ICARCV), Singapur, Singapore (2014)
12. Zehang Sun, Bebis, G., Miller, R.: Monocular precrash vehicle detection: features and classifiers. IEEE Transactions on Image Processing **15** (2006) 2019–2034
13. Alonso, D., Salgado, L., Nieto, M.: Robust vehicle detection through multidimensional classification for on board video based systems. In: IEEE International Conference on Image Processing. Volume 4., IEEE (2007) IV–321
14. Bergmiller, P., Botsch, M., Speth, J., Hofmann, U.: Vehicle rear detection in images with generalized radial-basis-function classifiers. In: Intelligent Vehicles Symposium., IEEE (2008) 226–233
15. Southall, B., Bansal, M., Eledath, J.: Real-time vehicle detection for highway driving. In: IEEE Conference on Computer Vision and Pattern Recognition., IEEE (2009) 541–548
16. Wen-Chung Chang, Chih-Wei Cho: Online Boosting for Vehicle Detection. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) **40** (2010) 892–902
17. Kowsari, T., Beauchemin, S.S., Cho, J.: Real-time vehicle detection and tracking using stereo vision and multi-view AdaBoost. In: 14th International Conference on Intelligent Transportation Systems (ITSC)., IEEE (2011) 1255–1260