2014 IEEE Intelligent Vehicles Symposium (IV)
June 8-11, 2014. Dearborn, Michigan, USA

Road Terrain Detection: Avoiding Common Obstacle Detection Assumptions Using Sensor Fusion

Patrick Y. Shinzato[1] and Denis F. Wolf[1] and Christoph Stiller[2]

Abstract— **Obstacle detection is a fundamental task for Advanced Driver Assistance Systems (ADAS) and Self-driving cars. Several commercial systems like Adaptive Cruise Controls and Collision Warning Systems depend on them to notify the driver about a risky situation. Several approaches have been presented in the literature in the last years. However, most of them are limited to specific scenarios and restricted conditions. In this paper we propose a robust sensor fusion-based method capable of detecting obstacles in a wide variety of scenarios using a minimum number of parameters. Our approach is based on the spatial-relationship on perspective images provided by a single camera and a 3D LIDAR. Experimental tests have been carried out in different conditions using the standard ROAD-KITTI benchmark, obtaining positive results.**

I. INTRODUCTION

Over the last years, several intelligent vehicle applications that depend on obstacle detection have been presented in the literature. Nevertheless, those systems are still often limited to specific conditions like flat surface or geometric appearance. In some cases, they must be calibrated to detect obstacles with a minimum height or size. The most common application that use obstacle detection is tracking of objects, which is presented often as a specialized system to track or detect specific type of objects like cars, pedestrians, posts and trees or even curbstones. In most cases, statistical or machine learning techniques associated to temporal information are used in order to treat possible errors, sensor noise and limitations from detection methods.

In this work we present a road terrain estimation method that uses an obstacle detection technique that avoids the majority of common assumptions presented in the literature. Our obstacle detection method is an evolution of a work already presented on literature by [1], [2]. Our algorithm takes advantage from sensor fusion like accuracy of each point from 3D-lidar and creates a local spatial-relationship between all points using the 3D projection on perspective images.

The results obtained show robust classifications in several different scenarios and conditions. Our method is accurate enough to detect many types of unstructured objects like vegetations, trees and medium grass. Also, our system does not need to make assumptions about size, height or shape of obstacles allowing detecting cars, pedestrians, posts and most importantly, curbstones. Therefore, our road terrain estimation system is capable of detecting the road area without geometric parameters definitions.

[1]Patrick Y. Shinzato and Denis F. Wolf are with Mobile Robotic Laboratory, Institute of Mathematics and Computer Science, University of Sao Paulo - ICMC-USP, Sao Carlos, SP Brazil [shinzato, denis]@icmc.usp.br

[2]Christoph Stiller is with the Department of Measurement and Control Systems, Karlsruhe Institute of Technology, Karlsruhe, Germany stiller@kit.edu

II. RELATED WORK

Road terrain estimation is a well know research topic for vision approaches, see [3] for a review. Most often, algorithms try to find edges as road delimiters (curbstones, lane-markings, edges between asphalt and grass) in order to detect the drivable space ahead of vehicle [4], [5]. These approaches are often limited to well-maintained roads as they make use of some image processing techniques that need threshold parameters. Some works based on machine learning learn texture and color in a previous training step, where a human manually annotates the road area [6] [7]. These vision approaches can easily fail if they are evaluated in scenarios that differ from those on the training data or if assumptions like an always navigable bottom part of the image are not true.

For approaches that use 3D-lidars, a planar road assumption has been imposed in [8] and a local convexity feature has been introduced in [9] to distinguish road surface from objects in lidar data. Also, the road estimation can be based on intensity measurement processing in order to detect lane-markings [10], detection of curbs, and berms [11] as road delimiting. Or even estimation of roughness for segmentation and then detection of road delimiters [12]. These approaches are based on data structures such as grid or voxels, they are limited due the need of parameters definitions like intensity threshold value, cell size, and minimum height difference. Also, grid approaches do not work properly for distant objects, since few points hit the same cell and loose accuracy depending on the cell size. In order to avoid these problems, many approaches accumulate few scans (temporal information) before performing the classification.

Most presented approaches have many assumptions or conditions to work well. The most important common characteristic in these works is the previous set up of parameters that helps achieve good results but decrease the robustness from their system. In this work, we developed a road estimation technique that relies on only two constant parameters manually selected, without temporal information and without any previous information like maps.

III. ROAD TERRAIN ESTIMATION

Our method is based on fusion of a sparse and unstructured 3D point clouds and images. It requires a previous calibration step to estimate intrinsic and extrinsic parameters that make

Fig. 1. Results from our obstacle detection method in 2D image world. Red represents edges between two points classified as *obstacle*, green represents edges between *non-obstacles* points, and blue represents edges between *non-obstacle* points and *obstacle* points.

possible to transform a 3D point from real world in the 2D image coordinate. The main idea of our method is to use spatial-relationship in image perspective view combined to real 3D metric values to determine if a point corresponds to an *obstacle* or not. After that, polar histograms are used to generate a confidence map that represents the road area in the image.

The entire system follows a sequence of five processing steps. The first one is the fusion step, where each 3D point is projected in the image. The second step generates a graph that creates a local spatial-relationship between all points. The third step is a classification of the points as *obstacle* or *non-obstacle*. The fourth step generates several polar histograms that estimate the free space area. Finally, the fifth step creates a confidence map combining all free space areas determined by the polar histograms created in previous step.

A. Sensor Fusion

As presented in [13], we used intrinsic and extrinsic parameters to transform a 3D point $p_l = (x_l, y_l, z_l, 1)^T$ relative to 3D LIDAR coordinate frame to a 3D point $p_c = (x_c, y_c, z_c, 1)^T$ in camera coordinate frame using the Equation 1:

$$p_c = \mathbf{R}^0_{rect}\mathbf{T}^{cam}_{velo}p_l, \tag{1}$$

where $\mathbf{R}^0_{rect}$ is the rotation matrix from raw-image-camera to rectified-image-camera and $\mathbf{T}^{cam}_{velo}$ is the transformation matrix defined as:

$$\mathbf{T}^{cam}_{velo} = \begin{pmatrix} \mathbf{R}^{cam}_{velo} & \mathbf{t}^{cam}_{velo} \\ 0 & 1 \end{pmatrix}, \tag{2}$$

where $\mathbf{R}^{cam}_{velo}$ and $\mathbf{t}^{cam}_{velo}$ are, respectively, rotation matrix and translation vector that transforms a point from 3D LIDAR coordinate frame to the camera coordinate frame. After applying Equation 1 to all points from 3D LIDAR sensor, we can remove any point p_c such that not satisfy the Equation 3:

$$p_c[z_c] > 0 \tag{3}$$

In order to obtain the screen coordinates (u,v), we apply the projection matrix $\mathbf{P}^{(i)}_{rect}$ from the i^{th} image-plane in all 3D points p_c, as shows the Equation 4:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} x/w \\ y/w \\ w/w \end{pmatrix} = \mathbf{P}^{(i)}_{rect}\begin{pmatrix} x_c \\ y_c \\ z_c \\ 1 \end{pmatrix}, \tag{4}$$

After applying Equation 4 to all remaining points, we can remove points such that coordinates (u,v) are not inside the image. The image can be defined as $\mathbf{I}(m,n)$, where m is number of lines and n is number of columns of image. The remaining points compose a set of a new type of point $p = (x_c, y_c, z_c, u, v)^T$.

B. Graph Generation

In this step, it is created an undirected graph $G = \{P,E\}$, where P is the set of all nodes that represents remaining points $p = (x_c, y_c, z_c, u, v)^T$ (from previous step) and E is the set of all valid edges. This step generates edges that connects all nearest neighbours for each node from G using the coordinates (u,v) in the image plane and the planar *Delaunay Triangulation* [14].

The *Delaunay Triangulation* properties guarantee that there is not any other point between points from *Delaunay's* edges. A post-processing step eliminates all edges (p,q) that not satisfies Eq. 5:

$$||p-q|| < \varepsilon, \tag{5}$$

where $||p-q||$ is the vector length in 3D values and ε is the maximum length for any edge. All edges created in this step represent the local spatial-relationship between two points and are be used in the obstacle classification step.

C. Obstacle Classification

The obstacle classification is a evolution of [1] subsequently used by [15] and others, most used in stereo vision processing. Using the graph G from previous step, our approach differentiates *obstacle* from *non-obstacle* nodes using only a spatial condition represented by the Equation 6 for each pair of nodes connected by edge. This method classifies points as obstacles without any height threshold

value. The result image (Fig. 1) from this step is used to generates the polar histograms in next step.

Given the graph $G = \{P, E\}$, for each node $p \in P$ exists a subset $Q(p) \subset P$ of nodes that represents its neighborhood. The node p is classified as *obstacle* if the function $obst()$ returns a positive value and *non-obstacle* if it returns zero. The function $obst()$ is defined in Equation 6:

$$obst(p) = \sum_{n \in Q(p)} \begin{cases} 1 & \text{if } \frac{-(p[y_c]-n[y_c])}{||p-n||} > cos(\theta) \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where $(p[y_c] - n[y_c])$ corresponds to the difference from height values between these two points, θ is a constant, $||p-n||$ is vector length from vector $(p-n)$ and $Q(p)$ can be expressed as $Q1(p) \cup Q2(p)$ where $Q1(p) = \{p_1 \mid (p, p_1) \in E\}$ and $Q2(p) = \{p_2 \mid (p_1, p_2) \in E \wedge p_1 \in Q1(p)\}$.

D. Multiples Free Space Detection

Since our obstacle classification is accurate enough to detect curbs, it can be used to estimate the road terrain area delimited by them. Our system creates several polar range histograms in the image world with origin in bottom center of image. The difference between all polar histograms is the coordinate v from origin point. This allows our system to estimate the road area without geometric parameters.

The creation of one polar histogram is obtained using the coordinates (u, v) from all *obstacle* points to calculate the 2D distance in pixels (image coordinates) from the origin point. Each bin stores the closest obstacle point on a respective angle. This histogram represents the free area in the image (Fig. 2).

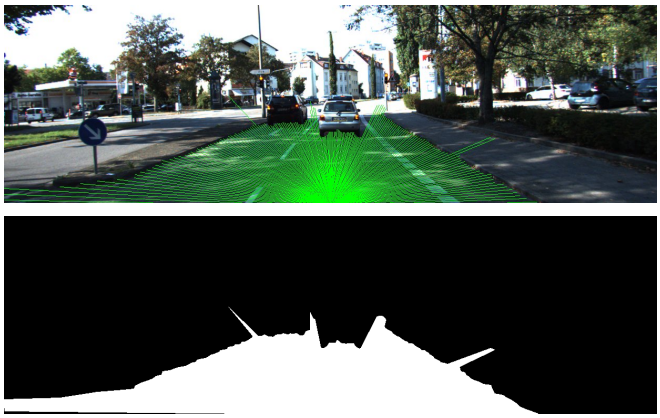

Fig. 2. **Upper Image**: Polar range histogram with origin in the center of the image. Each green line represents one bin from histogram, that shows the first *obstacle* point in the respectively direction. **Bottom Image**: shows the area covered by one histogram.

E. Road Area Estimation

Our road estimation approach calculates the confidence map $\mathbf{R}$ combining the area covered by all polar histogram with Equation 7:

$$\mathbf{R}(u, v) = \frac{1}{H} \sum_{i=1}^{H} img^{(i)}(u, v), \quad (7)$$

where (u, v) are coordinates in the image space, H is the number of histograms and $img^{(i)}$ is an image generated from histogram (i). In $img^{(i)}$, the area covered by the histogram has the value 1.0, while the remaining pixels has the value zero.

Finally, the Otsu's threshold method [16] is used to remove areas with low costs from our confidence map $\mathbf{R}$. The result of our road estimation method can be viewed in Fig. 3.

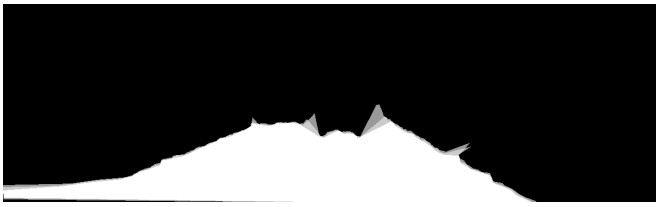

Fig. 3. Result of our road estimation method as confidence map to Fig. 2. The combination of several histograms improves classification results in road border area.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In order to validate the approach described in the previous sections, we used experimental data from the *AnnieWAY* platform that is available online on KITTI Benchmark [13]. More specifically we used the ROAD-KITTI Dataset that includes calibration parameters, ground truth images, and scripts for evaluation. This benchmark is useful because provides a comparison of systems with different algorithms and sensors.

As in [17], we adopted the evaluation of road area estimation results in metric 2D space. Therefore, we transform the ground truth images and our classification data in BEV. This avoids the perspective effect that has bias to results from near range. The images in BEV representation covers -10m to 10m in lateral direction and 6m to 46m in longitudinal direction. Using a resolution of 0.05m/px, these images have 800×400 pixels.

TABLE I

RESULTS [%] OF ROAD AREA EVALUATION FOR θ = 77.

	F-me	AVP	PRE	REC	FPR	FNR
UM	83.64	71.95	76.19	92.70	12.16	7.30
UMM	89.55	85.27	86.87	92.41	14.48	7.59
UU	81.99	70.41	74.92	90.54	10.29	9.46

Our general evaluation aims to check the influence of the θ threshold parameter from Equation 6. For this, our system was evaluated for θ values varying from 30 to 89 degrees. The charts in Fig. 4 show the F-measure (F-me), average precision (AVP), precision (PRE) and recall (REC) of our method for all training dataset. Each curve represents a set (*Urban Marked*, *Urban Multiple Marked*, *Urban Unmarked*). The X-axis corresponds to the θ variation and the Y-axis is the evaluation varying from 0 to 100. Also, Fig. 5 shows the false positive (FPR) and false negative rates (FNR). All these charts show that the best θ for this dataset is between 75.0 degrees and 80.0 degrees. Based on this analysis, we selected

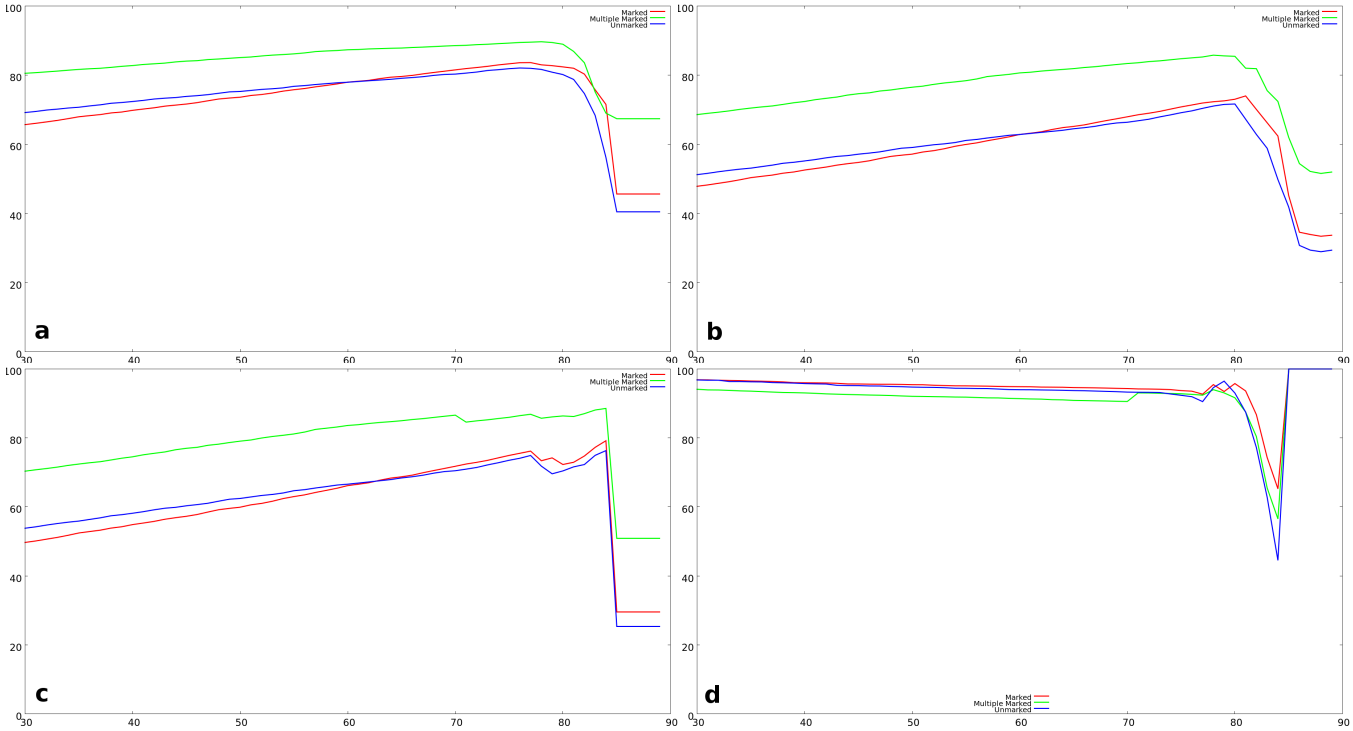

Fig. 4. Charts (a),(b),(c),(d) show respectively **F-measure, Average Precision, Precision** and **Recall** of our method for all training dataset. Each curve represents one set, *Urban Marked* (Red), *Urban Multiple Marked* (Green), *Urban Unmarked*) (Blue). The X-axis is the θ variation from 30 degrees to 90 degrees and the Y-axis is the evaluation values varying from 0 to 100.

the parameter $\theta = 77.0$, which maximizes the classification results. All metric evaluation proposed in [17] has been adopted and is presented in Table I. Figures 6, 7 and 8 show, respectively, raw image, ground truth and our estimation results.

Since our system is free of geometric model assumptions, the results can be considered very satisfactory when compared to others approaches in the literature. Despite that all tree dataset are different, our method achieved an homogeneous result of F-measure of more than 80% for all datasets. This happens because our system achieved almost 100% classification success in regular conditions, but still got a high FPR in some situations that we discuss below.

As we use detected obstacles to delimit and consequently estimates the road, this approach achieves good results in scenarios where other vehicles appears and where curbstones are totally visible, as showed in Fig. 6. Even in scenarios where other cars appears near to our vehicle (as in Fig. 7), our system was able to detect the road with good accuracy. The fact that we are using a 3D-lidar sensor also guarantees that our system is robust to light condition variations and shadow effects, as we can see on Fig. 8.

Despite the good results obtained, the general FPR shows that our method is still too permissive when the road is not delimited by curbstones, as showed in Fig. 9. However, it is important to notice that even with short grass delimiting the road, as in Fig. 9(a), our obstacle classification was able to detect robustly the near grass as obstacle. We believe that temporal or color information would be able to improve our

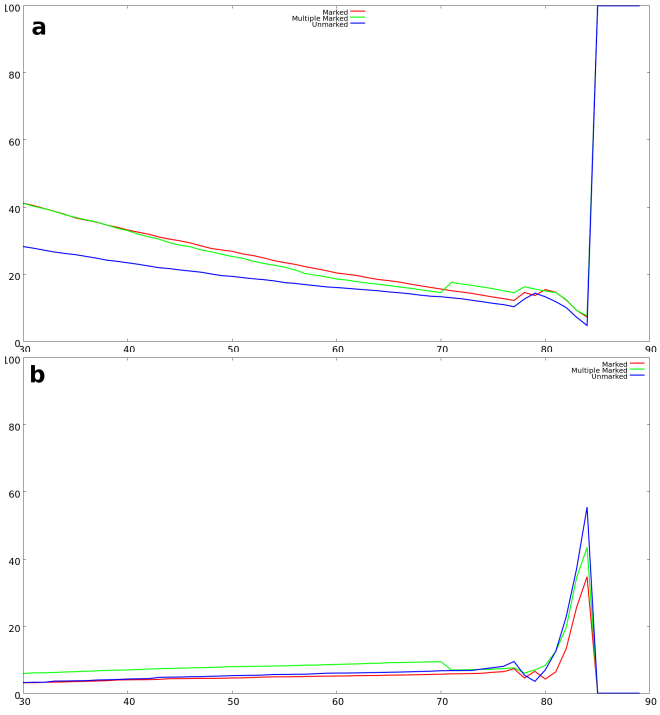

Fig. 5. Charts (a),(b) show respectively **false positive rate** and **false negative rate** of our method for all training dataset. Each curve represents one set, *Urban Marked* (Red), *Urban Multiple Marked* (Green), *Urban Unmarked*) (Blue).

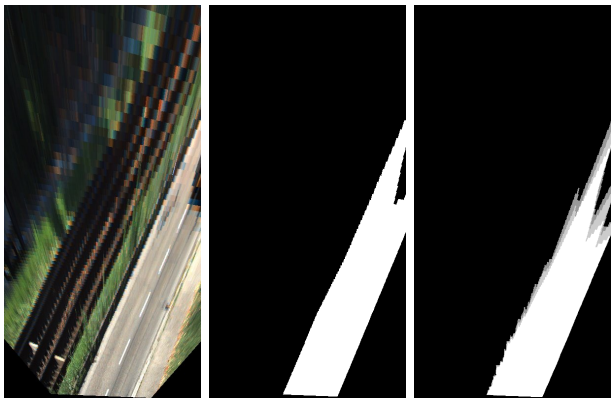

Fig. 6. From left to right, raw image, ground truth, our result. Results of our method in simple scenarios where curbstones are totally visible, we achieve almost 100% of right classification.

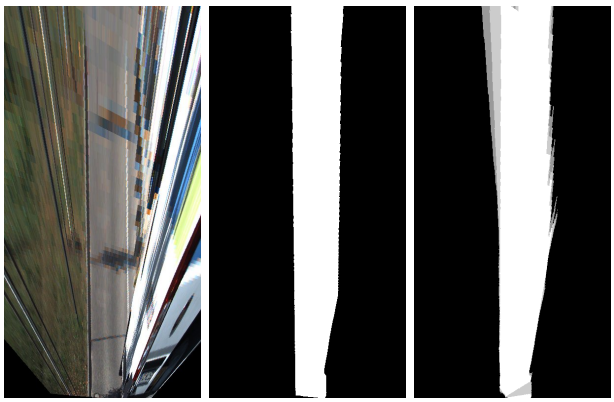

Fig. 7. From left to right, raw image, ground truth, our result. Scenarios with other vehicles near to our vehicle also was correctly classified.

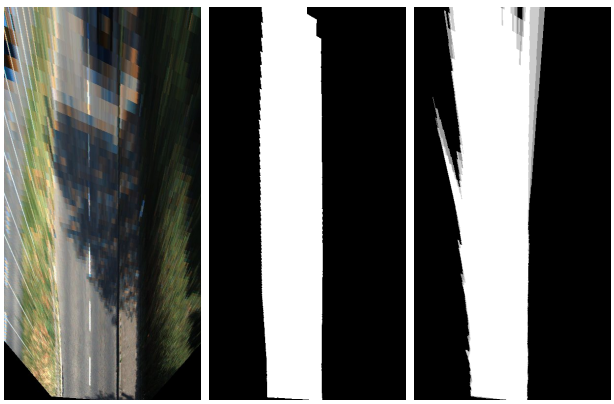

Fig. 8. From left to right, raw image, ground truth, our result. 3D-lidar sensor allow our system be robust to shadows.

system in cases like this. Another important fact about this case is the grey areas in our confidence map (Fig. 9(d)). These areas are result of the sum of costs from different road areas generated by our polar histograms. These areas can be very different when the obstacle detection fails due noisy from the sensor, the existence of a small object or a small holes on the road. The different origins from our polar histograms help to deal with situations where an obstacle detection fails and could increase the FNR.

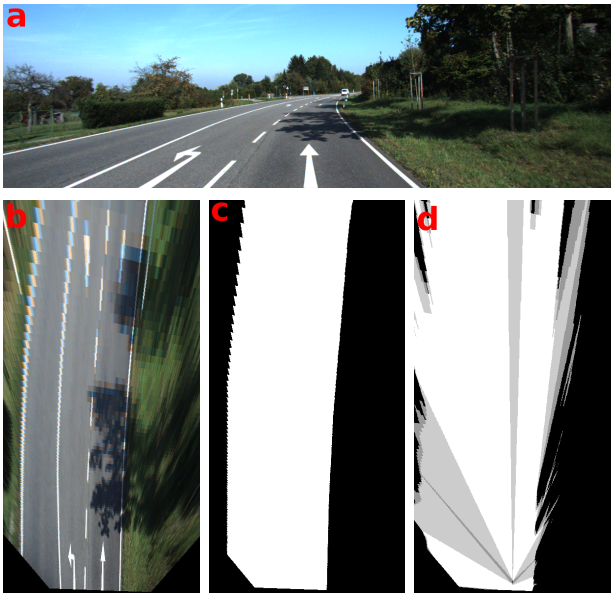

Fig. 9. (a) Raw image (perspective view) from scenario without curbstones on delimitations of road. (b) Bird eye view from raw image. (c) Ground truth in bird eye view. (d) Our result in bird eye view.

Despite the good appearance of our result (Fig. 10(a)), our system was not able to always detect the delimitation between the road and the sidewalk. Our system selected all area from origin of histograms until reaching the first evident obstacle, like cars or walls. Also, we believe that even temporal information will not help to detect these delimiters. In order to avoid to use previously obtained information like maps, we believe that only visual information like edges or texture could improve the classification in these cases.

V. CONCLUSIONS AND FUTURE WORK

Several intelligent vehicle applications that depend on obstacle detection have been already presented by research and industrial community. Nevertheless, those systems are still often limited to certain conditions and require several assumptions. This paper proposes an obstacle detection method that avoids those common assumptions and, based on our obstacle detection we also propose a road estimation method. We evaluated our approach using a standard benchmark dataset for road estimation. Our system was able to successfully detect the road (paved free space) in multiple scenarios without previous training, making assumptions about road geometry, or threshold values for height and size of obstacles. Also, our system does not have problems as

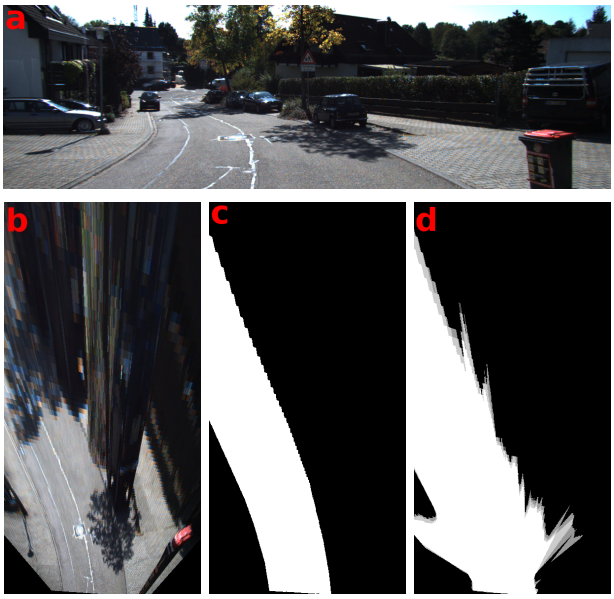

Fig. 10. (a) Raw image (perspective view) from scenario with very small curbstones between the road and the sidewalk, as parking entrance. (b) Bird eye view from raw image. (c) Ground truth in bird eye view. (d) Our result in bird eye view.

grid and voxel-based approaches like cell size parameter definition. As future work, we intend to replace the fusion by stereo processing and integrate visual information like edges, colors or texture. We also plan to extend our method to detect road-markers in order to approach the lane detection problem.

ACKNOWLEDGMENT

The authors acknowledge the support granted by FAPESP, processes 2010/01305-1 and 2011/21956-0.

REFERENCES

[1] L. Matthies, A. Kelly, T. Litwin, and G. Tharp, "Obstacle detection for unmanned ground vehicles: a progress report," in *Intelligent Vehicles '95 Symposium., Proceedings of the*, 1995, pp. 66–71.

[2] L. Matthies, T. Litwin, K. Owens, A. Rankin, K. Murphy, D. Coombs, J. Gilsinn, T. Hong, S. Legowik, M. Nashman, and Billibon, "Performance evaluation of ugv obstacle detection with ccd/flir stereo vision and ladar," in *IEEE ISIC/CIRA/ISAS Joint Conference*, 1998.

[3] A. Bar Hillel, R. Lerner, D. Levi, and G. Raz, "Recent progress in road and lane detection: a survey," *Machine Vision and Applications*, pp. 1–19, 2012. [Online]. Available: http://dx.doi.org/10.1007/s00138-011-0404-2

[4] C. Jung and C. Kelber, "A lane departure warning system using lateral offset with uncalibrated camera," in *Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE*, Sept. 2005, pp. 102–107.

[5] Y. He, H. Wang, and B. Zhang, "Color-based road detection in urban traffic scenes," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 5, no. 4, pp. 309–318, Dec. 2004.

[6] P. Y. Shinzato, V. Grassi Jr., F. S. Osório, and D. F. Wolf, "Fast visual road recognition and horizon detection using multiple artificial neural networks," in *Intelligent Vehicles Symposium*, 2012, pp. 1090–1095.

[7] C. Tan, T. Hong, T. Chang, and M. Shneier, "Color model-based real-time learning for road following," in *Intelligent Transportation Systems Conference, 2006. ITSC '06. IEEE*, Sept. 2006, pp. 939–944.

[8] F. Moosmann and C. Stiller, "Joint self-localization and tracking of generic objects in 3d range data," in *Proceedings of the IEEE International Conference on Robotics and Automation*, Karlsruhe, Germany, May 2013, pp. 1138–1144.

[9] F. Moosmann, O. Pink, and C. Stiller, "Segmentation of 3d lidar data in non-flat urban environments using a local convexity criterion," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, Xi'an, China, Jun. 2009, pp. 215–220.

[10] S. Kammel and B. Pitzer, "Lidar-based lane marker detection and mapping," in *Intelligent Vehicles Symposium, 2008 IEEE*, 2008, pp. 1137–1142.

[11] G. Zhao and J. Yuan, "Curb detection and tracking using 3d-lidar scanner," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*, 2012, pp. 437–440.

[12] M. Hselich, M. Arends, D. Lang, and D. Paulus, "Terrain classification with markov random fields on fused camera and 3d laser range data," in *5th European Conference on Mobile Robotics (ECMR), Proceedings of the*. rebro University, 2011, pp. 153–158.

[13] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *International Journal of Robotics Research (IJRR)*, vol. 32, pp. 1229–1235, September 2013.

[14] D. T. Lee and B. J. Schachter, "Two algorithms for constructing a Delaunay triangulation," *International Journal of Parallel Programming*, vol. 9, no. 3, pp. 219–242, Jun. 1980.

[15] A. Talukder, R. Manduchi, A. Rankin, and L. Matthies, "Fast and reliable obstacle detection and segmentation for cross-country navigation," in *Intelligent Vehicle Symposium, 2002. IEEE*, vol. 2, 2002, pp. 610–618 vol.2.

[16] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[17] J. Fritsch, T. Kuehnl, and A. Geiger, "A new performance measure and evaluation benchmark for road detection algorithms," in *International Conference on Intelligent Transportation Systems (ITSC)*, 2013.