2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)
Windsor Oceanico Hotel, Rio de Janeiro, Brazil, November 1-4, 2016

CaRINA Dataset: an Emerging-Country Urban Scenario Benchmark for Road Detection Systems

Patrick Y. Shinzato[1], Tiago C. dos Santos[1], Luis Alberto Rosero[1], Daniela A. Ridel[1], Carlos M. Massera[1], Francisco Alencar[1], Marcos Paulo Batista[1], Alberto Y. Hata[1], Fernando S. Osório[1] and Denis F. Wolf[1]

Abstract— **Road traffic crashes are the leading cause of death among young people between 10 and 24 years old. In recent years, both academia and industry have been devoted towards the development of Driver Assistance Systems (DAS) and Autonomous Vehicles (AV) to decrease the number of road accidents. Detection of the road surface is a key capability for both path planning and object detection on Autonomous Vehicles. Current road datasets and benchmarks only depict European and North American scenarios, while emerging countries have higher projected consumer acceptance of AV and DAS technologies. This paper presents a selected Brazilian urban scenario dataset and road detection benchmark consisting of annotated RADAR, LIDAR and camera data. It also proposes a novel evaluation metric based on the intersection of polygons. The main goal of this manuscript is to provide challenging scenarios for road detection algorithm evaluation and the resulting dataset is publicly available at `www.lrm.icmc.usp.br/dataset`.**

I. Introduction

Road traffic crashes are the leading cause of death among young people between 10 and 24 years old [17]. Most of these accidents occur when the driver is unable to maintain the vehicle control due to fatigue or external factors [1]. In recent years, both academia and industry have been devoted towards the development of Driver Assistance Systems (DAS) and Autonomous Vehicles (AV) to decrease the number of road accidents.

Detection of the road surface is a key for both path planning and object detection on Autonomous Vehicles (AV) [13]. It provides navigation boundaries for trajectory planning algorithms [3] and assists in the definition of regions-of-interest for object detection algorithms [18].

Current road datasets and benchmarks (such as the KITTI road benchmark [8], the DIPLODOC dataset [23] and the Cityscapes benchmark [4]) only depict European and North American scenarios. However, emerging countries, such as Brazil and China, have higher projected consumer acceptance of AV and DAS technologies [22] and lower quality roads. Therefore, it is of great importance to develop an emerging-country dataset and benchmark for road detection algorithms based on the commonly available sensors used in autonomous vehicle prototyping platforms, such as RADARs, 3D LIDARs and cameras [21].

The main objective of this manuscript is the proposal of a Brazilian urban scenario dataset and a benchmark for road detection. This dataset consists of annotated LIDAR, RADAR and camera information for several data sequences obtained at São Carlos city in the state of São Paulo. It also proposes a novel and simple evaluation metric based on the intersection of the detected road polygon and the reference polygon. This dataset is available online at `www.lrm.icmc.usp.br/dataset` and will provide opportunities to other research groups evaluate their proposed solutions in several challenging scenarios.

II. Related Work

KITTI benchmark suite [10] was developed by the Karlsruhe Institut für Technologie (KIT) from data collected with the AnnieWay research platform. It consists of several benchmarks for different applications, including stereo matching, odometry, object detection, object tracking and road detection. Currently, it is the main benchmark of computer vision algorithms for autonomous vehicles. Its road detection benchmark was developed by Fritsch and Kuchnl [8] and currently is divided in three sections: Unmarked roads, marked single lane roads and marked multi-lane roads. However, it lacks RADAR information, which will be intensively used for commercial AV and DAS applications, and its scenarios do not vary significantly, since it mostly consists of daytime urban and highway data collections around the city of Karlsruhe.

The DIPLODOC road stereo sequence dataset [23] was developed by the Fondazione Bruno Kassler (FBK) using a stereo camera mounted in a vehicle. A sequence of 865 image pairs ware captured on 2004 and the dataset was made available in 2013. However, this video sequence has low resolution for current standards (320x240 pixels) and is highly biased, since all images were obtained from a single, continuous, data collection.

The Cityscapes dataset [4] was jointly developed by Daimler, TU Dresden, TU Darmstadt and MPI Informatics. It consists of images captured during daytime, good and medium weather conditions, different times of the year in 50 different German cities and was made available in 2016. It contains 25000 images with semantics, instances or dense pixel annotations. This dataset extends on the current

This work was supported by FAPESP grant 2013/24542-7.

[1]Institute of Mathmatics and Computer Science, University of São Paulo, Avenida Trabalhador São-carlense, 400, São Carlos, Brazil {shinzato, tiagocs, massera, alencar, hata, fosorio, denis}@icmc.usp.br, {lrosero, danielaridel, marcos.batista}@usp.br

limitations of the KITTI benchmark, providing higher complexity and more diverse scenarios with richer annotations for algorithm comparison. However, it is still limited to German highways and cities.

Several other computer vision datasets designed or used for autonomous vehicles applications exists, some examples are as the INRIA pedestrian detection [5] and the MOT challenge benchmark for object tracking [14]. However, none of them address the road detection problem.

These datasets depict either European or North American scenarios in their majority. Therefore, most of the algorithm development for the field are oblivious to the challenging scenarios that exists in emerging countries, such as the example shown in Figure 1. Emerging country citizens are among the most susceptible early-adopters of AV technologies [22], with Brazil leading the consumer acceptance pool with 96% of the surveyed people stating they would trust an AV.

Fig. 1: Example of complex scenario (where the street has a pothole) from Brazil's roads.

Evaluation metrics from available road benchmarks have been designed to evaluate camera-based techniques and rely on pixel count accuracy. However, these metrics are not suitable for the evaluation of results obtained using other sensors, such as LIDAR, and do not directly represent the evaluated method performance on its applications. Therefore, a sensor agnostic metric capable of evaluating the performance for the main road detection application of path planning boundaries would provide a better comparison of effectiveness.

III. CaRINA Project

This study is part of the project entitled CaRINA (Intelligent Robotic Car for Autonomous Navigation), which began in 2010 with its first test platform, CaRINA 1 [6]. This vehicle was a small electric golf cart with automated steering and a small sensor suite, which consisted of monocular cameras and 2D-LIDARs. In 2011, the CaRINA 2 platform was acquired and modified for autonomous operations to enable the prototyping of the groups perception and control algorithms in higher speeds and more challenging urban scenarios. This vehicle, a Fiat Palio Adventure, was equipped with stereo cameras, a 32-beam 3D-LIDAR, RADARs, a GPS with RTK correction and a IMU.

In mid-2012, the first fully autonomous navigation tests were performed with this new platform and in late-2013 the vehicle performed its first autonomous navigation in real urban traffic with the permission and cooperation of the Municipal Secretariat of Transport and Traffic and the Municipal Secretariat of Sustainable Development, Science and Technology of São Carlos. Figure 2 shows a picture of the vehicle during this event. During the test, the vehicle had to track a previously mapped route on streets with good pavements and road sign conditions and to handle traffic without human intervention.

Fig. 2: CaRINA 2 public road tests performed in a well maintained city avenue.

The group has ongoing research on a variety of topics related to the CaRINA project, such as object detection [19], road detection [20], vehicle tracking [2], driver assistance and autonomous control [15] and simulations [11].

As a result of the ongoing prototyping on Brazilian urban scenarios, it was noticed there are challenging scenarios for autonomous vehicles perception systems, mostly due to infrastructure problems that may not be present on currently available datasets and benchmarks. Therefore, the development of a Brazilian dataset for autonomous vehicle perception algorithms has a potential to contribute to the evaluation and comparison of state-of-the-art methods such challenging situations.

IV. CaRINA-Road Dataset

This paper presents a Brazilian urban road scenario dataset, focusing in poorly maintained streets. These logs have been collected in good weather conditions, with adequate visibility and good lighting conditions with the exception of the presence of shadows in several sequences. It is online available at `www.lrm.icmc.usp.br/dataset`.

Fig. 3 presents the collected data coverage and the segments selected for the dataset. This acquisition has been performed on a neighborhood surrounding the university campus. Selected frames that compose the ground-truth are shown in red, while yellow represents frames where RTK GPS correction failed, the remaining frames are colored in blue. In its entirety, $7.1Kms$ were covered at an average speed of $\pm 5m/s$.

A. Sensors and Data Acquisition

For this study, the CaRINA 2 Vehicle platform was equipped with:

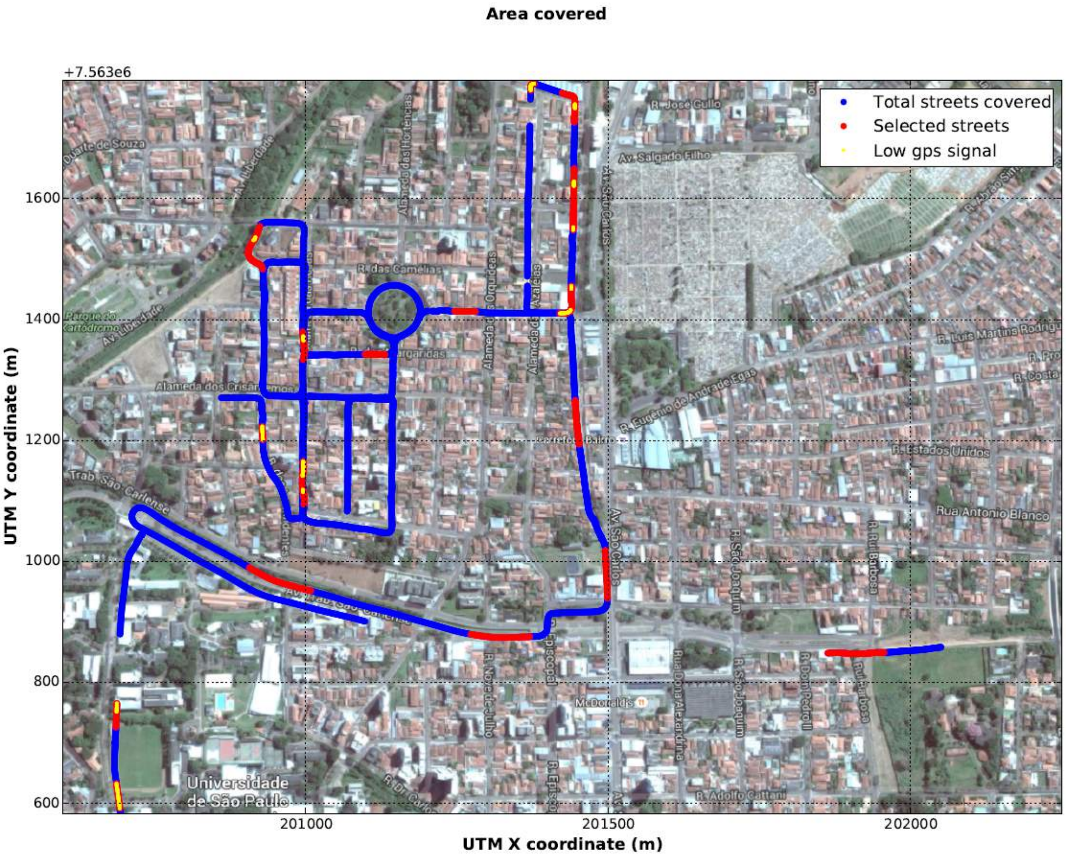

Fig. 3: Collected data coverage (blue) from a neighborhood surrounding the university campus of São Carlos; Select segments (red); and regions without RTK correction on GPS signal (yellow).

- 1 Bumblebee XB3 [1] Firewire stereo camera with 16 fps,1280×960 resolution, 3.8 mm focal length (66-deg horizontal field of view) and color image.
- 1 Velodyne HDL-32E 3D laser scanner [2] operating at 10Hz, $2cm$ accuracy, 32 Channels, $80m - 100m$ range, $700,000$ points per second, $360°$ horizontal field-of-view and $\pm 20°$ vertical field-of-view.
- 1 Commercial bi-mode MMW RADAR (76.5 GHz), Delphi ESR [3] operating at 20Hz with $\pm 10°$ field-of-view with range of $174m$ and $\pm 45°$ with range of $60m$.
- 1 Septentrio AsteRx2eH GPS [4] operating at 10Hz with RTK correction signals and two antenna.

RADAR has been mounted in the front bumper of the vehicle while other sensors, such as as camera, 3D-lidar and GPS, were mounted on top, as Fig. 4 shows. The stereo camera has three lenses, providing two baselines, $12cm$ and $24cm$, which are close to industrial standards [9], [7]. Hardware synchronization has not been used during data collection due to the belief that commercially available systems won't perform this function. Furthermore, sensor synchronization imposes significant limitations since every

[1]https://www.ptgrey.com
[2]http://velodynelidar.com/hdl-32e.html
[3]http://www.autonomoustuff.com/delphi-esr-9-21-21
[4]http://www.septentrio.com/products/gnss-receivers

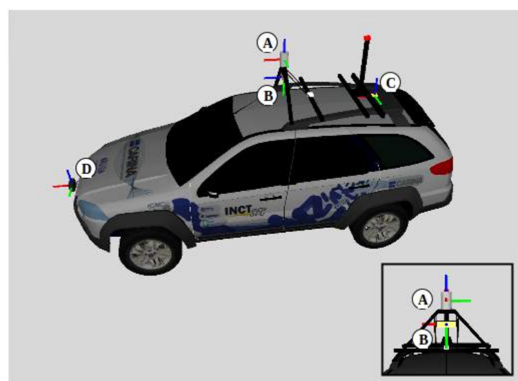

Fig. 4: **A:** Velodyne HDL-32E. **B:** Bumblebee XB3. **C:** Septentrio AsteRx2eH GP. **D:** Radar. Red arrow means x-axis, while green and blue arrows means y-axis and z-axis, respectively.

sensors will operate as fast as the slowest one (10Hz with our sensors). The unsynchronized sensors also enables the development of sensor fusion algorithms capable of accounting for these different frequencies and delays, similar to current localization solutions that integrate IMU, GPS and RTK correction.

B. Ground-Truth Generation and Evaluation Method

2D polygons have been chosen to represent the free-space on the road dataset ground truth. These polygons cover all road area from the ego-road and is specified in the 3D-LIDAR XY-plane coordinate frame. Vision-based algorithms require a transformation between camera coordinate frame and 3D-LIDAR coordinate frame. The transformation matrix is provided with each dataset sequence in order to stimulate and facilitate comparisons of approaches which use different sensor.

The ground-truth has been generated by projection of 3D points with $z < 0$ on a XY-plane for each frame, from which a set of points that represents the non-convex boundary of the road area was manually annotated. All polygons for a picture sequence are saved on one file, where each line contains the polygon for a frame. The camera ground-truth is automatically generated through a mask that creates the intersection of the polygon and the camera field-of-view region, the result is then saved in a separate text file in the same manner.

These points are used by an evaluation script to compare polygons generated by a submitted algorithms with the ground-truth. The output of a proposed solution must generate a frame in the same described format. To perform this comparison, the evaluation script calculates the intersection area divided by the union area of both polygons, i.e, the evaluation script compares the output polygon of a method with the ground-truth for each frame. This operation results in a value between 0 and 1, where 0 represents non-overlapping polygons while 1 represents identical polygons.

Ground-truth consists of data from 90 seconds of collection equally divided into three different categories: *Easy*, *Medium* and *Hard*. Frames with potholes, cobblestone or unpaved road are present in *Medium* and *Hard* categories only. Since the ground-truth is annotated on the 3D-LIDAR sensor data, it contains 10 readings per second. Therefore, there are 300 evaluation frames per category. The evaluation script selects the ground-truth of the closest (in time) 3D-LIDAR frame to a given image to generate its ground-truth, since our stereo camera operates at $16Hz$ and is not synchronized with the 3D-LIDAR. This approximation is plausible since the acquisition was performed at low speeds.

V. Emerging-Country Urban Environment and Challenges

Emerging countries are known by its poorly maintained roads, however there are several other aggravating factors on these environments. For example, there is no consistency for street layouts, materials or road markings, thus this study focuses on scenarios where handling these characteristics is necessary for good performance of the evaluated algorithm. These scenes are challenging for pattern recognition and machine learning methods since it differs significantly from current datasets where these methods are trained.

Fig. 5 presents a case where the road pavement ends and continues as a unpaved road containing both potholes and gravel. In this scenario, the 3D-LIDAR birds-eye-view shows a roughened surface on transition region of this street (the irregular rings in front the vehicle). The dataset consists of heterogeneous frames depicting marked (Fig. 7),

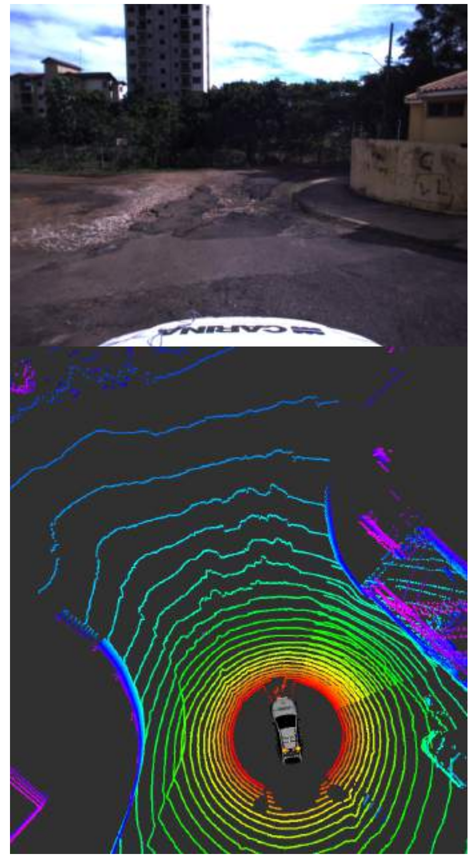

Fig. 5: End of a road pavement which continues as unpaved road, containing both potholes and gravel. A 3D-LIDAR view shows a roughened surface in this street (irregular rings data in front the vehicle).

unmarked (Fig. 8), paved, unpaved and cobblestone roads (Fig. 6), besides the appearance of several potholes (Fig. 9). Roundabouts and uncommon scenes, such as a narrow one-lane street (Fig. 10) and a road containing a short treetop which can collide with passenger vehicles (Fig. 11) are also included on the *Hard* sequences. These situations can disturb common navigation systems, such as 2D height-maps or occupancy-grids, since the vehicle needs to consider it as an obstacle due to the pose of the test platform sensors.

VI. Experiments

A previously proposed method for curb detection [12] was adapted for road detection and was evaluated with the proposed benchmark. Such method was based on the concept of compression of consecutive rings of the Velodyne sensor, proposed by [16]. Such compression between rings is proportional to the surface slope, for example, walls are associated to higher compressed rings compared to the ground surface. By obtaining the compression rates of curbs, this value is then used as threshold to classify sensor points as curbs. Essentially, a minimum and a maximum compression threshold provide the association to curbs, a robust regression is then used to filter the curb candidate points.

A convex hull is formed from the detected curb points. Sample points are extracted from the resulting curves of the

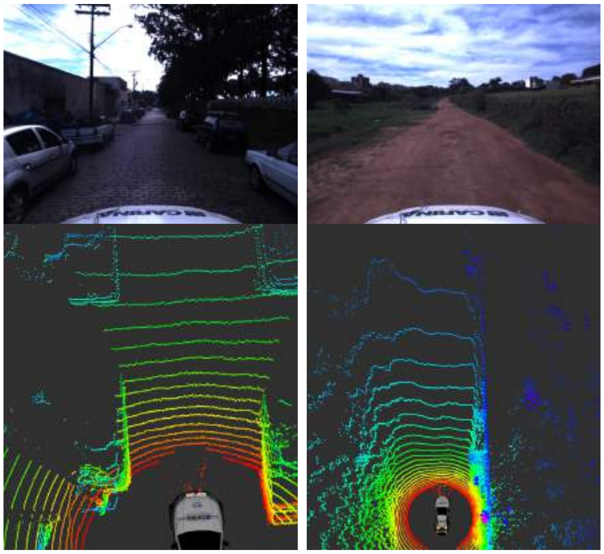

Fig. 6: Cobblestone and Unpaved Streets.

Fig. 7: Well-maintained paved and marked streets.

Fig. 8: Well-maintained paved and unmarked streets.

robust regression and used as the polygon boundaries. Figure 12 shows road detection for easy and hard datasets.

Since outer rings become sparser, only the lower 21 of 32 laser rings have been used. Therefore, the maximum detection distance is approximately 33 m. For detection performance evaluation, we employed the polygon intersection method described in Section IV.B. We performed the evaluation in the Easy, Medium and Hard datasets using three different compression thresholds. The obtained results are listed in TABLE I. The first two rows are thresholds obtained by gradient descent method and the third one was obtained empirically.

The evaluation shows the decreasing of metric results according to the difficulty of the dataset. From Easy to Hard datasets we observed a variation of at least 40% in the detection performance. Besides that, it is possible to check which threshold values deliver a better detection. In this case, we can notice that the threshold parameters of the first row

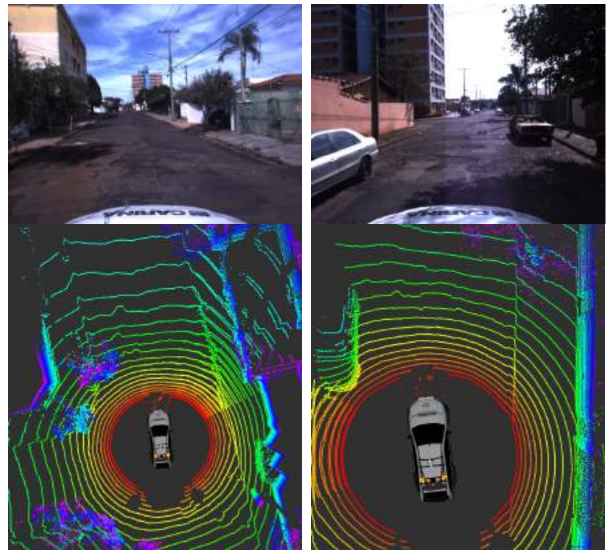

Fig. 9: Paved streets with potholes.

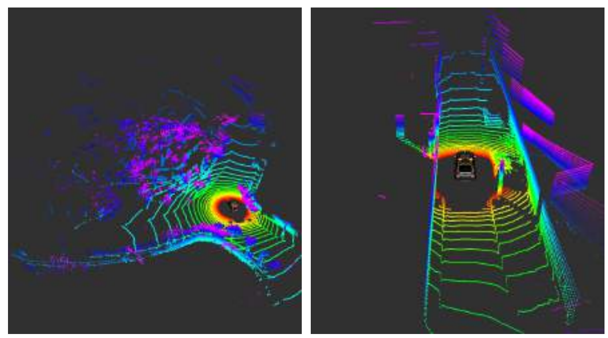

Fig. 10: Uncommon scene in road datasets, containing a roundabouts (left) and a narrow single lane road (right).

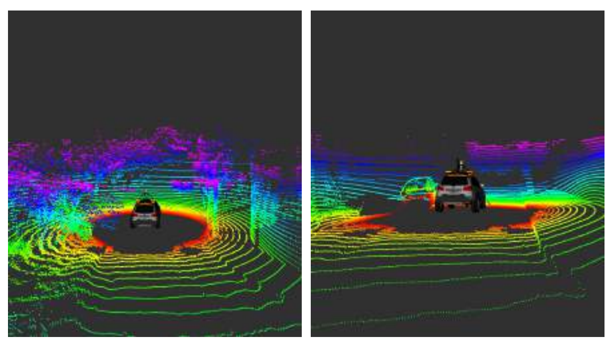

Fig. 11: Uncommon situation in road datasets, containing a short treetop which may collide with the vehicle sensors.

presented better results. However, as the metric values ranges from 0.0 to 1.0, this method detects less than the half of the actual road area.

VII. Conclusion

Current road datasets and benchmarks only depict European and North American scenarios, while emerging country citizens are the most susceptible early-adopters of AV technologies, with Brazil leading a consumer acceptance pool with 96% of the surveyed people stating they would trust and use an AV.

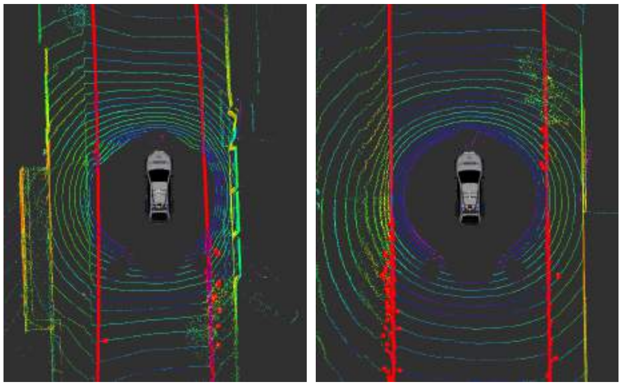

Fig. 12: Left and right images are respectively road detection results for easy and hard datasets. The red lines delimits the road boundary.

TABLE I: Quantitative results of curb detection by ring compression.

Compression Threshold	Easy	Medium	Hard
$[0.012462; 1.37543]$	0.416816	0.334410	0.252004
$[0.434093; 1.02254]$	0.380651	0.316736	0.252820
$[0.500000; 0.70000]$	0.370324	0.295512	0.220701

This paper proposes a selected Brazilian urban road scenario dataset focused in unmaintained, poor condition streets. This dataset provides challenging situations that exists in emerging countries scenarios to the development of autonomous vehicle perception algorithms. It consists of GPS, annotated LIDAR, RADAR and camera information for several data sequences obtained at a small neighborhood of São Carlos city in the state of São Paulo. A novel evaluation metric based on the intersection of the detected road polygon and the reference polygon is also proposed. We tested a previous work adapted to detect road area, and results shows that even in *Easy* situations, use only detection of curbs is not enough to correctly detect road in unmaintained roads.

The dataset is available online at `www.lrm.icmc.usp.br/dataset` and will provide the opportunity to other research groups to evaluate and compare their algorithms and methods in challenging scenarios, contributing towards the improvement of road detection algorithms state-of-the-art.

REFERENCES

[1] N. H. T. S. Administration *et al.*, "National motor vehicle crash causation survey: Report to congress," *National Highway Traffic Safety*

[2] F. A. Alencar, C. Massera Filho, D. A. Ridel, and D. F. Wolf, "Fast metric multi-object vehicle tracking for dynamical environment comprehension," in *2015 12th Latin American Robotics Symposium and 2015 3rd Brazilian Symposium on Robotics (LARS-SBR)*. IEEE, 2015, pp. 175–180.

[3] M. Brown, J. Funke, S. Erlien, and J. C. Gerdes, "Safe driving envelopes for path tracking in autonomous vehicles," *Control Engineering Practice*, 2016.

[4] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[5] N. Dalal and B. Triggs, "Inria person dataset," 2005.

[6] L. C. Fernandes, J. R. Souza, G. Pessin, P. Y. Shinzato, D. Sales, C. Mendes, M. Prado, R. Klaser, A. C. Magalhães, A. Hata, D. Pigatto, K. C. Branco, V. G. Jr., F. S. Osorio, and D. F. Wolf, "Carina intelligent

Administration Technical Report DOT HS 811 059, 2008.
robotic car: Architectural design and applications," *Journal of Systems Architecture*, vol. 60, no. 4, pp. 372 – 392, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1383762113002841

[7] U. Franke and S. Gehrig, "How cars learned to see," *Proceedings of 54th Photogrammetric Week*, pp. 3–10, 2013.

[8] 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.

[9] P. Furgale, U. Schwesinger, M. Rufli, W. Derendarz, H. Grimmett, P. Mühlfellner, S. Wonneberger, J. Timpner, S. Rottmann, B. Li, B. Schmidt, T. N. Nguyen, E. Cardarelli, S. Cattani, S. Brüning, S. Horstmann, M. Stellmacher, H. Mielenz, K. Köser, M. Beermann, C. Häne, L. Heng, G. H. Lee, F. Fraundorfer, R. Iser, R. Triebel, I. Posner, P. Newman, L. Wolf, M. Pollefeys, S. Brosig, J. Effertz, C. Pradalier, and R. Siegwart, "Toward automated driving in cities using close-to-market sensors: An overview of the v-charge project," in *Intelligent Vehicles Symposium (IV), 2013 IEEE*, June 2013, pp. 809–816.

[10] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *International Journal of Robotics Research (IJRR)*, 2013.

[11] A. E. Gómez, T. C. dos Santos, C. M. Filho, D. Gomes, J. C. Perafan, and D. F. Wolf, "Simulation platform for cooperative vehicle systems," in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, Oct 2014, pp. 1347–1352.

[12] A. Y. Hata and D. F. Wolf, "Feature detection for vehicle localization in urban environments using a multilayer lidar," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 2, pp. 420–429, Feb 2016.

[13] T. Kühnl, F. Kummert, and J. Fritsch, "Spatial ray features for real-time ego-lane extraction," in *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on*. IEEE, 2012, pp. 288–293.

[14] L. Leal-Taixé, A. Milan, I. Reid, S. Roth, and K. Schindler, "MOTChallenge 2015: Towards a benchmark for multi-target tracking," *arXiv:1504.01942 [cs]*, Apr. 2015, arXiv: 1504.01942. [Online]. Available: http://arxiv.org/abs/1504.01942

[15] C. Massera Filho and D. F. Wolf, "Driver assistance controller for tire saturation avoidance up to the limits of handling," in *2015 12th Latin American Robotics Symposium and 2015 3rd Brazilian Symposium on Robotics (LARS-SBR)*. IEEE, 2015, pp. 181–186.

[16] M. Montemerlo, J. Becker, S. Bhat, H. Dahlkamp, D. Dolgov, S. Ettinger, D. Haehnel, T. Hilden, G. Hoffmann, B. Huhnke, D. Johnston, S. Klumpp, D. Langer, A. Levandowski, J. Levinson, J. Marcil, D. Orenstein, J. Paefgen, I. Penny, A. Petrovskaya, M. Pflueger, G. Stanek, D. Stavens, A. Vogt, and S. Thrun, "Junior: The stanford entry in the urban challenge," *Journal of Field Robotics*, vol. 25, no. 9, pp. 569–597, 2008.

[17] W. H. Organization *et al.*, "Youth and road safety," 2007.

[18] D. Pfeiffer and U. Franke, "Towards a global optimal multi-layer stixel representation of dense 3d data." in *BMVC*, 2011, pp. 1–12.

[19] D. A. Ridel, P. Y. Shinzato, and D. F. Wolf, "Obstacle segmentation with low density disparity maps," in *7th Workshop on Planning, Perception and Navigation for Intelligent Vehicles in IROS 2015*, 2015.

[20] P. Y. Shinzato, D. F. Wolf, and C. Stiller, "Road terrain detection: Avoiding common obstacle detection assumptions using sensor fusion," in *2014 IEEE Intelligent Vehicles Symposium Proceedings*, June 2014, pp. 687–692.

[21] J. Wei, J. M. Snider, J. Kim, J. M. Dolan, R. Rajkumar, and B. Litkouhi, "Towards a viable autonomous driving research platform," in *Intelligent Vehicles Symposium (IV), 2013 IEEE*. IEEE, 2013, pp. 763–770.

[22] J. B. White, "The big worry about driverless cars? losing privacy," *Wall Street Journal*, 2013.

[23] M. Zanin, S. Messelodi, C. M. Modena, and F. B. Kessler, "diplodoc road stereo sequence."