3D ROAD CURB EXTRACTION FROM IMAGE SEQUENCE FOR AUTOMOBILE PARKING ASSIST SYSTEM

Véronique Prinet *JinSong Wang* *JongHo Lee* *David Wettergreen*

General Motors, R&D

Robotics Institute, Carnegie Mellon University

ABSTRACT

We extract 3D curb from video sequence, using a single camera equipped with fish-eye lens and located at the front/rear of the vehicle. The challenge in extracting curbs from images lies in their small size and their lack of texture. We show that by appropriately exploiting appearance features, 3D geometry, and temporal information, one can reliably detect and localize the curbs in the 3D scene. The main underlying assumption of our model is that the road surface is flat and that the curb is approximately orthogonal to the road plane. We collected nine videos with ground truth, under day-time sunny weather condition, up to 2m range. Our experimental results compare favorably wrt the current the state-of-the-art on our database —90% precision rate in average and over 85% accuracy in curb height estimation.

Index Terms— Advanced Driving Assist Systems, Video Processing, Computer Vision

1. INTRODUCTION

ADAS (Advanced Driving Assistance Systems) is often seen as an intermediate stage before reaching the ultimate goal of fully autonomous driving. ADAS functionality integrates a bunch of active safety features. Its main goal is to alert the driver of possible danger and prevent any collision or crash (eg, pedestrian/vehicle detection, lane departure warming, lane keeping assist). Most advanced systems can take partial control of the vehicle: speed and steering wheels are automatically adjusted, taking into account the surrounding environment (eg., automatic cruise control, lane change on demand, parking assist). These drastic changes in vehicle conception, manufacturing and use (and more are expected for the next two decades) are the results of three major advances: i) development of new, more powerful and cheaper sensors (camera, lidar, radar, ultra-sound); ii) relative maturity of computer vision and advanced signal processing techniques; iii) cloud-based storage capacity and efficient wireless communication/data transfer.

Curb detection is a problem that has been barely thoroughly addressed in the literature, thought it is one of the important features in ADAS or future autonomous driving systems. Detecting side curbs –that separate the road from the sidewalk– contributes to accurate vehicle positioning in urban areas. On the other hand, the detection of curbs in front/rear of the vehicle is crucial for applications such as parking assist/autonomous parking. The accurate localisation and range estimation of the curb is passed to the vehicle control system, which will in turn smoothly maneuver the vehicle so as to avoid possible shock with the front/rear curb.

Fig. 1. Various types of parking curb.

The challenge in extracting curbs from images lies in their small size[1], ie around 10-20cm high. Conversely, their 3D shape is pretty standard, and can be roughly modeled as a 2D step-function. Consequently, most current approaches rely on active sensor (Lidar) or stereo-camera, which make it possible the direct extraction of 3D information and further eases the processing.

In this work, we depart from traditional approaches and instead propose to extract 3D curb from video sequence, using a single camera equipped with fish-eye lens and located at the front/rear of the vehicle. We show that by appropriately exploiting appearance features, 3D geometry, and temporal information, one can reliably detect and localize the curbs in the 3D world. We chose not to rectify the images to perspective projection so as to speed-up the processing. The main underlying assumption of our model is that the road surface is flat and that the curb is approximately orthogonal to the road plane (Fig. 2). Our experimental results compare favorably wrt the current the state-of-the-art on our database of nine sunny-weather videos —90% precision rate in average and over 85% accuracy in curb height estimation.

The main contribution of this work is two-fold:

[1] A small spatial support most often is not sufficient to provide strong evidence of presence/absence of the object (eg., [1]).

- We introduce a general framework for front/rear view 3D curb extraction from image sequence. Based primarily on a binary classifier, it is robust to variabilities in curb appearance.

- We demonstrate that the curb's 3D shape (distance and height) can be efficiently retrieved directly from the fish-eye image, without rectifying it to perspective geometry. This indeed is a major advantage when computation time is an issue.

2. RELATED WORK

There has been significant progress in solving tasks such as outdoor scene image parsing and road scene understanding during the past decade. Most of those work aim at segmenting "stuffs" and detecting "things" in a holistic manner, ie by minimizing a global cost function that accounts for different object classes and constraints. Some relies on deep learning [2], on hierarchical probabilistic graphical model [3], other integrate prior knowledge [4, 5] or exploit multi-sensor fusion [6]. None of those general frameworks however cope explicitly with the *curb* class specifically. Indeed, curbs/kerbs, because of their small size and non-distinctive pattern, still require dedicated approach to be detected successfully.

One of the main characteristics of a curb is its geometry: it can be modeled, in first approximation, as a 2D step function. Hence the use of 3D sensors (Lidar or stereo-vision) seems natural. Thought existing work differ in their detailed implementation, the core idea is always to fit a 3D point cloud using planar or polynomial approximation, then to detect regions/lines of strong discontinuity along one direction. Those discontinuities are an indicator of the presence of a curb [7, 8, 9, 10, 11]. Apart from those 3d-modeling based approaches, few work develop image processing techniques from still image or video. Then, the overall methodology consists in classifying various road marks, e.g soft shoulder, curbs, guardrails, based on strong prior knowledge about road scenes [12, 13, 14].

Different from those previous work, we model a road curb using both visual cues and geometric characteristic, from a monocular camera. In addition, we leverage temporal information by exploiting observation's redundancy between consecutive frames.

3. OVERVIEW

Our approach exploits appearance features, 3D geometry, and temporal information, in order to detect and localize the curbs in the 3D world.

The main pipeline is as follows. The first stage aims at detecting and localizing *candidate* curb regions in the image. We detect candidate regions using a machine learning SVM classifier combined with a robust image appearance descriptor and sliding window strategy. Our observation is that the curb can be treated as an object category, whose appearance is mainly similar across images and across different types of curbs (to a certain extend) –see Fig 1. We chose Histogram of Gradient as image descriptor. The sliding window SVM returns a score per window, indicating the region where the curb is most likely to appear in the image. In order to finely localise the curb borders, we extract parallel edge lines within the candidate region. The result of which is a pair of curves, which might delineate the curb in the image.

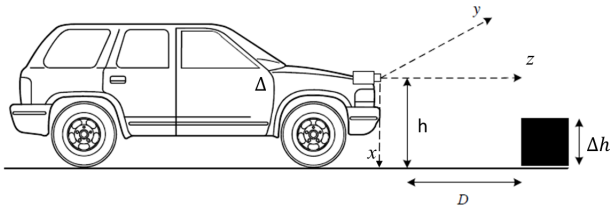


Fig. 2. Vehicle and scene geometry.

The second stage consist in computing the *geometry* of the candidate curb. Given a few assumptions about the scene and the camera, we can recover, from the fish-eye image directly, camera-to-curb distance and curb's height.

Finally, we leverage *temporal information* to filter out false detection (resp. to recover missed detection). A candidate curb border detected in the current frame will appear in the next frame at a position determined by the only camera motion. It is then a simple task to define a Kalman filter to track the pair of curves over time and potentially remove thoses which are inconsistent over time.

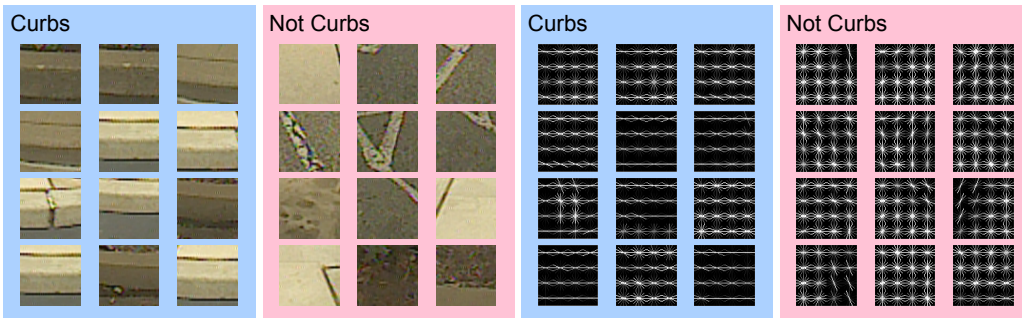
We further detail each step of the algorithm in the following section.

4. APPROACH

4.1. Curb detection and localization

Feature Extraction We chose Histogram of Gradient as image descriptor [15]. HOG is computed locally at every patches/windows of a regular grid. We extract HOG at two scales, using a two-level pyramid representation of each image [16]. Examples of curbs and non-curb patches and their HOG features are depicted in Figure 3. As an alternative to multi-scale strategy, we could have used an adaptive-size window [17].

Classifier A binary linear support vector machine (SVM) is applied to classify each patch [18, 19]. The SVM model is fed with the HOG descriptor —agnostic to the descriptor scale, a single SVM model is learned. Overlapping patches enable to vote for the curb/non-curb category. A single connected candidate region in each frame is then extracted by applying a threshold on the voting score of each patch.



(a) Samples of curb and non-curb patches. (b) Samples of curb and non-curb HOG representation.

Fig. 3. Gradient orientation is a very strong cue to characterize and represent the curb pattern.

Fine localization The borders of the curb (ie the top and bottom discontinuities of the 2D step function) enable to delineate the exact area of the curb. To localise those borders, we extract Canny edge points from the candidate region, after having smoothed the image with a Gaussian filter. From those strong edge points, we fit (up to five) curves (second order polynomials) using a sequential Ransac algorithm[2]. Among the possible extracted curves, we select the most probable pair based on a simple heuristic, and label each as "bottom edge'" or "top edge", so as to form couples of points $\{(p_i^b, p_i^t)\}_{i=1..N}$.

4.2. Range and height estimation

Geometric analysis enables us to estimate depth and elementary 3D information. A few assumptions about the scene are required (see Fig. 2): i) planar ground road, ii) known camera height wrt ground plane, h, iii) camera optical axis parallel to the road plane, iv) curb plane orthogonal to the ground plane. In addition, we assume that the camera has been calibrated[3]. Given the camera parameters, a point $p = (u, v)$ on the distorted fish-eye image can be mapped onto its unit-norm bearing vector $b = (\hat{u}, \hat{v}, \hat{w})$ [20] —the bearing vector defines the optical ray associated with point p. Note that coordinates (u, v) are here expressed w.r.t the image center (u_0, v_0). Thus, given a point P^g on the road ground plane that projects in p^g in the image, the camera-to-curb distance, D^g, writes:

$$D^g = \left| h \, \frac{\hat{w}^g}{\hat{u}^g} \right| , \tag{1}$$

where $\hat{w}^g$ is the coordinate along axis $\mathcal{Z}$ of the bearing vector in the camera-center reference system, and $D^g > 0$.

In practice, since the points of the curb's bottom-border detected in the image, $\{p_i^b\}$, are also located on the road plane in the scene, we first compute D_i^b from these points. Using assumption iv) ($D_i = D_i^b = D_i^t$), we can then deduce the height of the curb Δh for each pair of points (p_i^b, p_i^t): $\Delta h_i = D_i \frac{\hat{u}_i^t}{\hat{w}_i^t} - h$.

[2]Given the fact that the fish-eye images are not geo-corrected a priori, straight (curb) lines in the real scene appear as curves in the image.

[3]We use a calibration toolbox which can be used for catadioptric and fish-eye cameras up to 195 degrees field of view [20].

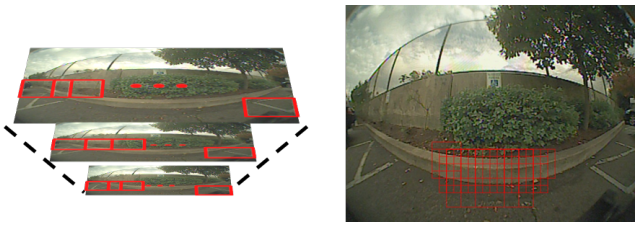
4.3. Temporal filtering

So far, we have considered that the candidate regions were correctly detected and that fine curb localisation was accurate. Temporal filtering intervenes as a post-processing step, so as to eliminate possible false-positives or to recover false-negatives due to possible erroneous detection or localisation. Ideally, if one knew the camera pose, one could define a Kalman filter whose state transition matrix would be defined by the known camera motion [21]. In practice, we do not have knowledge of the camera pose parameters (neither do we try to estimate them) and simply assumes a "smooth" monotonous change of the camera-to-curb distance, and a constant curb height over time. This leads us to define a Kalman filter whose state variables are the D_is and Δh_is and whose transition matrix is restricted to the identity matrix. This heuristic is justified –and works well experimentally– in scenarios where the vehicle moves at low speed (which is precisely the type of scenarios we are interested in, i.e parking), but would indeed be inappropriate for fully unconstrained vehicle motions.

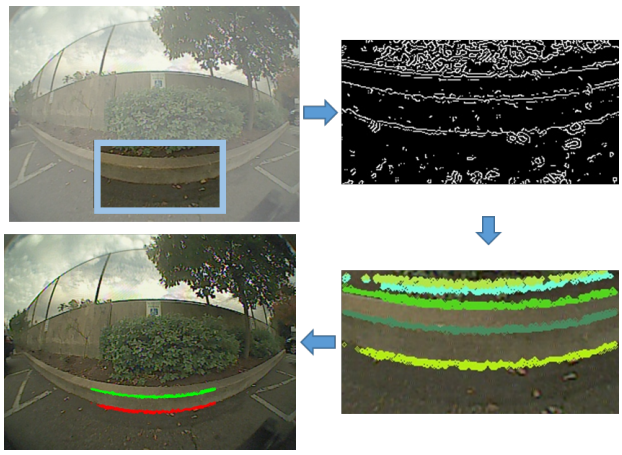
5. EXPERIMENTS

5.1. Experimental setting

The camera is mounted on the front vehicle bumper, at a position around 70cm above the ground. The image resolution is 640x480pix, recorded at 10 frames per second. The fish-eye lens has 180 degrees of horizontal field of view. We collected nine video sequences. All acquisitions correspond to a "parking scenario", with a maximum vehicle velocity around 15mph. The acquisitions are performed under "sunny weather", without particular illumination challenge. The camera intrinsic parameters are estimated off-line using dedicated matlab toolbox. The frames are not rectified to perspective geometry.



(a) Curb detection. Left: multi-scale local features. Right: sliding window SVM candidate detection.



(b) Curb Localisation. Clockwise: detected candidate region. Edge extraction. Curve fitting. Extracted bottom and top curb edges.

Fig. 4. Candidate detection and fine localization

Video frames were manually annotated, marking curb-regions and non-curb-regions patches. Patches are further used to train the SVM classifier. We did not have any means to measure a frame by frame ground-truth camera-to-curb distance[4]; instead we simply use the curb height that we measured manually on each scene. Evaluation is performed using frames for which the camera to curb range is maximum two meters.

5.2. Quantitative results analysis

Figure 4 illustrates the intermediate stages of the curb candidate detection and fine localization procedure.

We report precision-recall rates for each video sequence in Table 1. We marked as True Positive (TP) the images for which our approach detected correctly either the top edge or the bottom edge of the curb (i.e, within a few pixel from the manual labeling). We counted as False Positive (FP), the images in which our approach failed to detect correctly both curb edges. Images in which none of the curb edges were detected are marked as False Negatives (FN). We tested the algorithm based on a leave-one-out strategy. For example, we trained the SVM classifier using data-subset 2-9, then tested the approach on video 1. As expected, we observed a precision-recall rate higher at short range (> 90 % in average at 0.25-1m) than at larger range (recall rate of 0.7% at 1-2m).

We report curb height estimation error in Table 2 (6.5% error in average), and standard variation. We computed the curb height by averaging the estimated values from all TPs and FPs frames.

Rigorous quantitative comparison with other existing methods was not possible due to the lack of public benchmark. Quantitative results were nevertheless reported in, e.g [14, 22]: our experimental results compare favorably.

	0.25 - 1 meter		1 - 2 meters	
Dataset	Precision	Recall	Precision	Recall
1	1.000	0.818	1.000	0.800
2	1.000	1.000	1.000	0.672
3	1.000	0.996	1.000	0.768
4	0.980	0.969	0.739	0.405
5	0.957	0.856	0.917	0.898
6	1.000	1.000	1.000	0.970
7	0.960	0.923	1.000	0.857
8	0.457	0.615	0.450	0.450
9	0.864	0.792	0.875	0.226
Average	0.959	0.943	0.922	0.704

Table 1. Detection performance based on different ranges

Dataset	Est.(STD.)	GT	% Error
1	15.2 (1.6)	14	8.8
2	19.5 (1.2)	18	8.0
3	22.5 (1.4)	20	12.8
4	14.9 (2.1)	14	6.9
5	13.6 (1.7)	14	2.6
6	13.5 (2.4)	14	3.2
7	13.7 (2.1)	14	1.8
8	19.1 (3.4)	18	6.4
9	17.3 (2.0)	16	8.4

Table 2. Height (in cm) estimation performance

6. CONCLUSION AND FUTURE WORK

While we demonstrate in this work that one can successfully extract 3D curb from front view monocular camera, the future of ADAS technology lies in the joint exploitation of multiple sensors mounted on each side of the vehicle. An exciting possibility for future work would be to extend the approach to multi-view or multi-sensor joint processing.

Acknowledgement This work was supported by *General Motors-Carnegie Mellon Connected and Autonomous Driving Collaborative Research Laboratory (CAD-CRL)*. The authors thanks the anonymous reviewers for their constructive comments.

[4]Range ground truth could be obtained using a Lidar, or by marking the road with a accurate and fine ruler –tools that we did not have at disposal.

7. REFERENCES

[1] Jia Deng, Jonathan Krause, Michael Stark, and Li Fei-Fei, "Leveraging the wisdom of the crowd for fine-grained recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2015.

[2] A. Tandon J. Kiske W. Son J. Pazhayampallil M. Andriluka P. Rajpurkar T. Migimatsu R. Cheng-Yue F. Mujica A. Coates A. Ng B. Huval, T. Wang, "An empirical evaluation of deep learning on highway driving," in *arXiv:1504.01716v3*, 2015.

[3] S. Gould, T. Gao, and D. Koller, "Region-based segmentation and object detection," in *NIPS*, 2009.

[4] S.L. Wang, S. Fidler, and R. Urtasun, "Holistic 3d scene understanding from a single geo-tagged image," 2015.

[5] J. yao, S. Fidler, and R. Urtasun, "Describing the scene as a whole: Joint object detection, scene classification and semantic segmentation," 2012.

[6] W.Q. Huang and X. Gong, "Fusion based holistic road scene understanding," in *arXiv:1406.7525v1*, 2014.

[7] R. Aufrère, C. Mertz, and C. Thorpe, "Multiple sensor fusion for detecting location of curbs, walls, and barriers," in *Intelligent Vehicles Symposiumitle*, 2003.

[8] O. Gallo, R. Manduchi, and A. Rafii, "Robust curb and ramp detection for safe parking using the canesta tof camera," in *Computer Society Conference on Computer Vision and Pattern Recognition Workshop on Time of Flight based Computer Vision*, 2008.

[9] F. Oniga and S. Nedevschi, "Curb detection for driving assistance systems: A cubic spline-based approach," in *Intelligent Vehicles Symposium*, 2011.

[10] J. Siegemund, D. Pfeiffer, U. Franke, and W. Forstner, "Curb reconstruction using conditional random fields," in *Intelligent Vehicles Symposium*, 2010.

[11] R. Turchetto and R. Manduchi, "Visual curb localization for autonomous navigation," in *International Conference on Intelligent Robots and Systems*, 2003.

[12] Vladimir Haltakov, Heidrun Belzner, and Slobodan Ilic, "Scene understanding from a moving camera for object detection and free space estimation," in *Intelligent Vehicles Symposium*, 2012.

[13] A. Seibert, M. Hahnel, A. Tewes, and R. Rojas, "Camera based detection and classification of soft shoulders, curbs and guardrails," in *Intelligent Vehicles Symposium*, 2013.

[14] M. Kellner, U. Hoffmann, and S. Neumaier, "Multi-cue, model-base detection and mapping of road curb features using stereo vision," in *IEEE International Conference on Intelligent Transportation Systems*, 2015.

[15] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2005.

[16] P. J. Burt and E. H. Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, 1983.

[17] Y. Ono, A. R. Junaidi, and T. Kuroda, "Adaptive window search using semantic texton forests for real-time object detection," in *ICIP*, 2013.

[18] H. Drucker, C. Burges, L. Kaufman abd A. Smola, and V. Vapnik, "Support vector regression machines," in *Advances in Neural Information Processing Systems*, 1996.

[19] Vladimir Vapnik, *The Nature of Statistical Learning Theory*, Springer Verlag, 2000.

[20] D. Scaramuzza, A. Martinelli, and R. Siegwart, "A flexible technique for accurate omnidirectional camera calibration and structure from motion," in *International Conference on Computer Vision Systems*, 2006.

[21] R.E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, 1960.

[22] A. Seibert, M. Hahnel, A. Tewes, and R. Rojas, "Camera based detection and classification of soft shoulder, curbs and guardrails," in *IEEE Intelligent Vehicles Symposium (IV)*, 2013.