A Self-Adaptive Approach for Curbstone/Roadside Detection based on Human-Like Signal Processing and Multi-Sensor Fusion

Thomas Michalke, Robert Kastner, Jannik Fritsch, Christian Goerick

2010

Preprint:

This is an accepted article published in IEEE Intelligent Vehicles Symposium (IV). The final authenticated version is available online at: https://doi.org/[DOI not available]

A Self-Adaptive Approach for Curbstone/Roadside Detection based on Human-Like Signal Processing and Multi-Sensor Fusion

Thomas Michalke[1], Robert Kastner*, Jannik Fritsch◇, Christian Goerick◇

◇Honda Research Institute Europe GmbH
D-63073 Offenbach, Germany
{jannik.fritsch, christian.goerick}@honda-ri.de

*Darmstadt University of Technology
Institute for Automatic Control
D-64283 Darmstadt, Germany
robert.kastner@rtr.tu-darmstadt.de

***Abstract*— Newly emerging, highly complex Advanced Driver Assistance Systems (ADAS) fuse the output of various system modules (e.g., lane detection, object classification). Such knowledge fusion is realized in order to gain additional information of the environment allowing for complex system tasks as path planning, the active search for specific objects and task-specific analysis of the environment. As part of our previous work, we realized a highly generic type of such ADAS using biological principles. The present contribution offers a novel approach for the detection of curbstones and elevated roadsides in inner-city that relies on biological principles taking inspiration from the human neural signal processing. The gathered results can be fused to an ADAS in order to improve the quality of various other system percepts and allow additional system tasks.**

Keywords: curbstone detection, driver assistance, robust path identification, lane detection

I. INTRODUCTION

Following the significant progress in computer hardware combined with its decreasing costs, a new class of driver assistance systems has emerged. These so-called Advanced Driver Assistance Systems (ADAS) typically combine numerous system modules into highly interlinked system architectures that allow for complex system tasks, as path planning or active collision avoidance (see, e.g., [1], [2]). Typically these system architectures are rather static in terms of possible system tasks and supported environments. Different from that, in our earlier work a dynamic, task-dependent tunable ADAS was developed (see, e.g., [3], [4]) that relies on biological principles. Among other things this system comprised a subsystem for marked and unmarked road detection, an environmental map as well as a generic image preprocessing stage for the detection of basically any object class. As part of this human-like ADAS, a subsystem for the detection of curbstones and elevated roadsides was developed that relies on biological principles. This system will be described and evaluated in this contribution.

In general, a robust approach for the detection of curbstones and elevated roadsides would improve various other system percepts in an ADAS (e.g., in a more precise environmental map a pedestrian could be assigned to the sidewalk, the vision processing could be guided to suspicious objects near the detected roadside, a precise localization in a digital map would be possible). Furthermore various system tasks in inner-city become possible (improvement of autonomous parking maneuvers, improved collision avoidance in inner-city or more precise analyzation of intersections).

As the following Section 2 will show, only few dedicated approaches for curbstone and elevated roadside detection exist that mainly suffer from different limitations. Most of them rely on a single sensor approach, disregarding the full potential of sensor fusion. As opposed to that, the approach described in Section 3 relies on the fusion of different sensor modalities using biologically inspired methods. In Section 4 the capabilities of the proposed approach are assessed, after which the contribution is summarized.

II. RELATED WORK

Only few dedicated approaches for the detection of curbstones and elevated roadsides exist. Typically these approaches rely on a single sensor for the generation of 3D data. Based on this 3D data a height map of the environment can be generated and suspicious edges are detected and refined.

More specifically, in [5] a laser scanner is used to derive 3D data of the environment. The derived differences in height of neighboring scan points are thresholded, filtered and a line model is fitted. The laser scanner used in this application is marked by a relatively low frame rate, which restricts the maximum supported velocity. Furthermore, the range of robust detection of curbs is restricted due to the growing influence of noise with increasing distance to the ego vehicle.

Another more sophisticated approach is presented in [6]. For generating dense 3D data, a stereo camera setup is used. The authors propose the computation of a so-called Digital Elevation Map (DEM), which is a height map of the scene as viewed from above. On the DEM, an edge detection, filtering and spline fitting is realized. Although the presented DEM is very noisy, the gathered results appear to be robust. As the presented results show, only curbs that are near to the vehicle can be detected. This is due to the growing influence of inaccuracies (e.g., the quantization error) with increasing distance to the ego vehicle (see [7] for a comprehensive treatment of this issue).

[1]After finishing his Ph.D. in a cooperation between Honda Research Institute Europe GmbH and the Darmstadt University of Technology, the author now works for Daimler AG (email: thomas_paul.michalke@daimler.com). The system presented in this contribution was realized as part of the Ph.D. thesis.

Other approaches fuse different sensor outputs in order to improve the detection robustness. For example, in [8] a laser sensor is combined with a monocular camera. The single-layer laser sensor is used to detect the position where the curbstone cuts the sensed layer. This point is used as starting point for the image processing. More specifically, a static edge detector is applied starting from the laser-sensed curbstone. Since, in some cases the appearance of curbstones only slightly differs from the road (i.e., virtually no edge is present in the captured image) this approach is marked by a restricted robustness. Another multi-sensor approach is described in [9]. Here stereo data and vision data is fused using probabilistic methods. The described approach realizes a late fusion of the road detection results of all present sensors. More specifically, curbstones are detected by specific edge filters and elevated roadsides by the stereo sensor. No early information fusion between the detection results takes places, limiting the achievable performance gain from sensor fusion.

Newly emerging multilayer laser scanners offer novel possibilities for the detection of curbstones and road boundaries, since such sensors combine high frame rates with a high accuracy 3D scan (see e.g., [10] for typical high precision applications of such sensors). However, it is important to note the relatively high costs of this sensor type.

As opposed to that, stereo cameras are getting affordable and technologically sound (e.g., robust solutions for calibrating the cameras exist) and require a comparatively low amount of space in the vehicle. Still, as described above sufficient accuracy for curb detection can only be gained in the first few meters from the ego-vehicle. To solve this dilemma, our contribution proposes a specific fusion of 3D and vision data. Vision data is marked by its high information density. The vision-based detection of curbstones and lane borders is possible even at large distances. Still, curbstones and lane borders can have various appearances, which makes the design of a generic appearance-based model difficult. As described in the following, in the here proposed system stereo data is used to detect curbstones near the vehicle. At greater distances a complete switch to vision takes place. The detection result of the stereo cue is used to adapt a vision template of the present curbstone/elevated roadside, while relying on biologically inspired approaches for signal processing.

III. SYSTEM

The proposed overall system concept for the adaptive detection of curbstones and elevated roadsides is depicted in Fig. 1. After giving a rough overview of the major processing steps, all system modules are described in detail.

The proposed system consists of two major parts. First a stereo-based detection step, second a vision-based detection step that is initialized and modulated by the results of the first. The stereo-based curbstone detection relies on dense 3D data coming from a stereo camera setup. Based on that, a specific height map (the so-called Digital Elevation Map) is computed and temporally integrated in order to reduce noise.

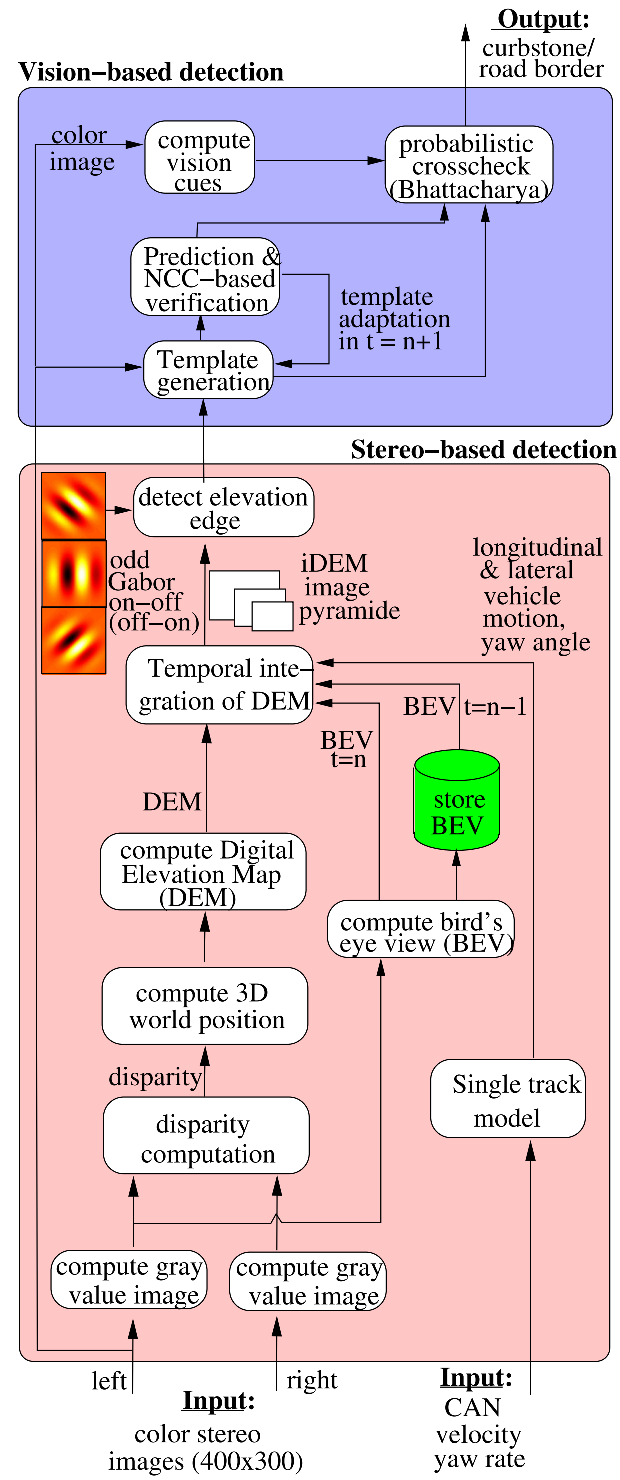

Fig. 1. System architecture.

On this thereby extended Digital Elevation Map, a specific biologically motivated edge filtering is applied resulting in a robust curbstone detection up to a distance of 9 meters from the vehicle. Based on the stereo detection result a vision template of the roadside is generated and adapted. Using this template, the course of the roadside is detected via means of computer vision. The second detection step stops, when the

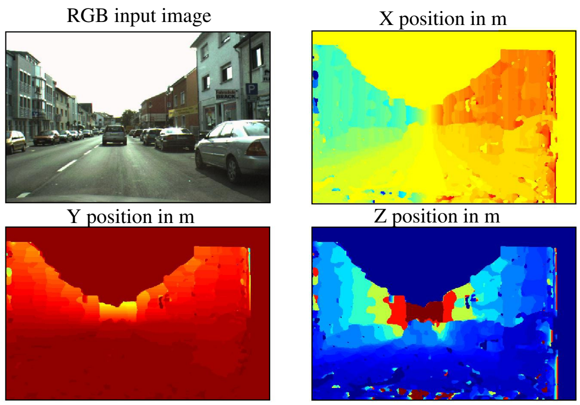

Fig. 2. Dense 3D world positions for all image pixels based on stereo.

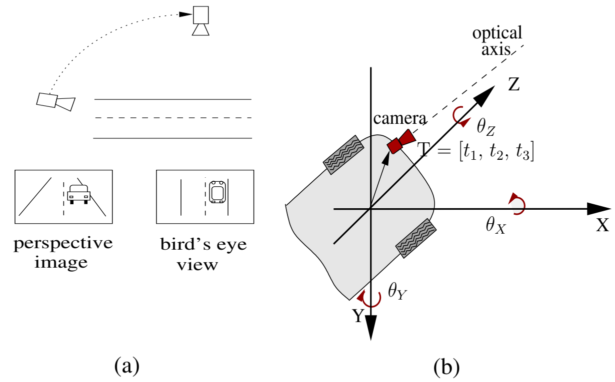

Fig. 3. (a) Visualization of the bird's eye view, (b) Coordinate system and position of the camera (car is heading in Z-direction).

statistical properties of the road and nonroad area (on both sides of the road) change abruptly. This change typically marks an erroneous vision-based curbstone detection far in the distance or an object that occludes the roadside.

In the following, all system modules are described in detail. The system uses pairs of color images of 400x300 pixels captured by a stereo camera setup mounted in the car as input information. After an image undistorsion and rectification step (see [11]) as well as a transformation into grayscale, a dense disparity $D(u, v)$ is computed using correspondence search with a probabilistic matching algorithm (see [12] for details). Based on the disparity image the 3D world position for all image pixels can be computed using:

$$Z_{\text{stereo}}(u, v) = \frac{f_u B}{D(u, v)} + t_3 \quad (1)$$

$$Y_{\text{stereo}}(u, v) = \frac{Z(v - v_0)}{f_v} + t_2 \quad (2)$$

$$X_{\text{stereo}}(u, v) = \frac{Z(u - u_0)}{f_u} + t_1. \quad (3)$$

With: B... basic distance between cameras principal point
f_u, f_v... normalized focal length [in pixels]
$D(u, v)$... disparity
u_0, v_0... principal point of the left camera
t_1, t_2, t_3... translational camera offset.

Please refer to Fig. 2 for a visualization of the used dense stereo data and Fig. 3b for the here applied coordinate system. In the following step, the stereo maps $Z_{\text{stereo}}(u, v)$, $Y_{\text{stereo}}(u, v)$, $X_{\text{stereo}}(u, v)$ are unrectified (i.e., the prior rectification before disparity computation is neutralized) to make them comparable to the input image in terms of pixel position. Now, the Digital Elevation Map (DEM) is computed, which is a metric height map of the scene as viewed from above (see [13] for details). However, as dedicated testing has revealed, the plain DEM is too noisy for a direct elevation edge detection (see Fig. 4c). Therefore different from [13] and hence as a novel approach a temporal integration procedure is applied on the DEM. More specifically, based on a single track model that uses the vehicle velocity and yaw-rate from the CAN bus, the longitudinal and lateral vehicle motion as well as yaw angle is estimated. Furthermore, from the left camera image a bird's eye view representation (see [14] for background information) is computed for all input images. The bird's eye view (BEV) is a metric representation of the scene as viewed from above that results from remapping the gray scale input image (see Fig. 3a and 4b). Using the BEV of the current and previous image the motion of the vehicle is computed using Normalized Cross Correlation (NCC). In order to reduce the computational costs of the NCC, the estimated motion of the vehicle from the single track model is used to define an anchor point for the correlation template in the current BEV. Based on this procedure, the vehicle motion of the previous 10 time frames is determined and stored. The DEMs of the previous 10 time frames are shifted accordingly and superimposed resulting in the integrated DEM (iDEM). As Fig. 4d shows the iDEM is marked by much less noise as the DEM in Fig. 4c (see [15] for technical details on the temporal integration procedure).

Finally, on the iDEM an elevation edge detection is realized. More specifically, odd Gabor filters of the orientations 45, 90 and 135 degree are applied on three scales for reducing the computational costs, while allowing the detection of different edge widths. For that, on the iDEM a three level Gaussian image pyramid is computed. The applied Gabor filters are biologically motivated image filters that were shown to exist in the vision pathway of the mamal brain (see [16] for the biological and [17] for a theoretical background on this filter type). For the right/left roadside odd Gabor filters with off-on/on-off contrast type selectivity are used (refer to Fig. 5 and 4e for an example and [18] for technical details on this specific decomposition technique for Gabor filters). By decomposing the Gabor filters, a simple and effective way for side-specific filtering of elevated edges on the right and left roadside becomes possible.

Dedicated evaluation has shown that (for our camera setup) from a distance of about 9 meters on a curb detection

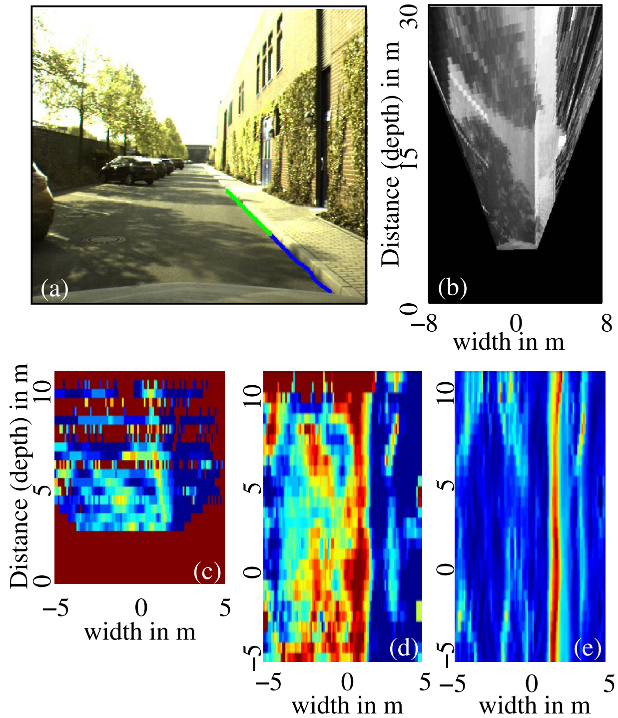

Fig. 4. Exemplary inner-city scenario (visualization for right curbstone only): (a) Captured image with detected right curbstone (blue line: stereo-based, green line: vision-based), (b) Bird's eye view, (c) DEM, (d) Temporal integrated DEM, (e) Odd Gabor off-on filtered.

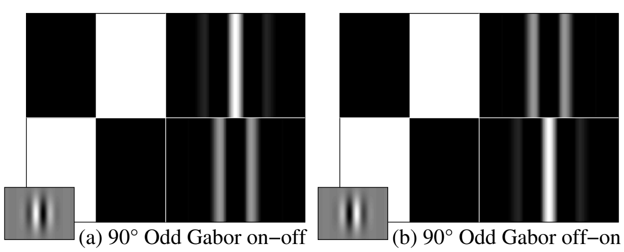

Fig. 5. Application of 0 degree odd Gabor filter kernels on simplified road border images: (a) Kernel for on-off contrasts (left road border), (b) Kernel for off-on contrasts (right road border).

on stereo starts to get noisy and cannot run robustly. Therefore, stereo is used as cue for roadside detection for the first 9 meters distance from the vehicle only.

As stated by [16], also psychophysical evidence exists that marks stereo as the primary depth cue of the human for the first few meters only. After that other (monocular) vision features take over. In accordance to that, the processing is switched to vision-based edge detection. More specifically, the last stereo-based elevation edge measurement is used to adapt a vision template using the data of the left color image. After a prediction step (a segment-wise linear line model is applied) the template is used to verify and correct the real position of the roadside using NCC. In case of a low NCC value a probabilistic crosscheck is done. For that 5 vision cues are computed on the left color image: the hue and saturation channel of the HSI color space, as well as the edge density on the hue, saturation and intensity. All 5 features have been shown to be robust cues for unmarked road detection (see [7] for technical details on these features). In order to determine, if a vision-based detected roadside segment is valid, image patches left respectively right of the roadside segment (subsequently called "road patch" P_{road} and "nonroad patch" P_{nonroad}) are compared to their template counterparts T_{road} and T_{nonroad}. The named templates are derived from the valid roadside region that was previously determined using the DEM. The comparison between the road/nonroad patches and their templates is realized by applying the distance measure $\delta(P_k, T_k)$ with $k \in \{\text{road}, \text{nonroad}\}$ that is based on the Bhattacharya coefficient $\gamma(H_i^{P_k}, H_i^{T_k})$ (a measure for determining the similarity between two histograms) calculated on the histograms $H_i^{P_k}$ and $H_i^{T_k}$ of the image patches of all $N = 5$ vision cues:

$$\begin{aligned} \delta(P_k, T_k) &= \sum_{i=1}^{N} \sqrt{1 - \gamma(H_i^{P_k}, H_i^{T_k})} \qquad (4) \\ \gamma(H_i^{P_k}, H_i^{T_k}) &= \sum_{\forall u,v} \sqrt{H_i^{P_k}(u,v) H_i^{T_k}(u,v)}. \end{aligned}$$

In case $\delta(P_k, T_k)$ is bigger than the maximum distance between all previously found (valid) DEM road/nonroad patches, the newly found roadside segment is invalid and the algorithm stops. A big distance $\delta(P_k, T_k)$ can result from a detection error (typically happening far in the distance) or from an object that occludes and interrupts the roadside. The thereby gathered information can be used in our biologically motivated ADAS to guide the attention. In case the roadside is interrupted in the vicinity of the camera-carrying ego-vehicle, the attention system could be set to analyze the specific image region more closely in order to rule out a potential danger coming from a so far unknown object on the road.

In the following, an abundant evaluation of the described algorithm will allow the assessment of its capabilities and robustness.

IV. RESULTS

With [7] a robust unmarked road detection system for inner-city application was presented. Based on that, [15] describes a temporal integration system for unmarked road detection results. More specifically, in [15] road patches detected by systems as [7] are shifted according to the vehicle ego-motion, overlaid, and thresholded in order to gather more robust temporally integrated road segments. Since the analysis of the remaining road detection errors in the results of [15] has shown that in challenging situations the curbstone can be surpassed by the detected road segment, we extended the system and added the here presented curb detection system as final postprocessing step. More specifically, all road segments surpassing the detected curbstones were corrected (clipped). In order to evaluate the benefit of the here presented system, we used an inner-city stream (210 images with hand-labeled groundtruth) and the same evaluation methods as in [15]. The evaluation was realized on road detection results with and without temporal integration of road segments. For the evaluation we hence adopt the following Equations:

$$\text{Completeness} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (5)$$

$$\text{Correctness} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (6)$$

$$\text{Quality} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}. \quad (7)$$

The Equations define different groundtruth-based measures, which were taken from [19] (with pixels being True Positive (TP), False Negative (FN), False Positive (FP)).

On a descriptive level, the Completeness states, based on given ground truth data, how much of the present road was actually detected. The Correctness states how much of the detected road is actually road, in order to avoid classifying all pixels as road leading to a Completeness of 100%. The Quality combines both measures, since between the Completeness and Correctness a trade-off is possible. Based on this, the Quality measure should be used for a comparison, since it weights the FP and FN pixels equally. For a more detailed analysis, the Completeness and Correctness can be evaluated and thereby determining what exactly caused a difference in Quality.

Road detection approaches	**# test images**	**Correct-ness**	**Comple-teness**	**Quality**
System [7] without temp. integ. of road segments	210	83.9%	77.1%	66.4%
Curbstone postproc., without temp. integ. of road segments	210	**95.8%**	**76.7%**	**73.7%**
System [15] with temp. integ. of road segments	210	80.2%	88.7%	72.3%
Curbstone postproc., with temp. integ. of road segments	210	**94.9%**	**88.1%**	**83.8%**

TABLE I

EVALUATING THE INFLUENCE OF DETECTED CURBSTONES ON UNMARKED ROAD DETECTION RESULTS WITH AND WITHOUT TEMPORAL INTEGRATION.

As Tab. I shows using the curbstone detection results in a postprocessing step improved the detection Quality from 66.4% to 73.7% (without applying temporal integration) respectively from 72.3% to 83.8% (with applying temporal integration). When analyzing Tab. I more closely, it can be perceived that postprocessing with curbstone detection results lowers the Completeness by a small extend. For both cases the Completeness is diminished: without temporal integration (from 77.1% to 76.7%) and with temporal integration (from 88.7% to 88.1%). More specifically, that means that some true-positively detected road segments were falsely corrected and hence clipped by applying the curbstone detection results. Still, the postprocessing step

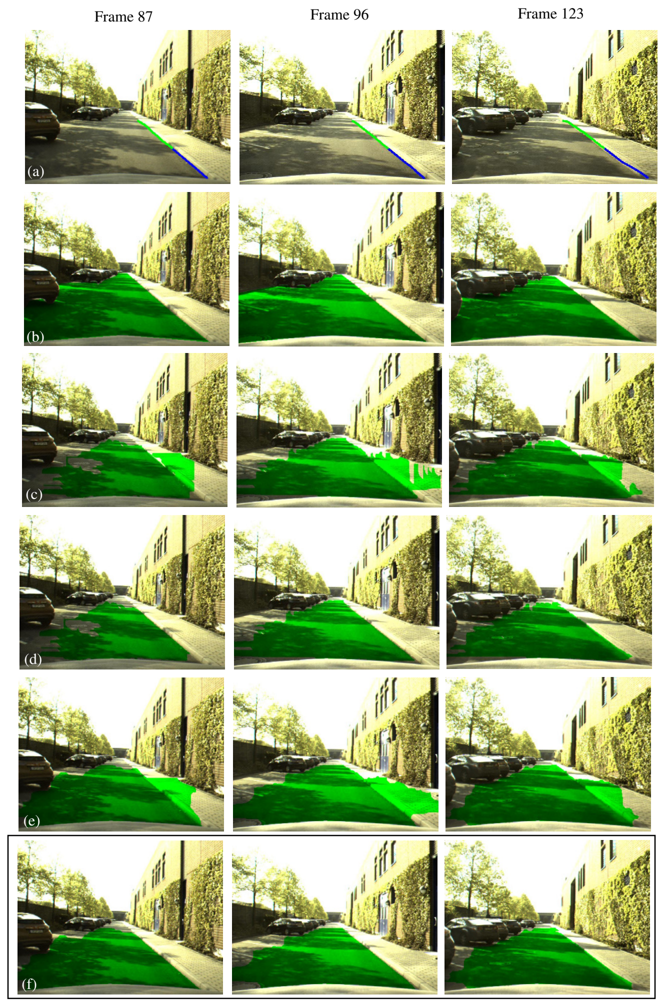

Fig. 6. Typical frames of inner-city stream: (a) Captured images with detected right curbstone (blue line: stereo-based, green line: vision-based), (b) Hand-labeled groundtruth road segments, (c) Detected road without temporal integration, (d) Incorporating curbstone detection result (without temporal integration), (e) Detected road with temporal integration, (f) Final detection result: Incorporating curbstone detection (with temporal integration)

corrects more false-positive pixels than it falsely corrects true-positive ones. For both cases the Correctness grows: without temporal integration (from 83.9% to 95.8%) and with temporal integration (from 80.2% to 94.9%). Summing up, the Correctness is overproportionally improved leading to the measured significant increase in Quality. In order to facilitate the empirical assessment of the gathered results, in Fig. 6 some typical frames of the used inner-city stream are depicted.

For our experiments we use a Honda Legend prototype car equipped with a mvBlueFox CCD color camera from Matrix Vision delivering images of 800x600 pixels, which are resized to 400x300 for the referenced ADAS sub-modules. The images are grapped with 10Hz, which is hence the processing rate our road detection module has at least to reach. The image data as well as the vehicle state data from the CAN bus is transmitted via LAN to several Toshiba Tecra A7 (2 GHz

Core Duo) running our RTBOS integration middleware [20] on top of Linux. As most of the system's sub-modules are already part of our ADAS running in real-time (e.g., stereo computation, bird's eye view, temporal integration architecture, single track model), we expect the remaining currently Matlab-implemented sub-modules to fulfill real-time requirements also.

V. SUMMARY

The paper describes a biologically inspired approach for detecting curbstones and elevated road borders. More specifically, image filter kernels were applied that are known to exist in the human vision pathway. Also a human-like multi-sensor fusion between stereo and various other vision cues is applied that allows the online adaptation of a road border respectively curbstone model. As a novel approach, our system applies a temporal integration step on the digital elevation map (DEM) for diminishing noise and increasing the range for stereo-based curbstone detection.

An incorporation of the presented system into the ADAS described in [3] would enhance its robustness (e.g., improvement of the environmental model) and would furthermore offer additional possibilities to modulate the system behavior (e.g., focus attention to image locations of abruptly ending curbstones).

REFERENCES

[1] S. Thrun *et al.*, "Winning the DARPA grand challenge," *Journal of Field Robotics*, 2006.

[2] C. Urmson *et al.*, "Autonomous driving in urban environments: Boss and the urban challenge," *Journal of Field Robotics Special Issue on the 2007 DARPA Urban Challenge, Part I*, vol. 25, no. 1, pp. 425–466, June 2008.

[3] T. Michalke, R. Kastner, J. Fritsch, and C. Goerick, "Towards a proactive biologically-inspired advanced driver assistance system," in *IEEE Intelligent Vehicles Symposium*, Xian, 2009.

[4] J. Fritsch, T. Michalke, A. Gepperth, S. Bone, F. Waibel, M. Kleinehagenbrock, J. Gayko, and C. Goerick, "Towards a human-like vision system for driver assistance." in *IEEE Intelligent Vehicles Symposium*, Eindhoven, 2008.

[5] R. Wang, B. Gu, T. Yu, and L. Gou, "Study on curb detection method based on 3d range image by laserradar," in *IEEE Intelligent Vehicles Symposium*, Las Vegas, 6 2005.

[6] F. Oniga, S. Nedevschi, and M. Meinecke, "Curb detection based on a multi-frame persistence map for urban driving," in *IEEE Conference on Intelligent Transportation Systems*, Beijing, 2008.

[7] T. Michalke, R. Kastner, M. Herbert, J. Fritsch, and C. Goerick, "Adaptive multi-cue fusion for robust detection of unmarked inner-city streets," in *IEEE Intelligent Vehicles Symposium*, Xian, 2009.

[8] R. Aufrere, C. Mertz, and C. Thorpe, "Multiple sensor fusion for detecting location of curbs, walls, and barriers," in *IEEE Intelligent Vehicles Symposium*, 2003.

[9] R. Danescu, S. Nedevschi, M. Meinecke, and T. To, "A stereovision-based probabilistic lane tracker for difficult road scenarios," in *IEEE Intelligent Vehicles Symposium*, Eindhoven, 2008.

[10] B. Roessler, K. Fuerstenberg, and U. Lages, "Laserscanner for multiple applications in passenger cars and trucks," in *Proceedings of AMAA 2006, 10th International Conference on Advanced Microsystems for Automotive Applications*, Berlin, 2006.

[11] J. Heikkila and O. Silven, "A four-step camera calibration procedure with implicit image correction," in *CVPR97*, 1997, pp. 1106–1112.

[12] V. Willert, J. Eggert, J. Adamy, and E. Koerner, "Non-gaussian velocity distributions integrated over space, time and scales," *IEEE Transactions on Systems, Man and Cybernetics B*, vol. 36, no. 3, pp. 482–493, 2006.

[13] F. Oniga, S. Nedevschi, M. Meinecke, and T. To, "Road surface and obstacle dectection based on elevation maps from dense stereo," in *IEEE Conference on Intelligent Transportation Systems*, Seattle, 2007.

[14] A. Broggi, "Robust real-time lane and road detection in critical shadow conditions," in *Proc. Int. Symp. on Computer Vision*. Parma: IEEE, 1995.

[15] T. Michalke, R. Kastner, J. Fritsch, and C. Goerick, "A generic temporal integration approach for enhancing feature-based road-detection systems," in *IEEE Intelligent Transportation Systems Conference*, Beijing, 2008.

[16] S. Palmer, *Vision Science: Photons to Phenomenology*. MIT Press, 1999.

[17] R. Trapp, "Stereoskopische korrespondenzbestimmung mit impliziter detektion von okklusionen," Ph.D. dissertation, University of Paderborn Germany, 1998.

[18] T. Michalke, J. Fritsch, and C. Goerick, "Enhancing robustness of a saliency-based attention system for driver assistance," in *The 6th International Conference on Computer Vision Systems (ICVS), Santorini, Greece, 2008. Lecture Notes in Computer Science, Springer*, no. 5008, 2008, pp. 43–55.

[19] P. Lombardi, M. Zanin, and S. Messelodi, "Unified stereovision for ground, road and obstacle detection," in *IEEE Intelligent Vehicles Symposium*, 2005.

[20] A. Ceravola, F. Joublin, M. Dunn, J. Eggert, and C. Goerick, "Integrated research and development environment for real-time distributed embodied intelligent systems," in *Proc. Int. Conf. on Robots and Intelligent Systems*, 2006, pp. 1631–1637.