Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems (ITSC 2013), The Hague, The Netherlands, October 6-9, 2013

A New Performance Measure and Evaluation Benchmark for Road Detection Algorithms

Jannik Fritsch, Tobias Kühnl, Andreas Geiger

Abstract— **Detecting the road area and ego-lane ahead of a vehicle is central to modern driver assistance systems. While lane-detection on well-marked roads is already available in modern vehicles, finding the boundaries of unmarked or weakly marked roads and lanes as they appear in inner-city and rural environments remains an unsolved problem due to the high variability in scene layout and illumination conditions, amongst others. While recent years have witnessed great interest in this subject, to date no commonly agreed upon benchmark exists, rendering a fair comparison amongst methods difficult.**

In this paper, we introduce a novel open-access dataset and benchmark for *road area* and *ego-lane* detection. Our dataset comprises 600 annotated training and test images of high variability from the KITTI autonomous driving project, capturing a broad spectrum of urban road scenes. For evaluation, we propose to use the 2D Bird's Eye View (BEV) space as vehicle control usually happens in this 2D world, requiring detection results to be represented in this very same space. Furthermore, we propose a novel, behavior-based metric which judges the utility of the extracted *ego-lane* area for driver assistance applications by fitting a *driving corridor* to the road detection results in the BEV. We believe this to be important for a meaningful evaluation as pixel-level performance is of limited value for vehicle control. State-of-the-art road detection algorithms are used to demonstrate results using classical pixel-level metrics in perspective and BEV space as well as the novel behavior-based performance measure. All data and annotations are made publicly available on the KITTI online evaluation website in order to serve as a common benchmark for road terrain detection algorithms.

I. Introduction

Recent years have witnessed a strong increase of image processing functions in Advanced Driver Assistance Systems (ADAS) for passenger cars. While holistic traffic scene understanding might still be a dream of the future [1], specialized systems such as *lane keeping assistance* have become a standard ADAS component that is available in most new car models. While such systems are already commercially available [2], they are targeted at marked roads with smooth curvatures, limiting their applicability to highways and highway-like roads.

In order to provide support on unmarked roads that are common in rural areas and inner-city, a number of publications have proposed road detection algorithms that avoid the need for lane marking detection. Early works focused on the overall *road area* as it is more easy to extract. For example, the physical property of the road being flat has been used in a variety of approaches [3], [4]. However, this requires the road area to be limited by sufficiently elevated structures as well as accurate depth information which for stereo cameras is only available in the close range. Consequently, many approaches put higher emphasis on appearance cues such as the color and texture of the road area [5], [6], [7], [8], [9], [10], [11], [12]. These visual properties of the road area have been used for estimating the overall road shape [12] or for segmenting the complete road area [5], [6], [8], [10].

Unfortunately, most of the existing approaches have been evaluated on different datasets, prohibiting a fair performance comparison. Furthermore, most of the existing work does not distinguish road areas on a semantic level, e.g., ego-lane vs. opposing lane, which is crucial for ADAS and autonomous driving. Only recently, methods for detecting lane markings and curbstones have been combined explicitly [13], [14], [15] or implicitly [16] to achieve city lane detection. The final target of such approaches is the identification of the *ego-lane*, independent of the boundary type or lane shape.

In order to turn research efforts into actual driver assistance systems, their performance has to be evaluated and the strengths and weaknesses of the approaches have to be clearly identified, requiring suitable evaluation measures. Often, pixel-based evaluation measures, inspired from the vision community, have been directly applied in the image domain. We believe that the evaluation of road detection algorithms requires a stronger focus on the target application: Any driving maneuver or vehicle control is performed in the metric 2D space on the road. Consequently, the result of any road detection algorithm will need to be provided or transformed into such a spatial representation which is appropriate for vehicle control. The metric road space can be represented in the so-called Bird's Eye View (BEV) [17] by assuming a flat world for the transformation from the perspective image to the BEV space. We argue that any road detection algorithm should be evaluated in the BEV. While previous attempts at pixel-based evaluation [18] could be configured to mimic this requirement by weighting the pixels in the perspective image, the weighting would need to be adaptive as the 2D world position of a pixel varies non-linearly with its location in the image.

Instead, in this paper we apply classical pixel-based evaluation measures directly in the BEV. Furthermore, as pixel-level annotation does not provide suitable answers to questions like *'Where are the ego-lane boundaries?'* or

J. Fritsch is with the Honda Research Institute Europe GmbH, Offenbach am Main, Germany. Jannik.Fritsch@honda-ri.de

T. Kühnl is with the Honda Research Institute Europe GmbH, Offenbach am Main, Germany, and with the Research Institute for Cognition and Robotics, Bielefeld University, Bielefeld, Germany. TKuehnl@cor-lab.uni-bielefeld.de

A. Geiger is with the Max Planck Institute for Intelligent Systems in Tübingen and with the Karlsruhe Institute of Technology, Germany. andreas.geiger@tue.mpg.de

978-1-4799-2914-613/$31.00 ©2013 IEEE

'Can the driver safely continue driving?', we additionally fit a *hypothesis* for a possible driving path to the road detection result in the BEV. This *driving corridor* hypothesis represents an abstraction from the pixel level and allows for a *behavior-based* evaluation of road detection algorithms.

A second contribution of this paper is the introduction of a benchmark for the evaluation of road detection algorithms. We present three novel and highly challenging datasets derived from the KITTI autonomous driving project [19], capturing a variety of inner city scenarios including unmarked and marked lanes, as illustrated in Fig. 2. Our KITTI-ROAD dataset and benchmark are made publicly available together with the proposed behavior-based performance measure, classical pixel-level evaluation metrics as well as an automatic evaluation service on the KITTI website[1]. We believe this to be an important step towards investigating the pros and cons of different approaches on the same basis and to foster novel research and progress in this field.

The outline of this paper is as follows: Section II reviews existing evaluations measures. The new KITTI-ROAD benchmark dataset is introduced in Section III. Section IV reviews classical pixel-based performance metrics and introduces the proposed behavior-based performance measure. Evaluation results for the pixel-based and behavior-based performance measures using a simple baseline and several road detection algorithms are presented in Section V. The paper concludes in Section VI.

II. Related Work

For evaluating *road area* and *ego-lane* detection approaches a variety of evaluation measures have been used. They can be partitioned into two groups, metrics that directly operate on the perspective image pixels and metrics that are applied in the BEV (see also Fig. 1).

Segmentation-based approaches (Fig. 1a) usually perform pixel-level evaluation in the perspective space. Metrics include the classical true positive (TP) and false positive (FP) rates on the pixel/patch level [20], [21], [22], the accuracy [6] as well as precision/recall and the derived F-measure [7], [10], [18]. In order to capture also traffic participants, Alvarez et al. [18] propose to incorporate vehicle detections into the evaluation measure. More recently, also the evaluation of pixel-level correctness in the BEV space has been carried out [16].

Approaches providing lane boundaries [23], [24], see Fig. 1b, are traditionally evaluated via the distance of the estimated lane marking to the ground truth borders in the image. By allowing a flexible margin for counting successful border candidates, TP and FP rates can be obtained [24], [25], [26], [27]. The metric deviations of the borders of the road using a segmentation approach in the BEV space are evaluated in [8]. Besides the lane boundaries, the unoccupied lane length (Fig. 1c) is important for ADAS and has been evaluated in [9], where the distance to the preceding vehicle is combined with lane information. Another way of measuring the road detection performance is the overall corridor width (Fig. 1d). While the boundary positions provide the width only implicitly and might be subject to some underlying lane model, the width is especially relevant for inner-city driving with sudden congestions. Evaluation measures that focus on the lane width have been presented in [8], [16].

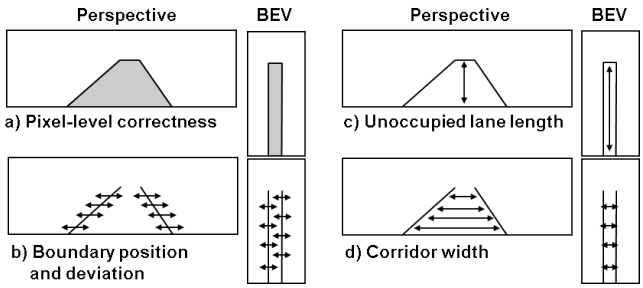


Fig. 1: Visualization of state-of-the-art evaluation metrics.

Each of the outlined evaluation measures has its advantages. Especially the pixel-level measures on the perspective image domain are intensively used to evaluate and improve image processing algorithms. While this results in a direct mapping between pixel processing and pixel evaluation, pixel-based evaluation in the BEV space is much more suitable for any type of driver warning or vehicle control.

Consequently, we propose to focus on evaluations in the BEV domain. Towards this goal, we also present a novel behavior-based measure (Section IV-B) that better accounts for the ultimate task of navigation. We further present a large dataset of challenging scenarios with accurately labeled ground truth for *road area* and *ego-lane* detection. While existing approaches are often evaluated on different (often non-public) datasets, we make our data and evaluation methodology public and target a fair and comprehensive comparison of state-of-the-art methods that can be permanently accessed via our evaluation server. We bootstrap our benchmark with three recent approaches and draw initial conclusions. We believe that our efforts will spur novel interest on this subject and motivate researchers to evaluate their methods on a common basis.

III. The KITTI-ROAD Dataset

The KITTI-ROAD dataset consists of 600 frames (375×1242 px) extracted from the KITTI dataset [19] at a minimum spatial distance of 20m. The recordings stem from five different days and contain relatively low traffic density, i.e., the road is often completely visible. All data (color stereo images, Velodyne laser scans and GPS information) is made available on the KITTI website. We split the data into three sets (see Table I), each representing a typical road scene category in inner city. Fig. 2 depicts some example images. For each category we created a training set of $\sim$100 annotated images and a test set of $\sim$100 images with held-out annotations for evaluation via our website. Results on the test set can be evaluated using the KITTI evaluation server.

[1] http://www.cvlibs.net/datasets/kitti/

978-1-4799-2914-613/$31.00 ©2013 IEEE

TABLE I: Dataset statistics of the KITTI-ROAD dataset.

abbreviation	# train	# test	description
UU	98	100	urban unmarked
UM	95	96	urban marked two-way road
UMM	96	94	urban marked multi-lane road
URBAN	289	290	all three urban subsets

Fig. 2: Example test images from the different categories of the KITTI-ROAD dataset. Note the high variability in our dataset.

Fig. 3: Example polygonal annotation of *road area* (blue) and *ego-lane* (green) in an image from the UM dataset.

For each image we manually annotate the *road area*. In addition, for the UM dataset we also annotate the *ego-lane*[1]. Initial annotation has been carried out in the perspective view (see Fig. 3). We have further refined all annotations in the BEV space, which allows for a higher precision, e.g., by considering a constant road width for distant locations that differ only by a couple of pixels in the perspective image. Note, however, that obtaining an exact pixel-level annotation is a difficult task as many ambiguities remain such as road area which is visible underneath a car or leaves covering the road. For easy access, the annotation files are made available as label images in BEV space (see Section V). A development kit containing functions to map between the BEV space and the perspective image is available on the website.

IV. Performance Evaluation

In order to judge the quality of road detection algorithms for use in automotive applications, we propose to carry out all evaluations in the BEV. In this paper we also provide evaluations in the perspective space in order to illustrate the differences stemming from the choice of evaluation space. We assume that detection results are available as confidence maps (or binary maps) in either space. The transformation between image domain and BEV is obtained by fitting a road plane to the disparity maps via RANSAC[2]. The pixel-based evaluation is carried out for both *road area* and *ego-lane* detection results while the proposed behavior-based performance metric is only applicable to *ego-lane* detection results.

A. Classical Pixel-based Metrics

Similar to [7], [10], [18], we employ the F-measure derived from the precision and recall values (Eq. 1-3) for the pixel-based evaluation. We make use of the harmonic mean (F1-measure, $\beta = 1$), while an unbalanced F-measure using a different weighting of precision and recall could also be applied. In addition, we evaluate accuracy [6].

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

$$F\text{-measure} = (1 + \beta^2)\,\frac{\text{Precision Recall}}{\beta^2\,\text{Precision} + \text{Recall}} \tag{3}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

For methods that output confidence maps (in contrast to binary road classification), the classification threshold τ is chosen to maximize the F-measure, yielding F_{max}:

$$F_{\text{max}} = \underset{\tau}{\text{argmax}}\ \ F\text{-measure} \tag{5}$$

Furthermore, in order to provide insights into the performance over the full recall range, the average precision (AP) as defined in [28] is computed for different recall values r:

$$\text{AP} = \frac{1}{11} \sum_{r \in 0, 0.1, \ldots 1} \max_{\tilde{r}:\tilde{r} > r} \text{Precision}(\tilde{r}) \tag{6}$$

Considering both measures provides insights into an algorithm's optimal (F_{max}) and average (AP) performance. Precision-recall curves are employed to compare different algorithms over the complete range of confidence values.

Note that we explicitly refrain from any weighting of boundary pixels, as we consider the pixel-based metric to give only a rough indication of performance. Depending on the targeted application, we expect that other metrics, like the one proposed in the next section, capture the actual performance much better.

B. Behavior-based Metrics

Classical pixel-based metrics measure the quality for all pixels of a class. As discussed earlier, this might not be adequate if we are interested in the performance with respect to the relevant goals (behavior) of a road vehicle, e.g., following the lane. While lane keeping assistance systems

[1]For the other datasets no *ego-lane* annotation could be provided as in UU no objective annotation guidelines could be identified and UMM contains many ambigous situations due to ongoing lane changes when approaching a crossing with multiple turn lanes. Especially for UU, we expect that vehicle control has to operate directly on an abstraction of the *road area* detection.

[2]The development kit contains a conversion tool to map perspective results into BEV for upload to the webserver.

978-1-4799-2914-613/$31.00 ©2013 IEEE

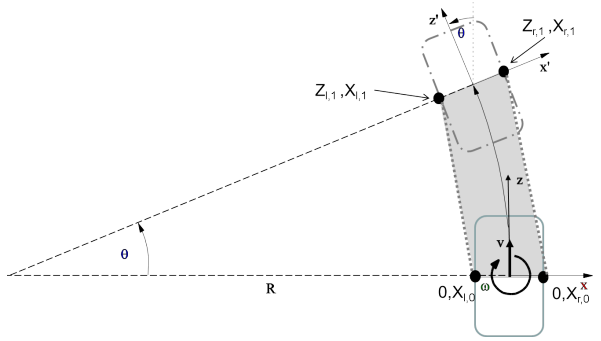


Fig. 4: Evaluation of a single track model for one time step Δt with driven corridor depicted in gray. Note that we represent the road coordinate system using (x, z)-coordinates, similar to the coordinate system of a forward facing camera when viewed from above.

require a very detailed lane shape detection due to the high velocities on highways, in this paper, we argue that for supporting lane keeping behavior in city traffic it is sufficient if an algorithm provides an *driving corridor* hypothesis within the annotated ground truth boundaries.

The basic concept for the behavior-based metric is to fit a number of corridor hypotheses to the lane confidence map in BEV. A single track model is used to generate different driving corridor hypotheses. The area underneath each corridor hypothesis is used to calculate a fitness value by integrating the covered continuous-valued confidence values. The corridor hypothesis with the highest fitness value, i.e. covering the most confident *ego-lane* area, is selected as *driving corridor* hypothesis for comparison with the *ego-lane* ground truth. The next paragraphs describe this process in detail.

We generate a spatial corridor hypothesis as follows: A single track model (see Fig. 4) is initialized at time $t = 0$ with vehicle speed v_0 and wheel angle ρ_0. The initial vehicle orientation α is assumed to be $\alpha_0 = 0$. With the distance between the axles of the vehicle L_{vehicle} and using the single track model, we can calculate the vehicle's yaw rate ω_t using Eq. (7) and the turn radius R using Eq. (8). The yaw angle θ_t after Δt can be computed using Eq. (9) and the movement in longitudinal (Δz) and lateral (Δx) direction using Eq. (10). The overall vehicle orientation angle α_{t+1} is updated by adding the yaw angle θ_t in Eq. (11).

$$\omega_t = (v_t / L_{\text{vehicle}}) \sin(\rho_t) \tag{7}$$

$$R = L_{\text{vehicle}} / \tan(\rho_t) \tag{8}$$

$$\theta_t = \omega_t \Delta t \tag{9}$$

$$\begin{pmatrix} \Delta x \\ \Delta z \end{pmatrix} = \begin{pmatrix} R(1 - \cos(\theta)) \\ R \sin(\theta) \end{pmatrix} \tag{10}$$

$$\alpha_{t+1} = \alpha_t + \theta_t \tag{11}$$

The absolute vehicle location $(X_{\text{vehicle},t+1}, Z_{\text{vehicle},t+1})$ is obtained from Eq. (10) via integration. Given the single track model we represent the area covered by a vehicle with width W when moving forward using two polygons $P_{l/r}$ with points $P_{l/r} = (X_{l/r,i}, Z_{l/r,i})$, representing the left and right corridor boundary. They are derived as:

$$\begin{pmatrix} X_{\text{vehicle},t+1} \\ Z_{\text{vehicle},t+1} \end{pmatrix} = \begin{pmatrix} \cos(\alpha_t) & \sin(\alpha_t) & X_{\text{vehicle},t} \\ -\sin(\alpha_t) & \cos(\alpha_t) & Z_{\text{vehicle},t} \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta z \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} X_{l,t+1} \\ Z_{l,t+1} \end{pmatrix} = \begin{pmatrix} X_{\text{vehicle},t+1} - W/2 \cdot \cos(\alpha_t) \\ Z_{\text{vehicle},t+1} + W/2 \cdot \sin(\alpha_t) \end{pmatrix}$$

$$\begin{pmatrix} X_{r,t+1} \\ Z_{r,t+1} \end{pmatrix} = \begin{pmatrix} X_{\text{vehicle},t+1} + W/2 \cdot \cos(\alpha_t) \\ Z_{\text{vehicle},t+1} - W/2 \cdot \sin(\alpha_t) \end{pmatrix}$$

Note that by providing a behavior model consisting of a sequence of velocity values v_t and wheel angle values ρ_t we are able to generate arbitrary vehicle corridor hypothesis in the BEV space. A simple behavior model would be the assumption of constant velocity and wheel angle. While applicable to highway scenarios, inner city behaviors are much more diverse. This rises the question of how many different corridor hypotheses should be generated and matched to the confidence map. Allowing arbitrary parameters results in an exponential number of hypotheses rendering the evaluation intractable. On the other hand, choosing just a few elementary behaviors is not appropriate for city traffic maneuvers as S-curves cannot be represented, for example. Instead, we propose a multi-stage process for generating corridor hypotheses and matching them iteratively to the confidence map of an *ego-lane* detection algorithm.

We start with 2N+1 hypotheses representing basic models for N steering curves of increasing curvature to the left/right or for driving straight. The behavior models are predicted only for a short duration of $S\Delta t$ time steps, mimicking the minimal duration of a steering maneuver. The area of a hypothesis is used to integrate all covered confidence values, providing the fitness of this behavior model. At the end of each iteration, a new set of 2N+1 hypotheses is started, incorporating the vehicle orientation in the starting condition. This results in a total of (2N+1)(2N+1) hypotheses after $2S\Delta t$. Based on the accumulated fitness for each of these models, we prune the hypotheses and keep only the 2N+1 best ones. From the corridor hypothesis end points, the next iteration of hypothesis generation and pruning is carried out. Hypothesis matching is continued as long as at least one corridor element of the current iteration contains a sufficient percentage of successful detections ($\frac{\text{detections}}{\text{area}} > \text{DET}_{\text{Min}}$) and the end of the BEV space has not been reached. Finally, from all $2N + 1$ hypotheses, we select the hypothesis with the highest overall fitness as *driving corridor* detection. This hypothesis is then compared to the ground truth using the standard evaluation metrics introduced in (1)-(5) (see Section V-C). An example result is depicted in Fig. 5 (d/D).

It should be noted that the granularity of maneuvers $S\Delta t$ influences the quality of the corridor hypotheses. If the granulariy is too small, 'holes' in the confidence maps due to, e.g., pot holes result in a curvy track around these. On the other hand, if the granularity is too large, the corridor hypothesis does not follow curvy roads such as S-curves.

978-1-4799-2914-613/$31.00 ©2013 IEEE

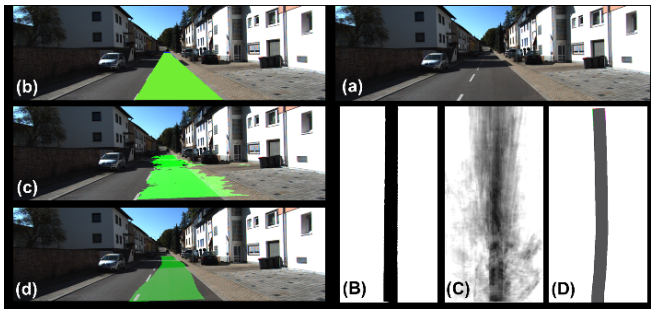


Fig. 5: Example of input image (a) and ground truth annotation in perspective (b) and BEV (B) space. The *ego-lane* confidence map (c/C) is the basis for fitting a track hypotheses (d/D). (Note: map in (c) is thresholded for better visualization). For a behavior-based evaluation, the hypothesis (D) is compared to the ground truth (B) in Section V-C.

V. Experimental Results

In order to demonstrate the different metrics, we provide pixel-based *road area* evaluation results on the complete URBAN KITTI-ROAD dataset using four different methods. The proposed behavior-based evaluation metric is demonstrated on the UM subset, with two *ego-lane* detections methods trained on this subset only.

A BEV representation covering -10m to 10m in lateral (x) direction and 6m to 46m in longitudinal (z) direction is used for evaluation. Per frame we recover the homography between the image and the road plane using RANSAC plane fitting on the 3D measurements of a Velodyne laser scanner which has been calibrated with respect to the cameras [29]. Using a resolution of 0.05m/px, this results in evaluation BEV 'images' of size 800×400 pixels.

A. Evaluated Methods

1) Baseline (BL): In order to provide a lower bound for the performance any road detection algorithm should achieve, we extract baselines for *road area* and *ego-lane* by averaging all ground truth road maps from the training set. This results in confidence maps indicating for each perspective/BEV location the confidence for beeing *road area* or *ego-lane*. These baselines can be viewed as scene priors similar to the one used as input to [10].

2) Geometric Context (GC): The first method we evaluate is geometric context from Hoiem et al. [30], which segments the image into superpixels and estimates a distribution over a set of discrete surface orientations for each superpixel from a single image using a sophisticated set of features, capturing local appearance, but also global cues such as vanishing points. We use the probability map for 'ground' to evaluate road detection performance.

3) Spatial Ray Classification (SPRAY): The second algorithm we evaluate is a two-stage approach that incorporates the spatial layout of the scene [16]. In a first stage, the system represents visual properties of the road surface, the boundary, and lane marking elements in confidence maps based on analyzing local visual features. From the confidence maps

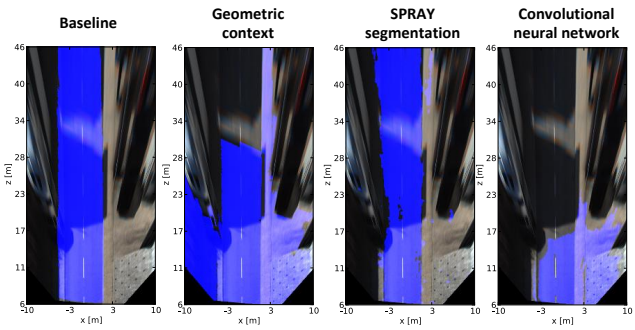


Fig. 6: Classification results for *road area* from Fig. 5.

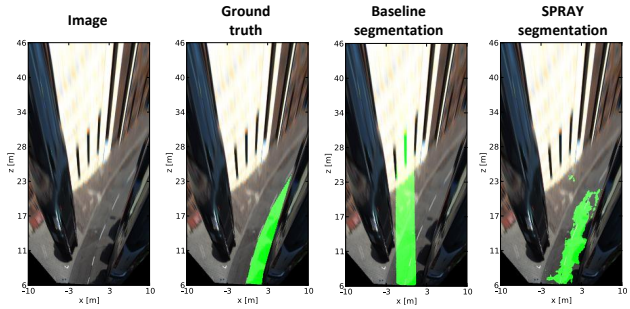


Fig. 9: Classification results for challenging *ego-lane* image.

which are converted into BEV space, SPatial RAY (SPRAY) features that incorporate spatial properties of the overall scene are computed. A boosting classifier trained using the KITTI-ROAD training set provides confidence values. The approach can learn the spatial layout of driving scenes as the features implicitly represent both local visual properties as well as their spatial layout. The method can be trained on both road terrain categories *road area* and *ego-lane*.

4) Convolutional Neural Network (CNN): The fourth method applies a convolutional neural network to label road scene images [6]. It includes a texture descriptor that learns a linear combination of color planes to obtain maximal uniformity in road areas in the test image. The final classification is obtained by combining acquired (offline) and current (online) image information. Note that this algorithm has not been retrained on the KITTI-ROAD dataset but uses the classifier trained on the original dataset [6], resulting in non-optimal performance. Finally, the weights of the color planes for each image have been obtained using the quadratic formulation detailed in [31].

Example processing results for the different evaluated methods for *road area* are depicted in Fig. 6-8 and for *ego-lane* in Fig. 9.

B. Classical Pixel-based Evaluation

The pixel-based evaluation is applicable to both classes, *road area* and *ego-lane*. Table II and Fig. 10 depict the *road area* evaluation results for the URBAN dataset in perspective and BEV space. For the UM dataset, Table III and Fig. 11 depict the *ego-lane* evaluation results (algorithms GC and CNN do not provide this result type).

978-1-4799-2914-613/$31.00 ©2013 IEEE

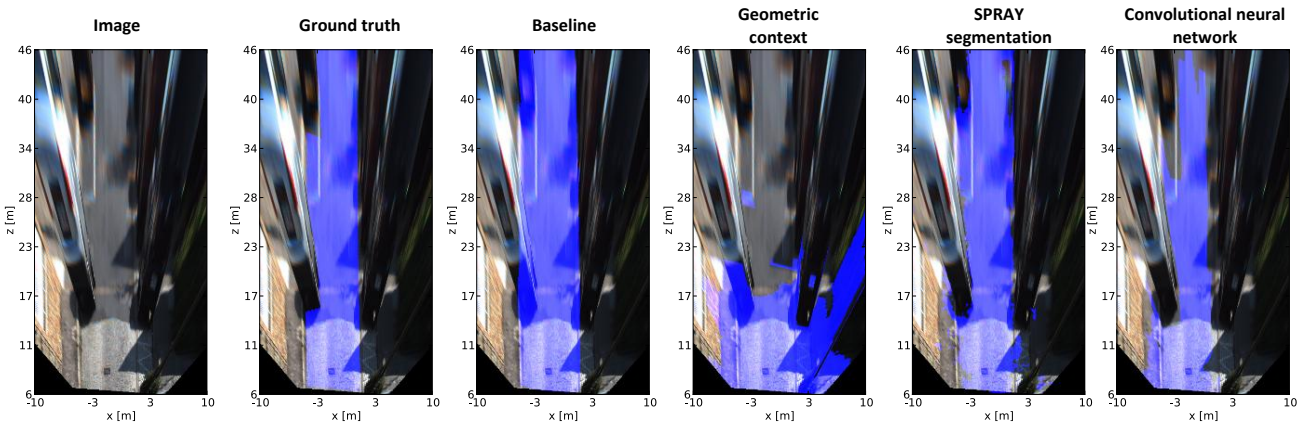


Fig. 7: Classification results for challenging UU *road area* image (Fig. 2 top left).

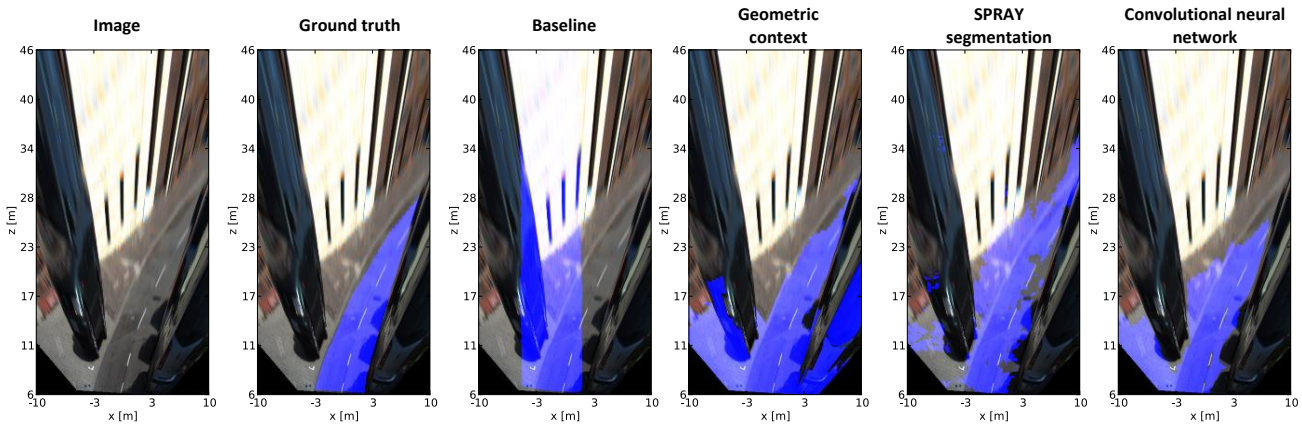
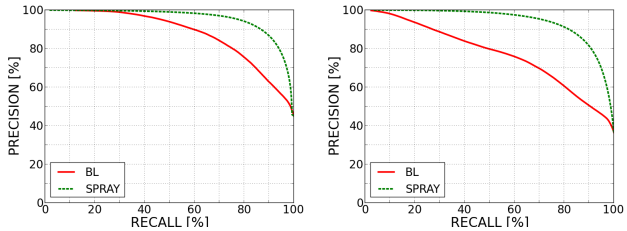


Fig. 8: Classification results for challenging UM *road area* image (Fig. 2 top right).

TABLE II: Results [%] of pixel-based *road area* evaluation.

URBAN - Perspective space						
	AP	$F_{\max}$	Prec.	Recall	Acc	FPR
BL	86.0	77.8	78.1	77.6	80.1	17.9
GC[30]	66.9	68.8	58.2	84.1	65.6	49.7
SPRAY[16]	92.7	88.5	88.5	88.4	89.6	9.4
CNN[6]	82.4	83.4	76.1	92.1	83.4	23.8
URBAN - Metric space						
	AP	$F_{\max}$	Prec.	Recall	Acc	FPR
BL	76.0	70.1	67.6	72.7	77.3	20.1
GC[30]	65.3	64.2	53.1	81.1	67.0	41.1
SPRAY[16]	90.9	86.3	86.7	85.8	90.0	7.6
CNN[6]	78.9	78.9	76.1	81.8	84.0	14.8

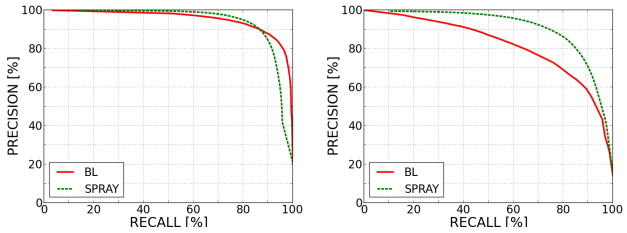


(a) Perspective evaluation (b) Metric evaluation

Fig. 10: Precision-Recall curves for URBAN *road area*.

TABLE III: Results [%] of pixel-based *ego-lane* evaluation.

UM - Perspective space						
	AP	$F_{\max}$	Prec.	Recall	Acc	FPR
BL	89.7	88.9	87.3	90.6	95.3	3.5
SPRAY[16]	90.5	88.3	90.7	86.0	95.2	2.3
UM - BEV space						
	AP	$F_{\max}$	Prec.	Recall	Acc	FPR
BL	78.1	74.4	72.6	76.2	92.5	4.8
SPRAY[16]	87.1	83.9	84.0	83.8	95.4	2.7



(a) Perspective evaluation (b) Metric evaluation

Fig. 11: Precision-Recall curves for UM *ego-lane*.

Note that the presented results are obtained on isolated images. While temporal information and context from, e.g., digital map data will lead to improved detection quality, here

978-1-4799-2914-613/$31.00 ©2013 IEEE

we target at evaluating a very diverse set of challenging images. Note that improving detection on single images will automatically lead to improved performance in the presence of a tracking stage as well.

C. Behavior-based Evaluation

The behavior-based evaluation (see Section IV-B) is performed for N=2 maneuvers ($\rho_0 = [0.05, 0.1], v_0 = 10m/s$) to each side and straight driving. It uses a maneuver duration of $S\Delta t = 300ms$ ($\Delta t = 50ms$), resulting in prototypical corridor primitives for urban driving. We require a corridor area of 50% of each primitive to be covered by positive detections ($\text{DET}_{\text{Min}} = 50\%$). The detections are determined with the threshold used for F_{max} in Table III. We use a track width of 2.2m (typical car width including safety margin) for evaluating the fitness of the individual hypotheses. An example corridor hypothesis is depicted in Fig. 12.

Crucial information for judging the applicability of this concept is the distance up to which corridor detection can be achieved reliably. This information can be extracted by comparing the corridor with the *ego-lane* ground truth and integrating all evaluation results (TP, FP) in the BEV space up to a certain distance[3]. This allows to calculate the precision, which drops if the FP number rises due to a corridor hypothesis leaving the ground truth. This captures mainly lateral deviations but also penalizes hypotheses extending beyond the ground truth. Table IV provides the precision values for different distances (measured from the camera position) that might be relevant for city safety applications. The precision is intended to capture the correctness of the driving width (see Fig. 1d). Fig. 13 (left) depicts the precision over the full distance range, revealing how the integrated precision drops for increasing distances. Note that the baseline ends at roughly 39m while the integrated precision is not affected by the many FNs beyond this distance.

In order to capture the longitudinal aspects (see Fig. 1c), the corridor and ground truth are laterally shrinked to result in a single path line. The line encodes whether there is a corridor / ground truth at a certain distance. Note that for calculating TP, we still require the matched corridor to have an overlap of 2.0m with the ground truth, i.e., we allow for 10% lateral mismatch in the longitudinal evaluation (FP otherwise). We use the F1 measure (3) as it balances between FP (too long corridor) and FN (too short corridor). Again, the values for relevant city distances are listed in Table IV and Fig. 13 (right) depicts the F1-measure over the full distance range, revealing how it drops for increasing distances.

Successful corridor matching is also captured in the hitrate, i.e., the percentage of corridors matched correctly up to the selected distance. The values in Table IV represent the fraction of corridors correctly matched to the ground truth *ego-lane* up to that distance, i.e., with less than 10% lateral mismatch over the complete distance range.

[3]In order to cope with invalid BEV cells close to the vehicle due to the imprecise annotation of the *road area* and *ego-lane* bottom points at the border of the perspective/BEV space and distortions from the dynamic transformation, the integration starts at 9m only.

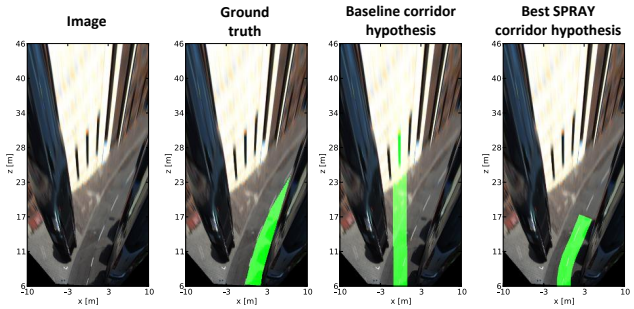


Fig. 12: Example of a corridor hypothesis result.

TABLE IV: Results of behavior-based *ego-lane* evaluation.

	Precision$_{lat}$ [%]			F1$_{long}$ [%]			Hitrate [%]		
	20m	30m	40m	20m	30m	40m	20m	30m	40m
BL	93.9	88.9	82.8	93.4	89.0	82.6	84.0	66.7	0.0
SPRAY	97.9	97.0	97.0	94.1	91.8	88.8	83.0	75.9	47.4

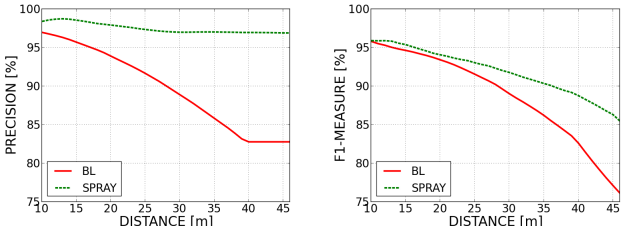


Fig. 13: Integrated lateral precision and longitudinal F1 measure over complete distance range.

D. Discussion

For all algorithms, the pixel-based evaluation results in perspective space are generally higher than in BEV. This is caused by the fact that the near range is more homogenous, i.e., easier to classify, and covers a larger area of the evaluated perspective pixels. In essence, this implies that the evaluation in perspective space is biased and does not reflect the actual performance at regions far away adequately. This is especially prominent in the UM *ego-lane* results in Fig. 11 where BL and SPRAY algorithm deliver rather similar results in perspective space but exhibit strong differences in BEV.

Another important issue is the influence of an unbalanced number of Positives and Negatives in the ground truth. The accuracy measure (Eq. 4) is dominated by the larger group and for the *ego-lane* BEV evaluation this results in extremely high accuracy values of 92.5% (BL) and 95.4% (SPRAY). The high accuracy is due to the large number of correctly classfied TNs, while the precision values have a similar range as for the *road area* on the URBAN dataset.

Using the concept of a *driving corridor hypothesis* in the behavior-based evaluation achieves higher precision values for both, BL and SPRAY. However, the BL precision drops much quicker over distance (see Fig. 13 left), indicating that the exact lateral *ego-lane* shape is missed in far distances. The F1 measure shows that both, BL and SPRAY, reach similar performance in estimating the length of the *ego-lane* in near distances but with stronger decrease for BL at far

978-1-4799-2914-613/$31.00 ©2013 IEEE

distances. However, the hitrate indicates that a substantial number of corridors is missed in both (>0.2m lateral deviation). For BL these are obviously curvy roads and all corridors extending beyond the BL corridor length (∼39m). It is interesting to note that the SPRAY algorithm is capable of extracting the correct corridor in half of the cases (47.4%) while reaching an integrated precision of 97% even at 40m.

An important characteristic of the KITTI-ROAD dataset is that it captures a uniform distribution across scenes, resulting in relatively low traffic density and roads that are mostly straight. Therefore, the baseline often achieves good results compared to the algorithms that have to cope with images containing strong sunlight and shadows. Especially for the UM evaluation, the SPRAY algorithm can consistently outperform the baseline only in the BEV space. As the SPRAY algorithm performs road detection based on features represented in BEV space, we believe its superior performance emphasizes the need for a stronger incorporation of spatial characterstics in road detection algorithms.

VI. Conclusion and future works

In order to stimulate further research, this paper proposes the KITTI-ROAD dataset with images of three challenging real-world city road types derived from the KITTI dataset. We argued for a pixel-based evaluation of *road area* and *ego-lane* in the BEV space in order to capture the fact that vehicle control happens in the 2D road environment. Furthermore, we introduced a novel behavior-based performance metric targeted at evaluating the quality of a highly relevant sub-class of road terrain, the *ego-lane*. The behavior-based measure gives an indication of the usefulness of an *ego-lane* detection approach and is not restricted to classic lane-marking detection methods. The KITTI-ROAD dataset as well as the classic and novel performance measures are made available on the KITTI website. A web interface enables other researchers to benchmark their road detection approaches on any one (or all) of the subsets, advancing the application of road and lane detection algorithms for future driver assistance systems.

References

[1] A. Geiger, M. Lauer, and R. Urtasun, "A generative model for 3d urban scene understanding from movable platforms," 2011.

[2] S. Ishida and J. Gayko, "Development, evaluation and introduction of a lane keeping assistance system," in *Proc. IEEE Intelligent Vehicles Symp.*, 2004, pp. 943 – 944.

[3] J. Siegemund, U. Franke, and W. Forstner, "A temporal filter approach for detection and reconstruction of curbs and road surfaces based on conditional random fields," in *Proc. IEEE Intelligent Vehicles Symp.*, 2011, pp. 637–642.

[4] M. Konrad, M. Szczot, and K. Dietmayer, "Road course estimation in occupancy grids," in *Proc. IEEE Intelligent Vehicles Symp.*, 2010, pp. 412–417.

[5] B. Wang and V. Fremont, "Fast road detection from color images," in *Proc. IEEE Intelligent Vehicles Symp.*, 2013, pp. 1209–1214.

[6] J. M. Alvarez, T. Gevers, Y. LeCun, and A. M. Lopez, "Road scene segmentation from a single image," in *ECCV 2012*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2012, vol. 7578, pp. 376–389.

[7] Y. Kang, K. Yamaguchi, T. Naito, and Y. Ninomiya, "Multiband image segmentation and object recognition for understanding road scenes," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1423–1433, 2011.

[8] T. Kuehnl, F. Kummert, and J. Fritsch, "Monocular road segmentation using slow feature analysis," in *Proc. IEEE Intelligent Vehicles Symp.*, 2011, pp. 800–806.

[9] T. Gumpp, D. Nienhuser, and J. M. Zollner, "Lane confidence fusion for visual occupancy estimation," in *Proc. IEEE Intelligent Vehicles Symp.*, 2011, pp. 1043–1048.

[10] J. M. Alvarez, T. Gevers, and A. M. Lopez, "3D scene priors for road detection," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2010, pp. 57–64.

[11] C. Wojek and B. Schiele, "A dynamic CRF model for joint labeling of object and scene classes," in *European Conference on Computer Vision*, vol. 5305, 2008, pp. 733–747.

[12] U. Franke, H. Loose, and C. Knoeppel, "Lane recognition on country roads," in *Proc. IEEE Intelligent Vehicles Symp.*, 2007, pp. 99–104.

[13] M. Enzweiler, P. Greiner, C. Knoeppel, and U. Franke, "Towards multi-cue urban curb recognition," in *Proc. IEEE Intelligent Vehicles Symp.*, 2013, pp. 902–907.

[14] A. Seibert, H. Haehnel, A. Tewes, and R. Rojas, "Camera based detection and classification of soft shoulders, curbs and guardrails," in *Proc. IEEE Intelligent Vehicles Symp.*, 2013, pp. 853–858.

[15] R. Danescu and S. Nedevschi, "Probabilistic lane tracking in difficult road scenarios using stereovision," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 2, pp. 272–282, 2009.

[16] T. Kuehnl, F. Kummert, and J. Fritsch, "Spatial ray features for real-time ego-lane extraction," in *Proc. IEEE Intelligent Transportation Systems Conf.*, Anchorage, Alaska, USA, 2012, pp. 288–293.

[17] H. A. Mallot, H. H. Bulthoff, J. Little, and S. Bohrer, "Inverse perspective mapping simplifies optical flow computation and obstacle detection," *Biological Cybernetics*, vol. 64, pp. 177–185, 1991.

[18] J. M. Alvarez and A. Lopez, "Novel index for objective evaluation of road detection algorithms," in *Proc. IEEE Intelligent Transportation Systems Conf.*, 2008, pp. 815–820.

[19] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the KITTI vision benchmark suite," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2012, pp. 3354–3361.

[20] Q. Wu, W. Zhang, and B. V. K. V. Kumar, "Example-based clear path detection assisted by vanishing point estimation," in *Proc. IEEE Int Robotics and Automation (ICRA) Conf*, 2011, pp. 1615–1620.

[21] P. Shinzato, V. Grassi, F. Osorio, and D. Wolf, "Fast visual road recognition and horizon detection using multiple artificial neural networks," in *Proc. IEEE Intelligent Vehicles Symp.*, 2012, pp. 1090–1095.

[22] C. Guo, S. Mita, and D. McAllester, "Robust road detection and tracking in challenging scenarios based on markov random fields with unsupervised learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1338–1354, 2012.

[23] M. Serfling, R. Schweiger, and W. Ritter, "Road course estimation in a night vision application using a digital map, a camera sensor and a prototypical imaging radar system," in *Proc. IEEE Intelligent Vehicles Symp.*, 2008, pp. 810–815.

[24] K. Zhao, M. Meuter, C. Nunn, D. Muller, S. Muller-Schneiders, and J. Pauli, "A novel multi-lane detection and tracking system," in *Proc. IEEE Intelligent Vehicles Symp. (IV)*, 2012, pp. 1084–1089.

[25] R. Gopalan, T. Hong, M. Shneier, and R. Chellappa, "A learning approach towards detection and tracking of lane markings," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1088–1098, 2012.

[26] C. Guo, T. Yamabe, and S. Mita, "Robust road boundary estimation for intelligent vehicles in challenging scenarios based on a semantic graph," in *Proc. IEEE Intelligent Vehicles Symp.*, 2012, pp. 37–44.

[27] A. Linarth and E. Angelopoulou, "On feature templates for particle filter based lane detection," in *Proc. IEEE Intelligent Transportation Systems Conf.*, 2011, pp. 1721–1726.

[28] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (VOC) challenge," *Int. J. of Computer Vision*, vol. 88, no. 2, pp. 303–338, June 2010.

[29] A. Geiger, F. Moosmann, O. Car, and B. Schuster, "A toolbox for automatic calibration of range and camera sensors using a single shot," 2012.

[30] D. Hoiem, A. A. Efros, and M. Hebert, "Recovering surface layout from an image," *Int. J. of Computer Vision*, vol. 75, no. 1, pp. 151–172, Oct. 2007.

[31] J. M. Alvarez, M. Salzmann, and N. Barnes, "Learning appearance models for road detection," in *Proc. IEEE Intelligent Vehicles Symp.*, 2013, pp. 423–429.