Honda Research Institute EU
Honda Research Institute Europe GmbH
https://www.honda-ri.de/

Monocular Road Terrain Detection by Combining Visual and Spatial Information

Jannik Fritsch, Tobias Kühnl, Franz Kummert

2014

Preprint:

This is an accepted article published in IEEE Transactions on Intelligent Transportation Systems. The final authenticated version is available online at: https://doi.org/[DOI not available]

Monocular Road Terrain Detection by Combining Visual and Spatial Information

Jannik Fritsch, *Member, IEEE,* Tobias Kühnl, and Franz Kummert, *Member, IEEE*

***Abstract*—For future driver assistance systems and autonomous vehicles, the road course, i.e., the width and shape of the driving path, is an important source of information. In this contribution we introduce a new hierarchical two-stage approach for learning the spatial layout of road scenes. In the first stage, base classifiers analyze the local visual properties of patches extracted from monocular camera images and provide metric confidence maps. We use classifiers for road appearance, boundary appearance, and lane-marking appearance. The core of the proposed approach is the computation of SPatial RAY (SPRAY) features from each metric confidence map in the second stage. A boosting classifier selecting discriminative SPRAY features can be trained for different types of road terrain and allows to capture the local visual properties together with their spatial layout in the scene. In this contribution, the extraction of road area and ego-lane on inner city video streams is demonstrated. Especially the detection of the ego-lane is a challenging semantic segmentation task showing the power of SPRAY features, because on a local appearance level, the ego-lane is not distinguishable from other lanes. We have evaluated our approach operating at 20 Hz on a GPU on a publicly available dataset, demonstrating the performance on a variety of road types and weather conditions.**

***Index Terms*—Automotive computer vision, Road terrain detection, Lane detection, Spatial scene analysis.**

I. Introduction

FUTURE Advanced Driver Assistance Systems (ADAS) are expected to further decrease the number of traffic accidents accompanied by an increase of driving comfort. One very important source of information for any ADAS is the road course, i.e., the width and shape of the driving path. Knowing the complete road terrain allows to predict where the ego-vehicle and other traffic participants will probably move to and where other road users, e.g. cars and pedestrians, can potentially appear [1]. Additionally, the driver can be provided with relevance information ('potential dangerousness') of perceived road users by determining the relative location of road users to the estimated road terrain. Note that the term *road terrain* is used here as general term for different sub-categories like, e.g., *road area* and *ego-lane*.

In current commercially available ADAS, ego-lane information is derived based on lane marking detection methods assuming roads with clearly visible lane markings. Consequently, the operation range of these ADAS like, e.g., *Lane Keeping Assistance* [2] is usually restricted to highway-like situations with certain conditions, e.g., a low curvature of the lane and good visual quality of the lane-markings. However, the robust recognition of the driving path on arbitrary roads - including unmarked roads - will be needed for future ADAS operating in more complex traffic situations, especially in inner city and on rural roads [3]. Some typical example scenes are depicted in Fig. 1.



Fig. 1. Demo images showing the challenges in the urban dataset: situations with low curbstones, unmarked lanes with clear shade contours, and with scattered shades.

If there are no explicit road boundaries (curbstones, lane markings, a.s.o.) detectable due to, e.g., parking cars on the side occluding them, current systems based on delimiter detection [4], [5], [6], [7] are not working. Furthermore, also the appearance of the road itself, i.e., the color and texture of the asphalt, is strongly varying and makes appearance-based road segmentation [8], [9], [10] highly challenging. Most importantly, the separation of the *road area* into different semantic categories (like, e.g., *ego-lane* vs. *non-ego-lane*) is not straightforward using segmentation-based approaches as the visual appearance of the lanes on the local level is rather identical.

The novelty of the presented approach is the improvement of appearance-based classification by incorporating the spatial layout of the scene. To achieve this goal, the system first represents visual properties of the road surface, the boundary, and lane marking elements in confidence maps based on analyzing local visual features. This is similar to other state-of-the-art approaches [11], [12], [13]. The resulting confidence maps are processed further in the second stage operating in the metric space, i.e., the perspective confidence maps are transformed into a Bird's-Eye-View (BEV). In this metric space SPatial RAY (SPRAY) features are extracted that incorporate spatial

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

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

F. Kummert is with the Faculty of Technology, Bielefeld University, Bielefeld, Germany Franz@techfak.uni-bielefeld.de

DOI:10.1109/TITS.2014.2303899 © 2014 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.

properties of the road environment. After this two-staged extraction process, a decision for a road terrain type is taken by applying a boosting classifier that is trained on annotated ground truth data. The proposed approach learns the visuo-spatial properties of driving scenes as the features implicitly represent both local visual properties and their spatial layout.

The proposed two-stage analysis process considering the spatial layout of a scene is useful for any classification task exhibiting a clear spatial correspondence between visual properties at different metric locations. For the task of road terrain detection, this allows the separation of visually similar parts of *road area* into the *ego-lane* and the remaining part of the road. The main novelties of the presented approach are:

- Proposal of spatial ray features for analyzing spatial properties in metric space and over arbitrary distances.
- Individual modeling of visual and spatial characteristics allows separating weather/lighting influences from size variations in road types (highway vs city).
- Approach can be optimized for different types of road terrain (independent of the road delimiter type) by providing appropriately labeled ground truth.

The article is organized as follows: In Section II we review related work on road segmentation and ego-lane detection. Section III provides an overview of the proposed system approach. Section IV describes the base classification of local visual properties and Section V introduces the extraction of spatial properties from the scene using the proposed SPRAY features. The classification of road terrain based on the resulting feature vectors is outlined in Section VI. The evaluation of the proposed approach for *road area* and *ego-lane* detection on real-world data from a publicly available dataset is presented in Section VII. The contribution is concluded in Section VIII.

II. Related Work

For the task of vision-based road segmentation a variety of approaches have been developed, making different assumptions about road terrain:

A. Extraction of Road Delimiters

Identifying the lane by detection of lane markings has a long history (see, e.g., [14]). Today's State-of-the-art approaches extract the delimiting elements of the driving space either for the detection of the *ego-lane* (see, e.g., [4], [5], [15], [6], [16], [7]). or for the detection of the complete *road area* (see, e.g., [17], [18], [19]). Features for these models are extracted from longitudinal road structures like lane markings or road boundary obstacles (e.g. curbstones, barriers) by visual processing. This is mainly based on color and edge appearance [5], [15], [6], [7], 3D information from stereo processing [5], [15], [17], [19] or Structure From Motion [18]. From the extracted features, the road/lane shape can be tracked using different road shape models (see, e.g., [20]). However, especially for inner city the applicability of these approaches is limited because road delimiters cannot be detected that easily (missing/bad lane markings, parked cars occluding curbstones, very low curbstones, ...).

B. Segmentation of Road Area

Instead of delimiter-based road/lane detection, the properties of the complete road surface can be applied for the detection process. Complementary to using 3D information for delimiter detection, the physical property of the road flatness has been used in a variety of approaches [17], [20], [21], [22]. However, this requires the road area to be limited by sufficiently elevated structures.

Consequently, many recent approaches are based on visual properties like, e.g., the mainly gray and untextured asphalt region [9], [10], [11], [12], [13], [23], [24], [25]. These visual properties of the road area have been used for estimating the overall road shape [25] or for segmenting the complete road area [11], [13]. Pixel based classifications, using Conditional Random Fields (CRF), can also be used to segment the complete field of view into individual elements, including the road surface [23]. However, classifying the visual appearance on a local scale only can lead to many ambiguities. Therefore, in [12] it was shown that incorporating a pixel's larger visual context by using multi-scale grid histograms increases the detection quality of all classes.

This contribution introduces a related idea where the focus lies on capturing specific spatial configurations of visual features in a larger metric representation instead of the directly surrounding image parts. Including such a bottom-up spatial context allows for detecting semantic categories like ego-lane, which has previously only been done with delimiter-based approaches.

C. Inclusion of Scene Context

In order to further enhance the robustness of (bottom-up) classification decisions, top-down scene context is often included as explicit prior. This context can be extracted directly from image data like, e.g., the vanishing point [26], the location of horizon line, and the scene category [27]. Additionally, external information sources like map data [18] or data from other sensors [28] can be incorporated. The proposed approach can take advantage of such context information indicating different road types with distinct spatial properties (e.g., width of city road vs. highway). This can be realized by model switching in the training/classification stage and we assume in the following that such a symbolic context information of the road type is available.

III. System Overview

The overall system depicted in Fig. 2 consists of two stages for achieving road terrain detection:

- Several base classifiers capturing local visual appearance properties
- A road terrain classifier based on SPRAY features capturing visuospatial properties

RGB images from the camera are fed into each of the base classifiers. Each base classifier provides one metric confidence map in the 2D BEV driving space for one specific visual property as outlined in Section IV. We propose to use three base classifiers: 'base road classifier', 'base boundary classifier', and 'base lane marking classifier'.

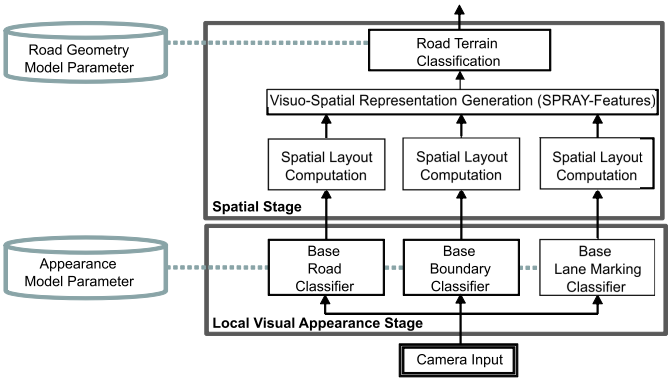


Fig. 2. System block diagram showing the overall system architecture. Note that all base classifiers include preprocessing and inverse perspective mapping to provide metric confidence maps to the spatial layout computation. Training has to be done separately to capture appearance variations and road geometry characteristics.

The SPRAY feature generation described in Section V is carried out for each base classifier and captures spatial aspects of the confidence map. This is done for all base classifiers at the same pre-defined positions in the metric maps which we call *base points*. For each base point all the individual SPRAY feature vectors computed from the different base classifiers are concatenated to obtain a single feature vector. This feature vector forms the input for the road terrain classification of this base point as outlined in Section VI. Performing the classification of many base points and subsequent interpolation realizes a segmentation of the input image.

IV. Base Classification

The task of each of the three base classifiers is the analysis of the input image to extract specific appearance properties and to generate a metric map of confidence values. The base road classifier is specialized to generate high confidences on road-like area and low confidences on non-road terrain. The base boundary classifier is specialized on detecting boundaries between the road-like area and adjacent regions like, e.g., sidewalks, traffic islands, turf. It generates low confidences on road and non-road areas and high confidences at locations that correspond to boundaries. The lane marking classifier is introduced to detect lane markings as these are structurally similar to boundary elements (e.g., curbstones) but can have a varying semantic meaning: they form part of the *road area* but represent the boundary of the *ego-lane*. Its task is to generate a confidence map having high confidences at locations corresponding to lane markings and low confidences on all other terrain (e.g. road-like area, sidewalk, traffic island).

For the base lane marking classifier a dark-light-dark transition detection is applied. The method is similar to standard techniques (see, e.g., [6], [29]) and tuned to obtain only few false negative while having a lot of false positive detections. Multiple filter kernels are applied on the luminance channel mapped into the BEV. Subsequently, thresholding is applied on each filter result to select image regions corresponding to lane markings with a certain width and exhibiting the typical dark-light-dark illumination transition. The final confidence map is

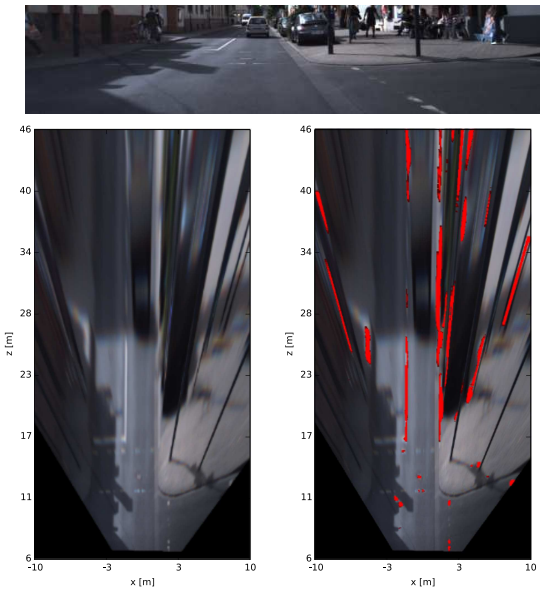


Fig. 3. An example image (left) with the associated transformation to BEV (middle) and the result of base lane marking classification as provided in BEV space (right).

obtained by summing up the binary results and normalizing. This confidence map typically shows high confidences at locations corresponding to lane markings and low confidences on low textured road terrain (e.g., *road area* or sidewalk). An input image with its BEV representation and the base lane marking classification result is depicted in Fig. 3.

The processing for base road and base boundary classifier consists of four steps: Patch extraction, feature extraction, classification, and mapping to metric space. After a patch is extracted from the input image and features are calculated on this patch, a classifier assigns a confidence to this patch. A perspective confidence map is obtained by interpolating patch confidences, indicating for every pixel position how well it matches the appearance model. The perspective confidence map is transformed into the BEV using inverse perspective mapping [30] and subsequently interpolated in order to provide a smooth metric confidence map. This map is the basis for the SPRAY feature calculation described in later Sections and is used in this Section directly for classification. We will use the classification of *road area* as example, as it is intuitive to understand.

We will start by looking at the feature extraction for the base classifiers. For this we first review slow feature analysis as it forms a specific feature type that differs from the standard color and texture features that are outlined shortly in later subsections. All of these features are used as input for the classifier. Evaluation results on a benchmark dataset demonstrate the performance of the individual features.

A. Slow Feature Analysis of temporal signals

Slow Feature Analysis (SFA) is an unsupervised learning technique which provides invariant representations [31]. During the training the algorithm performs an optimization in order to obtain a static transformation from a highly varying

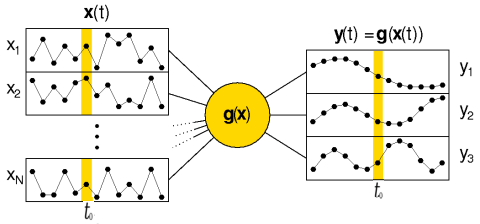


Fig. 4. Schematics of the optimization problem solved by slow feature analysis. Timestep t_0 marked in gray, illustrating the instantaneous transformation.

multidimensional temporal input signal to a slowly-varying output signal. This concept is illustrated in Fig. 4. For vision-based tasks, rapidly changing sensory inputs, namely the pixel values, encode behaviorally relevant visual information like class membership only indirectly. In our patch-based classification system the temporal signal corresponds to the change of pixel values x_i in the image space. By spatially shifting a patch over image-areas that belong to one class, we obtain varying pixel values representing within-class variation. The pixel values represent the function values of the noisy input signal.

In order to easily separate road and non-road input signals in feature space, we need a transformation that creates output signals with low variance from arbitrary input signals belonging to one class. This can be achieved with SFA because it creates a class specific representation for our type of input signals. Additionally it can be used for order reduction, because in general a specified number of slow features can be found that are able to distinguish inputs from different classes [32]. Mathematically we search the set of functions $g_j(x)$ that is generating the slowest varying output functions $y_j(t)$ from a multidimensional input signal $x(t)$:

$$y_j(t) = g_j\left(x(t)\right) \tag{1}$$

Given (1) we can formulate an optimization problem: Finding the transfer function $g_j(x)$ that minimizes the temporal variance of the output signals $\Delta(y_j)$:

$$\Delta(y_j) = \left\langle \dot{y}_j^2 \right\rangle_t \tag{2}$$

We require uncorrelated output signals, having equal variance and zero mean, which leads to the constraints in (3)-(5). Here, (3) forces the output signals to be decorrelated and (4)-(5) exclude trivial solutions.

$$\forall i < j : \left\langle y_i \cdot y_j \right\rangle_t = 0 \tag{3}$$

$$\left\langle y_j^2 \right\rangle_t = 1 \tag{4}$$

$$\left\langle y_j \right\rangle_t = 0 \tag{5}$$

In (2)-(5) $\langle f \rangle_t := \int_{t_0}^{t_1} \frac{1}{t_1 - t_0} f(t)\,\mathrm{d}t$ means averaging the function f over time. For further information on SFA see [31].

B. Training of SFA for Feature Extraction

As mentioned in the previous Subsection, we need to extract patches and serialize the pixel values into signals needed for SFA training. We realize this spatial image sampling by using predefined paths which define how the point of patch

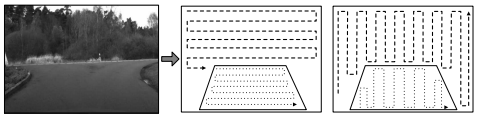


Fig. 5. Example of spatial patch sequence extraction for SFA training: on the left the horizontal path and on the right the vertical path is illustrated. The paths are partitioned into road (dotted) and non-road (dashed) sections.

extraction moves over the image plane. There are two paths, one horizontal p_{hor} and one vertical path p_{ver}, as illustrated in the example in Fig. 5. The advantage of using two paths is that the higher variability of the spatial input signal increases the likeliness of finding a useful transformation.

Given a patch $P_i = f(c_i, a_P)$, defined by its center $c_i = [c_{i,u}\,, c_{i,v}]$ and size $a_P = [a_{P,u}\,, a_{P,v}]$, we can sequentially extract patches by shifting the center c_i along a path $p_{\text{hor|ver}}$ with a constant step size s_p which results in a spatial signal $x_{\text{SFA}}(k_t)$. The spatial index k_t corresponds to t from (1). A signal $x_{\text{SFA}}(k_t)$ is a $d_k \times d_x$ matrix, where d_k describes the number of samples and $d_x = a_P \cdot 3$ is the input dimension of a color image patch. In order to minimize the temporal variance for each class, temporal signals corresponding to road $x_{\text{SFA,R}}(k_t)$ and non-road $x_{\text{SFA,NR}}(k_t)$ are extracted, as it is illustrated in Fig. 5. With ground truth information, given by a binary matrix (road = true), the assignment for every patch along the path can be found by thresholding the number of patch-pixels belonging to the road class. Every patch containing more than 50% of true pixels in the ground truth is interpreted as belonging to the road class. By applying this processing to every training image we can train a model (linear SFA), defined by the transfer function $g(x(k_t))$ (cf. Section IV-A). For training the SFA-TK Toolbox [33] is used.

With the trained transformation, we are now able to extract a slowly varying output signal $y_{\text{SFA}}(k_t)$ for every input image patch P_i. The signal $y_{\text{SFA}}(k_t)$ has the dimension n_{slow} ($n_{\text{slow}} \leq d_x$) which is the number of slow features. In principle it should be sufficient to use the first slowest feature to separate the slowly varying road signal $x_{\text{SFA,R}}(k_t)$ from the rapidly varying non-road signal $x_{\text{SFA,NR}}(k_t)$. But due to noise and additional influences like changes in illumination and appearance (road markings, different surface colors), the classification results improve for multiple slow features (see Section IV-E).

C. Color and Texture Features

Besides the SFA features, we obtain color features based on RGB color space. Initial experiments comparing features from HSV, LUV, and RGB color space for base road classification have shown that RGB features perform best. We use 6 color features for each patch by calculating mean and variance on each color channel. In addition, the gradient within the patch is calculated. For this, the patch is divided in half, once horizontally and once vertically. For each half, the mean and variance is calculated and subtracted from the other half, giving a total of 2 mean gradients and 2 variance gradients per color channel, i.e., 12 additional gradient features. Furthermore, in order to obtain texture features we apply the 2D Walsh Hadamard transform [34] on greyscale images. It has

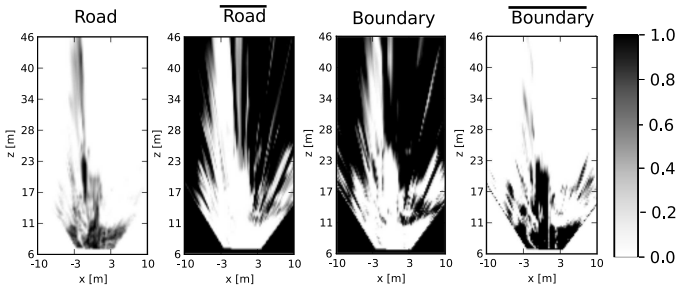


Fig. 6. Result of the base road and base boundary classification for the scene depicted in Fig. 10. From left to right the positive and negative part of the confidence values is depicted for each base classifier. Dark points denote high confidence of the classification.

already been shown that Walsh Hadamard features calculated on color images are useful for road segmentation [35], but we did not find a substantial decrease when only using graylevel images.

D. Classification

For each image patch the feature extraction generates a feature vector $x_{\text{Class},1}$ consisting of SFA features, color features, and Walsh Hadamard texture features. This is used for training a feature model and later in the classification stage to asses the confidence of a new image patch belonging to the base road or base boundary category.

1) Training of GentleBoost: We use the GentleBoost classification method [36] for patch classification, as it has been shown to be very useful for road appearance feature selection and classification [37]. The algorithm generates a sequentially weighted set of weak classifiers that build a strong classifier in combination. In every iteration of the procedure, according to the current distribution of weights on the input signal, the method attempts to find an optimal classifier. After training is finished, the classifier can generate a confidence value $y_{\text{Class},1}$ for a given feature vector, indicating whether the corresponding center position c_i of a patch is likely belonging to the base class or not.

2) Processing phase: In the processing phase, patches are extracted along a horizontal path across the complete image and the feature vector $x_{\text{Class},1}$ is computed. With the trained classifier, we obtain patch-based confidences $y_{\text{Class},1}$. Next, each confidence $y_{\text{Class},1}$ is mapped onto the perspective image plane at the patch center position c_i and an image-based confidence map $conf_1(u,v)$ is created by applying a 2D bi-linear interpolation. The image-based confidence maps $conf_1(u,v)$ are converted to metric confidence maps using inverse perspective mapping [30] and subsequent interpolation. This classification procedure is carried out for the base road and base boundary classifiers, exemplary results of the base classification in the metric space are depicted in Fig. 6. The Figure depicts for each base classifier the positive and negative part of the confidence values provided by the GentleBoost classifier.

E. Evaluation of Base Classifiers

For evaluating the base classifiers in isolation, we use a benchmark dataset with a total of 100 images containing a wide variety of road shapes and weather conditions. A 20-fold cross validation is performed for all evaluations in this Section. The images with 1280×1024 pixels (focal length 5.4 mm) are cropped to 1280×271 pixels to eliminate sky regions and engine hood.

1) Individual Features: In initial experiments, we verified that global color space normalization (zero mean and unit variance for every channel of the RGB image) improves the quality for all features (see also [35], [38]). Therefore, we perform normalization of the images before feature extraction. For patch extraction, a patch size a_P of 21×21 pixels and a step size of $s_p = 10$ pixels is used. The parameter settings for feature extraction have been varied in the experiments. For classifying the resulting features, a GentleBoost algorithm (100 iterations, 4 tree splits) is used. For evaluation we apply the "quality" measure (6) (see also [39]) because it considers all errors, also including false positives FP. We measure both, perspective image space quality Q^P and metric space quality Q^M.

$$Q = \frac{\sum_{i=1}^{n} TP}{\sum_{i=1}^{n} (TP + FN + FP)} \cdot 100\% \tag{6}$$

The metric representation is defined for a range of -10m to 10m in x direction (lateral) and 6m to 46m in z direction (longitudinal). The origin of the coordinate system is under the center of the vehicle's rear axis on the road. This means the metric representation starts roughly 3m in front of the vehicle bumper. With a resolution of 5cm, the metric representation contains 400×800 data points.

For comparing the proposed classifiers with a naive approach, Table I contains the baseline quality generated from ground truth annotations. By summing up all binary ground truth maps and normalizing the result, we obtain a probability map for road likelihood at the individual locations in image/BEV space. Using a threshold of 0.5 provides the baseline classification result. This baseline is, therefore, essentially a scene prior similar to the one used in [27]. All results in Table I are discussed in the following paragraphs. The training quality $Q_{\text{train}}^{P|M}$ measures how well the resulting classifier can handle the *known* training data. The test quality $Q_{\text{test}}^{P|M}$ measures how well *unknown* data can be classified.

SFA Features: Evaluations have been performed by varying the number of features n_{slow} presented to the GentleBoost classifier. The perspective quality Q_{test}^{P} continues to increase for higher number of features but the metric quality Q_{test}^{M} reaches a maximum for 30 slow features.

Walsh-Hadamard Features: The evaluation shows that perspective and metric quality increases with growing feature order but for higher feature orders the benefit becomes smaller.

RGB Features: While already the simple 6 RGB features show a good performance, a further increase is obtained by adding the 12 gradient values for the sub-patches. Note that both color feature sets achieve better performance than any SFA and WH feature set alone.

2) Feature Combination: A further performance increase on the benchmark set can be achieved by combining the features. Table II shows the evaluation results for different subsets of feature types as well as for two full combinations.

TABLE I
PERFORMANCE OF INDIVIDUAL FEATURES ON BENCHMARK

	Q^P_{train}	Q^P_{test}	Q^M_{train}	Q^M_{test}
Baseline	-	73.0	-	40.7
# SFA-features				
3	59.7	53.6	52.5	50.1
10	76.2	71.6	62.4	59.4
20	78.0	73.0	63.5	59.6
30	78.8	73.3	**64.2**	**59.9**
40	**79.2**	**73.9**	64.1	**59.9**
# WH-features				
16	73.5	71.2	57.0	54.5
36	76.0	73.5	59.7	56.5
64	76.8	74.2	60.5	57.2
100	**77.7**	**75.0**	**61.1**	**57.4**
# RGB-features				
6	79.0	74.9	62.8	60.0
18	**81.7**	**77.4**	**65.4**	**62.1**

TABLE II
FEATURE COMBINATIONS ON BENCHMARK DATASET.

Feature type	# features	Q^P_{train}	Q^P_{test}	Q^M_{train}	Q^M_{test}
Baseline	2	-	73.0	-	40.7
RGB & SFA	18+30	84.4	78.3	68.0	63.1
WH & SFA	64+30	83.3	77.5	67.3	62.0
RGB & WH	18+64	84.5	79.6	68.8	64.5
RGB & WH & SFA	18+64+30	86.1	**80.2**	**70.3**	64.9
RGB & WH & SFA	18+64+20	86.1	80.1	70.2	**65.0**
RGB & WH & SFA	18+100+20	**86.3**	**80.2**	70.2	64.6

The results show that each different feature type contributes to a performance increase. Picking the classifier with individual feature parameterization that performed best in isolation is, however, not always a good choice for the overall classifier. In combination with other classifiers it can lead to overfitting of the joint classifier. For example, a set of 30 SFA features performed best in isolation, in combination with RGB and WH features the use of 20 SFA features is sufficient and even results in a better performance. A similar effect is observable for WH features, where 64 features seem to be sufficient.

The base classifier configuration for the overall approach are 20 slow features $y_{\text{SFA},20}(k_t)$, a set of 64 WH texture features and 18 RGB features resulting in a feature vector $x_{\text{Class},1}$ with 102 elements.

V. SPRAY FEATURE EXTRACTION

This Section describes our approach to incorporate spatial information based on the confidence maps representing the visual information of the current scene for the task of road terrain classification. In [40] it has been shown that spatial features can be beneficial for classifying biological cell shapes. Features extracted at different locations relative to a base point have also been used for body part recognition [41]. In order to apply these concepts to the task of road terrain detection a ray-like feature approach inspired by [40] has been developed (see also [42]).

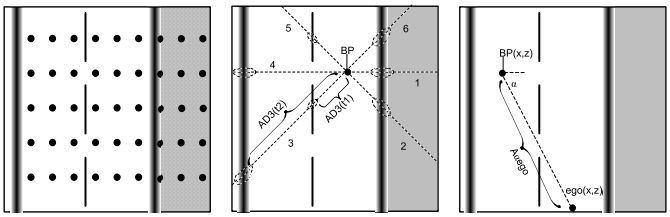


Fig. 7. Distribution of base points in metric space (left), the SPRAY feature generation procedure illustrated for one base point (middle), and the ego SPRAY feature (right).

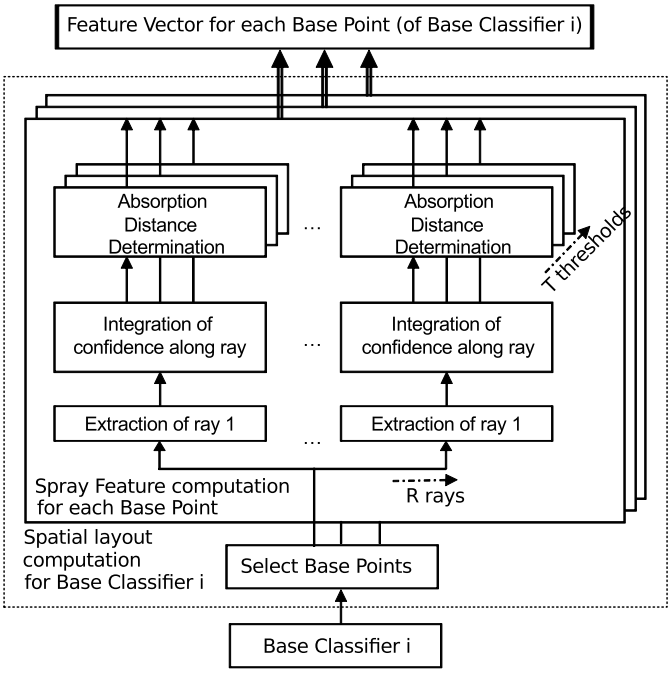


Fig. 8. Block diagram showing the processing steps of spatial feature generation applied for each base point.

A. Individual SPRAY features

The SPRAY feature generation process is carried out for a number of base points (BP). The distribution of base points is defined as a metric grid, as shown in Fig. 7 (left) illustrating a metric view of a two-lane road with lane markings in the center and curbstones on left and right side. Here, dark colors indicate high confidences of an idealized base boundary classifier.

The spatial layout of the confidence map is captured at each individual base point by radial vectors, which are called rays. Fig. 8 shows the processing steps for feature generation at a single base point. A ray R_α starts from a specific base point with angular orientation α and ends at the border of the metric representation. The example in Fig. 7 (middle) shows a base point with six rays numbered clockwise. To extract a defined number of feature values f from a ray, the integral of the confidence values $A_\alpha(\rho)$ along the ray R_α is computed (7). This integral can be interpreted as absorption of confidences along the ray.

$$A_\alpha(\rho) = \int_0^\rho R_\alpha(\gamma)\,\mathrm{d}\gamma \tag{7}$$

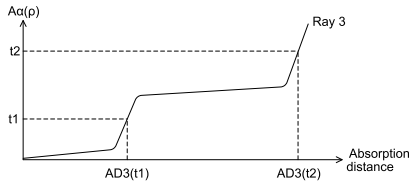


Fig. 9. Integral over the confidences (absorption) for the third ray from Fig. 7. Two SPRAY features $AD_3(t_1)$ and $AD_3(t_2)$ are obtained, which reflect in this case the distance to the lane-marking and the left road border.

$$f_{\text{SPRAY},\alpha}(t_i) = AD_\alpha(t_i) = \underset{\rho}{\operatorname{argmin}} \left(\rho \,|\, A_\alpha(\rho) > t_i\right) \quad (8)$$

By defining a certain number T of absorption thresholds t_i, the absorption distances $AD_\alpha(t_i)$, i.e., the locations where the integral value reaches a certain threshold t_i, are obtained as SPRAY features (8). The graph in Fig. 9 shows a principal sketch of the integral over the third ray of Fig. 7 (center). The thresholds t_1 and t_2 lead to absorption distances that correspond to the distances from the base point to a lane marking $AD_3(t_1)$ and the left curbstone $AD_3(t_2)$ (see also Fig. 7 (center)). The generation of the SPRAY features for a number of R rays with orientation α_r is performed for each base point.

For one specific base point in a confidence map, R rays result in $N_{AD} = R \times T$ absorption distances. These absorption distances, i.e., SPRAY features, serve as input for the road terrain classification. Note that this feature generation is done independently for every confidence map. All SPRAY features for one base point are finally merged to a single feature vector (see Fig. 8).

B. Ego SPRAY Features

In addition to the individual SPRAY features we also use an ego SPRAY feature f_{ego} which provides additional information about a base point's position relative to the ego-lane. For this, a specific ray sent from a base point (x_{BP}, z_{BP}) to the vehicle ego-position (x_{ego}, z_{ego}) in the metric representation is used. The integrated confidence values along the ray are used directly as feature value. This concept is depicted in Fig. 7 (right). The feature value $f_{ego} = A_{\alpha_{ego}}$ can be obtained with (7) after obtaining the angle α_{ego} between BP and ego-position:

$$\alpha_{ego} = \arctan\left(\frac{z_{BP} - z_{ego}}{x_{BP} - x_{ego}}\right) \quad (9)$$

In contrast to the standard SPRAY features the orientation α_{ego} is changing for different base points. This is beneficial for encoding ego-lane specific spatial properties.

VI. Road Terrain Classification

The SPRAY features introduced in the previous Section can be used for classifying a specific road terrain category with dedicated visual and spatial properties. Here, we focus on two types of road terrain that are especially relevant for processing challenging city scenes having both, explicit (lane-markings or curbstones) and implicit (unmarked road) delimiters:

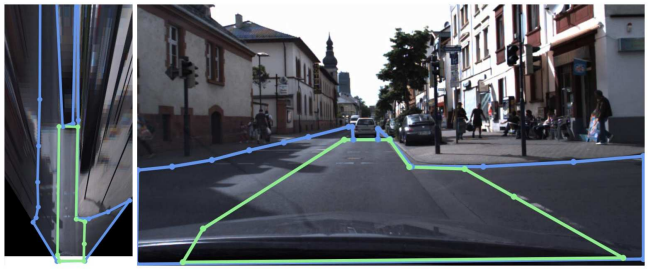


Fig. 10. Example of ground truth for *road area* (blue) and *ego-lane* (green) terrain types annotated in perspective space (left) and transformed in metric space (right).

- *road area*: the complete space where a vehicle can normally drive, i.e., the composition of all lanes
- *ego-lane*: the lane to be followed by the ego-vehicle

Being able to extract the *ego-lane* is an important feature of our approach as *ego-lane* information is already used for commercial ADAS like lane keeping assistance on marked (highway) roads. Our target is to enable this support also for rural roads and inner city driving.

1) Choosing SPRAY parameters: Selecting proper angles for ray vectors R_α and absorption thresholds t_i is highly significant for the performance of the system. In Fig. 9, ideal example thresholds are depicted encoding the distance to relevant scene elements. On real data, however, the confidence maps are noisy and the scene geometry varies, making the choosing of suitable thresholds more challenging.

Based on experiments, the SPRAY feature generation uses eight ray orientations $\phi = [-30, 0, 30, 90, 150, 180, 210, 270]$ ($\phi = 0$ is to the right, counting clockwise) for which five suitable absorption thresholds $th = [3, 10, 30, 70, 120]$ have been identified. Together with one ego SPRAY feature we obtain a total of $N_{AD} = 41$ SPRAY features for each base point for a single confidence map. We calculate for each confidence map one SPRAY feature vector per base point. Since base road and base boundary classifier each provide positive and negative confidence maps and base lane marking classifier a positive confidence map only, this gives a total of $N_{Total} = 5 \times N_{AD} = 205$ SPRAY features per base point. Base points on the metric map are distributed on a grid with a step size of S_{step} pixels.

2) Training: For training we again apply a GentleBoost classifier with 100 decision trees as weak learners (4 tree splits). Ground truth for the road terrain target class (see Fig. 10) is needed to train the road terrain classifier. The training is performed on base classifier confidence maps obtained on test data. This assures that the SPRAY features used for training are representative for unseen data. Consequently, any data set has to be split into three parts: 1) base classifier training. 2) base classifier testing & road terrain classifier training. 3) road terrain classifier testing.

3) Processing: Once the road terrain classifier is trained, the system outputs for each base point the confidence for the trained category. Exemplary segmentation results after interpolating and thresholding of the confidence values for *road area* and *ego-lane* are given Fig. 11.

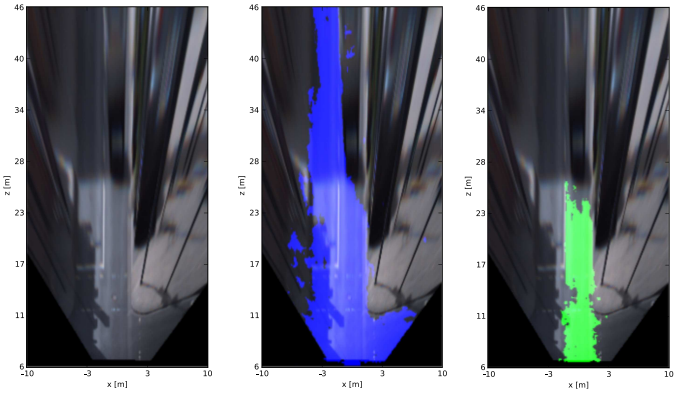


Fig. 11. Result of the road terrain classification showing the BEV representation of Fig. 10 (left) and the classification result for *road area* (middle) and *ego-lane* (right). The corresponding base classifier results are those shown in Fig. 6.

VII. Evaluation

For the evaluation of the overall approach we measure the *road area* and *ego-lane* detection performance on inner-city streams recorded while driving through the inner city of Offenbach, Germany. In order to judge the quality of the proposed approach for use in an automotive application, we will carry out evaluations in the perspective space as well as in the metric bird's-eye-view. For a comparison to state-of-the-art, the reader is referred to the evaluation[1] on the KITTI-ROAD benchmark [43].

A. Input Data & Parameter Settings

Input data consists of three rounds of driving at three different weather conditions (overcast, sunny, and mixed), giving a total of nine rounds. The round track covers a mixture of main roads and small streets present in a German city (Offenbach). The images (1280×1024, camera focal length 5.4 mm) were recorded at 20 Hz frame rate. For training and evaluation the images are cropped excluding the sky and the hood of the ego-vehicle (see Fig. 1), resulting in images with 1280×271 pixels. Depending on traffic, each round took 9-15 minutes and resulted in a total of 10000-18000 frames. Annotations of *road area* and *ego-lane* have been done manually, labeling one frame for every 8 seconds[2]. This resulted in a total of 742 annotated frames (66-114 frames per round). The metric representation covers -10m to 10m in lateral (x) direction and 6m to 46m in longitudinal (z) direction (see Fig. 11), resulting in a BEV with 800×400 px.

The whole system is implemented to run on a GPU using OpenCL. On an NVidia GTX 580, the base classifiers process the RGB images and extract confidence maps ($\sim$12ms) which are handed to the SPRAY feature calculation ($\sim$16ms) and the final road terrain classification ($\sim$17ms). The complete processing pipeline requires $\sim$45ms, sufficiently fast for achieving 20 fps.

[1]Results are available at *http://www.cvlibs.net/datasets/kitti/eval_road.php*

[2]The used image data and annotations can be obtained by sending an e-mail with subject 'InnerCity dataset' to *Jannik.Fritsch@honda-ri.de*.

B. Evaluation Measures

For the evaluation of *road area* and *ego-lane* detection results a variety of evaluation measures can be used. Similar to [12], [27], [44], we employ the F-measure derived from the precision and recall values (10-12) for the pixel-based evaluation. As there is no concrete application, 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. Furthermore, the accuracy (13) is a measure frequently used in road segmentation tasks [11].

$$\text{Precision} = \frac{TP}{TP+FP} \tag{10}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{11}$$

$$F\text{-measure} = (1+\beta^2)\,\frac{\text{Precision Recall}}{\beta^2\,\text{Precision} + \text{Recall}} \tag{12}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \tag{13}$$

The threshold TH for SPRAY feature classification (i.e., obtaining TP,FP,TN,FN) is chosen such that a maximal F-measure (F_{max}) is obtained (14).

$$F_{\text{max}} = \underset{\text{TH}}{\text{argmax}}\, F\text{-measure} \tag{14}$$

Furthermore, in order to provide insights into the performance over the full recall range, the average precision (AP) as defined in [45] is computed for different recall values r:

$$\text{AP} = \frac{1}{11} \sum_{r \in 0, 0.1, \ldots 1} \max_{\tilde{r}:\tilde{r}>r} \text{Precision}(\tilde{r}) \tag{15}$$

Considering both measures provides insights into an algorithm's optimal (F_{max}) and overall (AP) performance. A graphical impression of the overall performance can be gained using precision-recall curves. Traditionally, such evaluations have been carried out in the complete perspective space [11], [12], [44]. In order to focus the evaluation on the driving task, we limit the evaluation of the perspective space to the area covering the BEV road space (ignoring sky, a.s.o.) and, more importantly, we also apply all metrics in BEV space.

C. Evaluation Results

For each weather condition, the first round is used for training the base classifiers and the second round is used for generating confidence maps of unseen data with the trained base classifiers. On these confidence values the SPRAY features are trained. The third round is then used for testing the full road terrain classification on unseen data. By iterating through the three rounds of each weather condition we perform a three-fold cross validation.

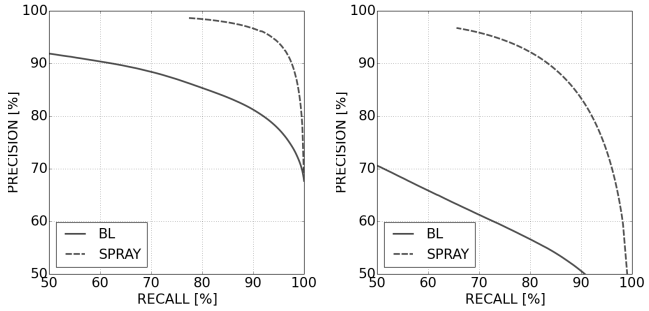
In a first experiment, the contribution of the individual base classifiers combined with SPRAY features for *road area* detection is evaluated (see Table III). Similar to Section IV we generate the baseline from ground truth. For comparison, the performance of the best base classifier feature combination from Section IV is shown, this represents the performance level of a purely appearance-based classification approach.

TABLE III
RESULTS OF ROAD AREA DETECTION ON INNER CITY DATASET.

	Q^M_{train}	Q^M_{test}
Baseline		43.0
base classifier C18WH64SF20	65.6	62.1
SPRAY features trained on		
base road	80.0	70.2
base boundary	79.4	71.6
lane markings	45.9	40.3
base road + base boundary	84.3	73.6
all base classifiers	**85.4**	**74.5**

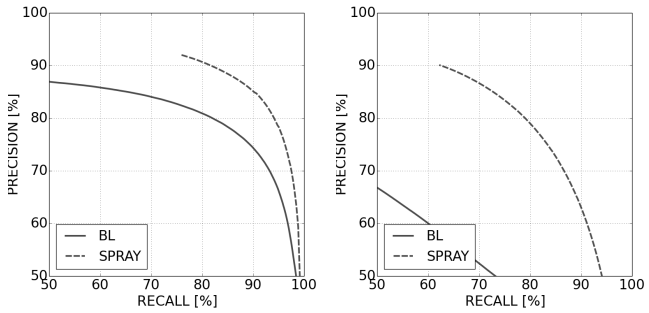
TABLE IV
RESULTS OF PIXEL-BASED EVALUATION.

perspective road area							
	AP	F_{max}	Prec.	Recall	Acc	FPR	Q_{test}
BL	89.1	85.6	79.4	92.8	78.9	50.4	74.8
SPRAY	95.6	94.5	94.0	95.0	92.5	12.8	89.5
metric road area							
	AP	F_{max}	Prec.	Recall	Acc	FPR	Q_{test}
BL	70.0	66.3	56.4	80.5	68.1	39.7	49.6
SPRAY	89.8	87.0	87.1	86.9	89.9	8.2	77.0
perspective ego-lane							
	AP	F_{max}	Prec.	Recall	Acc	FPR	Q_{test}
BL	80.1	81.7	76.4	87.7	90.2	9.0	69.1
SPRAY	85.2	87.6	84.7	90.6	93.6	5.4	77.9
metric ego-lane							
	AP	F_{max}	Prec.	Recall	Acc	FPR	Q_{test}
BL	61.7	60.3	56.6	64.6	92.5	4.8	43.2
SPRAY	78.9	79.5	79.6	79.4	96.4	2.0	66.0



(a) Perspective evaluation (b) Metric evaluation

Fig. 12. Precision-Recall curves for *road area*.



(a) Perspective evaluation (b) Metric evaluation

Fig. 13. Precision-Recall curves for *ego-lane*.

The proposed SPRAY features always reach a higher detection quality using either a single base road/boundary classifier or a combination of base classifiers. Note that lane marking detection in itself is not satisfying, but as additional cue it leads to a further improvement.

In a second experiment, the evaluation is performed using the combination of all base classifiers for the two road terrain types *road area* and *ego-lane* (Table IV). Here, a more powerful configuration of the GentleBoost classifier was chosen (250 decision trees, 4 tree splits). The respective precision-recall curves are depicted in Fig. 12 and Fig. 13. The pixel-based evaluation results (AP, F-measure) in perspective space are generally higher than in metric space. 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 pixels. In essence, this implies that the evaluation in perspective space is biased and does not reflect the actual performance in far distance adequately. Another important issue is the influence of the unbalanced number of Positives and Negatives in the ground truth. The accuracy measure (13) is dominated by the larger group and especially for the *ego-lane* evaluation this results in extremely high accuracy values. This is due to the large number of correctly classified Negatives, while the precision values for the few Positives are in a similar range as for *road area*.

Exemplary segmentation results are depicted in Fig. 14 and Fig. 15. Fig. 14 shows that even with a monocular camera, the separation of *road area* from areas separated by low curbstones can be achieved, although there is a tendency for FP on parking areas and sidewalks that look similar to the road. The *road area* does exhibit FN if there are many scattered shades.

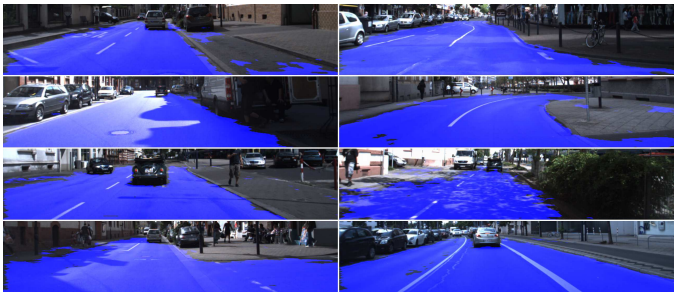


Fig. 14. Exemplary results for *road area* for the images from Fig. 1.

The *ego-lane* results in Fig. 15 show similar FP and FN characteristics. Note the ability of the SPRAY features to separate the own lane from other scene parts even in the absence of lane markings.

All results are obtained using three-fold cross-evaluation. Inspecting the results for the individual rounds/folds exhibits a good repeatability, i.e., the individual values for AP and F_{max} are within $\pm$ 1% of the overall result. Note that the presented classification results are obtained on single frames without including temporal information. A temporal integration of the metric results by considering the ego-motion (see, e.g., [46]) would lead to improved evaluation results. Furthermore, using a road sign detection method [47] as additional information source in the base road classification would allow to further

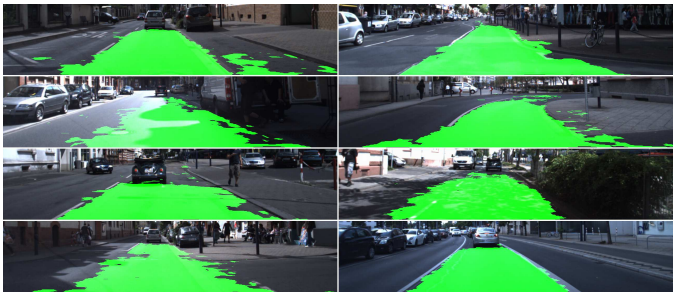


Fig. 15. Exemplary results for *ego-lane* for the images from Fig. 1.

raise the quality of the overall system. This stems from the fact that the base lane marking classifier cannot detect road signs well as it is optimized for lane structures. This results in road signs partly getting high base boundary confidences that are not counteracted by base road confidences, as road signs are visually different from typical road appearance. As a result, road signs are often classified as non-road although they make up part of the road.

From the pixel-based evaluation results obtained on individual images we argue that the metric performance of *road area* is sufficient for using it as context cue. For example, it can be used to support/boost car detection results at positions close/on the *road area* [1]. Besides using the pixel-level results directly as a kind of support map, also more abstract representations of the extracted information are required for real applications. For example, the pixel-based evaluation is not suitable for discussing the benefit for a lane-oriented ADAS application which classically requires the lane boundary positions instead of a segmented area. In order to achieve such a kind of representation, the confidence maps in BEV can be filtered. Subsequently, some kind of clothoid lane model (see Section II) or a behavior-oriented driving model [43] can be fitted to the data. This abstraction from the low-level pixel classification is then also well suited for performing temporal integration based on the application requirements.

VIII. Conclusion

In this paper, we introduced a novel approach combining visual and spatial information to enhance local classification decisions. The proposed SPRAY features capture the geometric characteristics of road environments over larger spatial areas. Through applying machine learning techniques for training the classifiers, the approach can be tuned for extracting different road terrain types from road scenes. The evaluation of *road area* and *ego-lane* detection on a dataset with several weather conditions has shown that this approach can handle various scenes such as roads without lane-markings and varying asphalt appearances. The pixel-based evaluation carried out in the metric BEV space provides the first step towards bridging the gap between image processing and vehicle control. The second step is the identification of suitable algorithms for generating parametric representations that can be used for vehicle control in ADAS, which we are currently investigating. We plan to extend our approach for application in different road scenarios (e.g., highway and inner-city) and to perform an automatic switching for these different scenarios. This is necessary mainly for improved performance of the SPRAY features, as different road environments exhibit different spatial properties (lane width, minimal curve radius). Through these extensions the presented two-stage approach will become applicable to most road scenes encountered during arbitrary day-time driving.

Acknowledgment

The authors gratefully acknowledge the anonymous reviewers for their comments as well as the colleagues at the Honda Research Institute Europe GmbH for many fruitful discussions.

References

[1] T. Weisswange, B. Bolder, J. Fritsch, S. Hasler, and C. Goerick, "An integrated adas for assessing risky situations in urban driving," in *Proc. IEEE Intell. Veh. Symp.*, 2013.

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

[3] C. Guo, J. Meguro, Z. Kojima, and T. Naito, "CADAS: a multimodal advanced driver assistance system for normal urban streets based on road context understanding," in *Proc. IEEE Intell. Veh. Symp.*, 2013, pp. 228–235.

[4] J. McCall and M. M. Trivedi, "Video-based lane estimation and tracking for driver assistance: survey, system, and evaluation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 1, pp. 20–37, 2006.

[5] R. Danescu and S. Nedevschi, "New results in stereovision based lane tracking," in *Proc. IEEE Intell. Veh. Symp.*, 2011, pp. 230–235.

[6] 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.

[7] A. Linarth and E. Angelopoulou, "On feature templates for particle filter based lane detection," in *Proc. IEEE Intell. Transp. Syst. Conf.*, 2011, pp. 1721–1726.

[8] G. K. Siogkas and E. S. Dermatas, "Random-walker monocular road detection in adverse conditions using automated spatiotemporal seed selection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 527–538, 2013.

[9] J. M. Alvarez and A. M. Lopez, "Road detection based on illuminant invariance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 1, pp. 184–193, 2011.

[10] 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.

[11] 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.

[12] 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.

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

[14] M. Bertozzi and A. Broggi, "Real-time lane and obstacle detection on the GOLD system," in *Proc. IEEE Intell. Veh. Symp.*, 1996, pp. 213–218.

[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] J. Choi, J. Lee, D. Kim, G. Soprani, P. Cerri, A. Broggi, and K. Yi, "Environment-detection-and-mapping algorithm for autonomous driving in rural or off-road environment," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 974–982, 2012.

[17] 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 Intell. Veh. Symp.*, 2011, pp. 637–642.

[18] M. Darms, M. Komar, and S. Lueke, "Map based road boundary estimation," in *Proc. IEEE Intell. Veh. Symp.*, 2010, pp. 609–614.
[19] 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 Intell. Veh. Symp.*, 2012, pp. 37–44.
[20] M. Konrad, M. Szczot, and K. Dietmayer, "Road course estimation in occupancy grids," in *Proc. IEEE Intell. Veh. Symp.*, 2010, pp. 412–417.
[21] A. Wedel, U. Franke, H. Badino, and D. Cremers, "B-spline modeling of road surfaces for freespace estimation," in *Proc. IEEE Intell. Veh. Symp.*, 2008, pp. 828–833.
[22] M. Okutomi and S. Noguchi, "Extraction of road region using stereo images," in *Proc. Fourteenth Int Pattern Recognition Conf*, vol. 1, 1998, pp. 853–856.
[23] 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.
[24] T. Gumpp, D. Nienhuser, and J. M. Zollner, "Lane confidence fusion for visual occupancy estimation," in *Proc. IEEE Intell. Veh. Symp.*, 2011, pp. 1043–1048.
[25] U. Franke, H. Loose, and C. Knoeppel, "Lane recognition on country roads," in *Proc. IEEE Intell. Veh. Symp.*, 2007, pp. 99–104.
[26] 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.
[27] 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.
[28] 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 Intell. Veh. Symp.*, 2008, pp. 810–815.
[29] T. Veit, J.-P. Tarel, P. Nicolle, and P. Charbonnier, "Evaluation of road marking feature extraction," in *Proc. 11th Int. IEEE Conf. Intelligent Transportation Systems ITSC 2008*, 2008, pp. 174–181.
[30] 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.
[31] L. Wiskott and T. Sejnowski, "Slow feature analysis: Unsupervised learning of invariances," *Neural Computation*, vol. 14, no. 4, pp. 715–770, 2002.
[32] M. Franzius, N. Wilbert, and L. Wiskott, "Invariant object recognition with slow feature analysis." in *Proc. Conf. on Artificial Neural Networks (ICANN)*, ser. Lecture Notes in Computer Science, vol. 5163. Springer, 2008, pp. 961–970.
[33] P. Berkes, "SFA-TK: Slow feature analysis toolkit for matlab (v.1.0.1)," 2003. [Online]. Available: http://itb.biologie.hu-berlin.de/berkes/software/sfa-tk/sfa-tk.shtml
[34] B. Fino and V. Algazi, "Unified matrix treatment of the fast walsh-hadamard transform," *Computers, IEEE Transactions on*, vol. 100, no. 11, pp. 1142–1146, 1976.
[35] Y. Alon, A. Ferencz, and A. Shashua, "Off-road path following using region classification and geometric projection constraints," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2006, pp. 689–696.
[36] Y. Freund and R. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, pp. 119–139, 1997.
[37] Y. Sha, X. Yu, and G. Zhang, "A feature selection algorithm based on boosting for road detection," in *Proc. Conf. Fuzzy Systems and Knowledge Discovery*, vol. 2, 2008, pp. 257–261.
[38] T. Kuehnl, "Road terrain detection for advanced driver assistance systems," Ph.D. dissertation, University of Bielefeld, 2013.
[39] T. Michalke, R. Kastner, M. Herbert, J. Fritsch, and C. Goerick, "Adaptive multi-cue fusion for robust detection of unmarked inner-city streets," in *Proc. IEEE Intell. Veh. Symp.*, 2009, pp. 1–8.
[40] K. Smith, A. Carleton, and V. Lepetit, "Fast ray features for learning irregular shapes," in *Proc. IEEE Int Computer Vision Conf*, 2009, pp. 397–404.
[41] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2011, pp. 1297–1304.
[42] T. Kuehnl, F. Kummert, and J. Fritsch, "Spatial ray features for real-time ego-lane extraction," in *Proc. IEEE Intell. Transp. Syst. Conf.*, 2012, pp. 288–293.
[43] J. Fritsch, T. Kuehnl, and A. Geiger, "A new performance measure and evaluation benchmark for road detection algorithms," in *Proc. IEEE Intell. Transp. Syst. Conf.*, 2013, accepted.
[44] J. M. Alvarez and A. Lopez, "Novel index for objective evaluation of road detection algorithms," in *Proc. IEEE Intell. Transp. Syst. Conf.*, 2008, pp. 815–820.
[45] 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, Jun. 2010.
[46] T. Michalke, R. Kastner, J. Fritsch, and C. Goerick, "A generic temporal integration approach for enhancing feature-based road-detection systems," in *Proc. IEEE Intell. Transp. Syst. Conf.*, 2008, pp. 657–663.
[47] T. Wu and A. Ranganathan, "A practical system for road marking detection and recognition," in *Proc. IEEE Intell. Veh. Symp.*, 2012, pp. 25–30.



Jannik Fritsch (M 09) received the Dipl.-Ing. degree in electrical engineering from Ruhr University Bochum, Bochum, Germany, in 1996 and the Ph.D. degree and the venia legendi (Habilitation) in computer science from Bielefeld University, Bielefeld, Germany, in 2003 and 2012, respectively.

In 1998, he joined the Applied Computer Science Group, Bielefeld University. In 2004, he joined the EU Project COGNIRON (The Cognitive Robot Companion), where he headed the integration efforts for the Key Experiment Robot Home-Tour. Since 2006, he has been a Principal Scientist with the Honda Research Institute Europe GmbH, Offenbach am Main, Germany. His research interests include image processing methods for environment perception, spatial representations, and cognitive system concepts for intelligent automotive systems.



Tobias Kühnl received the Dipl.-Ing. degree in electrical engineering and information technologies from the Technical University of Darmstadt, Darmstadt, Germany, in 2010 and the Dr.-Ing. degree in computer science from Bielefeld University, Bielefeld, Germany, in 2013.

Since 2010, he has been with the Research Institute for Cognition and Robotics (Cor-Lab), Bielefeld University, working in the fields of road terrain detection for advanced driver assistance systems, in collaboration with the Honda Research Institute Europe GmbH, Offenbach am Main, Germany. His further research interests include computer vision systems, image processing, and machine learning.



Franz Kummert (M 91) received the Dipl.-Ing. and Ph.D. (Dr.-Ing.) degrees in computer science from the University of Erlangen-Nrnberg, Erlangen, Germany, in 1987 and 1991, respectively, and the venia legendi (Habilitation) in computer science from Bielefeld University, Bielefeld, Germany, in 1996.

From 1987 to 1990, he was with the Pattern Recognition Group (Institut fr Informatik, Mustererkennung), University of Erlangen-Nrnberg. Since 1991, he has been with the Applied Informatics Group (Angewandte Informatik), Bielefeld University, where he has been an Applied Professor in pattern recognition since 2002, and the Dean of Studies of the Faculty of Technology since 2003. He has published various papers in these fields, and he is the author of a book on the control of a speech-understanding system and on the automatic interpretation of speech and image signals. His fields of research include speech and image understanding and the applications of pattern understanding methods to natural science domains.

Prof. Kummert has been a member of the Senat of Bielefeld University since 2004.