Adaptive Multi-Cue Fusion for Robust Detection of Unmarked Inner-City Streets

Thomas Michalke, Robert Kastner, Michael Herbert, Jannik Fritsch, Christian Goerick

2009

Preprint:

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

Adaptive Multi-Cue Fusion for Robust Detection of Unmarked Inner-City Streets

Thomas Michalke◇, Robert Kastner⋆, Michael Herbert⋆, Jannik Fritsch◇, Christian Goerick◇

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

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

***Abstract*— First vision-based approaches for detecting the drivable road area on unmarked streets were introduced in recent years. Although most of these visual feature-based approaches show sound results in scenarios of limited complexity, they seem to lack the necessary system-inherent flexibility to run in complex cluttered environments under changing lighting conditions. Our proposed architecture relies on four novel approaches that make such systems more generic by autonomously adapting important system parameters to the environment. As the presented results show, the approach allows for robust road detection on unmarked inner-city streets without manual tuning of internal parameters. The described system was implemented in C relying on the Intel Performance Primitives library and proved its real-time capability. It will be a sub-module of an advanced driver assistance architecture, which runs in real-time on a test vehicle.**

Keywords: driver assistance, robust path identification, lane detection

I. INTRODUCTION

The importance of driver assistance systems for further decreasing the number of traffic accidents is a widely acknowledged fact. The growing complexity of tasks, which these Advanced Driver Assistance Systems have to handle, leads to complex systems that use information fusion from many sensory devices and incorporate processing results of multiple modules. One important field of interest for said systems are applications like, e.g., the "Honda Intelligent Driver Support System" [1] supporting the driver to stay in the lane and to maintain a safe distance from the car in front. Other systems focus on collision avoidance based on autonomous steering and braking (see, e.g., [2]) as well as path-planning even in unstructured environments (see, e.g., [3]). All these applications need a robust detection of the drivable road area. The more safety-relevant applications become, the more the required quality of the detected drivable road area must be improved. As "drivable road area" we define the space in front, which the car can move on safely in a physical sense, but without taking symbolic information into account (e.g., one-way-street, traffic signs).

In this paper, we present a robust system approach for detecting the drivable road area on unmarked roads in inner-city. The proposed system reliably detects the road in complex scenarios by adapting its internal parameters autonomously. Unlike other approach no scene-dependent manual adaptation of system parameters is required.

II. RELATED WORK

Initial approaches for lane detection on marked roads date back to the 1990s (see [4] for an overview of the early approaches). These to date commercially available systems are restricted to marked roads with a course predictable by a clothoid lane model that is also used for road construction of motor-ways. In recent years, the field of research for road detection has shifted to unmarked country roads and inner-city streets. To this end, current prototype systems evaluate and fuse different visual features. In the following, the structure of such visual feature-based systems is analyzed. It is shown that despite the large number of existing road detection systems some important techniques for increasing the road detection robustness are not considered so far.

Image training regions: Current road detection approaches often use street training regions in front of the car in order to parameterize the probability distributions that describe the road feature characteristics (e.g., [5], [6], see also Fig. 3a). Only very few approaches partially incorporate information of non-road image regions to improve road detection (e.g., [7], [8]). However, to our knowledge no approach uses the full potential of non-road information, e.g., for the autonomous adaptation of internal system parameters and the dynamic online assessment of the cue quality, as it is done in our system.

Features: Typical visual features for road detection in state-of-the-art systems are: texture (edge density) on the intensity [9], [7], stereo disparity [10], [6], HSI color [7], [5], [6], or depth from Lidar / Radar [11]. Many system approaches use the feature edge density (structure) on the intensity map. However, edge density on further feature maps is so far not considered. To our knowledge, no approach uses the edge density on color maps for road detection. During the evaluation of our system, we experienced the edge density on color maps as a very robust cue for detecting the road.

Feature granularity: Numerous system approaches rely on probabilistic methods for classifying street and non-street pixels (e.g., [7], [12]). Such iconic (i.e., pixel-based) approaches do not include information of the surround of a pixel, but handle all pixels independently. Nevertheless, discontinuities in feature maps often contain important information that allow improved scene decomposition (e.g.,

curbstones separating road from sidewalk). Other approaches stress the importance of region-based information and use region growing or vertical filling (e.g., [5], [13]). Such approaches are often sensitive to changes in lighting conditions causing large gradients in the feature maps (e.g., shadows on the road). Both, iconic and region-based approaches have important advantages that partially compensate their drawbacks. However, to our knowledge, no system approach for road detection uses both approaches to the same extend.

Road modeling: Many of the recent feature-based systems use road models of varying complexity that support the feature-based road detection (e.g., [7], [12] use clothoids, [10] distinguishes between left, right, and straight street course, [14] uses second order polynomials). For country roads and highways such approaches seem to yield sound results. Nevertheless, as further discussed in Section III, we claim that said rigid street models are not flexible enough to robustly run on inner-city streets that often show abrupt changes in their course as well as occlusions of significant parts of the drivable road area. However, some kind of road model seems to be necessary in order to improve robustness of the road detection. This dilemma can be resolved by relying on a generic and flexible road model that makes only simple assumptions about the course of the road. One of the few system approaches that follows this idea is presented in [15]. The authors point out that the road area typically covers between 30 to 85% of the image. The feature thresholds are adapted in order to reach this ratio. Unfortunately, the proposed approach is restricted in its flexibility, since the ratio is set offline without constantly adapting it to match the current characteristics of the scene.

To sum up, existing state-of-the-art road detection systems are marked by a limited flexibility, which restricts their application to country roads and highways. In order to allow reliable road detection in more complex inner-city scenarios, we propose four novel techniques to enhance robustness and system-inherent flexibility by enabling adaptation to the environment. To our knowledge, a combination of these techniques has not been used for road detection before.

In detail, these techniques are:

- Using street *and* non-street training regions (see Fig. 3) that both adapt the feature probability distributions,
- Using edge density (structure) feature, computed on the HSI hue and saturation maps,
- Combining iconic *and* region-based feature processing,
- Fusing feature-based road detection with a dynamic and generic road model.

In the following section, details of our road detection system embedding these four techniques are given. The presented system approach is not restricted to inner-city streets, but was tested on country roads and highways as well.

III. SYSTEM DESCRIPTION

In the following, the realized system architecture for unmarked road detection is described (see Fig. 1). It relies on our four novel techniques that enhance the system-inherent flexibility. After giving a rough overview on the individual processing steps, all system modules are described in detail.

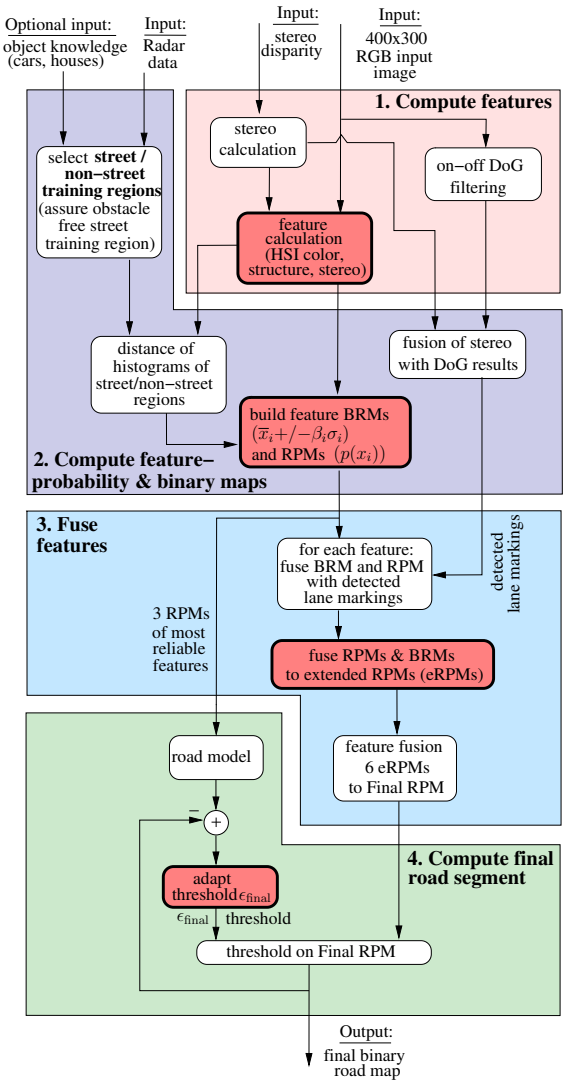

Fig. 1. System overview: Adaptive road detection system (bold modules contain novel techniques).

Our system takes RGB images, stereo disparity (from two parallel cameras), and Radar data as input. Knowledge about previously detected objects in the scene can be used as optional input. The system detects the road based on six robust features that are evaluated and fused in a probabilistic way. For this step, street and non-street training regions are defined in the input image. In parallel, the system detects present lane markings with a biologically motivated filter approach. The lane markings are fused with the detected road segments. In the final step, a binary road map is computed relying on a road model that adapts itself to the environment.

Next, the system is described in more detail. In the first step, different features are calculated on the 400x300 pixel RGB input images. The features we use are saturation and hue of the HSI color space (see, e.g., [16] for details on HSI). Furthermore, we apply the structure tensor in Eq. (1) (with W being a 9x9 region around the current pixel) to compute the edge density E_j (see Eq. (2)) on the hue, saturation, and

intensity of the HSI color space. Typically, the edge density of these feature channels is different for the road and the rest of the scene, which makes it a reliable feature.

$$A_j(u,v)=\begin{bmatrix} \Sigma_W(G_u*F_j)^2 & \Sigma_W(G_u*F_j)(G_v*F_j) \\ \Sigma_W(G_v*F_j)(G_u*F_j) & \Sigma_W(G_v*F_j)^2 \end{bmatrix} \quad (1)$$

with $j \in \{\text{hue, saturation, intensity}\}$

$$G_u(u,v) = -\frac{u}{2\pi\sigma^4} exp(-\frac{u^2+v^2}{2\sigma^2})$$

$$G_v(u,v) = -\frac{v}{2\pi\sigma^4} exp(-\frac{u^2+v^2}{2\sigma^2})$$

$$E_j(u,v) = \frac{det(A_j(u,v))}{trace(A_j(u,v))} \quad (2)$$

Furthermore, vision-based stereo data is used as feature. For computing stereo vision the camera images are rectified in order to facilitate the correspondence search between the two camera images (i.e., the images are remapped, virtually aligning the two camera coordinate systems with the world coordinate system). The thereby necessary intrinsic (i.e., internal camera properties, like the focal length and the principal point) and extrinsic (i.e., external camera properties, like camera angles and offsets) camera parameters were determined using the freely available calibration toolbox [17]. The toolbox was applied on a calibration scene similar to the one described in [18]. There is no dynamic change of the camera pitch angle, since on the one hand the input images are pitch-corrected using a correlation-based method similar to [19]. On the other hand, we assume a flat road, which is present in most inner-city environments. When using the system in an urban environment, the course of the road and hence the camera angles could be estimated using a surface model (e.g., a hyperplane, please refer to [20] for details). The image rectification assures that the camera angles (including the static pitch angle) will not influence the stereo results. The correspondence search (see [21]) yields a disparity map. Based on that, three dense maps containing the 3D-world position for all image pixels can be obtained (see Fig. 2c for the gathered depth map). The stereo data is remapped using the measured camera angles in order to have the pixel positions of stereo maps and the image comparable.

The stereo maps are postprocessed for solving the problem of missing disparity values near to the car (see Fig. 2b). More specifically, during the computation of the stereo disparity no correspondence search is possible at image regions near to the car, since this would come at the cost of high computation time. We solve this problem by searching line-wise for high horizontal gradients in the bird's eye view of the camera image (for information on this representation see [4]) taking only the area directly in front of the car (e.g., first 10 meters) into account (see example in Fig. 2a). It is assured that no objects are present in the said area based on radar and low vertical gradients in the bird's eye view. The area between the found gradients, which mark the road borders, is assumed to be road. The image regions in bird's eye view representation are mapped to the perspective image with a pin hole camera

MODALITY	Cue #	VISUAL ROAD DETECTION FEATURE
Color	1	Hue
	2	Saturation
Structure	3	Edge density on Hue
	4	Edge density on Saturation
	5	Edge density on Intensity
Stereo	6	Height of objects in scene

TABLE I

USED VISUAL FEATURES FOR UNMARKED ROAD DETECTION.

model, which includes the determined intrinsic and extrinsic camera parameters (e.g., static camera angles). Based on the perspectively mapped road regions the three stereo maps are corrected assuming a perfectly flat plane (see corrected depth map in Fig. 2c). Since only the region directly in front of the car is corrected, the error induced by a non-flat road plane is considered as small. However, to eliminate this error the estimated camera angles coming from the optional surface model could be included into the pin hole camera model.

Tests have shown that huge shadows on the road result in poor stereo quality, since the correspondence search gets difficult on dark, noisy image regions. This supports using more cues that are to some extend more invariant to shadows (e.g., HSI color space). Altogether, our system relies on six different cues for road detection (see Tab. I for an overview).

In the second step, binary road maps (BRM) and road

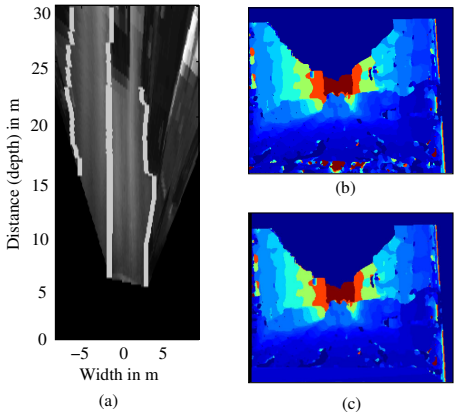

Fig. 2. (a) Gradient-based road search on the bird's eye view, (b) Missing disparity values near to the camera vehicle induce false and missing depth values, (c) Corrected depth map.

probability maps (RPM) for the six feature maps are computed. The BRMs are binary maps that hold "1" for pixels belonging to the detected street and zero for the rest. The six BRMs are calculated with a region-growing algorithm, by which region-related feature properties are incorporated. Opposed to that, the six RPMs contain continuous probability values that assess the "road-likeness" of the feature values for all pixels independently. Both map types rely on the same normal distribution, see Eq. (13) and (14). The parameters of the normal distribution are calculated using a street and at least 2 non-street training regions (see Fig. 3a). Please note that the stereo training region needs to be set beyond the regions of corrected height values (see Fig. 2c). The training regions are adapted dynamically depending on the scene. For example, it is assured that no obstacle is within the training region by incorporating Radar data. Furthermore, the size of

(a) (b) (c)

Fig. 3. (a) Visualization of street and non-street training regions, Structuring element for region growing for (b) Left image half, (c) Right image half.

the street training region is set proportionally to the velocity of the ego vehicle, exploiting that typically no near obstacles exist during fast driving, e.g., on highways. The street and non-street training regions are chosen by considering the height map of the scene derived from the stereo disparity map (see [20] for more details) and existing knowledge about objects in the scene. In the following, the computation of the BRMs and RPMs is described in more detail.

For computing the BRMs a region-growing algorithm, which connects continuous regions in the feature maps is applied (i.e., the surround of a pixel is evaluated). The latter approach is done, in order to get crisp borders between the road and the sidewalks that often have road-like features. The region-growing algorithm recursively sets all pixels that are adjacent to the currently known street segment in BRM$_i$ to "1", when the corresponding pixels in the feature map i are within the confidence interval (see Eq. (3) and Eq. (4)). The region-growing algorithm starts from the road-training region. The normal-distribution-based confidence interval in Eq. (3) uses the feature thresholds $\overline{x}_i$ +/- $\beta_i\sigma_i$, which are independently calculated for all five visual features. Here, the parameter $\overline{x}_i$ is the mean and σ_i the standard deviation of the normal distribution calculated on the street training region. The parameter β_i is introduced to adapt the confidence interval to the current scene properties.

$$\overline{x}_i - \beta_i\sigma_i < x_i < \overline{x}_i + \beta_i\sigma_i \quad \forall i = 1..5 \tag{3}$$

$$\text{with } \beta_i = 4d_i(H_{s_i}, H_{n_i}) \; \forall i = 1..6$$

$$\overline{x}_6 - \epsilon_Y(v) < x_6 < \overline{x}_6 + \epsilon_Y(v) \tag{4}$$

$$\text{with } \epsilon_Y(v) = \beta_6(\sigma_6 - \sigma_q(v_{\text{train}})) + \sigma_q(v) \tag{5}$$

$$d_i(H_{s_i}, H_{n_i}) = \sqrt{1 - \gamma_i(H_{s_i}, H_{n_i})} \; \forall i = 1..6 \tag{6}$$

$$\gamma_i(H_{s_i}, H_{n_i}) = \sum_{\forall x} \sqrt{H_{s_i}(x) H_{n_i}(x)} \; \forall i = 1..6 \tag{7}$$

Different from $\overline{x}_i$ and σ_i, which are calculated on the street training region alone, the threshold parameter β_i changes dynamically depending on the characteristics of the street *and* non-street training regions. More specifically, the parameter β_i, which influences the feature thresholds, is calculated from d_i (see Eq. (6)). The parameter d_i is the distance between the two histograms H_{s_i} and H_{n_i} of the street and non-street training regions for the i=1..6 features. The measure d_i is based on the Bhattacharya coefficient $\gamma_i(H_{s_i}, H_{n_i})$ in Eq. (7), which assesses the similarity of two histograms. Based on β_i the confidence interval in Eq. (3) is adapted. The larger the difference between street versus non-street areas on a feature map is, the bigger the confidence interval. The region growing results are morphologically postprocessed with two different structuring elements for the left and right image half (see Fig. 3b and c) following the typical course of roads in a perspective image.

Different from the five visual cues (hue, saturation, and the three edge density maps), the normal distribution of the stereo height Y also depends on the measured distance to the car. This is empirically plausible, since Y is a function of the stereo disparity $D(u, v)$ and the relative influence of the quantization error of $D(u, v)$ (measured in pixels) grows the smaller $D(u, v)$ and hence the bigger the distance of a road segment is to the car. Hence, the part σ_q of the standard deviation of the stereo height cue that is induced by the quantization error of $D(u, v)$ increases with growing distance to the car. In order to mathematically assess the error propagation of the quantization error of disparity $D(u, v)$ to the stereo height Y their functional relation is required. The stereo height $x_6 = Y$ can be computed using Eq. (8) (with B as the horizontal distance between the stereo cameras, h the camera height, v the vertical pixel position and v_0 the vertical principal point of the camera). Equation (10) defines the standard deviation σ_D of the disparity (measured in pixels), which is induced by the quantization error (the step size Δg is set to 1 pixel). For computing the propagated standard deviation σ_q (required in Eq. (5)), we use Eq. (11) (refer to [16]), which describes how the standard deviation of a random variable (here the disparity $D(u, v)$) is propagated through a function (here $Y(D)$). We are interested in the disparity on road surface D_{surf} alone (see Eq. (9), gathered after reforming Eq. (8) with Y=0). Hence, D_{surf} defines the position at which Eq. (11) is linearized. Here, the vertical pixel position v is a parameter of the distribution. For the quantization-error-induced standard deviation of the height cue Y, we finally find Eq. (12). The hyperbolic form of Eq. (12) confirms the made empirical assumptions. Based on that, the confidence interval ϵ_Y for the stereo height Y Eq. (4) includes the standard deviation $\sigma_q(v)$ that is adapted depending on the current vertical image position v of the current pixel in focus (see Eq. (5)). Besides adding $\sigma_q(v)$ in Eq. (5), the standard deviation σ_6, computed on the training region on the Y map, needs to be corrected by $\sigma_q(v_{\text{train}})$ present at the vertical image position v_{train} of the training region. As result, we now have 6 BRMs for 6 features.

$$x_6 = Y = \frac{B \cdot (v - v_0)}{D(u, v)} - h \tag{8}$$

$$D_{\text{surf}}(v) = \frac{B \cdot (v - v_0)}{h} \tag{9}$$

$$\sigma_D = \frac{\Delta g}{\sqrt{12}} = \frac{1}{\sqrt{12}} \tag{10}$$

$$\sigma_q(v) \approx \sigma_D \left| \frac{dY}{dD} \right|_{D(u,v)=D_{\text{surf}}(v)} \tag{11}$$

$$\approx \sigma_D \left| -\frac{B \cdot (v - v_0)}{[D_{\text{surf}}(v)]^2} \right|$$

$$\sigma_q(v) \quad \approx \frac{1}{\sqrt{12}} \left| \frac{h^2}{B \cdot (v - v_0)} \right| \tag{12}$$

Additionally to the region-based processing for calculating the BRMs, a pixel-based (iconic) processing for computing the RPMs is done (i.e., each pixel is handled independently from its surrounding). All pixel values x_i receive a probability value $p(x_i)$ (see Eq. (13) and (14)), which results in 6 independent Road Probability Maps (RPMs) for the 6 features. The probability distribution for the stereo-based height cue Y Eq. (14) assumes the mean height zero $\overline{x}_6 = 0$ and adapts $\sigma_q(v)$ during the computation of RPM_6 and BRM_6 dependent on the vertical pixel position v. The approach assumes a normal distribution of the six features in the street training region and beyond. As described, for x_6 a position-dependent variance was introduced. We verified the assumed normal distribution for all features independently based on the Kolmogorov-Smirnov (KS) test of goodness of fit with its Lilliefors extension [22]. Due to restrictions in space, the gathered results are not presented in the paper.

$$p(x_i) \quad = \quad e^{-\frac{(x_i - \overline{x}_i)^2}{2\sigma_i^2}} \qquad \forall i = 1..5 \tag{13}$$

$$p(x_6) \quad = \quad e^{-\frac{x_6^2}{2[\sigma_6 - \sigma_q(v_{\text{train}}) + \sigma_q(v)]^2}} \tag{14}$$

In the third step, the computed BRMs and RPMs are fused with the detected lane markings. More specifically, the RPMs for all features are set to a high probability for the detected lane markings. The lane-marking detection is done with the biologically motivated Difference of Gaussian (DoG) kernel (see Fig. 4a), which takes the receptive fields of neurons in the retina as a role model. The DoG filter kernel is adapted to be selective to bright structures on a dark background, the so-called on-off contrasts without reacting to dark structures on a brighter background. Figure 4c shows the filter response on the inner-city frame shown in Fig. 4b. All image regions with on-off contrasts, that have a height within the confidence interval Eq. (4), and that are below the horizon are detected as being lane markings (see Fig. 4d). The separation between on-off and off-on contrasts reduces the number of false positive road marking detections. For example, in [23] the prefiltered images still contain lane marking unspecific off-on contrasts (e.g., traffic signs in front of a bright sky). Such off-on contrasts are filtered out in our approach to improve the road detection performance.

The six iconic RPMs and their respective BRMs are combined by multiplication, which leads to six extended RPMs (eRPM), see Eq. (15). Based on this, the advantage of probability-based computation is preserved. At the same time, discontinuities in the feature maps can be detected. As a result, the advantages of both approaches are combined.

Next, all eRPMs are fused resulting in the final RPM

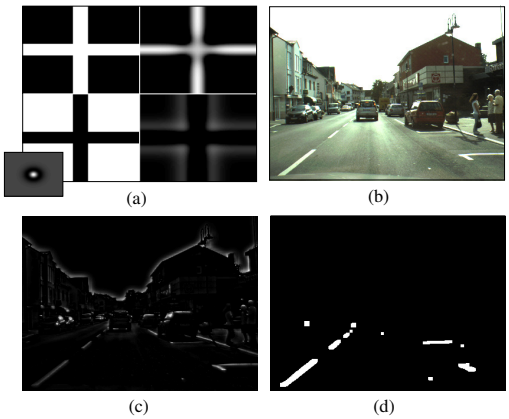

Fig. 4. a) On-off Difference of Gaussian (DoG) filtering on two test images with on-off and off-on contrast (left) as well as the respective filter responses (right), b) Inner-city test frame, c) On-off DoG filter response for bright contrasts on a dark background (with lane markings popping out), d) Detected lane markings (after fusion of DoG and object height from stereo).

(fRPM) using the geometric mean (see Eq. (16)).

$$\text{eRPM}_i = \text{RPM}_i \text{BRM}_i \quad \forall i = 1..6 \tag{15}$$

$$\text{fRPM} = \left(\prod_{i=1}^{6} \text{eRPM}_i \right)^{1/6} \tag{16}$$

$$\text{Final Road Map}(u, v) = \begin{cases} 1 & \forall\ \text{fRPM}(u, v) > \epsilon_{\text{final}} \\ 0 & \text{else} \end{cases} \tag{17}$$

In the forth and final step, the Final Road Map is determined by applying a threshold ϵ_{final} to the fRPM (see Eq. (17)). The threshold ϵ_{final} is set dynamically based on the correlation results of the three currently most reliable features maps, in order to get a prediction of the current relative size of the road versus the rest of the image. For these three features the currently best HSI color feature (hue or saturation), the best structure feature (structure on hue, saturation, or intensity), as well as stereo are selected. For the selection process the Bhattacharya coefficient $\gamma_i(H_{s_i}, H_{n_i})$ is evaluated (see Eq. (7)), by which the separability of street versus non-street histograms H_{s_i},H_{n_i} can be assessed.

Hence, the technique relies on a simple road model (expected fraction of the road area in the current image, termed road-to-image-ratio). No assumptions are made regarding the current position of the road in the image. As our evaluation results in Section IV show, it is of crucial importance to adapt the said expected fraction dynamically to the current scene. This dynamic adaptation enables the system to run robustly in complex scenes, as in inner-city scenarios.

For adapting ϵ_{final} the control loop depicted in Fig. 5 is used. The threshold ϵ_{final} is adapted by a gradient method based on Eq. (22). In the following, the applied procedure is described in detail. It uses the BRMs of the three most reliable feature maps A, B, and C that are combined to the road reference map (i.e., feature product R that represents the expected road area), depicted in Fig. 5. The four binary maps are summed up with Eq. (18), which results in four scalar values $S_{\{A,B,C,R\}}$. The values $S_{\{A,B,C,R\}}$ represent the integral number of pixels detected as road for the three

feature maps and the road reference map.

$$S_X = \sum_{\forall(u,v)} \text{BRM}_X(u,v) \quad \text{with} \quad X \in \{A,B,C,R\} \tag{18}$$

$$\kappa = \frac{\frac{S_R}{S_A} + \frac{S_R}{S_B} + \frac{S_R}{S_C}}{3} \tag{19}$$

Then, the parameter κ (see Eq. (19)) is calculated. It represents the mean percentage with which the three most reliable feature maps correspond to the road reference map R. The larger κ is, the more the features match to each other, i.e., the more similar the three features maps are. The degree of similarity of these features gives a hint about what to expect from the remaining cues and can hence be used to adapt ϵ_{final}. The Final Road Map is computed (see Eq. (17), where ϵ_{final} is set to a typical initial value for bootstrapping) and summed up yielding the scalar value S_{FRM} (see Eq. (20)).

$$S_{\text{FRM}} = \sum_{\forall(u,v)} \text{Final Road Map}(u,v) \tag{20}$$

$$\frac{1}{\kappa} < \frac{S_{\text{FRM}}}{S_R} < 1.2\frac{1}{\kappa} \tag{21}$$

$$\epsilon_{\text{final}}(t) = \begin{cases} \alpha^- \epsilon_{\text{final}}(t-1) & \text{when } \frac{S_{\text{FRM}}}{S_R} < \frac{1}{\kappa} \\ \alpha^+ \epsilon_{\text{final}}(t-1) & \text{when } \frac{S_{\text{FRM}}}{S_R} > 1.2\frac{1}{\kappa} \end{cases} \tag{22}$$

Next, it is checked if the calculated scalar value S_{FRM} fulfills Eq. (21). If the inequality is fulfilled, the Final Road Map is valid. If not, ϵ_{final} is adapted incrementally based on Eq. (22) (with $\alpha^- < 1$ and $\alpha^+ > 1$), until the inequality is fulfilled. Equation (22) is motivated from the well-known RPROP approach. The step sizes α^+ and α^- are adapted using the well-known SuperSAB approach. The processing stops after 100 iterations at the latest.

In the following section, our system approach is evaluated based on an inner-city scenario.

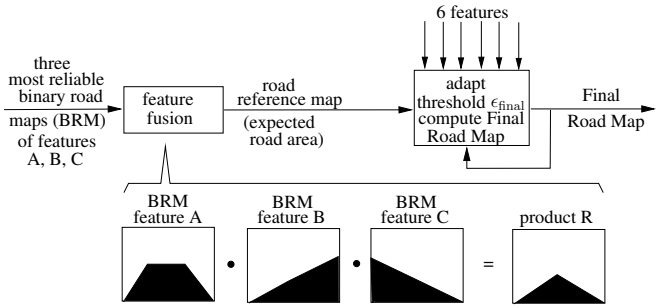

Fig. 5. Control loop to adapt the final road detection threshold ϵ_{final}.

IV. RESULTS

In the following, accumulated road detection results on 440 frames of an inner-city stream are presented. The performance gain reached by incorporating street and non-street training regions as well as the dynamic road model is assessed. In a final step, details of the needed computation time and our test vehicle are given. The inner-city result stream, the input images, and the manually annotated ground truth street segments are available in the internet [24] for benchmark testing.

In order to evaluate our system, we adopt the Equations (23), (24), and (25) on the resulting road segment. The equations define different ground-truth-based measures, which were taken from [10] (with pixels being True Positive (TP), False Negative (FN), False Positive (FP)).

$$\text{Completeness} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{23}$$

$$\text{Correctness} = \frac{\text{TP}}{\text{TP}+\text{FP}} \tag{24}$$

$$\text{Quality} = \frac{\text{TP}}{\text{TP}+\text{FP}+\text{FN}} \tag{25}$$

On a descriptive level the Completeness states, based on given ground-truth data, how much of the present road was actually detected. The Correctness states how much of the detected road is actually road to avoid classifying all as road leading to a Completeness of 100%. The Quality combines both measures, since between the Completeness and Correctness a trade-off is possible. Based on this, the Quality measure should be used for a comparison, since it weights the FP and FN pixels equally. For a more detailed analysis the Completeness and Correctness state what exactly caused a difference in Quality. The necessary ground-truth data was produced by accurate manual annotation of the road in the 440 images.

In order to evaluate the novel techniques, the three measures were calculated on the detected road segments of 440 image frames for three system instances. The first instance is our system as proposed in Section III with all four novel techniques running. The second system instance is equivalent to the first but runs with a constant road-to-image-ratio (i.e., with a rigid road model). The third system is equivalent to the first but uses no non-street training regions, which makes the confidence interval thresholds less adaptive to the environment ($\beta_i = const.$, see Eq. (3)).

We used 220 frames of our inner-city scenario as training data for the two competitive, less adaptive systems in order to tune the road-to-image-ratio of the second system and confidence interval factors β_i of the third system. The accumulated results in Tab. II show that all three systems have a similar performance in terms of Quality on the training data.

Road detect. approaches	# training images	Correctness	Completeness	Quality
Our system	-	**96%**	**75%**	**73%**
Without non-street training areas	220	96%	73%	71%
With rigid road model	220	88%	84%	75%

TABLE II

COMPARISON OF OUR ROAD DETECTION SYSTEM WITH TWO COMPETITIVE SYSTEMS RUNNING WITHOUT TWO OF THE PROPOSED NOVEL TECHNIQUES ON TRAINING IMAGES.

On training images, the highest Quality is reached by the second competitive system that uses a rigid road model. The accumulated results for the training sequence are plausible, since both competitive systems were tuned to run with good performance on the training images. Different from that, our system adapts itself to the environment based on the four described techniques. Therefore, for our system no manual tuning to the training sequence was done.

For the actual evaluation, the two competitive systems were run on consecutive parts of the remaining stream (in sum 220 images) that were used for testing. In a direct comparison between our system and the rigid road model system we could gain the results depicted in Tab. III. Table IV shows the results of the comparison between our system and the system without non-street training regions. In both cases, our system significantly outperforms the competitive systems in terms of Quality (75% compared to 68% and 69% compared to 50%). These results confirm the gain of the system-inherent adaptation capabilities offered by the proposed four techniques.

Road detect. approaches	# test images	Correctness	Completeness	Quality
Our system	120	**97%**	**77%**	**75%**
With rigid road model	120	77%	85%	68%

TABLE III

COMPARISON OF OUR SYSTEM AND AN EQUIVALENT SYSTEM WITH A RIGID ROAD MODEL ON A TEST STREAM WITH NARROW STREET.

Road detect. approaches	# test images	Correctness	Completeness	Quality
Our system	100	**99%**	**68%**	**69%**
Without non-street training region	100	99%	50%	50%

TABLE IV

COMPARISON OF OUR SYSTEM AND AN EQUIVALENT SYSTEM WITHOUT NON-STREET TRAINING REGION FOR A SHADY TEST STREAM.

As Tab. III and IV reveal, the Correctness of found street segments is high, which means a small number of false positive street pixels are found. However, the gathered results show that the detection performance varies between frames. This is due to the changing content of the training region in front of the car. Thereby, the system possibly adapts to local characteristics present in the current training region that might differ from the current global road characteristics. Furthermore, local illumination changes that depend on the current view angle and lighting conditions influence the detection performance. To solve this, the temporal integration method introduced in [20] is used for tracking. The method fuses the current and previous detected road segments based on correlation on the bird's eye view.

Road detect. approaches (BEV: bird's eye view)	**# test images**	**Correctness**	**Completeness**	**Quality**
No temp. integration	440	98.1%	61.5%	60.5%
Temp. integration, BEV	440	**95.2%**	**94.1%**	**89.9%**

TABLE V

PERFORMANCE GAIN BASED ON TEMPORAL INTEGRATION.

As the results in Tab. V show, the highest Quality (89.9% enhancing the 60.5% of the initial street detection algorithm) is reached with temporal integration based on our algorithm. The initial road detection approach without temporal integration has the highest Correctness with 98.1%, but this comes to the cost of reduced Completeness of merely 61.5%. Applying temporal integration decrease this value from 98.1% to 95.2%, but it increases the Completeness disproportionately (from 61.5% to 94.1%).

For further evaluation Fig. 6 shows typical results of our system compared to the two competitive systems and the ground-truth data, based on four sample frames of the inner-city stream (overall stream available at [24]). As can be seen, our system performs better in complex and shady scenes.

For the experiments we use a Honda Legend prototype car equipped with a mvBlueFox CCD color camera from Matrix Vision delivering images of 800x600 pixels at 10Hz, which is hence the processing rate our road detection module should approximately reach. The image data as well as the laser and vehicle state data from the CAN bus is transmitted via LAN to several Toshiba Tecra A7 (2 GHz Core Duo) running our RTBOS integration middleware [25] on top of Linux. The road detection component together with other driver assistance components (see, e.g., [26]) are implemented in C using an optimized image processing library based on the Intel IPP [27]. Table VI shows the computational demands of different submodules of the presented approach running on one of these laptops. The overall computation time of our road detection system currently amounts to 123.5 ms (8.1 Hz), which allows real-time processing on our prototype vehicle.

M	**T**	Used RAM [in MB]	**Comp. time [in ms] (frame rate [in Hz])**
-	-	185	93.5 (10.7)
X	-	203	101.0 (9.9)
-	X	214	105.0 (9.5)
X	X	233	**123.5 (8.1)**

TABLE VI

COMPUTATION TIME (M - INCLUDING DETECTION OF LANE MARKINGS, T - INCLUDING TEMPORAL INTEGRATION APPROACH).

V. SUMMARY

This paper describes a visual feature-based road detection system, which relies on four novel techniques (usage of street *and* non-street image training regions, edge density computed on the hue as helpful feature, iconic *and* region-based feature processing, usage of a dynamic and generic road model) that allow robust road detection by adapting system parameters to the environment. The described techniques could help similar systems to get more flexible and independent of environmental conditions. Based on the proposed system, results are obtained that allow building safety-relevant algorithms, like, e.g., trajectory planning and active collision avoidance even in cluttered inner-city scenarios.

In our future work, we plan to incorporate the proposed approach in our biologically motivated driver assistance system (described in [26], [28]).

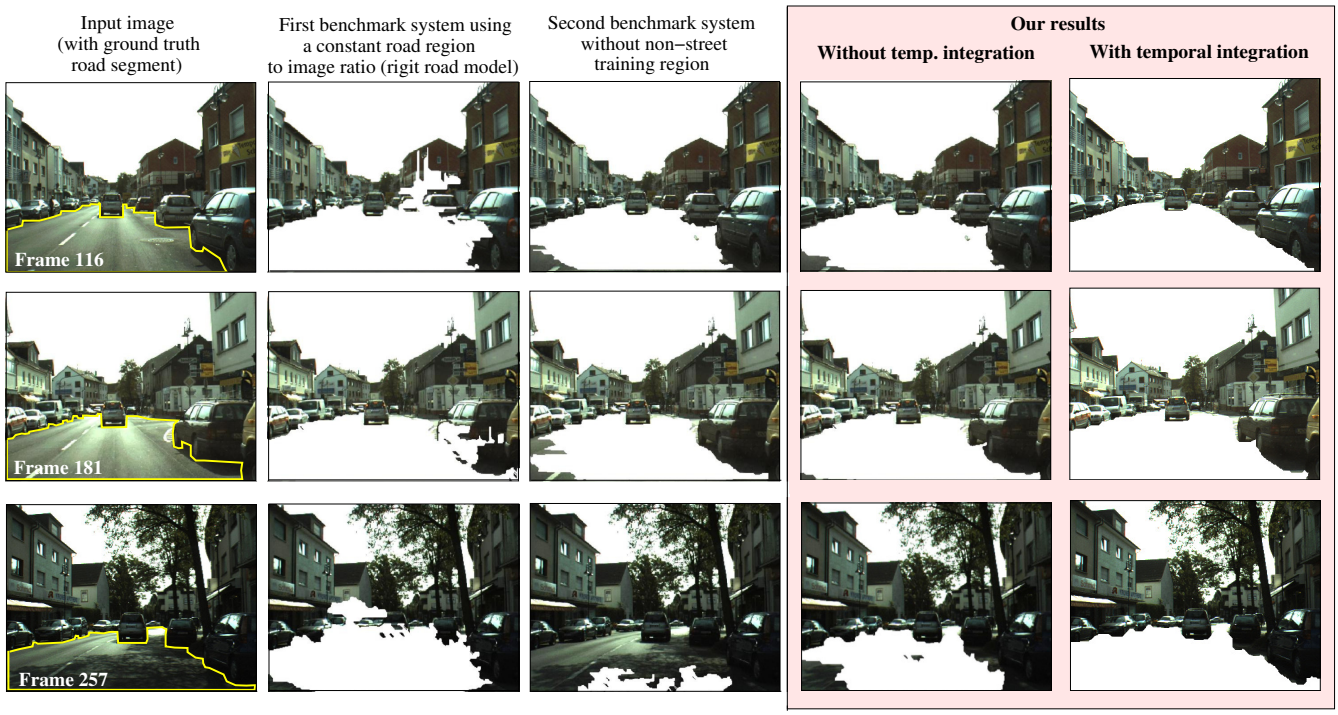

Fig. 6. Example images of the benchmark inner-city stream (First column: Input image with ground-truth road segment, Second column: First benchmark system with rigid road model, Third column: Second benchmark system without non-street training segment, Last columns (highlighted): Resulting road segment of our system improved by temporal integration.

REFERENCES

[1] M. Ikegaya, N. Asanuma, S. Ishida, and S. Kondo, "Development of a lane following assistance system," in *Int. Symp. on Advanced Vehicle Control*, Nagoya, 1998.

[2] M. Schorn, U. Stahlin, A. Khanafer, and R. Isermann, "Nonlinear trajectory following control for automatic steering of a collision avoiding vehicle," in *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2006.

[3] T. Dang, S. Kammel, C. Duchow, B. Hummel, and C. Stiller, "Path planning for autonomous driving based on stereoscopic and monoscopic vision cues," in *IEEE Proceedings of the 2006 American Control Conference*, 2006, pp. 191–196.

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

[5] C. Rotaru, T. Graf, and J. Zhang, "Extracting road features from color images using a cognitive approach," in *IEEE Intelligent Vehicles Sysposium*, 2004.

[6] N. Soquet, D. Aubert, and N. Hautiere, "Road segmentation supervised by an extended vdisparity algorithm for autonomous navigation," in *IEEE Intelligent Vehicles Sysposium*, 2007.

[7] U. Franke, H. Loose, and C. Knoeppel, "Lane recognition on country roads," *Intelligent Vehicles Symposium, 2007 IEEE*, pp. 99–104, 13-15 June 2007.

[8] N. Apostoloff and A. Zelinsky, "Robust vision based lane tracking using multiple cues and particle filtering," in *IEEE Intelligent Vehicles Symposium*, 2003.

[9] T. Hong, T. Chang, C. Rasmussen, and M. Shneier, "Road detection and tracking for autonomous mobile robots," in *Proceedings of SPIE Aerosense Conference*, 2002.

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

[11] C. Rasmussen, "Combining laser range, color and texture cues for autonomous road following," in *IEEE International Conference on Robotics and Automation*, Washington DC, 2002.

[12] O. Ramstroem and H. Christensen, "A method for following unmarked roads," in *IEEE Intelligent Vehicles Sysposium*, 2005, pp. 650–655.

[13] M. Y. Chern and S. Cheng, "Finding road boundaries from the unstructured rural road scene," in *16th IPPR Conference on Computer Vision, Graphics and Image Processing*, 2003.

[14] M. Sotelo, F. Rodriguez, and L. Magdalena, "Virtuous: Vision-based road transportation for unmanned operation on urban-like scenarios," in *IEEE Transactions on Intelligent Trans. Systems*, vol. 5, no. 2, 2004.

[15] X. Lin and S. Chen, "Color image segmentation using modified hsi system for road following," in *IEEE International Conference on Robotics and Automation*, 1991.

[16] B. Jaehne, *Digital Image Processing*. Springer, 2005.

[17] J.Y.Bouguet, "Camera calibration toolbox for matlab," 2007, http://www.vision.caltech.edu/bouguetj.

[18] T. Marita, F. Oniga, S. Nedevschi, T. Graf, and R. Schmidt, "Camera calibration method for far range stereovision sensors used in vehicles," in *IEEE Intelligent Vehicles Sysposium*, 6 2007, pp. 356–363.

[19] A. Broggi and P. Grisleri, "A Software Video Stabilization System for Automotive oriented Applications," in *Procs. IEEE Vehicular Technology Conference*, Stockholm, Sweden, June 2005.

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

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

[22] W. H. Lilliefors, "On the Kolmogorov-Smirnov test for normality with mean and variance unknown." in *Journal of the American Statistical Association*, vol. 62, 1967, pp. 399–402.

[23] T. Luo-Wai, "Lane detection using directional random walks." in *IEEE Intelligent Vehicles Symposium*, Eindhoven, 2008.

[24] BenchmarkData, 2008, http://www.rtr.tu-darmstadt.de/~tmichalk/IV2009_RoadDetectionSystem/.

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

[26] T. Michalke, A. Gepperth, M. Schneider, J. Fritsch, and C. Goerick, "Towards a human-like vision system for resource-constrained intelligent cars," in *Int. Conf. on Computer Vision Systems*, Bielefeld, 2007.

[27] Intel, "Integrated Performance Primitives," 2006, http://www.intel.com/cd/software/products/asmo-na/eng/perflib/ipp/302910.htm.

[28] T. Michalke, R. Kastner, J. Adamy, S. Bone, F. Waibel, M. Kleinehagenbrock, J. Gayko, A. Gepperth, J. Fritsch, and C. Goerick, "An attention-based system approach for scene analysis in driver assistance," *at - Automatisierungstechnik*, vol. 56, no. 11, pp. 575–584, 2008.