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

Towards a Proactive Biologically-inspired Advanced Driver Assistance System

Thomas Michalke, Robert Kastner, 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]

Towards a Proactive Biologically-inspired Advanced Driver Assistance System

Thomas Michalke[◇], Robert Kastner[⋆], 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*— Driver assistance functionalities on the market are getting more and more sophisticated, which will lead to integrated systems that fuse the data of multiple sensors (e.g., camera, Photonic Mixer Device, Radar) and internal system percepts (e.g., detected objects and their states, detected road). One important future challenge will be to find smart solutions in system design that allow an efficient control of said integrated systems. A promising way for achieving this is to get inspiration from known signal-processing principles in the human brain. This paper presents a biologically motivated Advanced Driver Assistance System (ADAS) that uses the generic principle of attention as common front-end of all visual processes. Based on the attention principle an early task-dependent pre-selection of interesting image regions is done, which decreases scene complexity. Furthermore, internal information fusion increases the system performance (e.g., the attention is used to improve the object tracking; road-detection results improve the attention). Based on streams of a challenging traffic scenario it is shown how the system builds up and verifies its environment-related expectations relying on the attention principle. The ADAS is controlled by a central behavior control module that tunes submodules and parameters. The behavior control module has a simple structure, but still allows for robustly performing various tasks, since the complexity is distributed over the system in form of local control loops mimicking human cognition aspects.**

Keywords: advanced driver assistance system, system control, scene decomposition, tracking

I. INTRODUCTION

The goal of realizing Advanced Driver Assistance Systems (ADAS) can be approached from two directions: either searching for the best engineering solution or taking the human as a role model. Today's ADAS are engineered for supporting the driver in clearly defined traffic situations like, e.g., keeping the distance to the forward vehicle. While it may be argued that the quality of an engineered system in terms of isolated aspects, e.g., object detection or tracking, is often sound, the solutions lack necessary flexibility. Small changes in the task and/or environment often lead to the necessity of redesigning the whole system in order to add new features and modules, as well as adapting how they are linked. In contrast, biological vision systems are highly flexible and are capable of adapting to severe changes in the task and/or the environment. Hence, one of our design goals on our way to achieve an "all-situation" ADAS is to implement a biologically motivated, cognitive vision system as perceptual front-end of an ADAS, which can handle the wide variety of situations typically encountered when driving a car. Note that only if an ADAS vision system attends to the *relevant* surrounding traffic and obstacles, it will be fast enough to assist the driver in real time during all dangerous situations.

In order to realize such a cognitive vision system we have developed a robust attention sub-system [1] that can be modulated in a task-oriented way, i.e., based on the current context. The attention sub-system is a central component of the overall vision system, which realizes a temporal organization of different visual processes. Its architecture is inspired by findings of human visual system research (see, e.g., [2]) and organizes its different functionalities in a similar way. In a first proof of concept, we have shown that a purely saliency-based attention generation can assist the driver during a critical situation in a construction site by performing autonomous braking [3], [4].

Our previous work concentrated mainly on saliency-based attention (see [1]) and the creation of a generic system, which allows the dynamic modulation of modules and links between modules (see [3], [4]). This contribution focuses on ways to control the designed cognitive system in order to go beyond classical, reactive driver assistance systems. The goal is to develop a cognitive system that proactively plans and builds up expectations of the environment. The expectations are verified autonomously by adapting internal system processes. As we will show, a generic system structure is the key aspect for accomplishing such a proactive ADAS. Low-complexity system control strategies are sufficient to reach the targeted system behavior, because the control complexity is distributed over the system, e.g., in form of local loops. The realized system is tested on real-world data. A test stream is accessible in the internet (see [5]).

II. RELATED WORK

Recently, the topic of researching intelligent cars is gaining interest as documented by the DARPA Urban Challenge [6] and the European Information Society 2010 *Intelligent Car Initiative* [7] as well as several European Projects like, e.g., Safespot or PReVENT.

Regarding vision systems developed for ADAS, there have been few attempts to incorporate aspects of the human visual system into complete systems. In terms of complete

vision systems, one of the most prominent examples is a system developed in the group of E. Dickmanns [8]. It uses several active cameras mimicking the active nature of gaze control in the human visual system. However, the processing framework is not closely related to the human visual system. Without a tunable attention system and with top-down aspects that are limited to a number of object-specific features for classification, no dynamic preselection of image regions is performed. A more biologically inspired approach has been presented by Färber [9]. This publication as well as the German Transregional Collaborative Research Centre "Cognitive Automobiles" [10] address mainly human inspired behavior planning, whereas our current work focuses more on task-dependent perception aspects and their control.

More specifically, in the center of our work is a computational model of the human attention system that determines the "how" and "when" of scene decomposition and interpretation. Attention is a principle that was found to play an important role in the human vision processing as a mediator between the world and our actual perception [11]. Somewhat simplified, the attention map shows high activation at image positions that are visually conspicuous, i.e., that pop out (bottom-up attention) or that are important for the current system task (top-down attention). Derived from the first computational attention model [12], which showed only bottom-up aspects, some more recent models have been developed that also incorporate top-down information (see, e.g., [1], [13], [14], [15]). Please refer to [1] for a comprehensive comparison between multiple the state-of-the-art attention systems and our computational attention model.

A vision system approach in the vehicle domain that also includes an attention system and that hence is somewhat related to the here presented ADAS is described in [16]. Published after our work (see, e.g., [17]), the approach allows for a simple bottom-up attention-based decomposition of road scenes but without incorporating object or prior knowledge. Additionally, the overall system organization is not biologically motivated and therefore not as flexible as the here proposed system.

To our knowledge, in the car domain no biologically motivated large scale systems exists that allows proactive planning and the verification of expectations followed by an appropriate tuning of system processes.

III. SYSTEM DESCRIPTION

The proposed overall architecture concept for robust attention-based scene analysis is depicted in Fig. 1. It consists of five major parts: "what" pathway, "where" pathway, a part executing "static domain specific tasks", a part allowing "environmental interaction", and a "system control module".

The distinction between "what" and "where" processing path is somewhat similar to the human visual system where the dorsal and ventral pathway are typically associated with these two functions (see, e.g., [2]). Among other things, the "where" pathway in the human brain is believed to perform the localization and tracking of a small number of objects. In contrast, the "what" pathway considers the detailed analysis of a single spot in the image. Nevertheless, an ADAS also requires context information in form of the road and its shape, generated by the static domain specific part. Furthermore, for assisting the driver, the system requires interfaces for allowing environmental interaction (i.e., triggering actuators). The system control module relies on numerous internal system percepts as input and numerous system parameters for controlling the system states and behavior. In order to allow an understanding of the proposed system control strategies a rough system description is given (for more details on these system modules refer to [4]).

A. The "what" pathway

Starting in the "what" pathway the 400x300 pixel color input image is analyzed by calculating the saliency map S^{total}. The saliency map S^{total} results from a weighted linear combination of $N = 130$ biologically inspired input feature maps F_i (see Eq. (1)). More specifically, we filter the image using among others, Difference of Gaussian (DoG) and Gabor filter kernels that model the characteristics of neural receptive fields, measured in the mammal brain. Furthermore, we use the RGBY color space [13] as attention feature that models the processing of photoreceptors on the retina.

The top-down (TD) attention can be tuned (i.e., parameterized) task-dependently to search for specific objects. This is done by applying a TD weight set w_i^{TD} that is computed and adapted online, based on Eq. (2), where the threshold $\phi = K_{conj} Max(F_i)$ with $K_{conj} = (0, 1]$ (see Fig. 3a for a visualization). The weights w_i^{TD} dynamically boost feature maps that are important for our current task or object class in focus and suppress the rest. The bottom-up (BU) weights w_i^{BU} are set object-unspecifically in order to detect unexpected potentially dangerous scene elements. The parameter $\lambda \in [0, 1]$ (see Eq. (1)) determines the relative importance of TD and BU search in the current system state. For more details on the attention system please refer to [1]. It is important to note that the TD weights (calculated using Eq. (2)) are dependent on the features present in the background (rest) of the current image, since the background information is used to differentiate the searched object from the rest of the image [13]. In plain words, the system takes the current scene characteristics (i.e., its features) into account in order to determine the optimal TD weight set, which shows a maximum performance in the current frame.

Now, we compute the maximum on the current saliency map S^{total} and get the focus of attention (FoA, i.e., the currently most interesting image region) by generic region-growing-based segmentation on S^{total}. In the following, with the FoA a restricted part of the image is classified using a state-of-the-art object classifier that is based on neural nets [18]. This procedure (attention generation, FoA segmentation and classification) models the saccadic eye movements of mammals, where a complex scene is scanned and decomposed by sequential focusing of objects in the

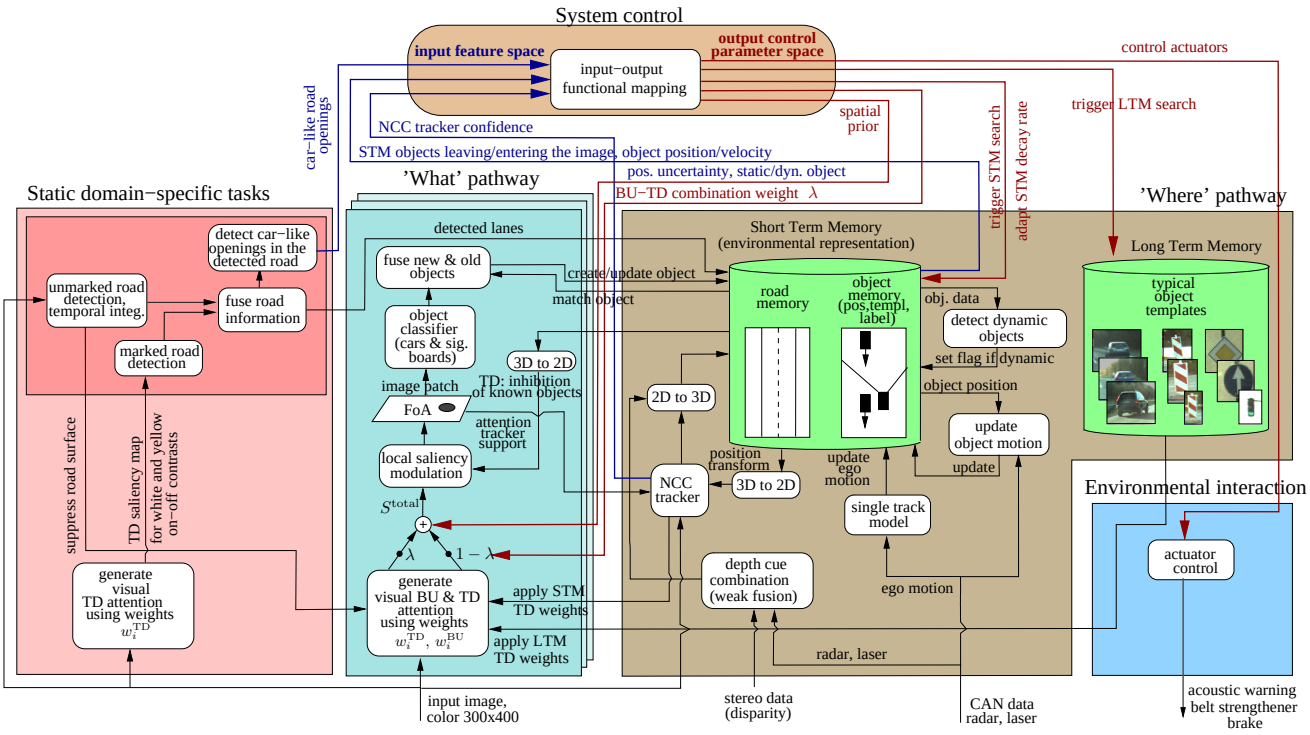

Fig. 1. Biologically motivated system structure for active, attention-based scene analysis.

central 2-3° foveal retina area of the visual field.

$$S^{\text{total}} = \lambda \sum_{i=1}^{N} w_i^{\text{TD}} F_i + (1-\lambda) \sum_{i=1}^{N} w_i^{\text{BU}} F_i \qquad (1)$$

$$w_i^{\text{TD}} = \begin{cases} \frac{m_{RoI,i}}{m_{rest,i}} & \forall \frac{m_{RoI,i}}{m_{rest,i}} \geq 1 \\ -\frac{m_{rest,i}}{m_{RoI,i}} & \forall \frac{m_{RoI,i}}{m_{rest,i}} < 1 \end{cases} \qquad (2)$$

$$\text{with} \quad m_{\{\text{RoI,rest}\},i} = \frac{\sum_{\forall x,y \in \{\text{RoI,rest}\}} F_i(x,y)}{\text{size region } \{\text{RoI,rest}\}}$$

$$\text{and} \quad F_i(x,y) = \begin{cases} F_i(x,y) & \forall (x,y), F_i(x,y) \geq \phi \\ 0 & \text{else} \end{cases}$$

Internal information fusion processes improve the performance of system modules. For example, the detected road (see Section III-C) is fused as context information into the attention system. More specifically, the road is suppressed in all feature maps F_i before fusing them in the overall saliency S^{total}. This procedure makes the saliency map S^{total} sparse and improves the TD weight quality. Additionally, TD-links are used for the modulation of the attention based on detected car-like openings in the found drivable road segment (see module "static domain-specific tasks" path in Fig. 1). This car-like openings are detected by searching for car-sized openings in the road segment (see [4] for details).

B. The "where" pathway

The next step is the fusion between the newly detected object and the already known ones. The result will be further processed in the "where" pathway and stored in the short term memory (STM). The objects in the STM are then suppressed in the current saliency map to enable the system to focus on new objects. The principle of suppressing known objects was proved to exist in the human vision system and is termed inhibition of return (IoR), [19].

All known objects are tracked using a 2D tracker that is based on normalized cross correlation (NCC). The tracker gets its anchor (i.e., the 2D pixel position where the correlation-based search for an object will be started in the new image) from a Kalman filter based prediction on the 3D representation taking the ego motion of the camera vehicle and tracked object into account (see [4] for details).

A comparison between the current Kalman fused 3D object position and the predicted object position (derived from the measured vehicle ego motion) allows the classification of detected objects as static/dynamic (see [4] for details).

For all dynamic (i.e., moving) and therefore potentially dangerous objects in the scene an additional attention-based tracker support is realized, in order to solve a typical problem appearance-based trackers suffer from (i.e., a tracker type that relies on the comparison of image patches). Said trackers depend strongly on the quality of the object template (i.e., the image patch the tracker has to relocalize in the current image). In Fig. 2a the functional description of a NCC tracker is visualized. Here, the template is fix, which leads to a decreasing tracking performance. The object gets lost

quickly. This is caused by the fact that in the vehicle domain the appearance of tracked objects quickly changes (due to changes in illumination, view angle, or varying scale), which makes an adaptation of the template necessary. A typical approach for template adaptation is shown in Fig. 2b, where the template is reset based on the previous tracking result. Using this procedure incremental errors of the tracker are accumulated. The detected region drifts away from the object and gets finally lost. More advanced approaches for template adaptation exist that adapt the initial template by model-based image transformations in order to compensate the scale variance and change of view angle (see, e.g., [20]). Said methods perform robustly, but require specific and complex algorithms. In the here presented system, a novel approach for template adaptation is proposed that relies on already system-immanent approaches. As visualized in Fig. 2c based on the previous template and Eq. (2) a TD weight set and a TD saliency map is computed for the previous image frame. The maximum of this map marks the new position of the object template used for tracking in the current frame. In other words, the initial template position is derived from the typical feature characteristics present in the previous template. As opposed to Fig. 2b, the template adaptation and object redetection is organized separately, thereby preventing the accumulation of incremental errors.

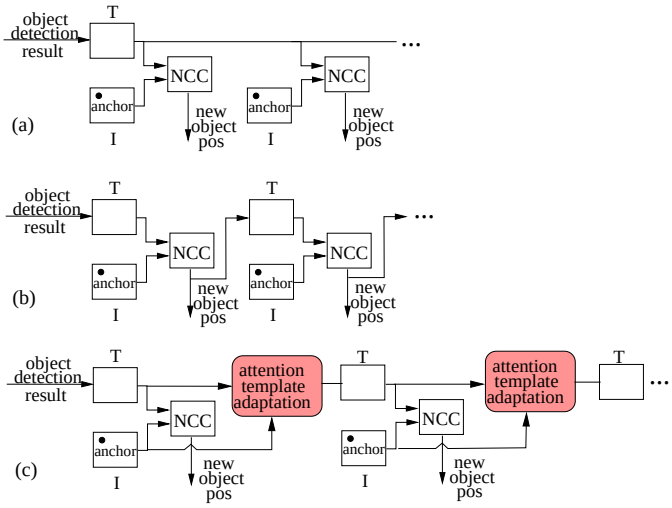

Fig. 2. Functional description of appearance based tracking approaches (with Normalized Cross Correlation NCC, image I, and template T): (a) Fixed template, (b) Continuous template adaptation based on the previous tracking result, (c) Continuous template adaptation based on saliency maximum (TD weight set computed using the previous template).

Coming back to the system description, in case the NCC tracker is able to re-detect the object in 2D pixel coordinates, the 3D position in the representation is updated using four different depth cues for the transformation of 2D pixels into 3D world coordinates. More specifically, our system uses stereo data, radar data, depth from object knowledge, and depth from bird's eye view (as described in detail in [3], [4]).

From a representational point of view, the "where" pathway of our system consists on the one hand of the STM, that stores all properties of sensed objects in a 3D representation and on the other hand of a long term memory (LTM) that stores the generic properties of object classes. The LTM is filled offline with typical patches and corresponding aggregated feature map activations $m_{RoI,i}$ for all supported object classes (see Eq. (2)). For evaluation purposes we use cars, reflection posts, and signal boards as LTM content, although our system is not restricted to these object classes (see [1]). It is important to note that multiple LTM object classes are searched at the same time, which requires several "what" pathways running in parallel (see Fig. 1). In the default case, a specific "what" pathway searches for a generic LTM object class. This is done by computing the geometric mean of all TD weight sets of the LTM object class that were calculated based on Eq. (2). These weights tune the TD attention in the "what" pathway.

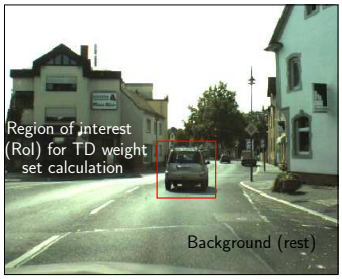

Fig. 3. Visualization of the object training region (RoI) for TD weight calculation against the background (rest).

As described above, in case the tracker has re-detected the object in the current frame the 3D representation is updated. In case the tracker looses the object, the system interrupts the processing in the specific "what" pathway and searches for the lost STM object in the following frames. This is realized by calculating a TD weight set that is specific to the lost STM object using Eq. (2). The object O_f found by the STM search is then compared to the searched object O_s by means of the distance measure $\delta(O_f, O_s)$ that is based on the Bhattacharya coefficient (a measure for determining the similarity between two histograms) calculated on the histograms of all N object feature maps $H_i^{O_f}$ and $H_i^{O_s}$ (see Eq. (3)).

$$\delta(O_f, O_s) = \sum_{i=1}^{N} \sqrt{1 - \gamma(H_i^{O_f}, H_i^{O_s})} \qquad (3)$$

$$\gamma(H_i^{O_f}, H_i^{O_s}) = \sum_{\forall x,y} \sqrt{H_i^{O_f}(x,y) H_i^{O_s}(x,y)}$$

C. Static domain specific tasks

The third major part of our system handles the domain specific tasks related to marked and unmarked lane detection. The marked lane detection is based on a standard Hough transform whose input signal is generated by our generic attention system. The TD attention weights used here boost white and yellow structures on a darker background (so called on-off contrast), to which the biological motivated DoG filter (see Section III-A) is selective. The yellow on-off

structures are weighted stronger than the white to allow the handling of lane markings in construction sites.

The state-of-the-art unmarked lane detection evaluates a street training region in front of the car and two non-street training regions at the side of the road. The features in the street training region (stereo, edge density, color hue, color saturation) are used to detect the drivable road based on dynamic probability distributions for all cues (see [21] for more details). A temporal integration procedure between the current and past detected road segments based on the bird's eye view is applied. The procedure is used to increase the completeness of the detected road by decreasing the number of false negative road pixels (refer to [22] for a comprehensive description of the temporal integration procedure). In the final step, a fusion between the detected marked and unmarked road segments is used to derive the currently drivable lanes.

D. Environmental interaction

The system can interact with the world via an actuator control module. Currently our ADAS implementation uses a 3 phase danger handling scheme depending on the distance and relative speed of a recognized obstacle. When an obstacle is detected in front at a rapidly decreasing distance, a visual and acoustic warning is issued and the brakes are prepared. In the second phase the brakes are engaged with a deceleration of 0.25 g followed by hard braking of 0.6 g in the third phase.

E. System control

The control module realizes a functional mapping of an input feature space of measured internal system-state variables and an output-parameter space for the modulation of the system behavior. In the following, the specific features and parameters are named, the current system uses for control.

Measurement of internal system-state variables (input feature space):

- (i_0) No condition,
- (i_1) Car-like road opening,
- (i_2) NCC tracker confidence (marking a lost object),
- (i_3) STM object leaving/entering the image,
- (i_4) Object position, object velocity.
- (i_5) Object position uncertainty,
- (i_6) Dynamic/static object.

Control parameters to influence/modulate system behavior (output-parameter space):

- (o_1) Actuator control (autonomous braking, acoustic warning, belt strengthener),
- (o_2) LTM search (for cars, signal boards, ...),
- (o_3) Trigger STM search (lost objects, saliency tracking support),
- (o_4) STM decay rate (number of objects in STM),
- (o_5) BU-TD combination weight λ,
- (o_6) Spatial prior (position, sharpness).

In the following, instances of the functional mapping are listed that control the multiple parallel "what" pathways, the actuators, as well as the STM data. The functional mapping between input and output is visualized by the symbol $\Rightarrow$. The task represented by each instance is set in parentheses at the end.

- 1. (i_0) $\Rightarrow$ (o_2) LTM search for cars, (o_5) $\lambda = 0.5$ (search for cars),
- 2. (i_2) NCC tracker confidence for a car below threshold $\Rightarrow$ (o_2) Interrupt LTM search, (o_3) redetect lost object using TD attention, (o_6) Set spatial prior, (o_5) $\lambda = 1$ (redetect lost cars),
- 3. (i_0) $\Rightarrow$ (o_2) LTM search for signal boards, (o_5) $\lambda = 0.5$ (search for signal boards),
- 4. (i_2) NCC tracker confidence for a signal board below threshold $\Rightarrow$ (o_2) Interrupt LTM search, (o_3) redetect lost object using TD attention, (o_6) Set spatial prior, (o_5) $\lambda = 1$ (redetect lost signal boards),
- 5. (i_4) Potential collision $\Rightarrow$ (o_1) Trigger danger handling (collision mitigation),
- 6. (i_6) Dynamic object leaving the field of view $\Rightarrow$ (o_5) $\lambda = 0$, (scene exploration, search dynamic objects),
- 7. (i_6) Dynamic object reentering the field of view $\Rightarrow$ (o_5) $\lambda = 1$, (o_4) Number of STM objects = 1, (o_6) Set spatial prior, (o_3) Saliency tracker support (track dynamic object),
- 8. (i_1) Car-like road opening detected $\Rightarrow$ (o_6) Set spatial prior in (o_2) for cars (analyze conspicuous image region).

IV. RESULTS

In Section IV-A we will evaluate different individual system modules that play the most important role in our cognitive ADAS architecture. In Section IV-B the overall system properties will be assessed. Based on a complex inner-city scenario it is shown how the system proactively plans and verifies expectations in order to allow a safe interaction with the environment (corresponding to the control instances 6 and 7 in Section III-E).

A. Evaluation of system modules

The results presented in [1] support the generic nature of the TD-tunable attention subsystem during object search. Following this concept, the task-specific tunable attention system can be used for scene decomposition and analysis, as it is shown exemplarily on two typical German highway scenes in Fig. 4. Moreover, we see the attention system as a common tunable front-end for the various other system tasks, e.g., as lane marking detection (see Section III-C). In the following the lane marking detection is qualitatively and quantitatively evaluated. Figure 5a shows a typical inner-city scenario with a lot of shadows, on which we tested the attention-based marked lane detection. For detecting the lane markings the bird's eye view is computed (see Fig. 5b). Lane marking-like contrasts (bright image regions on a darker background) are boosted by a DoG filter. Then a clothoid model-based approach for detecting the markings is used (see, e.g., [23], [24], [25] for related clothoid based approaches). Figure 5c depicts the DoG filter results without the described on-off/off-on separation. Since lane markings

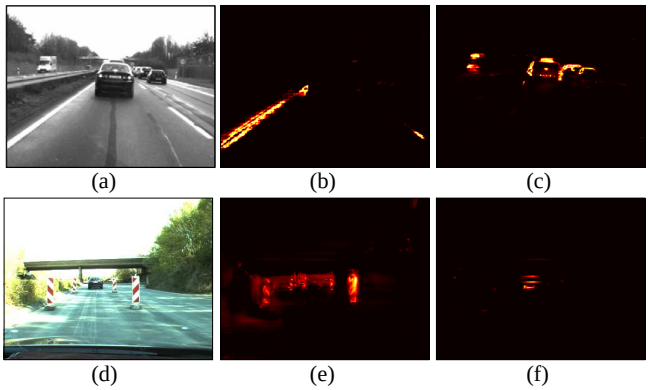

Fig. 4. Attention-based scene decomposition: (a) Highway scene, (b) TD attention tuned to lane markings, (c) TD attention tuned to cars, (d) Construction site (e) TD attention tuned to signal boards (f) TD attention tuned to cars

have a typical on-off contrast (white markings on a darker street), the on-off DoG filter results should be used, since these contain less false-positive activations (Fig. 5d). For example, in [26] the pre-filtered road image still contains the lane marking unspecific off-on contrasts (e.g., shadows on the road). Such off-on contrasts are filtered out in our marked street detection approach to improve the road detection performance. For a quantitative evaluation of the influence of the

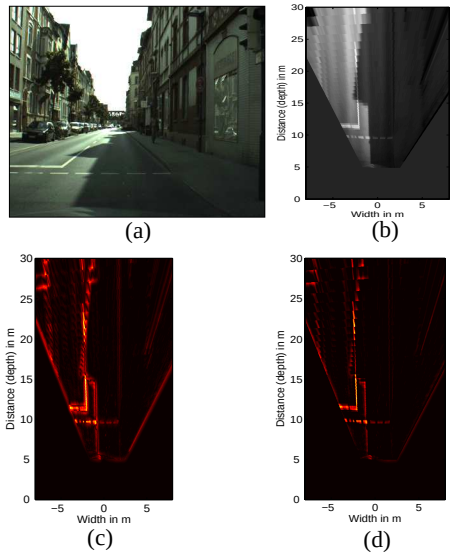

Fig. 5. Exemplarily performance gain of on-off DoG separation as preprocessing step of a lane marking detection system: (a) Shady input image, (b) Bird's eye view, (c) DoG result without on-off/off-on separation, (d) DoG result with on-off contrasts only (off-on contrasts are filtered out).

described on-off DoG separation the lane marking detection system gets a DoG edge image without on-off/off-on (please refer to Fig. 5c) and with on-off separation (as shown in Fig. 5d). The gathered results are summarized in Tab. I. The evaluation shows the improvement in accuracy of the detected offset (i.e., horizontal position of lane markings) and radius of the road based on manually labeled ground truth data consisting of 330 highway frames (see Fig. 6). In the following, the performance gain of the described attention-

Type of input data preprocessing	Mean relative error in offset MREO = $1/N \sum \frac{GT_{\text{offset}} - \text{offset}}{GT_{\text{offset}}}$	Mean relative error in radius MRER = $1/N \sum \frac{GT_{\text{radius}} - \text{radius}}{GT_{\text{radius}}}$
Without DoG on-off separation	4.46	80.87
With DoG on-off separation	**4.35**	**72.22**

TABLE I

MEAN RELATIVE ERROR OF OFFSET AND RADIUS OF THE DETECTED LANE MARKING.

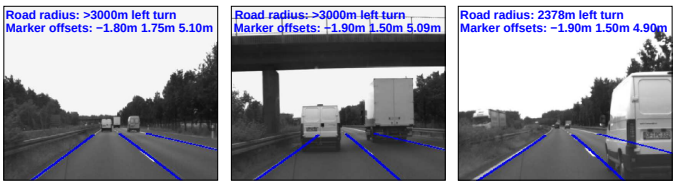

Fig. 6. Three sample images of the used evaluation scene with visualized lane marking detection results.

based-tracker-support is given based on the scenario shown in Fig. 7. In the scenario, a bicycle is tracked over 100 frames. For evaluation we use the measures defined in the Equ. (4), (5), (6), as well as the center accuracy. The Equ. define different ground-truth-based measures that are used here to assess the position and size of a tracked area in the image that contains an object. The measures are motivated from [27] (with pixels being True Positive (TP), False Negative (FN), False Positive (FP)).

$$\text{Completeness} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

$$\text{Correctness} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5)$$

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

On a descriptive level Completeness states, based on given ground-truth data, how much of the real object region is covered by the tracked and hence relocalized region. Correctness states how much of the relocalized region actually belongs to the object to allow a better assessment of large regions that show a high Completeness. Quality combines both measures, since between 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 center accuracy describes the mean absolute position error of the middle axis of the object region in pixels. The necessary ground-truth data was produced by accurate manual annotation of the bicycle region. As Tab. II shows, the applied saliency-enhanced tracking is superior to a classical NCC-based approach. Furthermore, a spatial prior that depends on the Kalman object position uncertainty improves the tracking result.

B. Evaluation of overall system performance

In order to qualitatively evaluate the presented control aspects, results in form of 4 sample frames of a test stream

Evaluation measure	NCC-based tracking	Saliency-enhanced tracking without spatial prior	**Saliency-enhanced tracking with spatial prior**
Completeness	0.23	0.60	**0.73**
Correctness	0.31	0.23	**0.37**
Quality	0.16	0.20	**0.30**
Center accuracy	22.05	20.5	**4.6**

TABLE II

EVALUATION OF OBJECT TRACKING ROBUSTNESS (BICYCLE STREAM).

are presented that show a complex real-world scenario (see Fig. 7). The test stream is accessible in the internet (see [5]). After a description of the visualized internal system percepts based on Fig. 7a, a detailed description of the gathered results is given. In the top row in Fig. 7a (from left to right) the object-unspecific bottom-up saliency, the top-down attention tuned to the tracked bicycle, and their combination (here with $\lambda = 1$ and a sharp spatial prior for the attention-based tracking support) is shown. In the bottom row left in Fig. 7a the input image including a visualization of the detected road area is shown. In the bottom row in the middle, the input image including the predicted vehicle trajectory (the longer the green region, the faster the camera vehicles moves), the detected dynamic object including its predicted trajectory (the red region codes a negative relative velocity: the longer the red region, the faster the dynamic object moves), and the area covered by the radar sensor (in magenta) is visualized. The bottom right image shows an environmental representation that visualizes the task-relevant dynamic object. An aura around the object codes the position uncertainty of the detected object.

In the scenario, while exploring the scene (with λ=0, no spatial prior set, and the STM holding up to 5 objects) the camera ego-vehicle detects a dynamic, i.e., moving, object based on the procedure described in Section III-B (see Fig. 7a). The object is tracked (the parameter λ is set to 1 allowing an attention-based tracking support, a spatial prior is set that depends on the object position uncertainty, as described in Section III-B). The ego-vehicle overtakes the bicycle,giving a blind spot warning. Based on the presented internal 3D representation the bicycle position is predicted linearily even while it is outside the field of view of the camera (see Fig. 7b, the growing position uncertainty is visualized by the growing object aura in the bird's eye view representation). Without any dynamic object in the field of view, λ is set to 0 again, the spatial prior is reset, and the object decay rate is reset to allow the tracking of up to 5 objects. All these adaptations support an object-independent scene exploration. Now, the ego-camera vehicle stops to turn right. The ADAS "remembers" the bicycle and gives a blind spot warning. The ego-vehicle waits for the bicycle to reappear. In order to allow a fast redetection, the top-down attention is tuned to the bicycle, setting a spatial prior with low sharpness. This allows its instantaneous redetection (see Fig. 7c). The object position is updated lowering the position uncertainty. The system tracks the object relying on the attention-based-tracker-support (see Fig. 7d). Summing up, the system at runtime builds up and verifies expectations of the environment, thereby autonomously tuning internal parameters and processes that improve and accelerate the system reaction. The processing described above relates to one "what" pathway (see Fig. 1), which concentrates on the detection, tracking, and prediction of one dynamic and hence potentially dangerous object. As visualized in Fig. 1, multiple "what" pathways run in parallel. For example, further "what" pathways handle the detection of cars, signal boards, and other traffic-relevant objects using the previously described LTM search (see Section III-B) that relies on object-specifically tuned saliency maps. The control of these pathways is realized by the functional mapping described in Section III-E. For each of these "what" pathways the object-specific LTM search can be interrupted by a STM search in case an object was lost during tracking (see [4]).

V. SUMMARY AND OUTLOOK

In this contribution, we presented an integrated, advanced driver assistance system that relies on human-like cognitive processing principles. The system uses a biologically motivated attention system as flexible and generic front-end for all visual processing. Based on top-down links modulating the attention task-dependently, the used internal 3D representation, a state-of-the-art object classifier, and a road recognition system, we realized a highly flexible and robust system architecture. As shown, simple control strategies are sufficient for the realized biologically inspired system to allow a safe system reaction in various scenarios. We currently port the described extensions from Matlab to C in order to integrate them in our existing online system [3] for evaluating them on our prototype vehicle. After the successful test of the low complexity control approach, in the next step, learning of the functional mapping between the measured input feature space and the output control parameter space will be in our focus. Also measuring and mimicking the reactions of an experienced driver is envisioned in the future. The introduced system contains first approaches towards such an efficient cognitive control concept. The central assumption, as proposed in this contribution, is that a robust (and as envisioned also learning) system requires a generic system structure with a high number of degrees of freedom for controlling the system reaction and measuring the system state.

REFERENCES

[1] T. Michalke, J. Fritsch, and C. Goerick, "Enhancing robustness of a saliency-based attention system for driver assistance," in *Int. Conf. on Computer Vision Systems*, 2008.

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

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

[4] 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.

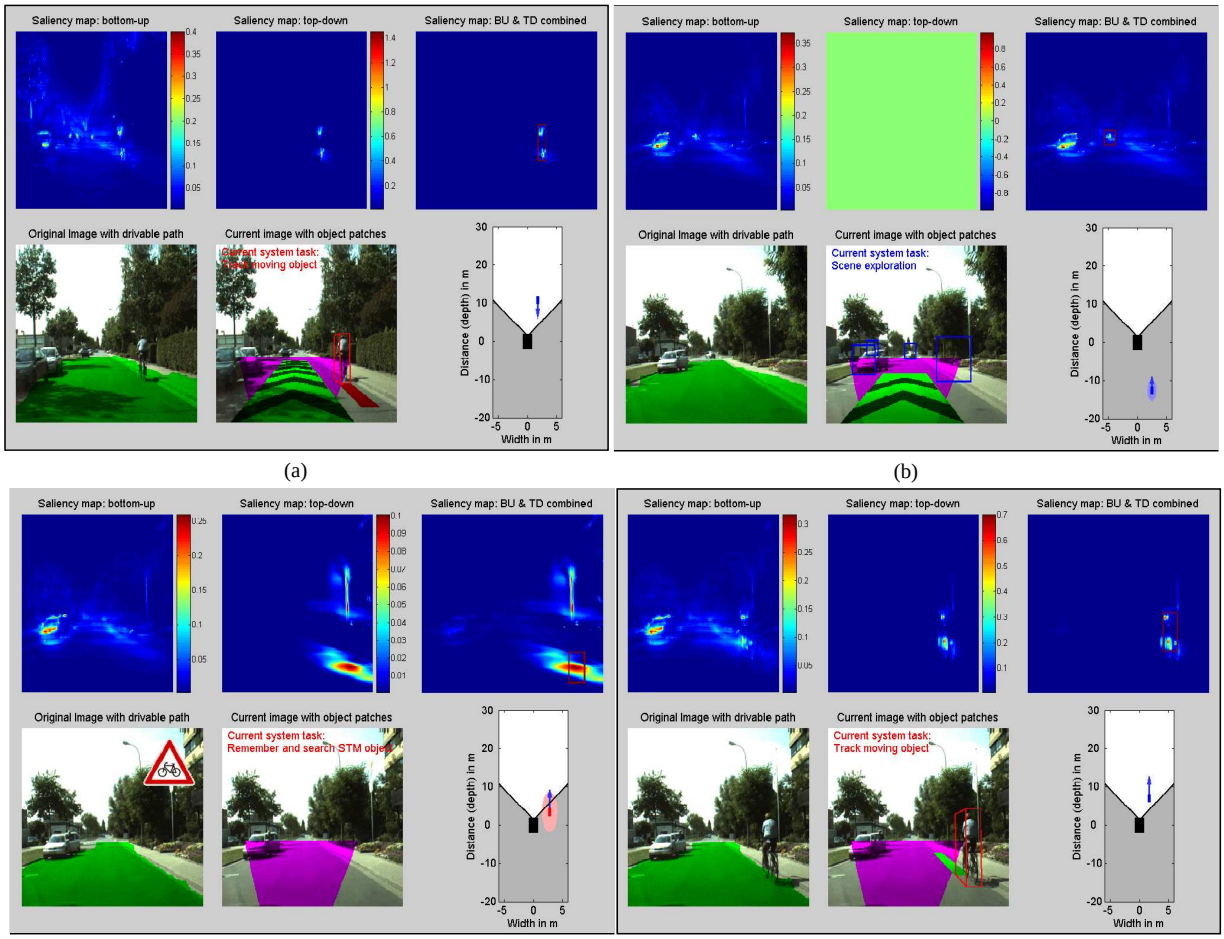

Fig. 7. Visualization of system states for bicycle stream: (a) After its detection the bicycle is tracked, ego-vehicle is closing in, (b) Blind prediction of bicycle, (c) Ego-car searching actively for the bicycle, waiting to turn right, (d) Bicycle redetected successfully, ego-vehicle turns right.

[5] WWW, 2009, http://www.rtr.tu-darmstadt.de/~tmichalk/IV2009_ADASControl/.

[6] *DARPA Urban Challenge*. [Online]. Available: http://www.darpa.mil/grandchallenge/

[7] WWW, *European commission information society 'Intelligent Car initiative*, 2007, http://ec.europa.eu/informationsociety/activities/intelligentcar/.

[8] E. Dickmanns, "Three-Stage Visual Perception for Vertebrate-type Dynamic Machine Vision," in *Engineering of Intelligent Systems (EIS)*, Madeira, Feb 2004.

[9] G. Färber, "Biological aspects in technical sensor systems," in *Proc. Advanced Microsystems for Automotive Applications*, Berlin, Mar 2005, pp. 3–22.

[10] C. Stiller, G. Färber, and S. Kammel, "Cooperative cognitive automobiles," in *IEEE Intelligent Vehicles Symposium*, 2007, pp. 215–220.

[11] D. Simons and C. Chabris, "Gorillas in our midst: Sustained inattentional blindness for dynamic events," *British Journal of Developmental Psychology*, vol. 13, pp. 113–142, 1995.

[12] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry," *Human Neurobiology*, vol. 4, no. 4, pp. 219–227, 1985.

[13] S. Frintrop, "Vocus: A visual attention system for object detection and goal-directed search," Ph.D. dissertation, University of Bonn, 2006.

[14] V. Navalpakkam and L. Itti, "Modeling the influence of task on attention," *Vision Research*, vol. 45, no. 2, pp. 205–231, 2005.

[15] Z. Aziz and B. Mertsching, "Visual search in static and dynamic scenes using fine-grain top-down visual attention," in *Lecture Notes in Computer Science*, vol. 5008, 2008, pp. 3–12.

[16] S. Matzka, Y. Petillot, and A. Wallace, "Proactive sensor-resource allocation using optical sensors," in *VDI-Berichte 2038*, 2008, pp. 159–167.

[17] 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.

[18] H. Wersing and E. Körner, "Learning optimized features for hierarchical models of invariant object recognition," *Neural Computation*, vol. 15, no. 2, pp. 1559–1588, 2003.

[19] R. M. Klein, "Inhibition of return," *Trends in Cognitive Science*, vol. 4, no. 4, pp. 138–145, April 2000.

[20] J. Eggert, C. Zhang, and E. Körner, "Template matching for large transformations," in *ICANN (2)*, 2007, pp. 169–179.

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

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

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

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

[25] E. Dickmanns and B. Mysliwetz, "Recursive 3-d road and relative ego-state recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 199–213, 1992.

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

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