Towards a Human-like Vision System for Resource-Constrained Intelligent Cars

Thomas Michalke[1], Alexander Gepperth[2], Martin Schneider[1], Jannik Fritsch[2], and Christian Goerick[2]

[1] Technische Universität Darmstadt, Institute for Automatic Control
D-64283 Darmstadt, Germany
{thomas.michalke,martin}@rtr.tu-darmstadt.de
[2] Honda Research Institute Europe GmbH, D-63073 Offenbach, Germany
{alexander.gepperth,jannik.fritsch,christian.goerick}@honda-ri.de

Abstract. Research on computer vision systems for driver assistance resulted in a variety of approaches mainly performing reactive tasks like, e.g., lane keeping. However, for a full understanding of generic traffic situations, integrated and more flexible approaches are needed. We present a system inspired by the human visual system. Based on combining task-dependent tunable visual saliency, an object recognizer, and a tracker it provides warnings in dangerous situations.

1 INTRODUCTION

The goal of building intelligent computer vision systems can be approached from two directions: either searching for the best engineering solution or taking the human as a role model. In the latter case, research results from other disciplines like, e.g., psychophysics or neurobiology, can be used to guide the vision system design. 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 the necessary flexibility. Small changes in the task and/or environment often lead to the necessity of redesigning the whole system. Considering the human vision system, nature has managed to realize a highly flexible system capable of adapting to severe changes in the task and/or the environment. Hence one of our main design goals is to implement a system able to accomplish new tasks without adding modules or changing the system's structure. Equally, we aim at the implementation of the underlying principles of the human vision system and not directly at engineering efforts to attain its measurable abilities. In other words, we intend to construct a generic vision system that can operate in the real-world by modulating and parameterizing submodules without being explicitly designed for specific tasks of a scenario.

Aiming at going beyond standard industrial computer vision applications, there is an increasing emphasis in the computer vision community on building so-called cognitive vision systems (i.e., systems that work according to human information processing principles) suitable for solving complex vision tasks. One

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

important cognitive principle is the existence of top-down links in the system, i.e., informational links from stages of higher to lower knowledge integration. Top-down links are believed to be a prerequisite for fast-adapting biological systems living in changing environments.

One particularly interesting dynamic environment for vision systems is ordinary traffic. Although, e.g., lane markings and traffic rules restrict the complexity of the driving task, the vision systems for driver assistance developed up to now are mainly capable of dealing with simple traffic situations. While this already resulted in specialized commercial products improving driving safety (e.g., the "Honda Intelligent Driver Support System" [1] to help the driver stay in the lane and maintain the right distance to the preceding car), the problem of developing a generic vision system for advanced driver assistance, i.e., capable of operating in all kinds of challenging situations, is still open.

One possible way to achieve this goal is to realize a task-dependent perception using top-down links. In this paradigm, the same scene can be decomposed in different ways depending on the current task. A promising approach is to use a high-performance attention system that can be modulated in a task-oriented way, i.e., based on the current context. For example, while driving at high speed, the central field of the visual scene becomes more important than the surrounding.

Aiming towards such a task-based vision system, this paper describes a vision architecture that is being developed as perceptual front-end of an Advanced Driver Assistance System (ADAS). The proposed system provides a framework that enables the task-dependent tuning of visual processes via object-specific weighting of input features of the attention system. The system generates an appropriate system reaction in dangerous situations (autonomous braking). Its architecture is inspired by findings in the human visual system and organizes the different functionalities in a similar way. For first proof of concept, we focus on assisting the driver during a critical situation in a construction site. For the analysis of the attention system, we evaluated the construction site scenario as well as a challenging inner-city traffic scene to illustrate the performance gain of the top-down approach in a more complex environment. The system has been implemented using a software framework for component integration and achieves real-time performance on a prototype car.

2 RELATED WORK

In the past, the human visual system has been examined in a large number of studies and the task-dependent nature of gazing has been proven in a variety of situations including steering a car. For example, the gaze of drivers in a virtual environment was examined in [2]. The results show that the performance in detecting stop signs is heavily modulated by context (i.e. top-down) factors and not only by bottom-up visual saliency. Endowing a vision architecture for an intelligent car with similar, task-based attention can result in a gain of performance with minimal additional resource requirements (see Section 6).

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

In most research on human visual attention the focus is on the bottom-up detection of salient features/objects in a scene (for a review see [3]). A well-known computational model for saliency calculation is the approach by Itti et al. [4] that is used in a number of implemented systems. Recently, this approach has been extended by various researchers to account for task-dependent aspects of visual attention (see, e.g., [5–7]) by applying dynamic weights to different processing stages. The tasks are often to find a specific object within a predominantly static indoor scene. A more complete view on a possible architecture for a visual system incorporating task-dependent visual attention is given by Navalpakkam and Itti [8]. However, they focus on giving results for a few specific aspects while some parts of the proposed architecture (especially the combination of top-down and bottom-up saliency) are not implemented or published yet.

Turning to the domain of vision systems developed for ADAS, there have been few attempts to incorporate aspects of the human visual system. One of the most prominent examples is a system developed in the group of E. Dickmanns [9]. 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 bottom-up attention system and top-down aspects that are limited to a number of object-specific features for classification, no dynamic preselection of image regions is performed.

To our knowledge, there are no integrated system architectures in the car domain that attempt to explicitly model aspects of the human visual system.

3 SYSTEM ARCHITECTURE

The overall architecture concept to realize task-based visual processing is depicted in Fig. 1. It contains a distinction between a 'what' and a 'where' processing path, similar to the human visual system where the dorsal and ventral pathway are typically associated with these two functions. Among other things, the 'where' pathway in the human brain is believed to perform the coarse tracking of a small number of objects that are relevant for the current task. This tracking is performed by the human visual system without focusing the eye gaze on individual objects to be tracked [10]. In contrast, the 'what' pathway considers the detailed analysis of a single spot in the image. In the human visual system this is intimately bound to the current eye gaze, as the human eye possesses a high resolution in the central 2-3° (foveal retina area) of the visual field only.

In our vision system the eye gaze is performed virtually as the camera mounted in the car has a constant resolution in the complete field of view. Changing the eye gaze is therefore equivalent to shifting the processing to another spot of the input image. This spot is analyzed in our 'what' pathway in full resolution while the whole image is analyzed in the 'where' path in lower resolution. Processing in these two pathways is believed to occur in parallel in the human brain, but their intertwining are as yet not known in too much detail. We here adopt the idea of continuously tracking a small number of objects in each

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

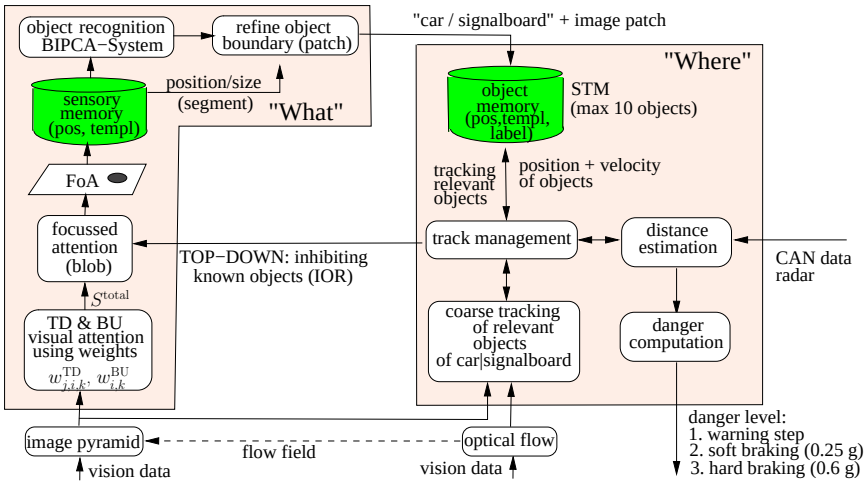

Fig. 1. Architecture concept.

image of the incoming visual stream to coarsely represent the current scene and at the same time acquiring more detailed information on one additional object.

The detailed organization of the two processing streams in our architecture concept is as follows: The input image is analyzed in the 'what' path for salient locations using a variety of visual features including orientation, intensity, and motion for saliency computation (see Section 4). By applying top-down information, image regions that contain known objects, i.e., that tracked in the 'where' path, are suppressed during the saliency computation. This suppression is also known as inhibition of return in the human visual system. A simple maximum search is used on the resulting saliency map to find the currently most salient point in the scene. The Focus of Attention [FoA] is determined by region growing on the overall saliency map using the most salient point as an anchor. This region is fed to the fast feedforward object recognition system BIPCA [11] that is trained here to recognize back views of cars as well as signal boards. The image region together with the object label is stored in the object memory in order to be tracked coarsely in subsequent images in the 'where' path. Before insertion, a check is performed to associate the new object to known objects already in the object memory. Concluding one iteration, for all objects in the object memory a 3D-world position estimation is calculated based on fusing measurements from laser and bird's eye view [12] using an Extended Kalman Filter.

All objects are constantly tracked through predicting the object position in image coordinates in the next image from the current relative velocity as extracted by the optical flow [13] and performing a local correlation for the refinement of the objects' position. The flow is extracted from downscaled images (128×64) where only half the vertical area is considered, cutting off the sky and the ego-vehicle. The tracked image regions allow to suppress these areas during FoA generation in the next images.

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

In case the prediction does not match (no good correlation found) the system will stop inhibiting the object in the 'what' pathway. Consequently, the attention will be focused on the missing object in one of the next images if the object is still present and salient. In this way, all objects being recognized and behaving as predicted are coarsely tracked while the 'what' attention is always focused on new objects and objects behaving unexpectedly.

The novelty of our architecture lies in the introduction of top-down aspects (like, e.g., task-dependent tunable attention generation via sets of weights and the simultaneous operation of inhibition of return predicted by coarse tracking) resulting in the ability to cope with highly dynamic traffic scenes using limited computational resources. The top-down tunable attention system is a key aspect of our ADAS, since such preprocessing will lead to a considerable reduction of scene complexity by restricting further processing steps to image regions that are interesting according to the current system task. Consequently, this sub-system will be described in more detail in the next Section.

4 VISUAL ATTENTION SUB-SYSTEM

A rough sketch of the visual attention sub-system is depicted in Fig. 2. It consists of a number of features that are extracted from the image on a set of different

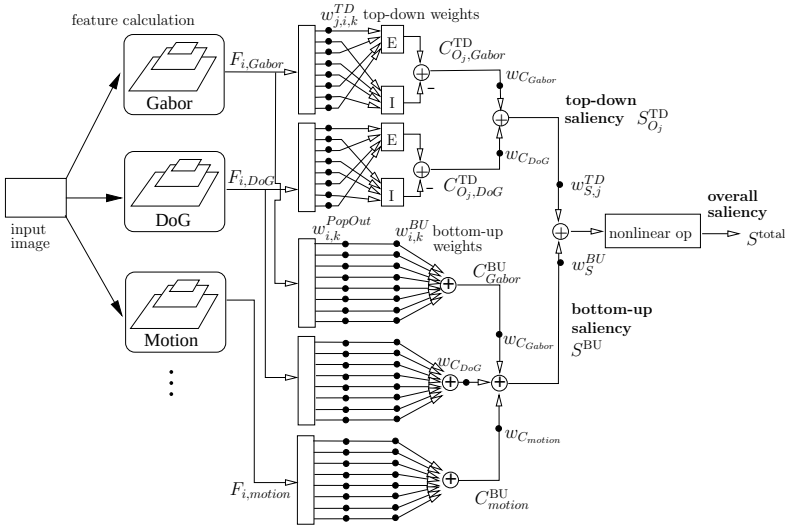

Fig. 2. Visual attention sub-system.

scales derived from a Gaussian image pyramid starting from 256 × 256 pixels. Currently we use the features *odd* and *even Gabor filters* in 4 orientations, *Difference of Gaussians filters* [DoG] and *motion* from differential images on 5

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

scales giving a total of 50 feature maps. The raw feature map responses are passed through a preprocessing step that consists of squaring, nonlinear noise suppression by a sigmoidal function, and normalization. In addition to combining these features to obtain a bottom-up saliency map [4, 14], we also compute top-down saliency maps using object-specific feature map weights. The object-specific weights are inspired by [5, 15] in the way the weights are obtained: During a supervised training stage, the feature map activations of an object are compared to the feature map activations in its surrounding. From this comparison, the relative importance of a feature (its signal-to-noise [SNR] ratio) can be determined. For each trained object O_j and feature channel $F_{i,k}$ we therefore get a top-down weight $w^{\mathrm{TD}}_{j,i,k}$ that is proportional to how well the feature channel i of feature type k is able to discriminate the object j from its surrounding:

$$w^{\mathrm{TD}}_{j,i,k} = \mathrm{SNR}^{O_j}_{i,k} \tag{1}$$

In order to emphasize matching features and suppress irrelevant features, separate maps for *excitation* E and *inhibition* I are constructed. Their combination leads to object-specific conspicuity maps $C^{\mathrm{TD}}_{O_j,k}$.

$$E^{\mathrm{TD}}_{O_j,k} = \sum_{i=1}^{108} w^{\mathrm{TD}}_{j,i,k} F_{i,k} \qquad \forall\, w^{\mathrm{TD}}_{j,i,k} \geq 1.0 \tag{2}$$

$$I^{\mathrm{TD}}_{O_j,k} = \sum_{i=1}^{108} \frac{1}{w^{\mathrm{TD}}_{j,i,k}} F_{i,k} \qquad \forall\, w^{\mathrm{TD}}_{j,i,k} < 1.0 \tag{3}$$

$$C^{\mathrm{TD}}_{O_j,k} = E^{\mathrm{TD}}_{O_j,k} - I^{\mathrm{TD}}_{O_j,k} \tag{4}$$

It is important to note that the performance gain of this approach compared to standard attention systems lies in the explicit inhibition of non-target regions. The conspicuity maps $C^{\mathrm{TD}}_{O_j,k}$ are combined to an object-specific top-down saliency map $S^{\mathrm{TD}}_{O_j}$ by using feature type specific weights w_{C_k} that are proportional to the confidence one can assign to the feature type k in the current scene. This could be done dynamically depending on, e.g., the current weather or lighting conditions.

$$S^{\mathrm{TD}}_{O_j} = \sum_{k=1}^{2} w_{C_k} C^{\mathrm{TD}}_{O_j,k} = w_{C_{Gabor}} C^{\mathrm{TD}}_{O_j,Gabor} + w_{C_{DoG}} C^{\mathrm{TD}}_{O_j,DoG} \tag{5}$$

In addition, we also calculate a biased bottom-up saliency map by adding all feature maps weighted with their specific bottom-up weights $w^{\mathrm{BU}}_{i,k}$:

$$S^{\mathrm{BU}} = \sum_{k=1}^{3} w_{C_k} C^{\mathrm{BU}}_{k} \qquad \text{with} \qquad C^{\mathrm{BU}}_{k} = \sum_{i=1}^{108} w^{\mathrm{PopOut}}_{i,k} w^{\mathrm{BU}}_{i,k} F_{i,k} \tag{6}$$

As $w^{\mathrm{BU}}_{i,k}$ we choose a set of weights that shows good performance for most environments. In the object-unspecific bottom-up path no inhibition takes place, since its purpose is to evaluate the general unspecific saliency of a scene. The

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

individual bottom-up feature maps are additionally preprocessed by a pop-out operator that globally amplifies maps with a small number of maxima and attenuates maps with many maxima [4]. The pop-out operator multiplies the feature maps with a dynamic factor $w_{i,k}^{\mathrm{PopOut}}$ computed at runtime. The factor is inversely proportional to the number of pixels that are near the maximum of the feature map. Additionally, $w_{i,k}^{\mathrm{PopOut}}$ is decreased by a factor of 4 with decreasing scale level in the image pyramid to maintain the comparability of scales. By applying this operator, the bottom-up path is designed to amplify feature maps that show few maxima, i.e., that are sparse. In consequence, feature maps containing image regions that pop out are boosted. It is of crucial importance that the top-down feature maps do not pass a similar pop-out step, since by tuning the top-down weights, we aim at finding objects based on feature conjunctions. The individual feature map responses for the searched objects might only reach medium values, whereas the combination of all relevant maps will lead to a strong response in the resulting saliency map. This explicit differentiation is not made in other top-down attention systems, which can lead to a performance loss.

The overall saliency map S^{Total} is calculated by linearly combining the normalized top-down and bottom-up saliency maps depending on the current task of the ADAS. Subsequently, a nonlinear operator is applied to cut off negative values before the overall saliency map is passed on to the FoA generation (see previous Section). For weighting the maps we currently use sets of weights for signal boards and cars that were calculated in a supervised training step. It is envisioned in later versions of our ADAS to calculate these weights dynamically at runtime to track and even learn new objects.

5 EXPERIMENTAL SETUP

Technical setup: 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. The image data as well as the laser and vehicle state data from the CAN bus are recorded. The recorded data is used during offline evaluation. For the online version, all data is transmitted via LAN to two Toshiba Tecra A7 (2 GHz Core Duo) running our RTBOS integration middleware [16] on top of Linux. The individual RTBOS components are implemented in C using an optimized image processing library based on the Intel IPP [17].

Scenario: In order to evaluate the proposed system in a challenging situation, we concentrate on typical construction sites on highways. This situation is quite frequent and a traffic jam ending exactly within a construction site is a highly dangerous situation: due to the S-curve in many construction sites, the driver will notice a braking or stopping car quite late, see Fig. 3a). 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, 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.

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

Test data for training and evaluation: In order to gain sufficient training data, we recorded image sequences during normal highway traffic including construction sites as well as visually complex scenes from driving in inner cities. For evaluating the vision system, we recorded data in an exemplary construction site on a private driving range.

6 RESULTS

For the evaluation of the proposed top-down attention system we use streams from a construction site and from driving in a city (data in parenthesis). Fig. 3 shows an example of the sparseness of the top-down saliency maps tuned to signal boards (cars) as well as the derived FoAs.

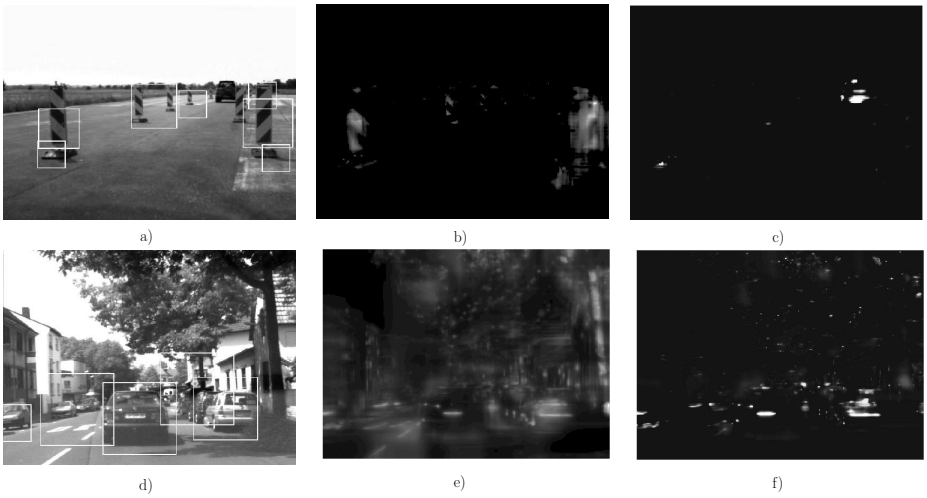

Fig. 3. Output of attention system for construction site and inner-city streams, a)Unsegmented FoAs, tuned to signal boards, b)TD saliency signal boards, c)TD saliency cars, d)Unsegmented FoAs, tuned to cars, e)BU saliency, f)TD saliency cars.

The improvements in selectivity gained by using top-down attention are measured by counting the number of FoA hits and misses of traffic-relevant signalboards (cars) on 280 (240) images of construction site (inner city) streams. We classify a target as traffic-relevant when it is situated within a maximum distance of 50 meters (based on the bird's eye view representation [12]). For evaluation we count an FoA as a hit if at least half of the object is within the FoA. An FoA is considered as a miss when a non-target is found although traffic-relevant and still undetected targets are in the scene. FoAs that are generated on border regions of already found but only inadequately suppressed targets are counted as a miss as well. The completeness is defined as the ratio of undetected targets

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

that have left the image to detected targets. We carried out this analysis by using either only bottom-up (BU) saliency or the combination of top-down (TD) and BU saliency (see Tab. 1). The streams contain 56(42) traffic-relevant signal boards (cars).

Scene	Saliency modulation TD, BU	Hits	Misses	Completeness
construction site	BU	**116**	**27**	84 %
	BU + TD (tuned to signal boards)	**126**	**11**	89 %
inner city	BU	**130**	**48**	90 %
	BU + TD (tuned to cars)	**138**	**22**	90 %

Table 1. Results of evaluation of attention system.

The obtained results show a high completeness for both saliency modulation methods. However, the number of misses in the pure BU driven case for signal boards (cars) is about 2.5 (2) times higher than in the TD supported case. Hence, the application of TD aspects increases the probability of FoAs containing objects that are relevant to the current task. This leads to fewer iterations of the whole system to accomplish an exhaustive scene decomposition. In other words, computationally more demanding processing steps (like, e.g., object recognition) will work on prefiltered data of higher relevance.

For a proof of concept, we trained the classifier to distinguish cars from non-cars (clutter). A set of image segments generated by our vision system during online operation was used for training. It contains 3000 roughly quadratic image patches scaled to a size of 64x64 pixels, and was divided into the classes 'car' (300 patches) and 'clutter' (2700 patches) by visual inspection. Car segments contain complete back-views of cars (at any position) which must be at least half as large as the patch in both dimensions. At equal false positive and true negative rates, an error of 5 % was obtained on an equally large test.

We tested the warning generation offline on 5 construction site streams showing the setting depicted in Fig. 3a. In all streams, the ADAS was able to recognize and track the car from a distance between 65 and 40 meters. Our system is operating online providing tracking results at a rate of 10 Hz. Currently, we are parameterizing the coupling between the 3D world position of a detected obstacle and the phases of the danger handling scheme in order to perform autonomuous braking experiments.

7 CONCLUSIONS AND FUTURE WORKS

The contribution introduced an integrated vision architecture for ADAS, which realizes cognitive principles. Encouraging results obtained from the application of an attention system that can be modulated in a task-oriented top-down style were presented. The system is working online performing the described autonomous braking functionality on a Honda Legend prototype car. Our future

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

work will concentrate on adding further top-down aspects to the system to make it more flexible and dynamic (e.g., gist that incorporates scene knowledge).

8 ACKNOWLEDGMENTS

The authors gratefully acknowledge the reviewers' comments and the support from Sven Bone, Falko Waibel, and Dr. Jens Gayko from Automobile Advanced Technology Research, Honda R & D Europe, for obtaining training data and demonstrating the system online on a prototype car.

References

1. Ikegaya, M., Asanuma, N., Ishida, S., Kondo, S.: Development of a lane following assistance system. In: Int. Symp. on Advanced Vehicle Control, Nagoya (1998)
2. Shinoda, H., Hayhoe, M.M., Shrivastava, A.: What controls attention in natural environments. Vision Research (41) (2001) 3535 – 3546
3. Wolfe, J.M., Horowitz, T.S.: What attributes guide the deployment of visual attention and how do they do it? Nature Reviews Neuroscience **5**(6) (2004) 495–501
4. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. **20**(11) (1998) 1254–1259
5. Frintrop, S., Backer, G., Rome, E.: Goal-directed search with a top-down modulated computational attention system. In: Lecture Notes in Computer Science. (2005) 117–124
6. Hawes, N., Wyatt, J.: Towards context-sensitive visual attention. In: Proceedings of the Second Int. Cognitive Vision Workshop, Graz, Austria (2006)
7. Goerick, C., Wersing, H., Mikhailova, I., Dunn, M.: Peripersonal space and object recognition for humanoids. In: Proc. Int. Conf. on Humanoid Robots. (2005)
8. Navalpakkam, V., Itti, L.: Modeling the influence of task on attention. Vision Research **45**(2) (2005) 205–231
9. Dickmanns, E.: Three-Stage Visual Perception for Vertebrate-type Dynamic Machine Vision. In: Engineering of Intelligent Systems (EIS), Madeira (2004)
10. Cavanagh, P., Alvarez, G.: Tracking multiple targets with multifocal attention. Trends in Cognitive Sciences **9** (2005) 350–355
11. Wersing, H., Körner, E.: Learning optimized features for hierarchical models of invariant object recognition. Neural Computation **15**(2) (2003) 1559–1588
12. Broggi, A.: Robust real-time lane and road detection in critical shadow conditions. In: Proc. Int. Symp. on Computer Vision, Parma, IEEE (1995)
13. Willert, V., Eggert, J., Adamy, J., Koerner, E.: Non-gaussian velocity distributions integrated over space, time and scales. IEEE Transactions on Systems, Man and Cybernetics B **36**(3) (2006) 482–493
14. Koch, C., Ullman, S.: Shifts in selective visual attention: towards the underlying neural circuitry. Human Neurobiology **4**(4) (1985) 219–227
15. Navalpakkam, V., Itti, L.: Optimal cue selection strategy. In: Advances in Neural Information Processing Systems, Vol. 19, Cambridge, MA, MIT Press (2006) 1–8
16. Ceravola, A., Joublin, F., Dunn, M., Eggert, J., Goerick, C.: Integrated research and development environment for real-time distributed embodied intelligent systems. In: Proc. Int. Conf. on Robots and Intelligent Systems. (2006) 1631–1637
17. Intel: Integrated Performance Primitives (2006) `http://www.intel.com/cd/software/products/asmo-na/eng/perflib/ipp/302910.htm`.

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de