Computer Vision and Artificial Intelligence On-Road

B. Lamprecht and S. Fuchs
Department of Informatics Systems
Transportation Informatics Group
Alpen-Adria-Universität Klagenfurt
bernhard.lamprecht@uni-klu.ac.at
simone.fuchs@uni-klu.ac.at

Abstract
This paper gives an overview of the application of artificial intelligence (AI) in the field of computer vision (CV) in the context of road traffic. The visual perception of the environment out of vehicles leaded to first vision-based driver support systems, such as lane keeping assistants. For efficient solutions, an exclusive computer vision based approach is not enough. Hybrid approaches try to exploit the advantages of both disciplines (CV and AI) for more robust and self-adaptive vision-enabled intelligent systems in the car.

Introduction
First cars become commercially available which are equipped with cameras at different locations. Cameras adjusted to the driving direction observe the immediate area in front of the car. If an emerging obstacle is detected, the necessary driving manoeuvre can be initiated. Cameras mounted on the outside mirrors keep track of the blind-spot and warn the driver in case of an overseen vehicle passing by. Computer vision in the field of intelligent transportation systems inspired researchers for new intelligent solutions that save lives and money. According to a study of the European commission in the year 2000 more than 40.000 people were killed and another 1.7 million people were injured on European roads. Furthermore, accidents amount for EUR 45 billion each year. Indirect costs are even estimated at EUR 160 billion [16]. CV in combination with concepts from AI enables machines to percept the environment through cameras like human beings and deduce intelligent recommendations to the driver. This makes car driving safer and thus saves lives and money.

This paper shows where CV meets AI in the road traffic context. While in some terminology CV is seen as a part of AI, due to learning and pattern recognition, we do understand CV and AI as basically independent disciplines but with strong interrelationships. A semantic traffic scene interpretation of a road crossing does not necessarily require cameras to sense the current situation [13].

If cameras are used for traffic and environmental monitoring, CV serves AI with sensor values, in other words there exists a *vertical relationship*. In some cases AI supports CV to parameterize algorithms for more robustness and automatic adaptation to new environmental conditions. This represents a *horizontal relationship* between both disciplines.

Object Recognition
A vision-based ADAS (Advanced Driver Assistance System) continuously monitors the environment of the vehicle with cameras. Identified objects in the car's surrounding are the system's input. Object recognition is a fundamental task for the usage of vision enabled systems on-road. An intelligent system needs to know what the object is and with whom it interacts. "Object recognition, which involves the classification of objects into one of many a priori known object types, and determining object characteristics such as pose, is a difficult problem." [12, p. 1]

Three basic paradigms for object recognition are distinguished [12]: Bayesian Statistics, Neural Networks and Expert Systems. Common to Bayesian and Neural approaches is learning. Both paradigms start with a system that needs to be trained from samples. A neural based pedestrian detection, for instance, is trained from an image database of pedestrians in different poses and of different sizes. After finishing the training, the system can be applied to real-world scenarios to classify and recognize pedestrians based on the learned samples from the database. Obviously, the more representative the trainings samples are, the better the overall reliability of the system will be for operation under real world conditions. Expert Systems for object recognition work in a different manner. The classification of an object does not depend on the knowledge/resemblance from/to previous learned samples, but rather on the knowledge about the object itself. A pedestrian has a certain size, shape and behaviour, for instance. This real-world knowledge is used to classify an object as a pedestrian. All three paradigms (Bayes, Neural and Expert Systems) have advantages and disadvantages under certain environmental and situational circumstances. For object recognition on the road (cars, pedestrians etc.), hybrid systems which integrate methods from all three paradigms show the most reliable results [17].

The pedestrian detection system in [10] uses support vector machines (SVMs) to classify pedestrians. The system integrates features measured over time to improve the object recognition quality. The decision that a recognized object is a pedestrian is determined through the incorporation of information about dynamic gait and leg position, amongst others. In [9] a traffic sign recognition system is presented. After the extraction of the traffic sign edges, the sign is classified with the help of neural networks.

Adaptive and robust computer vision algorithms are today's challenge for the application to road traffic scenarios. In the subsequent sections we will give an overview of principles where methods of AI and CV are applied together to scene interpretation of road traffic scenarios for more reliable systems.

Context and world modelling
Exploiting context has become more and more important in different technical and scientific disciplines. The inclusion of global knowledge into the system is used to improve and extend its functionality. Figure 1 shows a rectangular object on the left side. Without involving the surrounding context information (that we are on a road) the identification of the rectangular object as a lorry would have been almost impossible. In [15] three types of context are distinguished: physical, photogrammetric and computational context. The physical context covers information about the visual world that is independent of any particular set of image acquisition conditions. Weather conditions and geography infor-

mation, for instance, belong to the physical context. The road traffic context is part of this class. If it rains then the recognition algorithm needs a rain-removal filter to improve accuracy. The geographic information supports lane detection, because the system can guess in advance where the road is going to appear in the image. The photogrammetric context encompasses information surrounding the acquisition of the image currently under study. Beside camera parameters, location and orientation this category additionally includes date and time of the image acquisition. From the time information we can deduce lightning conditions and use different algorithms for day- or night-time (dusk/dawn). Computational context covers among others information about the internal state of processing. Urgency is an example for this category. Due to limited computational power different recognition tasks are prioritized. In a car it is advantageous to recognize near cars in front earlier than those that are far away.

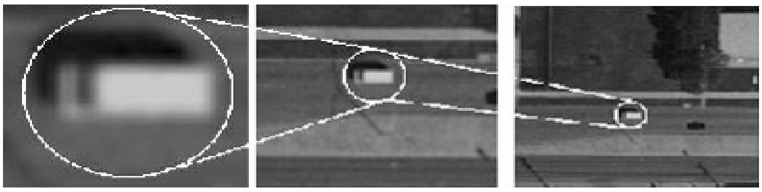

Figure 1: Exploiting context for object recognition [1]

Context information helps to resolve ambiguities in the image understanding process. Without including global knowledge into the recognition process the rectangular object from Figure 1 on the left, just to take an example, could also be a zoomed view of an eraser on the desk.

In [14] an approach is proposed where a camera-equipped robot maintains a spatial and conceptual hierarchy of the environment. The conceptual model represents the real world knowledge of the robot (e.g. a flat). A recognized object from the visual system is matched against the model to have an anchoring from the real world to the model. With reasoning over the model new knowledge is deduced. For instance, "*we have detected a sofa, so we are very likely standing in the living room*". In the scenario of road traffic such a model-based method can be used for improvement of object recognition. A car driving in front of us has the shape of a rectangle, more or less. Due to the knowledge that the license plate is somewhere in the boundary rectangle of the car, the CV algorithm can use this information to narrow down the search space of the recognition process for the license plate.

Learning and Adaptation

The CogVis [11] project (Cognitive Vision) was a European Union funded collaborative project which ran from 2001 to 2004 to study the design of cognitive vision systems. In this context a "cognitive vision system" is defined as a system that uses visual information to achieve learning and adaptation.

Vision-based road following systems in cars detect the road in front of it and keep the vehicle on the road it travels on. A robust vision-based road following system must handle difficult road conditions, e.g. unstructured roads, unclear road edges, non-homogeneous road appearance, arbitrary road shape, poor and inconsistent lightning conditions, paved and unpaved roads. In [5] a vision system is proposed that is capable of dealing with many of these limitations. The system uses color classification and learning to construct and use multiple road and background models. Every frame is segmented into road and background through probabilistic models. Based on this information the models are constructed and learned independently of road shape during the driving task.

In [6] Ross and Kaelbling present initial results of a systematic approach to learning and object segmentation from motion. They state that motion segmentation of objects is a more primitive task then the detection of object boundaries by static image cues, because motion information provides a plausible supervision signal for learning the static boundary detection task. On a moving platform there is a strong interrelationship between successive frames captured from a camera in driving direction. The closer we come to an object the better we can recognize what the object is. This effect offers the opportunity to learn from previous frames and design an adaptive algorithm to classify objects in the future by self-supervised learning from history.

Tracking of objects in video streams requires following the object in a continuous stream of successive frames. An object recognized in a frame needs to be found and recognized again in successive frames. If the object is moving, it will appear in successive frames at different locations. A possible approach to track the object is to first recognize it in every frame of the video stream, determine the position and generate the object's trajectory out of this data. A physical possible object must satisfy the following spatio-temporal constraints according to [7]:

a) *Exclusivity:*
   an object cannot be in more than one place at the same time;
b) *Continuity:*
   an object’s movement must be continuous (i.e. it cannot instantaneously ‘jump’ from one place to another).

The continuity constraint enables computer vision algorithms to reduce the search-space where they have to look for the tracked object in successive frames, because the object has to be somewhere in the local surrounding of the last position.

A learning algorithm for real-time vehicle tracking is shown in [4]. It uses an improvement of a feature selection method known from object detection. A good feature necessary for object detection is not necessarily a good one for object tracking. The challenge is to learn online which features have the best tracking power.

Conclusion

Intelligent vision-enabled systems for the usage on-road are capable of perceiving the environment like human-beings and react accordingly. The fundamental requirements are robustness and adaptability for the reliable application to real-world road traffic scenarios. While exclusively CV-based approaches focus on low-level image processing tasks, the extension of CV with concepts of AI leads to high-level processing and more intelligent systems. The inclusion of context information, world modelling and learning methods is the enabling factor for robust, real-time vision systems in highly changing environments, like on the road.

References

[1] Christopher O. Jaynes, **Seeing is Believing: Computer Vision and Artificial Intelligence**, *ACM Crossroads 3.1, http://www.acm.org/crossroads/xrds3-1/vision.html, 16.11.2006*

[2] Frank Dellaert and David Bruemmer, **Semantic SLAM for Collaborative Cognitive Workspaces**, *AAAI Fall Symposium Series 2004: Workshop on The Interaction of Cognitive Science and Robotics: From Interfaces to Intelligence, 2004*

[3] Peter Meer and Charles V. Stewart and David E. Tyler, **Robust computer vision: an interdisciplinary challenge**, *ACM Computer Vision and Image Understanding, vol. 78, issue 1, pp. 1-7, 2000*

[4] Daniel Withopf and Bernd Jähne, **Learning Algorithm for Real-Time Vehicle Tracking**, *Proceedings of the IEEE ITSC 2006, IEEE Intelligent Transportation Systems Conference Toronto, Canada, September 17-20, 2006*

[5] Ceryen Tan, Tsai Hong, Tommy Chang, and Michael Shneier, **Color-Model-Based Real-Time Learning for Road Following**, *Proceedings of the IEEE ITSC 2006, IEEE Intelligent Transportation Systems Conference Toronto, Canada, September 17-20, 2006*

[6] Michael G. Ross and Leslie Pack Kaelbling, A Systematic Approach to Learning Object Segmentation from Motion, *NIPS Workshop on Open Challenges in Cognitive Vision*, *Vancouver, 2004 (NIPS-03).*

[7] Bennett, Magee, Cohn and Hogg, **Using spatio-temporal continuity constraints to enhance visual tracking of moving objects** in: Lopez de Mantaras, R & Saitta, L (editors) ECAI 2004 Proceedings of the 16th European Conference on Artificial Intelligence, pp. 922-926 IOS Press. 2004.

[8] Galata, Cohn, Magee and Hogg, **Modeling Interaction Using Learnt Qualitative Spatio-Temporal Relations and Variable Length Markov Models**, *In Proc. European Conference on Artificial Intelligence (ECAI), ECAI'02, July 2002*

[9] Miguel Angel Garcia-Garrido, Miguel Angel Sotelo and Ernesto Martin-Gorostiza, **Fast Traffic Sign Detection and Recognition under Changing Lighting Conditions**, *Proceedings of the IEEE ITSC 2006, IEEE Intelligent Transportation Systems Conference Toronto, Canada, September 17-20, 2006*

[10] Amnon Shashua, Yoram Gdalyahu, Gaby Hayun, **Pedestrian Collision Avoidance Systems: A Survey of Computer Vision Based Recent Studies**, *Proceedings of the IEEE ITSC 2006, IEEE Intelligent Transportation Systems Conference Toronto, Canada, September 17-20, 2006*

[11] **European Research Network for Cognitive Computer Vision Systems**, http://www.ecvision.org, 15.11.2006

[12] J. K. Aggarwal, Joydeep Ghosh, Dinesh Nair and Ismail Taha, **A Comparative Study of Three Paradigms for Object Recogntion – Bayesian Statistics, Neural Networks and Expert Systems**, *Advances in Image Understanding, K. Bowyer and N. Ahuja, Eds, IEEE Computer Society Press, pp. 241-262, 1996*

[13] Sinem C. Ergen, Sing Yiu Cheung, Pravin Varaiya, Robert Kavaler, Amine Haoui, **Wireless Sensor Networks for Traffic Monitoring**, *IPSN 2005, April 2005*

[14] C. Galindo, A. Saffiotti, S. Coradeschi, P. Buschka, J.A. Fernández-Madrigal and J. González, **Multi-Hierarchical Semantic Maps for Mobile Robotics,** *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS), pp. 3492-3497. Edmonton, CA, 2005.*

[15] Strat, T.M., **Employing Contextual Information in Computer Vision**, DARPA93, pp. 217-229, 1993

[16] **CARE – European Road Accident Database**, *http://ec.europa.eu/transport/roadsafety/road_safety_observatory/care_reports_en.htm, 15.11.2006*

[17] Zehang Sun, George Bebis, Ronald Miller, **On-Road Vehicle Detection: A Review**, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 5, pp. 694-711, May, 2006.