The Mobile Revolution - Machine Intelligence for Autonomous Vehicles

Markus Enzweiler

Abstract: What started as a distant vision just a few decades ago is quickly becoming reality. Autonomous vehicles are about to be deployed on a large scale and will fundamentally change our transportation behavior. In this particular application, extreme demands on reliability and quality give rise to numerous problems and open issues that need to be jointly identified and addressed by both academia and industry.
In this article, we present an overview of the current state-of-the-art in the field of intelligent autonomous vehicles. We further discuss open problems and current research directions.

ACM CCS: Computer systems organization → Embedded and cyber-physical systems → Robotics; Computing methodologies → Artificial intelligence → Computer vision → Computer vision tasks → Scene understanding

Keywords: Computer vision, machine learning, robotics, intelligent vehicles, autonomous driving, scene understanding

1 Introduction

Machine intelligence, robotics and computer vision, formerly rather peripheral disciplines of computer science, are in fact already with us today and have a familiar embodiment - the modern vehicle. Systems that are currently available strongly couple interdisciplinary fundamental research with complex practical realizations.

The vision of autonomous vehicles in particular has a surprisingly long history with first prototypical implementations going back to the early 1980s [1]. What started then as a dream of pioneers such as Ernst Dickmanns is actually happening right now - we are on the verge of a mobile revolution with self-driving vehicles as its central foundation. The tremendous progress made in the last years has been sparked by the increased methodical and technically availability of better sensors, sophisticated algorithms, faster computers and more data.

However, we are not quite there yet. Autonomous systems make extreme demands on system performance, quality, availability, reliability and verification that significantly increase with the rising degree of automation. The ultimate design goal for autonomous systems is to mimic human behavior in terms of understanding and effortlessly acting within a dynamic human-inhabited environment. One could refer to this as the *Turing test for autonomous systems*. Although artificial sensors emulating the human sensory systems are nowadays widely available, current autonomous systems are still far behind humans in terms of understanding and interacting in real-world environments. The chief reason is the (theoretical and practical) unavailability of methods to reliably perform perception, recognition, understanding and action on a broad scale, i.e. not limited to isolated problems.

Following the classical perception-action cycle, the central problems evolve around three key questions posed from the perspective of an autonomous vehicle:

Figure 1: Mercedes-Benz S500 Intelligent Drive (*Bertha*) with well-integrated close-to-production sensors driving fully autonomously on open public roads [3].

Figure 2: Semantic scene model recovered using stereo vision-based scene understanding. Different colors encode distinct semantic object classes: road surface (brown), humans (red), vehicles (green), and buildings (blue) [4].

- What do I perceive and how can I interpret this?
- Where am I and what do I do next?
- How can I build up experience and learn?

2 Intelligent Robotics

Starting with initial careful steps and continuing up to today, robotics has been the research field in computer science most relevant to intelligent vehicles. It focuses on the fundamental building blocks of autonomous robots embedded in real-world environments, i.e. perception, (re-)action, learning, as well as interaction with the surroundings.

For **scene perception** and **scene understanding**, both previous and current prototypical autonomous vehicles make prominent use of artificial sensors such as radar, lidar, or camera systems to extract information from their environment [2]. Recent research trends have shown a steady shift towards favoring machine vision systems - mostly in a stereo setup [3] - that mimic the human visual system, which has developed to be our primary and most important sensory cue. Among the chief reasons for this trend is the tremendous amount of information that can be extracted from image data coupled with a very high generality in the recovered model. While beam-based sensors such as radar and lidar typically recover a very object-centric representation of a scene, i.e. in terms of position, size and velocity of objects, machine vision allows extracting a much richer representation including a concise semantic description of object and infrastructure types, e.g. pedestrian, vehicle, or road surface. Although simpler object-focused models have successfully been deployed for certain applications, e.g. adaptive cruise control, they are certainly not sufficient for ambitious applications such as autonomous vehicles.

At this point, current prototypical intelligent vehicles heavily leverage recent advances made in the computer vision community, particularly in the field of robust visual scene understanding. For most object classes present in real-world scenes there are no explicit models readily available that fully represent the structure and characteristics of that particular object category. This has spawned the use of machine learning and pattern recognition techniques that learn an implicit model of object categories from large sets of example images. Although tremendous progress has been made in this area, resulting in very detailed scene models recovered in real-time from real-world images [4], see Figure 2, there is still a great need for further research. Taking human recognition performance as a benchmark, current artificial systems paint a significantly inferior performance picture. Systems are plagued by the sheer number of different objects in a scene coupled with the high variability in object appearance, diverse environmental conditions and heavily cluttered backgrounds. At the same time, artificial sensory systems are subject to very high performance demands at low computational costs.

Intelligent machines do not only need to perceive and understand individual parts of a scene but are required to develop **situation awareness**. They must be able to combine possibly diverse sensory cues into a consistent higher-level model of the situation as a whole. Situation awareness involves not only an interpretation of current events but also a prediction of how both the environment will change and what effects the own actions will have in the future. To again draw the same comparison as above, human situation awareness heavily relies on knowledge, expectation, and experience of different situations resulting from a (possibly never-ending) learning process. This ability and versatility enables us to learn a large variety of tasks, to rapidly acquire new capabilities, and to make sensible decisions in most situations. Despite recent advances in artificial intelligence, such lifelong machine learning remains a largely unsolved problem.

Situation interpretation and awareness is the core necessity for a machine to **properly act in human environments**. Autonomous vehicles need to fit into human traffic without attracting any attention. They need to comply with the given rules, deduce the correct actions, plan ahead, and be able to act and react in a fraction of a second. In the event of failure of some of its components, a graceful system degradation is required, so that the vehicle is able to either continue to operate properly or bring itself to a safe operating state. This poses stringent requirements not only on the algorithmic system design but also on the corresponding hardware components.

Finally, besides understanding and acting, intelligent robotics also deals with **interaction**, i.e. the interface between the machine and the environment. Current research in this domain involves balancing sensory perception on the one hand with connectivity-based approaches on the other hand. The communication of intelligent vehicles with other traffic participants and traffic infrastructure constitutes an additional source of information that is mainly orthogonal to local sensory perception. Many experts believe that the adequate combination of those two domains is one key towards large-scale deployment of autonomous vehicles.

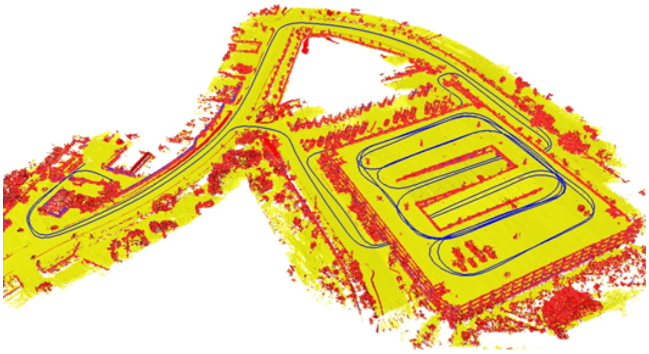

Figure 3: Example of self-localization within an on-line generated digital map of a parking area [5].

3 Digital Maps

Originally a rather minor sub-domain of robotics, digital maps have become a foundational component of many current prototypical autonomous vehicles. Such maps contain a wealth of information about static properties of the environment, such as the lane layout and the location of infrastructure elements, e.g. traffic signs, traffic lights, or pedestrian crossings. Within autonomous systems, maps are used to precisely determine the current position (self-localization) and to support path planning.

This level of importance put on digital maps in today's research community is rather astounding, given that humans are capable of navigating complex and dynamic environments without such explicit prior information. This is another clear indication that a significant perception gap exists between machines and humans that is currently attenuated with map-based information.

In view of the above, the community is currently engaged in issues dealing with quality requirements on self-localization and map content. Even more fundamental questions involve techniques for off-line and on-line map generation, for maintenance of maps, and for integrating and interconnecting sensory and map-based information on various system levels.

4 Verification and Validation

Major open issues that might put the large-scale deployment of intelligent autonomous vehicles into jeopardy involve **system verification and validation**. Here, system verification aims to ensure that the developed systems operate correctly as specified. System validation implies that the autonomous systems operate as the society expects. In this regard, similar requirements apply to software components, hardware components, and the integrated system as a whole.

One can deduce several key issues that can be summarized in a single question: How can we both determine and ensure reliability, availability, maintainability, trustworthiness, and safety of autonomous intelligent vehicles? Looking at current driver assistance systems on the market, established verification and validation strategies couple formal requirements on development processes and individual components with concepts to evaluate actual system realizations iteratively in operation on the road. Particularly the latter strategy requires immense time and effort, as the underlying systems are currently evaluated on a scale of millions of kilometers driven.

The fundamental problem with most current evaluation approaches is that they do not scale well with increased system complexity and system requirements, possibly requiring hundreds of millions of kilometers of testing to prove safety and reliability of fully autonomous intelligent vehicles. At this point, alternative evaluation concepts are needed with one promising candidate being system simulation. Simulation strategies that are actively discussed in the intelligent vehicles community today range from the low-level simulation of individual system components, e.g. sensors, up to the high-level simulation of the behavior of fully integrated systems in many different scenarios. Although being quite diverse, the underlying approaches share common requirements and characteristics: They need to deliver comprehensible, realistic, and reproducible results that are transferable to real-world situations without any restrictions.

5 Conclusion

Currently, there exists a variety of possible deployment and adoption scenarios for autonomous intelligent vehicles that differ considerably in time and scale. It is hence necessary to develop a clear understanding on what the current limits are that prevent deploying autonomous vehicles on a large-scale. Is it the previously mentioned perception gap? Or rather open issues with system evaluation? Is there a critical mass for large-scale deployment to become beneficial? How can trust in the systems be built up?

Potential ideas and hints to answer those questions could come from turning towards related disciplines of intelligent transportation systems other than cars. Autonomous trains, airplanes, and ships for example are already in operation on a considerable scale. Although the applications differ significantly at first glance, there are synergies that can be exploited. Such interdisciplinary relations need to be identified and concretized for the individual sub-communities to learn from each other.

Literature

[1] E. D. Dickmanns, B. Mysliwetz, and T. Christians. *An integrated spatio-temporal approach to automatic visual guidance of autonomous vehicles*. IEEE Transactions on Systems, Man, and Cybernetics, 20(6):1273–1284, 1990.

[2] J. Ziegler et al. *Making Bertha Drive - An Autonomous Journey on a Historic Route*. IEEE Intelligent Transportation Systems Magazine, 6(2):8–20, 2014.

[3] U. Franke, D. Pfeiffer, C. Rabe, C. Knöppel, M. Enzweiler, F. Stein, and R. G. Herrtwich. *Making Bertha See*. ICCV Workshop on Computer Vision for Autonomous Driving, 2013.

[4] T. Scharwächter, M. Enzweiler, U. Franke and S. Roth. *Stixmantics: A Medium-Level Model for Real-Time Semantic Scene Understanding*. European Conference on Computer Vision (ECCV), 2014.

[5] G. Grisetti, C. Stachniss, and W. Burgard. *Non-linear Constraint Network Optimization for Efficient Map Learning*. IEEE Transactions on Intelligent Transportation Systems, 10(3):428–239, 2009.

Dr. Markus Enzweiler received the MSc degree in computer science from the University of Ulm, Germany (2005), and the PhD degree in computer science from the University of Heidelberg, Germany (2011). Since 2010, he has been a research scientist at Daimler AG Research & Development in Sindelfingen, Germany, where he co-developed the Daimler vision-based pedestrian detection system which is available in Mercedes-Benz cars. His current research focuses on statistical models of object appearance with application to object recognition, scene understanding and autonomous driving in the domain of intelligent vehicles. He held graduate and PhD scholarships from the Studienstiftung des deutschen Volkes (German National Academic Foundation). In 2012, he received both the IEEE Intelligent Transportation Systems Society Best PhD Dissertation Award and the Uni-DAS Research Award for his work on vision-based pedestrian recognition. He is part of the team that won the 2014 IEEE Intelligent Transportation Systems Outstanding Application Award. In 2014, he was honored with a Junior-Fellowship of the Gesellschaft für Informatik.

Address: Daimler AG Research & Development, Environment Perception, 71059 Sindelfingen, Germany, E-Mail: markus.enzweiler@daimler.com