Lecture Notes in Mobility

Tim Schulze
Beate Müller
Gereon Meyer *Editors*

Advanced Microsystems for Automotive Applications 2016

Smart Systems for the Automobile of the Future

EPoSS
European Technology Platform
on Smart Systems Integration

VDI | VDE | IT

Springer

Lecture Notes in Mobility

Series editor

Gereon Meyer, Berlin, Germany

More information about this series at http://www.springer.com/series/11573

Tim Schulze · Beate Müller
Gereon Meyer
Editors

Advanced Microsystems for Automotive Applications 2016

Smart Systems for the Automobile of the Future

Springer

Editors
Tim Schulze
VDI/VDE Innovation + Technik GmbH
Berlin
Germany

Gereon Meyer
VDI/VDE Innovation + Technik GmbH
Berlin
Germany

Beate Müller
VDI/VDE Innovation + Technik GmbH
Berlin
Germany

ISSN 2196-5544 ISSN 2196-5552 (electronic)
Lecture Notes in Mobility
ISBN 978-3-319-44765-0 ISBN 978-3-319-44766-7 (eBook)
DOI 10.1007/978-3-319-44766-7

Library of Congress Control Number: 2016947919

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Automated vehicles are currently raising considerable attention of politics and industry alike. Although public funding of research and innovation on automated driving dates back as far as the 1980s, interest has been peaking very recently following the demonstration of advanced technologies by key players that seemingly is at the verge of market-readiness. In effect, fierce competition is unfolding between companies regarding the enabling technologies like sensors, powerful automotive ECUs and actuators. At the same time, governments are debating on the legal and infrastructural preconditions of automated driving and about how to harmonize the necessary investments. Yet, the recent first fatal incident with a beta-state self-driving vehicle remind us that the current technology is not yet capable of mastering all complex situations.

The current developments imply two major challenges: Firstly, research and innovation efforts need to be shifted from proof-of-concept to proof-of-safety on the system level of automated driving technology. For instance, the performance envelope of sensors, data fusion and object recognition systems has been pushed considerably in recent years, but does still not cover the complexity that a vehicle encounters in everyday life. Particularly for applications in urban environments and at higher levels of automation, perception of the driving environment is a challenging task still to be mastered in a robust fashion. Smart systems integration that applies a variety of vehicle and infrastructure based sensor systems and finally links it with big data analytics in the cloud will play an important role in this domain, with safety aspects as well as validation methodology becoming a prominent future focus.

Secondly, it is obvious that the economically viable large-scale rollout of driving automation requires involvement and agreement between a large and heterogeneous group of stakeholders encompassing the automotive, IT and telecom sectors as well as road infrastructure providers and public authorities. To this end, a first initiative was launched in Sept. 2015 by Commissioner Oettinger with the Round Table Automated Driving. The recently started Coordination and Support Action "Safe and Connected Automation of Road Transport (SCOUT)" will develop a

cross-sectorial roadmap regarding the implementation of high-degree automated driving in Europe, which will assist the ambition of the roundtable and may serve as a blueprint for research and innovation planning and regulatory actions in the coming years.

The International Forum on Advanced Microsystems for Automotive Applications (AMAA) has been exploring the technological foundations of connected, automated and electrified vehicles for many years. Consequently, the topic of this year's 20th anniversary edition of AMAA, held in Brussels on 22–23 September 2016, is "Smart Systems for the Automobile of the Future". The AMAA organisers, Berlin-based VDI/VDE Innovation + Technik GmbH together with the European Technology Platform on Smart Systems Integration (EPoSS), greatly acknowledge the support given for this conference, particularly from the European Union through the Coordination Actions "Safe and Connected Automation of Road Transport (SCOUT)" and "Action Plan for the Future of Mobility in Europe (Mobility4EU)".

The papers in this book, a volume of the Lecture Notes in Mobility book series by Springer, were written by leading engineers and researchers who have attended the AMAA 2016 conference to report their recent progress in research and innovation. The papers were selected by the members of the AMAA Steering Committee and are made accessible worldwide. As the organisers and the chairman of the AMAA 2016, we would like to express our great appreciation to all the authors for their high-quality contributions to the conference and also to this book. We would also like to gratefully acknowledge the tremendous support we have received from our colleagues at VDI/VDE-IT.

Berlin, Germany
August 2016

Tim Schulze
Beate Müller
Gereon Meyer

Supporters and Organisers

Funding Authority
European Commission

Supporting Organisations
European Council for Automotive R&D (EUCAR)
European Association of Automotive Suppliers (CLEPA)
Strategy Board Automobile Future (eNOVA)
Advanced Driver Assistance Systems in Europe (ADASE)
Zentralverband Elektrotechnik- und Elektronikindustrie e.V. (ZVEI)
Mikrosystemtechnik Baden-Württemberg e.V.

Organisers
European Technology Platform on Smart Systems Integration (EPoSS)
VDI/VDE Innovation + Technik GmbH

Steering Committee

Mike Babala, TRW Automotive, Livonia, MI, USA
Serge Boverie, Continental AG, Toulouse, France
Geoff Callow, Technical & Engineering Consulting, London, UK
Kay Fürstenberg, Sick AG, Hamburg, Germany
Wolfgang Gessner, VDI/VDE-IT, Berlin, Germany
Roger Grace, Roger Grace Associates, Naples, FL, USA
Klaus Gresser, BMW Forschung und Technik GmbH, Munich, Germany
Riccardo Groppo, Ideas & Motion, Cavallermaggiore, Italy
Hannu Laatikainen, Murata Electronics Oy, Vantaa, Finland
Jochen Langheim, ST Microelectronics, Paris, France
Günter Lugert, Siemens AG, Munich, Germany
Steffen Müller, NXP Semiconductors, Hamburg, Germany
Roland Müller-Fiedler, Robert Bosch GmbH, Stuttgart, Germany
Andy Noble, Ricardo Consulting Engineers Ltd., Shoreham-by-Sea, UK
Pietro Perlo, IFEVS, Sommariva del Bosco, Italy
Christian Rousseau, Renault SA, Guyancourt, France
Jürgen Valldorf, VDI/VDE-IT, Berlin, Germany
David Ward, MIRA Ltd., Nuneaton, UK

Conference Chair:
Gereon Meyer, VDI/VDE-IT, Berlin, Germany

Contents

Part I
Networked Vehicles & Navigation

Requirements and Evaluation of a Smartphone Based Dead Reckoning Pedestrian Localization for Vehicle Safety Applications

Johannes Rünz, Folko Flehmig, Wolfgang Rosenstiel and Michael Knoop

Abstract The objective of this paper is to propose a smartphone based outdoor dead reckoning localization solution, show its experimental performance and classify this performance into the context of Vehicle-to-X (V2X) based pedestrian protection systems for vehicle safety applications. The proposed approach estimates the position, velocity and orientation with inertial measurement unit (IMU) sensors, a global navigation satellite system (GNSS) receiver and an air pressure sensor without any restriction to pedestrians, like step length models or a relationship between smartphone orientation and walking direction. Thus, an application makes sense in handbags, trouser pockets or school bags. The dead reckoning localization filter was evaluated in measurements and compared with a highly accurate reference system. Furthermore, the requirements of measurement and modelling uncertainties in a pedestrian protection system with a low false-positive rate were derived and compared with the reference measurements. The results demonstrated that an appropriate use of the proposed system approach is only possible with more accurate positioning solutions from the GNSS receiver. According to this the necessity of differential GNSS methods was predicted.

J. Rünz (✉) · F. Flehmig · M. Knoop
Chassis Systems Control, Robert Bosch GmbH,
Robert-Bosch-Allee 1, 74232 Abstatt, Baden-Württemberg, Germany
e-mail: johannes.ruenz@de.bosch.com

F. Flehmig
e-mail: folko.flehmig@de.bosch.com

M. Knoop
e-mail: michael.knoop@de.bosch.com

W. Rosenstiel
Computer Engineering Department, Eberhard Karls Universität Tübingen,
Auf der Morgenstelle 8, 72076 Tübingen, Baden-Württemberg, Germany
e-mail: rosenstiel@informatik.uni-tuebingen.de

M. Knoop
Institute of Measurement and Control, Karlsruhe Institute of Technology,
Karlsruhe, Germany

T. Schulze et al. (eds.), *Advanced Microsystems for Automotive Applications 2016*, Lecture Notes in Mobility, DOI 10.1007/978-3-319-44766-7_1

Keywords Smartphone based pedestrian protection · Pedestrian dead reckoning · Vehicle safety

1 Introduction

To improve advanced pedestrian protection systems in future vehicles, it is necessary to obtain better knowledge about the pedestrian behavior and the perception to recognize pedestrians in the vehicle's environment [1]. Situations which include visually obstructed pedestrians or pedestrians that cannot be detected by the limited field of view (e.g. in curves), are especially challenging. A possible solution is the cooperative usage of the pedestrian's smartphone with its internal localization sensors and V2X technology to better protect the pedestrian. Besides, the modeling of the pedestrian's intention and behavior might be supported by information from the smartphone. General analysis about the possible system architectures, communication strategies as well as limitations and feasible benefits of pedestrians' smartphones as cooperative sensors are described in [2–5]. Moreover, an experimental proof of benefit for safety applications with bicycles using their smartphones is presented in [6]. The author states that the accuracy of the smartphone's global navigation satellite system (GNSS) receiver is improved with a map-matching algorithm and context knowledge.

As later described in this work, the requested low false-positive rates of pedestrian protection systems lead to strict requirements on the position accuracy of the pedestrian. A survey is given in [7], which shows the accuracy of smartphone GNSS sensors and compares them with a Differential GNSS (DGNSS) system. According to [8], the GNSS sensor is combined with a velocity and orientation estimation through step detection, a gyroscope and a magnetic field sensor in order to further improve the localization performance. A similar approach to improve the GNSS measurements in outdoor environments is used by [9]. The challenge of such approaches is that the orientation of the smartphone must be rigidly coupled with the walking direction, which is not given in its daily use. Thus, the sensor data fusion can deteriorate the accuracy of the GNSS measurements in wrong use cases.

Typical localization problems occur in the field of aerospace engineering with six degrees of freedom: three position and three orientation dimensions. In these surroundings, highly accurate acceleration sensors and gyroscopes are used to derive the position, velocity and orientation of an object. Furthermore in inertial navigation systems (INS) the determined position, velocity and orientation are corrected by GNSS measurements with e.g. a Kalman filter (KF). In [10] implementations of such systems are presented. These methods are then adopted to a drone equipped with MEMS acceleration and gyroscope sensors in [11].

In this work, methods from aerospace engineering are evaluated for INS by using the smartphone's internal sensors. The ambition is to show the localization performance of such a system and classify the results in the context of vehicle safety applications for pedestrians.

First, in Sect. 2 the dead reckoning localization filter structure with its propagation and observation models is described. Furthermore, in Sect. 3 the evaluation procedure with its measurement setup is presented. Finally, Sect. 4 shows the performance of the considered localization filter and classifies the results into the context of pedestrian protection systems.

2 Localization Estimation Filter

In this work, a general approach from the aerospace engineering described in [10, 11], without assumptions about the relationship between walking direction and smartphone orientation or step length, is used. In aerospace engineering, a filter that estimates the position, velocity and orientation, is called *navigation filter*. Because of the different meaning of "navigation" in the vehicle context, these filters will be called localization filters in this paper.

Figure 1 shows the reduced structure of the estimation filter. The output of the dead reckoning estimation filter consists of the estimated position $\hat{\boldsymbol{p}}^n_{bn}$, the estimated velocity $\hat{\boldsymbol{v}}^n_{bn}$ and the estimated orientation $\hat{\boldsymbol{q}}^n_{bn}$ given in quaternion representation. As inputs, an inertial measurement unit (IMU) with its measurements $\boldsymbol{f}^b_{ib}$ and $\boldsymbol{\omega}^b_{ib}$, a GNSS receiver and a barometer are used.

To propagate the position, velocity and orientation with an IMU two different methods are known. First the acceleration sensors can be mounted on a Gyro-stabilized gimbal platform, which leads to a fixed orientation of the acceleration sensors, to the earth. The other method is, that the acceleration is rigidly coupled with the body ("strap-down"), which leads to a more complex problem to interpret the measurements. The *strapdown algorithm* determines the updated position $\hat{\boldsymbol{p}}^n_{bn}$, velocity $\hat{\boldsymbol{v}}^n_{bn}$ and orientation $\hat{\boldsymbol{q}}^n_{bn}$ with the corrected measurements $\boldsymbol{f}^b_{ib}$

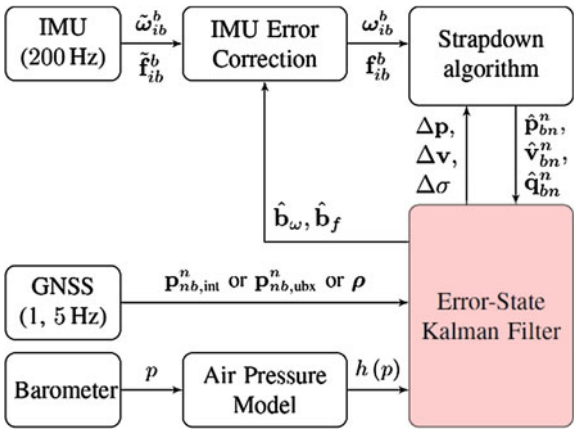


Fig. 1 Schematic representation of the localization filter structure

and $\boldsymbol{\omega}_{ib}^{b}$ from the rigidly coupled acceleration sensor and gyroscope of the smartphone. This is done by three integration steps and compensation of parasitic influences like the measured earth gravity. Because of the noncommutativity of rotations and the Nyquist Shannon theorem, the strapdown algorithm is sampled with a high frequency of 200 Hz. As a consequence of the strapdown algorithm, the measured acceleration in and the rotation rate around the navigation coordinate system axis (east-, north- and up-axis) can be directly determined. To have a long term stability and accuracy, the estimated values must be corrected with absolute position measurements. Therefore, an extended error-state KF uses GNSS and barometric pressure measurements to estimate the position errors $\Delta\boldsymbol{p}$, velocity errors $\Delta\boldsymbol{v}$ and orientation errors $\Delta\boldsymbol{\sigma}$, which are used to correct the absolute states $\hat{\boldsymbol{p}}_{bn}^{n}$, $\hat{\boldsymbol{v}}_{bn}^{n}$ and $\hat{\boldsymbol{q}}_{bn}^{n}$. Furthermore, the KF estimates the parameters $\boldsymbol{b}_{\omega}$ and $\boldsymbol{b}_{f}$ of the measurement error models.

It is also possible to observe the orientation with a magnetic field sensor and earth magnetic field assumption. Due to unpredictable disturbances by ferromagnetic materials in the environment, like vehicles, this method is not used in this paper.

In the following, the sensor error models, the propagation error-state model and the observations for correcting the KF error-state are presented.

2.1 Sensor Error Models and Impact of the Error Terms

The acceleration sensor should measure the specific force $\boldsymbol{f}_{ib}^{b}$ while the gyro measures the rotation rate $\boldsymbol{\omega}_{ib}^{b}$. The smartphone microelectromechanical systems (MEMS) sensors have time-variant bias sensor errors $\hat{\boldsymbol{b}}_{\omega}$ and $\hat{\boldsymbol{b}}_{f}$, which must be estimated by the KF. Furthermore, there are scaling errors $\hat{\boldsymbol{s}}_{\omega}$ and $\hat{\boldsymbol{s}}_{f}$, that are assumed as time-invariant. The error models are then given by

$$\boldsymbol{\omega}_{ib}^{b} = \mathrm{diag}(\hat{\boldsymbol{s}}_{\omega})\tilde{\boldsymbol{\omega}}_{ib}^{b} + \hat{\boldsymbol{b}}_{\omega} \quad \text{and} \tag{1}$$

$$\boldsymbol{f}_{ib}^{b} = \mathrm{diag}\left(\hat{\boldsymbol{s}}_{f}\right)\tilde{\boldsymbol{f}}_{ib}^{b} + \hat{\boldsymbol{b}}_{f}. \tag{2}$$

In this work, misalignment errors of the assembled IMU sensors are neglected.

The time-variant bias errors of MEMS IMUs lead to significant errors in the position propagation with the strapdown algorithm. Especially bias errors of the gyroscope result in wrong orientations that couple the gravity acceleration into the two integration steps for position determination. For this reason the performance of the strapdown algorithm is highly dependent on the bias errors estimation accuracy.

2.2 Error-State Model

The model has the error-state vector

$$\boldsymbol{x} = \left(\Delta\boldsymbol{p}, \Delta\boldsymbol{v}, \Delta\boldsymbol{\sigma}, \Delta\boldsymbol{b}_{\omega}, \Delta\boldsymbol{b}_{f}, \Delta c\boldsymbol{T}\right)^{T}, \tag{3}$$

where $\Delta\boldsymbol{b}_{\omega}$ and $\Delta\boldsymbol{b}_{f}$ are the errors of the IMU biases and

$$\Delta c\boldsymbol{T} = c(\delta \mathrm{t}_{\mathrm{rec}}, \delta \dot{t}_{rec})$$

is the GNSS receiver's clock error δt_{rec} and its drift $\delta \dot{t}_{rec}$, which are essential to consider the raw measurements of the GNSS receiver. The error propagation model is in-depth derived in [11] and given in its continuous form by

$$\dot{\boldsymbol{x}} = \begin{pmatrix} \mathbf{0} & \boldsymbol{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -(\boldsymbol{C}_{b}^{\hat{n}}\boldsymbol{f}_{ib}^{i}\times) & -\boldsymbol{C}_{b}^{\hat{n}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\boldsymbol{C}_{b}^{\hat{n}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{F}_{\Delta c\boldsymbol{T}} \end{pmatrix} \boldsymbol{x}, \tag{4}$$

where $\boldsymbol{I}$ is the 3×3 identity matrix, $\boldsymbol{C}_{b}^{\hat{n}}$ is the direct cosine matrix to rotate the IMU coordinates to the estimated navigation frame, $\boldsymbol{f}_{ib}^{i}\times$ is the skew-symmetric matrix representation of the vector cross product of $\boldsymbol{f}_{ib}^{i}$ and $\boldsymbol{F}_{\Delta cT}$ is given by

$$\boldsymbol{F}_{\Delta cT} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}. \tag{5}$$

The absolute states are always corrected with the estimated errors after every observation step of the KF. Afterwards, the error-states are reset to zero, so that a nonlinear error propagation is not necessary.

2.3 Observation Models

For correcting the determined position, velocity and orientation of the strapdown algorithm, the measurements of a GNSS receiver and a barometer are used.

Internally, a GNSS receiver measures the geometric distances to the satellites which are summed up with other error terms. In particular the clock error $c\boldsymbol{T}$ of the GNSS receiver produces a significant error. The measurement of the geometric distance, added by the clock error term, is referred to as pseudorange $\tilde{\rho}_i$ of satellite i. From a set of pseudoranges of different satellites, the GNSS receiver's position solution $\tilde{\boldsymbol{p}}_{GNSS}^{n}$ can be deduced.

In a localization filter, either the pseudoranges in a *tightly coupled filter* concept or the position solution in a *loosely coupled filter* concept can be used to improve the estimation. In the next sections, the different observation matrixes are presented.

2.3.1 Loosely Coupled GNSS Measurement

In a loosely coupled filter the position solution $\tilde{\boldsymbol{p}}^n_{GNSS}$ GNSS is used as an observation for the KF's correction step. The observation matrix $\boldsymbol{H}_{p,GNSS}$ is given in [11] by

$$\boldsymbol{H}_{p,GNSS} = (\boldsymbol{I} \quad \boldsymbol{0} \quad \boldsymbol{0} \quad \boldsymbol{0} \quad \boldsymbol{0} \quad \boldsymbol{0}), \tag{6}$$

where $\boldsymbol{I}$ is a 3×3 identity matrix and $\boldsymbol{0}$ are zero matrices. The advantage of using the position solution is the availability in every GNSS receiver and the low calculation effort as compared to a tightly coupled system.

2.3.2 Tightly Coupled GNSS Measurement

The tightly coupled observation matrix $\boldsymbol{H}_{\rho_i}$ is given by

$$\boldsymbol{H}_{\rho_i} = \begin{pmatrix} \boldsymbol{e}^n_{Sat,i} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & 1 & 0 \end{pmatrix} \text{ with} \tag{7}$$

$$\boldsymbol{e}^n_{Sat,i} = \frac{-(\boldsymbol{p}^n_{Sat,i} - \boldsymbol{p}^n_{bn})}{\sqrt{\left(\boldsymbol{p}^n_{Sat,i} - \boldsymbol{p}^n_{bn}\right)^T (\boldsymbol{p}^n_{Sat,i} - \boldsymbol{p}^n_{bn})}}. \tag{8}$$

The advantage of this method is that any number of pseudoranges helps to improve the absolute position accuracy of the localization filter, whereas the loosely coupled system typically needs four satellites to determine a position solution. Furthermore, the handling with obvious measurement errors is easier with the isolated pseudorange measurements.

2.3.3 Barometric Height Measurement

With the barometric formula

$$p(h) = p(h_0) \exp\left(-\frac{Mg\Delta h}{RT}\right), \tag{9}$$

with $p(h_0)$ is the static pressure, M is the molar mass of Earth's air, g gravitational acceleration, R universal gas constant for air, T is the standard temperature and Δh is the height difference to the height h_0, the observation matrix $\boldsymbol{H}_{\rho_u}$ can be derived to

$$\boldsymbol{H}_{\rho_u} = \left(0 \quad 0 \quad p(h_0)\frac{Mg\Delta h}{RT}\exp\left(-\frac{Mg\Delta h}{RT}\right) \quad \mathbf{0} \quad \ldots \quad \mathbf{0}\right). \tag{10}$$

To get consistent height measurements with the GNSS measurements, it is necessary to adopt the static pressure $p(h_0)$ online.

3 Methods

The presented localization filter was implemented on a Samsung Galaxy S3. In order to examine also a tightly coupled INS and the additional benefit of a more accurate GNSS receiver a commercial *u-blox M8N* receiver is connected to the smartphone by USB OTG.

The KF is implemented according to the numerical stable Bierman-Thronton [12] method. The scaling errors $\hat{\boldsymbol{s}}_\omega$ and $\hat{\boldsymbol{s}}_f$ are assumed to be time-invariant and are determined in a special calibration process before the measurements.

3.1 Reference Measurement System

The experiments were carried out with the setup shown in Fig. 2. A highly accurate GeneSys ADMA-G was used to give a reference. The ADMA-G can be mounted on the back of a pedestrian together with the GNSS antenna and the GSM receiver unit for DGPS corrections. Optionally, a smartphone can be mounted on a dedicated area to obtain a fixed orientation between the smartphone and the ADMA-G.

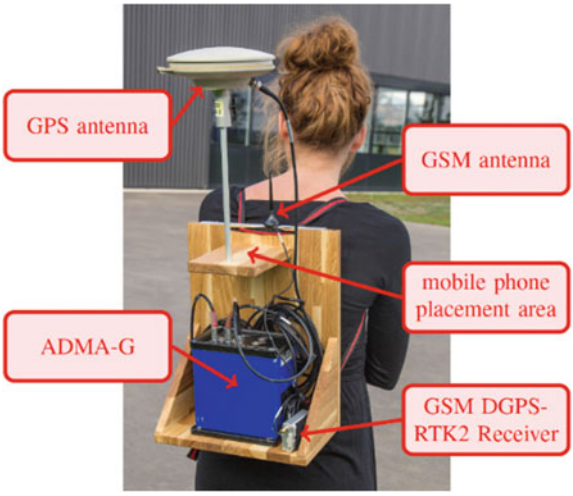


Fig. 2 Reference measurement setup with a high accuracy INS-System ADMA-G with closed-loop fiber-optic gyroscope and servo accelerometers

Table 1 Major differences between reference system GeneSys ADMA-G and smartphone sensors

	Smartphone	ADMA-G
Gyroscope type	MEMS	Closed-loop fiber gyroscope
Acceleration type	MEMS	3 servo accelerometers
GPS type	Single L1	Single L1, DPS-RTK2
Gyroscope accuracy	$\pm 250\,°\mathrm{h}^{-1}$	$1\,°\mathrm{h}^{-1}$
Acceleration accuracy	$\pm$10mg	1mg
Positioning accuracy		1.8 m. . .0.02 m

The ADMA-G itself is also an INS system with highly accurate fiber-optic gyroscope and servo accelerometers. The absolute position corrections are given by a Single L1, DPS-RTK2 GNSS receiver, which permits a position accuracy up to $\sigma_{p,min} = 0.02\,\mathrm{m}$ [13].

Table 1 demonstrates the important differences between the reference and the smartphone INS system. Primarily, the IMU sensors and the accurate DGPS system results in a higher accuracy. The smartphone is assembled with an integrated triaxial accelerometer and triaxial gyroscope of type InvenSense MPU-6500.

To minimize errors in the strapdown algorithm, mainly caused by the non-commutativity of the rotation, it is sampled with a high frequency of 200 Hz. The atmospheric pressure measurements are sampled with 1 Hz because of the low dynamics in height. The smartphone internal GNSS receiver supports a maximal sample frequency of 1 Hz, whereas the external u-blox M8N receiver is sampled with 5 Hz.

3.2 Measurement Environment

The measurements were done in a GNSS friendly environment (no or little multipath effects) on a cloudy day. The ADMA-G asserts a maximal position standard deviation of 0.03 m during the whole considered measurement.

Figure 3 shows the trajectory of the measurement. The pedestrian walks on a slightly curved way and does some 90° turns.

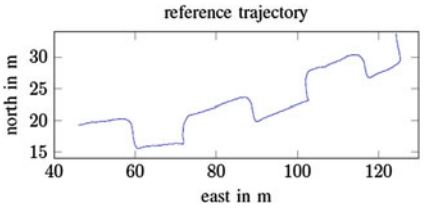


Fig. 3 The considered reference measurement trajectory measured with ADMA-G. The trajectory with some 90° turns was walked within 90 s on a slightly curved way in a GNSS-friendly environment

4 Results

4.1 GNSS Receiver and Method Comparison

Figure 4 shows the horizontal positioning errors of the different localization estimation filters. The best results are achieved with the position solution of the external u-blox M8N GNSS receiver with a RMS error of 0.74 m. The tightly coupled position solution has a RMS error of 1.194 m with the existence of significant outliers in comparison to the u-blox M8N loosely coupled solution. With a RMS error of 1.90 m, the smartphone's internal GNSS achieves the poorest positioning performance.

Due to the sensor data fusion, it is possible to compensate noise of the GNSS receiver with the IMU strapdown propagation. But if the GNSS receiver, like the smartphone's internal GNSS receiver, has got a constant offset compared to the reference, it is impossible to eliminate these errors through sensor data fusion or averaging. To guarantee a better absolute positioning performance it is necessary to have unbiased absolute position measurements e.g. usingDGNSS methods.

4.2 Velocity Accuracy

In particular for cooperative systems, as described in [14], the velocity of the pedestrian is interesting for the situation risk assessment. The position can be determined very quickly through the environment sensors, whereas the velocity must be determined over a few measurements.

With the u-blox M8N GNNS receiver and a loosely coupled localization filter architecture the overall RMS velocity error is 0.13 ms^{-1}. Later in this work, this value is used to assess the situation analysis.

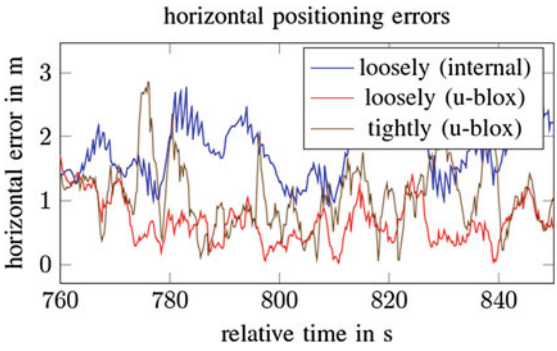


Fig. 4 Horizontal positioning error

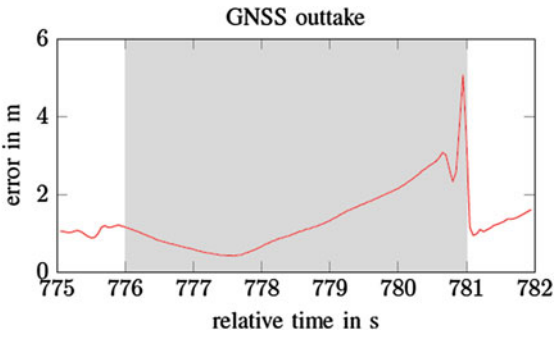


Fig. 5 Horizontal position error in a simulated 5 s GNSS outage during a 90° turn. During the shaded time interval there are no GNSS updates. The position is only determined with the IMU measurements through the strapdown algorithm

4.3 Simulated Short-Time GNSS Outage

In Fig. 5 GNSS outage of 5 s during a 90° turn is simulated. It can be noticed that the positioning error increases in this interval to a value of several meters. Thus, this shows that the IMU of the smartphone can be used for a short-time position prediction between GNSS updates. However, it is not possible to bridge longer GNSS outages or to give a useful position estimation.

The advantage of the strapdown position propagation is the immediate reaction response on changes in motion of the pedestrian. Furthermore, in comparison to the accuracy of [8], it can be observed that with a good GNSS availability the propagation accuracy of smartphone's IMU during a step interval is good enough, so that no further methods, like step detection to improve the velocity estimation, are needed.

4.4 Location Estimation Accuracy Requirements for Pedestrian Protection Systems

The parametrization of advanced driver assistant systems for pedestrian protection is always a compromise between an inadvertent false-positive rate r_{FP} and a required true-positive rate r_{TP} to achieve a system benefit. To determine these rates in a given situation, it is necessary to know the probability P_{FP} of a false-positive event and P_{TP} of a true-positive event. In addition the statistical occurrence rate r_{Sit} of such a situation must be known. This leads to

$$r_{FP} = P_{FP} r_{Sit} \quad \text{and} \tag{10}$$

$$r_{TP} = P_{TP} r_{Sit}. \tag{11}$$

In the following, the false-positive und true-positive probabilities P_{FP} and P_{FP} are analyzed in-depth.

First of all, the behavior of the pedestrian has to be modeled. In general the 2D position trajectory $\boldsymbol{p}_{pred}(t)$ of a pedestrian over a prediction interval $t \in [0, t_{pred}]$ can be expressed by

$$\boldsymbol{p}_{pred}(t) = \boldsymbol{p}_{meas} + t\boldsymbol{v}_{meas} + \int_0^t \int \boldsymbol{a}_{model}(t)dt\,dt, \tag{12}$$

where $\boldsymbol{p}_{meas}$ and $\boldsymbol{v}_{meas}$ are the position and velocity of the measurement, e.g. the localization filter, and $\boldsymbol{a}_{model}(t)$ is the predictive behavior model of the pedestrian. [14, 15] show possibilities to choose the prediction model $\boldsymbol{a}_{model}(t)$. If it is assumed that $\boldsymbol{p}_{meas}$, $\boldsymbol{v}_{meas}$ and $\boldsymbol{a}_{model}(t)$ are binomial normal distributed and time-invariant, the predicted position $\boldsymbol{p}_{pred}(t)$ is also binomial normal distributed with mean $\boldsymbol{\mu}_{pred}(t)$ and covariance $\boldsymbol{\Sigma}_{pred}(t)$

$$\boldsymbol{\mu}_{pred}(t) = \boldsymbol{\mu}_{pred} + t\boldsymbol{\mu}_{v_{meas}} + \frac{t^2}{2}\boldsymbol{\mu}_{a_{model}} \quad \text{and} \tag{13}$$

$$\boldsymbol{\Sigma}_{pred}(t) = \boldsymbol{\Sigma}_{pred} + t^2\boldsymbol{\Sigma}_{v_{meas}} + \frac{t^4}{4}\boldsymbol{\Sigma}_{a_{model}} \tag{14}$$

Based on this information, the front collision probability P_{coll} for a certain prediction time t_{pred} can be calculated as shown in [16]. It must be noted that the uncertainty of position, velocity and orientation of the ego vehicle must be taken into account for determining the collision probability P_{coll}.

In the following, for simplicity, a diagonal covariance matrix $\boldsymbol{\Sigma}_{pred}(t_{TTC}) = \sigma^2\boldsymbol{I}$ is assumed. Figure 6 shows the front collision probability P_{coll} of a pedestrian

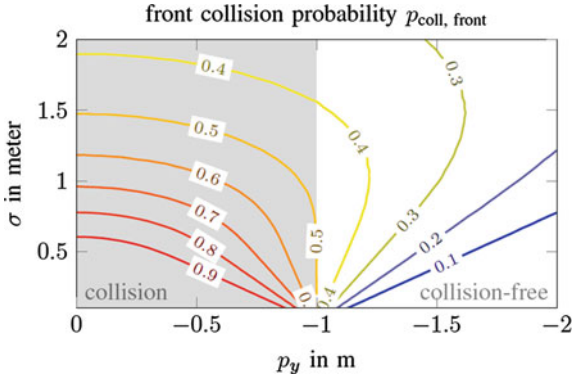


Fig. 6 Front collision probability p_{coll} respects to the standard deviation σ of the position measurement and the pedestrians crossing point p_y of the vehicle front line. It is assumed that the standard deviation σ is not growing with the time. In the shaded area there is a front collision with the vehicle. Only the half vehicle front line is represented with the middle of the vehicle at $p_y = 0\,\text{m}$

which crosses the vehicle front line at a given point p_y with a given standard deviation σ. The vehicle front covers a horizontal-axis from $p_y \in [-1, 1]$. The grey shaded area marks collisions where $p_y = 0\,\text{m}$ means a collision at the middle of the vehicle and $p_y = -1\,\text{m}$ is a collision with the right edge of the vehicle front. The white area shows the space free of front collisions. The figure displays that for an impact point e.g. at $p_y = -0.75$ m and a model and measurement based uncertainty of $\sigma = 1.2$ m, a front collision can only be predicted with a probability of about 55%.

When a system is triggered with a front collision probability $p_{coll,front}$ greater than a threshold value of p_{thres}, the probability of a system activation event p_{event} can be determined

$$p_{event} = \iint \iint_{\Omega} f_{meas} d^2\boldsymbol{v} \mathrm{d}^2\boldsymbol{p}, \tag{15}$$

where f_{meas} is the probability density function of the position and velocity measurement and Ω is the set of all measurement values which leads to a p_{coll} greater than the threshold value of p_{thres}. A positive event is a true-positive event if there is a collision with the vehicle, whereas a false-positive event is one without a collision with the vehicle front.

Figure 7 shows the positive event probability for a threshold value of $p_{thres} = 55\%$. As shown in Fig. 6, the horizontal-axis indicates the impact point of the pedestrian. Figure 7 shows that in order to get an acceptable number of false-positive events, it is necessary to have a low standard deviation σ. For $\sigma = 1\,\text{m}$, which would be a fair performance for GNSS based localization, the false-positive rate is still about 10% for $p_y = 2\,\text{m}$, e.g. if the pedestrian crosses the vehicle front line with 1 m distance to the vehicle which is a common situation in normal traffic. Therefore, such a great false-positive probability cannot be tolerated.

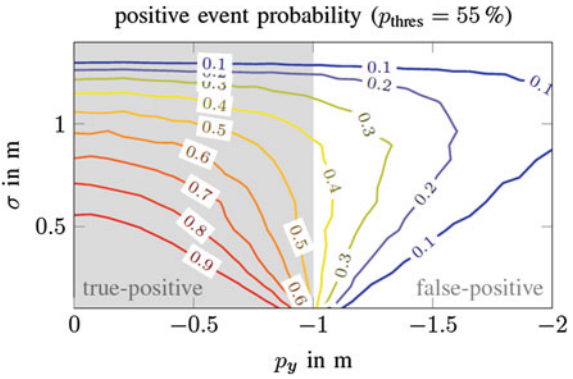


Fig. 7 False-positive and true-positive rate prediction with an assumed threshold front collision probability of $p_{coll,front} = 55\%$. The shaded area shows the true-positive rate, whereas the white area shows the predicted false-positive rate

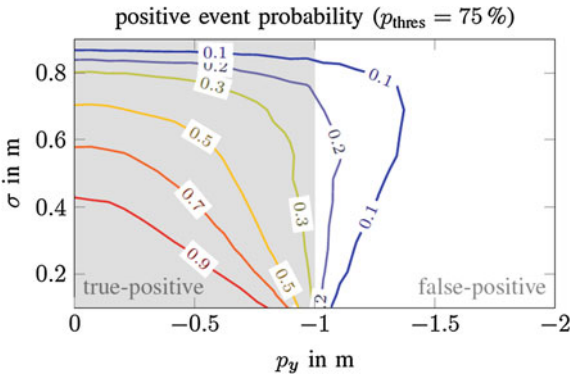


Fig. 8 False-positive and true-positive rate prediction with an assumed threshold front collision probability of $p_{coll,front} = 75\%$

Figure 8 shows the same content for a threshold of $p_{thres} = 75\%$. Assuming that a false-positive event in which a pedestrian crossing the front vehicle line with a distance of < 0.5 m is tolerable, the false-positive event probability looks better for all values of σ. However the requirement on the standard deviation σ to get a true-positive event probability by -0.75 m of 50% are a maximal value of $\sigma = 0.4$ m.

Furthermore, Figs. 7 and 8 also show impressively that a protection of pedestrians who impact the vehicle edge can only be protected if an increased false-positive rate of the system is tolerated.

The demonstrated loosely coupled localization filter with a Samsung Galaxy S3 and the u-blox M8N GNSS receiver has a standard deviation of 0.75 m ($t_{pred} = 1$ s) for the predicted pedestrian position with a neglected acceleration model $\boldsymbol{a}_{model}$. With an activation threshold of $p_{thres} = 75\%$ the true-positive rate is lower than 50%. Besides, one has to be aware that the pedestrian's acceleration model has a strong impact on the model uncertainties, which leads to a greater predictive model uncertainty $\boldsymbol{\Sigma}_{pred}(t_{TTC})$ and stronger requirements on the measurement accuracy. Research on the predictive model was made in [14, 15].

5 Discussion

In the context of situation assessment of a pedestrian protection system relevant parameters are the accuracies of the measured position, velocity and the behavior model to predict a collision and initiate the ego vehicle's best maneuver as well as the low false-positive rates with simultaneous high true-positive rates and earlier system triggering. Furthermore, a smartphone based cooperative pedestrian protection system has to have a low power consumption.

As derived in Sect. 4 the accuracy of presented localization filter has to be improved to protect pedestrians reliably. This is realizable by more accurate, especially unbiased, GNSS methods or a fusion with the ego vehicle's environment

sensors like video, radar or LiDAR. The estimated position accuracy is poor compared to the environment sensors, so that a fusion does not make sense. Whereas the estimated cooperative velocity can be used to initialize new object in the sensors' tracking filters and enables an earlier and reliable situation assessment. With the vehicle's sensors, the velocity components or tangential velocity must be derived by position measurements, which causes a settling time of the estimated velocity in the tracking filters. It has to be taken in consideration, that in a fusion approach, the problem is the object association because of the smartphone's low position accuracy and the impossible validation in the benefit scenarios, which are in particular situations where the pedestrian is visually obstructed or outside the field of view.

An advantage of the described method is, as described in Sect. 2, the availability of the IMU measurements in earth related coordinates (east, north and up), which could enable a better feature derivation for pedestrian behavior modeling.

In context of power consumption, the presented method generates small load on the communication channel to transmit the estimated position and velocity. Perhaps in a later system implementation it could be a pre-stage and a fundament of DGNSS based methods with a lower power consumption, communication load and provider of IMU's error model parameters.

6 Conclusions

In the context of smartphone based pedestrian protection systems for vehicles, this work evaluates different outdoor dead reckoning localization filters. The presented method has no restrictions of the position or the relative orientation to the walking direction of the smartphone. Thus, it allows pedestrians to carry the smartphone in trouser pockets, handbags or school bags.

Furthermore, due to the strapdown algorithm it is possible to determine the measured acceleration and rotation rate of the smartphone in earth fixed north, east and up direction with a high sampling frequency. This enables the usage for pedestrian behavior modeling and fast path prediction adaptations with V2X technology.

However, the paper shows also that further research and improvement on the used GNSS receiver and signal processing has to be done in order to ensure a better and unbiased localization accuracy. This can be effectuated by advanced GNSS techniques like computationally intensive Precise Point Positioning techniques or communication intensive DGNNS approaches. Especially DGNSS approaches between the ego vehicle and the smartphone GNSS receiver can enable better relative localization accuracy.

References

1. Gandhi T, Trivedi M (2007) Pedestrian protection systems: issues, survey, and challenges. IEEE Trans Intell Transport Syst 8(3):413–430
2. Flach A, David K (2009) A physical analysis of an accident scenario between cars and pedestrians. In: 2009 IEEE 70th vehicular technology conference fall (VTC 2009-Fall), pp 1–5, Sept 2009
3. Flach A, David K (2010) Combining radio transmission with filters for pedestrian safety: Experiments and simulations. In: 2010 IEEE 72nd vehicular technology conference fall (VTC 2010-Fall), pp 1–5, Sept 2010
4. Flach A, Memon A, Lau SL, David K (2011) Pedestrian movement recognition for radio based collision avoidance: a performance analysis. In: 2011 IEEE 73rd vehicular technology conference (VTC Spring), pp 1–5, May 2011
5. David K, Flach A (2010) Car-2-x and pedestrian safety. IEEE Vehi Technol Mag 5(1):70–76
6. Liebner M, Klanner F, Stiller C (2013) Active safety for vulnerable road users based on smartphone position data. In: Intelligent vehicles symposium (IV). IEEE, pp 256–261
7. Engel S, Kratzsch C, David K (2013) Car2pedestrian-communication: Protection of vulnerable road users using smartphones. In: Fischer-Wolfarth J, Meyer G (eds) Advanced microsystems for automotive applications 2013, ser. Lecture notes in mobility. Springer International Publishing, pp 31–41
8. Engel S, Kratzsch C, David K, Warkow D, Holzknecht M (2013) Car2pedestrian positioning: methods for improving GPS positioning in radio-based VRU protection systems. In: 6. Tagung Fahrerassistenzsysteme
9. Gabaglio V, Ladetto Q, Merminod B (2001) Kalman filter approach for augmented gps pedestrian navigation. GNSS, Sevilla
10. Grewal M, Weill L, Andrews A (2007) Global positioning systems, inertial navigation, and integration. Wiley
11. Wendel J (2011) Integrierte navigationssysteme—sensordatensensor, GPS und intertiale navigation. Oldenbourg Verlag, München
12. Bierman G (1977) Factorization methods for discrete sequential estimation. Academic Press, Inc.
13. Technical documentation—ADMA version 2x.6.2. GeneSys Elektronik GmbH, Technical report, June 2012
14. Westhofen D, Grundler C, Doll K, Brunsmann U, Zecha S (2012) Transponder- and camera-based advanced driver assistance system. In: Intelligent vehicles symposium (IV). IEEE, pp 293–298
15. Tiemann N, Branz W, Schramm D (2010) Predictive pedestian protection—situation analysis with a pedestrian motion model. In: Proceedings of the 10th International symposium on advanced vehicle control (AVEC)
16. Braeuchle C, Ruenz J, Flehmig F, Rosenstiel W, Kropf T (2013) Situation analysis and decision making for active pedestrian protection using bayesian networks. In: TÜVsüd, 6. Tagung Fahrerassistenz—Der Weg zum automatischen Fahren, Munich, 2013

Probabilistic Integration of GNSS for Safety-Critical Driving Functions and Automated Driving—the NAVENTIK Project

Robin Streiter, Johannes Hiltscher, Sven Bauer and Michael Jüttner

Abstract The NAVENTIK project will develop an automotive platform for computational demanding applications in the field of sensor data fusion and software defined radio. Based on this platform, the first component launched will be an automotive-grade GNSS (Global Navigation Satellite System) receiver that integrates state-of-the-art signal processing for lane level accurate navigation and that guarantees bounded false alarm rates. This is possible, thanks to a software-defined approach and the probabilistic integration of GNSS signal tracking algorithms on radio level. The explicit modelling of GNSS error sources and local signal degradation provide the basis for the proper Bayesian integration. The project will enable the first mass-market GNSS receiver based on a software-defined approach that is able to meet safety-critical requirements as it copes with false alarm specifications and safety related requirements.

Keywords GNSS · Localization · Automated driving · Safety requirements · Functional safety

R. Streiter (✉)
Technische Universität Chemnitz, Reichenhainer Str. 70,
Chemnitz 09126, Germany
e-mail: robin.streiter@naventik.de

J. Hiltscher · S. Bauer · M. Jüttner
NAVENTIK GmbH, Reichenhainer Str. 70, Chemnitz, Germany
e-mail: johannes.hiltscher@etit.tu-chemnitz.de

S. Bauer
e-mail: sven.bauer@naventik.de

M. Jüttner
e-mail: michael.juettner@naventik.de

T. Schulze et al. (eds.), *Advanced Microsystems for Automotive Applications 2016*, Lecture Notes in Mobility, DOI 10.1007/978-3-319-44766-7_2

1 Introduction to GNSS in Automotive Applications

The steadily increasing grade of automation in modern vehicle's driver assistance and automated driving functions has a considerable impact on the development of all subsystems involved. This applies especially to the functions derived from sensor data fusion and environmental perception, as the compliance to functional safety and false alarm specifications is hard to prove and guarantee. One approach to cope with these demanding requirements is the consistent modelling of the system from a probabilistic perspective, which includes signal-processing operations, the propagation of uncertainties, fusion of heterogeneous sensor data and the representation of the system state. This results in the so-called Bayesian framework as the fundamental basis for advanced signal processing that, if applied consistently, can satisfy false alarm specifications and safety issues by nature.

The NAVENTIK project provides a key enabling technology to leverage satellite based navigation for automated maneuvers and safety critical decisions in automotive applications. A combination of innovative low-level probabilistic signal processing algorithms will be implemented within a software defined GNSS receiver, in combination with advanced high-level data fusion approaches in order to derive a confidence estimate that is able to meet any safety requirements in urban areas.

There are different approaches that are able to detect multipath affected GNSS observations. An overview of these so-called Multipath Mitigation algorithms is given in [1]. A straightforward approach is identifying multipath by considering digital maps with modelled 3D buildings in order to validate the direct line of sight to each satellite. In the Bayesian framework, this approach is used to predict multipath affected GNSS observations [2]. Another algorithm for determining non-line-of-sight (NLOS) with the help of environmental knowledge is described in [3].

The next group is of approaches is of significant interest as it uses statistic tests and probabilistic filtering for the identification and mitigation of multipath. A known representative in this category is the *Receiver Autonomous Integrity Monitoring* algorithm (RAIM) or an extension called Probabilistic Multipath Mitigation (PMM) [4]. From a probabilistic perspective, the PMM algorithm is of major importance as it was able to show the benefits regarding the improvement of the integrity in different safety relevant automotive use cases.

All these approaches are implemented on observation level and suffer from the GNSS receiver's proprietary signal pre-processing, which is supposed to be not stringent from a probabilistic point of view. To a huge extent remaining errors of multipath mitigation algorithms can be reduced by a reimplementation of GNSS signal processing under the prerequisite of a probabilistically consistent signal tracking.

The idea of signal manipulation on radio level is called Software Defined Radio (SDR). In general, this approach is to substitute signal processing hardware on radio level by a software implementation. This technology is very demanding in terms of computing power and data throughput but has high potential from scientific point of

view as it allows a flexible implementation and validation of complex signal processing algorithms.

In the field of satellite-based navigation, it is called Software Defined GNSS. Complex algorithms like the a GNSS pseudorange error density tracking using a Dirichlet Process Mixture [5] or a Bayesian approach to multipath mitigation in GNSS receivers where implemented based on this technology. Another statistical signal tracking approach on this level with special consideration of tracking time-delays, amplitudes and phases is described in [6]. A Bayesian multisensor navigation, incorporating pseudorange measurements and a multipath model is presented in [7]. In order to meet these demanding computational requirements the NAVENTIK GNSS receiver will be implemented on an automotive grade system on chip that will be described in detail in the next chapter.

The NAVENTIK approach is to guarantee to not to discard any information, especially multimodalities in the measurement space. One major problem from multipath mitigation perspective is the evaluation of the auto correlation function for each satellite and the irreversible assignment of pseudoranges. The proposed approach uses the idea of a signal tracking implementation, as a derivative of probabilistic data association (PDA) in order to deal with the multimodal characteristics and outliers within the measurement space and to resolve these ambiguities within the Bayesian framework [8]. Thus, the NAVENTIK signal-tracking algorithm condenses multipath and NLOS effects as generic system properties and the Bayesian implementation improves the integrity of the system state in multipath and NLOS environments.

Figure 1 gives an illustrative idea of the integrity concept for one single point in time and its practical influence to applications are described in this chapter related to use cases. Figure 2 shows the metric of confidence within an urban scenario evolving over time. The blue plot represents the real 2-dimensional position error regarding the ground truth of the validation system. A reference GNSS System

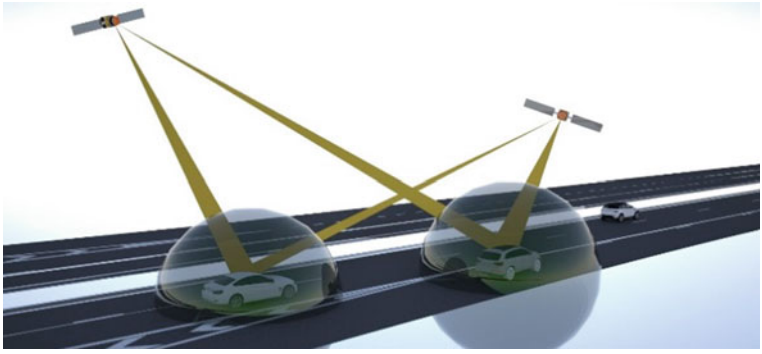


Fig. 1 High-level representation of the GNSS confidence level as a basis for safety-critical driving functions. The NAVENTIK GNSS receiver guarantees the vehicle is within the confidence level according to a given false alarm specification

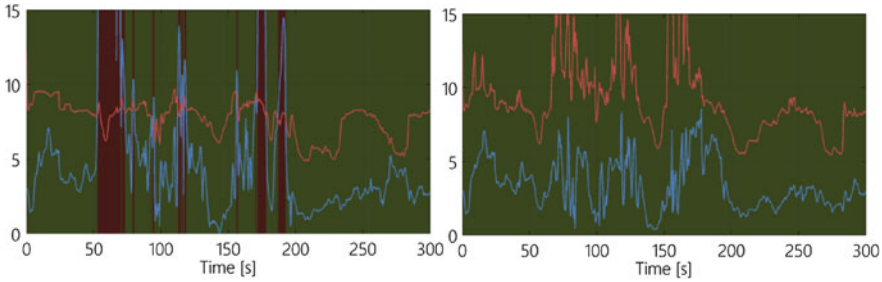


Fig. 2 Integrity of positioning data is guaranteed when the confidence interval (*red line*) covers the real position error (*blue line*). The *left* plot shows the violation of the integrity concept with a classical GNSS receiver. The *right* plot shows the result of the proposed system for the same situation without any disruption. Next to the improved positioning accuracy, the confidence measure is not violated

consisting of a Novatel Span with Real Time Kinematic and Inertial Measurement Unit provides the ground truth with a position accuracy within cm-level. The red plot shows the estimated confidence interval of the position. In the example, the confidence interval is set to 3σ, which means 99 % of all position candidates are supposed to be within the confidence interval. The left diagram shows a traditional GNSS receiver with average performance in an urban scenario. Next to the absolute position error (RMSE), it is evident that the assumption of 3σ integrity is not met at all. Areas where the true position error exceeds the estimated confidence results in a corrupted position estimate. The red areas shown in the diagram indicate those situations. Therefore, the traditional GNSS receiver is not legitimate for usage in safety critical applications. In contrast to the traditional receiver, the right diagram gives an indication about the performance of the NAVENTIK receiver, which will be implemented on the software-defined platform. On the one hand, the real 2-dimensional positioning error is improved due to the advanced signal tracking implementation and on the other hand, the estimated confidence covers the true position error for the entire sequence. This is the prerequisite for using the GNSS receiver in a safety critical context. Starting from this approach, we can computationally proof, that the GNSS receiver can meet dedicated false alarm specifications of a safety critical application.

This approach is a step towards the applicability of low cost satellite navigation for safety relevant applications in the automotive area, as it enables the computation of a reliable confidence interval even under degraded GNSS signal reception situations.

Unfortunately, the algorithms resulting from this strategy are demanding in terms of data throughput and computational power. Thus, the implementation on electronic control units (ECU), if possible at all, requires a very complex adoption process. Therefore systems-on-chip (SoC) are more and more in the focus for the integration in ECUs, as they can provide huge capacities especially for demanding

signal processing operations by employing DSPs (digital signal processor) and FPGAs (field programmable gate array) along with a flexible and reconfigurable system design and operating system.

2 Confidence Adaptive Use Cases

Figure 1 gives an indication about the estimated confidence interval that is provided by the NAVENTIK receiver. According to given false alarm specifications, the NAVENTIK system can guarantee the true position of the vehicle to be within the given confidence interval, which is represented by the sphere around the vehicles. Derived from this information a couple of use cases will be introduced within this chapter.

2.1 E-Call Extension

In case of an emergency situation, E-Call 2.0 activated by the involved vehicle sends information related to the accident to the emergency call center like speed level right before accident, number of passengers, damage report and a warning to surrounding vehicles. As the NAVENTIK system provides extended information about the position, especially about the likely distribution about affected roads, lanes and directions dedicated actions can be taken into account. The NAVENTIK receiver delivers the trusted and high integrity position information to enable E-Call to submit the precise location of the crashed vehicle. The following figure shows the additional information of the NAVENTIK system:

2.2 Active Navigation

Currently, navigation and routing is the strongest use case for GNSS in automotive mass-market applications. In the light of recent developments towards more integrated and automated driving, the consequent exploitation of GNSS for automated driving functions is expected as a prominent use case. Active navigation focuses on the realistic performance and requirements for an efficient exploitation of the NAVENTIK key technology. As an extension of the strongest and common applications, the NAVENTIK contribution towards the robust integration of GNSS data into safety critical driving functions has a huge potential from mass-market perspective Fig. 3.

Furthermore, active navigation does not rely on the penetration or existence of aiding technologies like communication or infrastructure and an implementation can be tailored to available technologies and resources. Derived from the generic

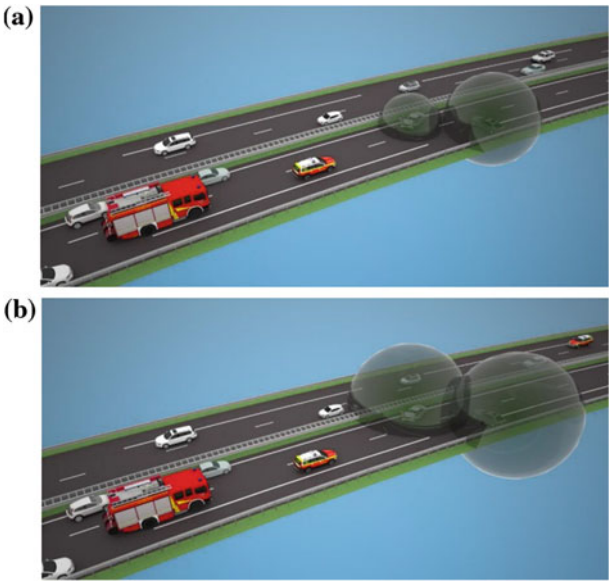


Fig. 3 **a** The confidence of the position estimate is high. The cars can be assigned to the road where the accident has happened and an ambulance can be sent in the correct direction on the highway. **b** The confidence of the position estimate is low and the position of the vehicles cannot clearly be assigned to the driving direction. Maybe it is helpful to send an ambulance for each direction to make sure no detour and additional time is needed

NAVENTIK approach, of integrating GNSS within any safety critical context, active navigation is a special implementation of this idea with the focus on a strong market potential with a very short time-to-market. The further development of conventional routing and navigation towards a more integrated and active assistance approach will leverage the awareness of GNSS based automated driving functions and passive navigation systems can be easily extended towards an automated mode, that reflects the current positioning performance and adds different modes of automation.

Figure 4 shows the extension of the classical "passive" navigation use-case to Active Navigation. If the accuracy of the position reflects lane or road level performance, the navigation system supports the driver with additional active steering and automated acceleration.

With decreasing performance, the degree of automation reduces accordingly, which means the steering wheel indicates only subtle and soft steering motions and driving instructions. The further limitation of the performance related with a very low confidence forces the system to switch back to conventional, uncritical turn-by-turn instructions. Finally, if GNSS is not available, the system gives very

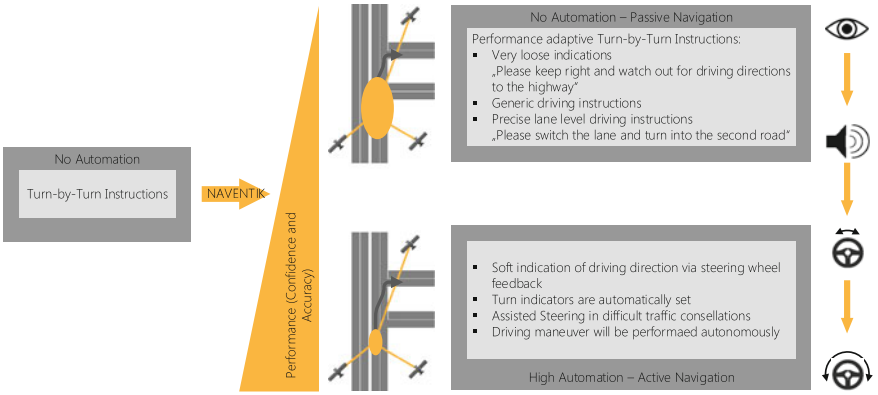


Fig. 4 Extension of the legacy navigation and routing application. Classical navigation systems do not offer any level of automation. Thanks to the confidence information, the system adopts different levels of automation depending on the positioning performance

rough estimations about the driving direction, only. That means the system can also influences the user acceptance by dramatically reducing the false alarm rate, not only in the safety critical context.

3 NAVENTIK Measures and System Architecture

The basic system architecture proposed is shown in Fig. 5. The positioning task is partitioned to a hardware and a software component. The hardware component will be implemented using FPGA resources; logic components are shown as squares and

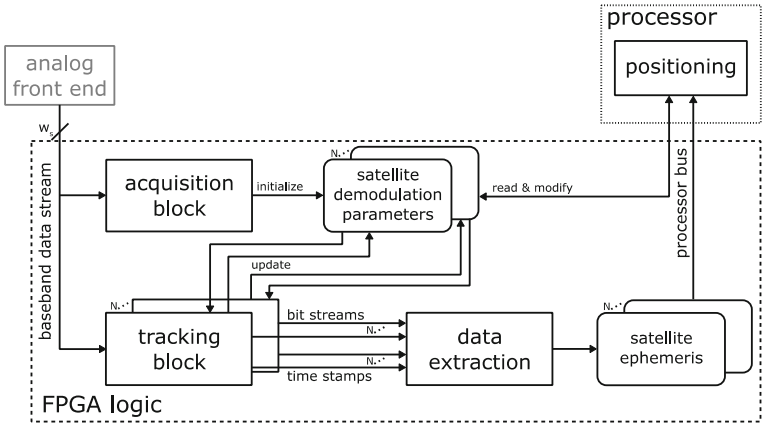


Fig. 5 NAVENTIK project system architecture

memory elements as rounded boxes. Carrier and code acquisition and tracking will be implemented as hardware blocks as they involve computing intense operations like Fourier transforms. The highly parallel FPGA hardware allows such functionality to be implemented very efficiently and with high throughput. Using fixed-point arithmetic the required functionality can be mapped to FPGA hardware efficiently.

Acquisition and tracking blocks work in concert, using the incoming sampled data stream from the analog frontend. The acquisition block continuously monitors the input data to detect newly available satellites. To increase resource efficiency only one acquisition block will be implemented which sequentially scans the different satellite codes. When the block has locked to a satellite's signal, it hands over the determined initial demodulation parameters (carrier frequency and code phase) to a tracking block, which adapts the demodulation parameters if necessary and extracts the transmitted bit stream. Each tracking block is capable of handling one satellite's signal. The tracking block also converts the high-speed sampled baseband data stream from the analog frontend to a low speed, oversampled bit stream. As the block is implemented using configurable hardware it can be modified to implement and study different mechanisms for multipath compensation addressing data demodulation. To compensate for processing delays, the tracking blocks also generate data collection time stamps required for pseudo range computation.

The data extraction block processes the demodulated data streams generated by the tracking blocks; it locks to the frames sent by the tracked satellites, checks whether data frames were received correctly and finally extracts the ephemeris data and inserts it into a data structure. One data extraction block is sufficient for processing the data received from all satellites in sequential order as the data streams have a very low speed.

The positioning algorithms, however, are not implemented in FPGA logic as they involve floating point logic and transcendental functions. Such operations are more conveniently handled by a general-purpose processor (GPP), shown in the upper right corner of Fig. 5. This can be an integrated processor as found in modern Programmable Systems on Chip (PSoC; e.g. Xilinx Zynq or Altera Cyclone/Arria devices); a processor IP-Core implemented using FPGA resources or even a distinct embedded processor. We consider implementing these algorithms on a GPP advantageous as software implementations are more flexibly adapted than hardware implementations and post processing—like probabilistic filtering–can be added and changed very easily to study the effects of different approaches. Ephemeris data extracted by the GNSS hardware blocks is accessible to software via the processor's memory map, which it is accessed via an interface to the processor peripheral bus. The architecture also allows the positioning software to alter the demodulation parameters, which are also memory mapped. This allows the software to alter settings and parameters of the tracking block, e.g. to change the precision of the code tracking unit.

Prospectively NAVENTIK aims at generalizing the proposed platform to enable software defined processing of arbitrary sensor data. This appears meaningful as sensors of any type usually perform internal pre-processing using fixed algorithms.

This is a necessary step to remove noise, distil the relevant information from the raw sensor data and reduce it to a manageable amount. Usually the pre-processing algorithm and its parameters are fixed within the sensor and not accessible by the user. We expect other types of sensors to benefit from adaptive statistical filtering in the same way as shown for GNSS. The mechanisms of sensor data filtering use a common set of algorithm building blocks which are numerically complex and computationally demanding, which hinders implementing them on common embedded hardware. To fill this gap, we aim at developing a more general version of the proposed GNSS SDR platform adaptable to a broad range of sensors.

Although the weak computational performance of embedded systems is tackled by integrating general-purpose programmable graphics processing units (referred to as GPGPU), there are strong reasons for employing a dedicated hardware platform for sensor data pre-processing. Sensor data pre-processing typically does not utilise floating-point spaces to their full capacity. Employing fixed-point hardware and a lower number of bits than provided by established floating-point standards can lead to implementations that are more efficient. Sensor data pre-processing employs a common set of functional blocks, which can be implemented as dedicated hardware elements, again altering processing efficiency. However, the biggest advantage over state-of-the-art general-purpose hardware is that a dedicated platform can provide hard real-time constraints, an important demand of safety-critical systems.

The proposed architecture of the proposed Software Defined Sensor system is shown in Fig. 6. The overall structure is similar to the system proposed for GNSS SDR. The hardware-software partitioning is identical; the complex pre-processing functionality is implemented as dedicated hardware. Based on the incoming raw sensor data stream (referred to as sensor baseband data), the hardware estimates initial settings for the first processing step, filtering and/or demodulation. This block extracts the relevant information from the incoming sensor baseband data and forwards it to the data generator. The data generator produces the actual

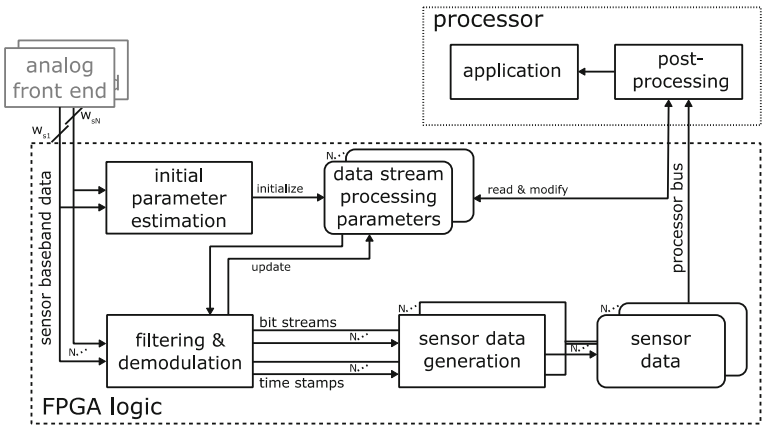


Fig. 6 Proposed system architecture of a general-purpose Software Defined Sensor system

measurements which can then be processed in software to alter flexibility. To be more general, sensor data processing is depicted as a post-processing step and the actual application working with the data. The post-processing step can include further, less demanding data refinement procedures. Furthermore, it can be used to assess the suitability of the automatically determined data stream processing parameters and adapt them if necessary. This enables applications to control the process of filtering the sensor baseband data, giving them the opportunity to adapt the mechanisms to varying environmental influences. We expect this approach to yield more precise sensor data by giving applications control over the pre-processing stage as they can integrate situational knowledge, leading to a situation aware tuning of pre-processing parameters rather than a fixed, good on average, model.

4 Conclusion

The NAVENTIK receiver is based on a software prototype that has been developed and verified in a wide range of use cases in the European research projects CoVeL (Cooperative Vehicle Localization) and GAIN (Galileo for Interactive Driving) within the 7th Framework Programme. The extension regarding confidence adaptive use cases in a safety related context will be implemented today, within the InDrive project in Horizon 2020. The NAVENTIK project is dedicated to the further development of the Prototype towards a mass-market product.

Starting from the software-defined approach for GNSS signal processing we are going to adopt a generalized platform design in order to complement requirements not only arising from the field of satellite navigation but also to extend the system by environmental perception sensors such as Radar and Lidar.

Those systems do also suffer from signal preprocessing which is, from a statistical perspective, always subject to strong losses of information. Clustering algorithms and noise reduction at the very early stage are violating the integrity of those sensors as well. Especially advanced tracking algorithms in nowadays ACCs (Adaptive Cruise Control) can benefit from the statistically correct representation within the measurement space especially where the weaknesses of existing systems, like tracking extended targets and obstructed objects are emerging. We can assume that tracking performance and overall reliability of those perception systems can be significantly improved by the proper integration within the NAVENTIK platform.

Acknowledgments This work is carried out within the NAVENTIK project. NAVENTIK is a research project co-funded by the Federal Ministry for Economic Affairs and Energy (BMWi) and the European Social Fund (ESF) within the EXIST Transfer of Research programme. The authors would like to thank all partners for supporting this work. Furthermore, for the evaluation and generation of ground truth data, precise real-time corrections provided by the ascos service (http://www.ascos.de) were used.

References

1. Groves PD, Jiang Z, Rudi M, Strode P (2013) A portfolio approach to NLOS and multipath mitigation in dense urban areas. In: Proceedings of the 26th international technical meeting of the satellite division of the institute of navigation
2. Obst M, Bauer S, Reisdorf P, Wanielik G (2012) Multipath detection with 3D digital maps for robust multi-constellation GNSS/INS vehicle localization in urban areas. In: Proceedings of the IEEE intelligent vehicles symposium
3. Peyret F, Betaille D, Carolina P, Toledo-Moreo R, Gomez-Skarmeta A, Ortiz M (2014) GNSS autonomous localization: NLOS satellite detection based on 3-D Maps. IEEE Robot Autom Mag 57–63
4. Bauer S, Streiter R, Obst M, Wanielik G (2015) Non-line-of-sight mitigation for reliable urban GNSS vehicle localization using a particle filter. In: 2015 18th IEEE international conference on information fusion (FUSION), Washington, DC, 6–9 July 2015
5. Viandier N, Marais J, Rabaoui A, Duflos E (2010) GNSS pseudorange error density tracking using Dirichlet process mixture. In: 2010 13th Conference on information fusion (FUSION), pp 1–7
6. Closas P, Fernandez-Prades C, Fernandez-Rubio J (2009) A Bayesian approach to multipath mitigation in GNSS receivers. IEEE J Sel Top Sign Process 3(4):695–706
7. Khider M, Jost T, Sanchez E, Robertson P, Angermann M (2010) Bayesian multisensor navigation incorporating pseudoranges and multipath model. In: Position location and navigation symposium
8. Streiter R, Bauer S, Wanielik G (2015) Probabilistic GNSS signal tracking for safety relevant automotive applications. In: 2015 18th IEEE International conference on information fusion (FUSION), Washington, DC, 6–9 July 2015

Is IEEE 802.11p V2X Obsolete Before it is Even Deployed?

Johannes Hiltscher, Robin Streiter and Gerd Wanielik

Abstract Years after publication of the IEEE 802.11p and IEEE 1609 standards for Intelligent Transportation Systems (ITSs), first production vehicles equipped with conforming communication hardware are about to become available. The standard's suitability for the hard real-time automotive environment has been debated intensively in recent years. Most publications use synthetic message sizes, while the comprehensive ITS-G5 standard allows for a performance evaluation of IEEE 802.11p in real scenarios. Realistic performance and scalability assessments of current automotive communication hardware can be derived from such an evaluation. Based on these we examine the suitability of the available standards for demanding hard real-time control tasks as cooperative adaptive cruise control (CACC).

Keywords Intelligent transportation systems · Cooperative awareness · ITS-G5 · IEEE 802.11p · IEEE 1609

1 Introduction

Vehicular communication is seen as one of the greatest advancements in the automotive domain, giving rise to technologies like autonomous driving and advanced driver assistance systems. Standardization efforts were completed several years ago, leading to the standards IEEE 802.11p covering the physical and IEEE 1609 covering the medium access (MAC) and networking layers (refer to [16] for

J. Hiltscher (✉) · R. Streiter · G. Wanielik
Professorship of Communications Engineering, Technische Universität Chemnitz, Reichenhainer Straße 70, 09126 Chemnitz, Germany
e-mail: johannes.hiltscher@etit.tu-chemnitz.de

R. Streiter
e-mail: robin.streiter@etit.tu-chemnitz.de

G. Wanielik
e-mail: gerd.wanielik@etit.tu-chemnitz.de

T. Schulze et al. (eds.), *Advanced Microsystems for Automotive Applications 2016*, Lecture Notes in Mobility, DOI 10.1007/978-3-319-44766-7_3

an overview). For the European market the European Telecommunications Standards Institute (ETSI) extended IEEE 1609 leading to ITS-G5. One major difference between the two standards is Distributed Congestion Control (DCC). This mechanism was introduced to better utilize the available bandwidth in high traffic conditions while retaining long communication distances. Altogether, ITS-G5 is much more comprehensive than IEEE 1609 and specifies a set of basic services besides the low levels of the communication protocol. Therefore, a detailed demand analysis can be performed which we consider valuable when deciding whether manufacturers should integrate next generation driver assistance systems like cooperative adaptive cruise control (CACC) into upcoming generations of vehicles or wait for other communication technologies to become available. In this contribution we examine the performance of an ITS-G5-conformant IEEE 802.11p vehicular network (VANET) in the context of Cooperative Awareness (CA). We simulate the operation of the service for different amounts of vehicles to determine the scalability of current vehicle-to-vehicle (V2V) communication standards. To the best of our knowledge, this is the most in-depth analysis of ITS-G5 according to the standard's actual demands and capabilities.

The remainder of this paper is structured as follows. We first give an overview of related research in Sect. 2. This section details our decision to perform another evaluation focused on ITS-G5. Next we give an overview of the relevant parts of the standard in Sect. 3. A custom simulation framework was developed for conducting this evaluation which is presented in Sect. 4 alongside with the methodology used. Our findings are presented and discussed in Sect. 5. The paper is concluded in Sect. 6 and an outlook to future work given.

2 Related Work

Numerous studies of the available standards for Dedicated Short-Range Communication (DSRC) have been published in recent years. One of the first publications examining the viability of IEEE 802.11p VANETs is [4]. The evaluation is performed in a Manhattan Grid using 500 byte messages. The author concludes that IEEE 802.11p is not well suited for VANETs due to high packet collision probabilities. An evaluation with a scope similar to the one of this publication, the viability of IEEE 802.11p for CACC applications, is presented in [2]. Although the authors state the 400 byte messages they employ for simulation encompass an ETSI Decentralized Environmental Notification Message (DENM), CA Message (CAM) and a beacon signal, this is not comprehensible. Furthermore, dissemination frequencies are very high and the DCC mechanism not taken into account. An extensive study of IEEE 802.11p network characteristics is presented in [19]. An analysis of the DCC mechanism is performed in [1] for packet sizes of 100 and 300 bytes in a network of 300 and 600 Intelligent Transportation System (ITS) Stations (ITS-Ss) arranged in a Manhattan Grid. The authors claim to exchange CAMs, however, this seems to only describe the periodic message

generation. The paper concludes that DCC does not effectively limit channel load and packet collisions. Another evaluation of the DCC mechanism performed with message sizes from 100 to 600 bytes is presented in [23] and has similar findings.

It can be concluded that previous studies consistently find IEEE 802.11p to be suboptimal for vehicular environments. However, it is the only standardized VANET technology currently available. Therefore it would be worthwhile to assess its scalability to decide whether it provides a viable intermediary solution until better ones become available. The presented publications also lack to define a realistic scenario. Packet sizes range from 100 to 600 Bytes and are, like the number of ITS-Ss, often not comprehensible. To close this gap, we perform a performance analysis of IEEE 802.11p using the ITS-G5 CA service as an application scenario.

3 The ETSI ITS-G5 Standard

The European market has its own DSRC standard, defined by the ETSI. The physical layer is IEEE 802.11p with minor differences in frequency allocation. The upper layers are derived from IEEE 1609, extending it with additional mechanisms like DCC. ITS-G5 also introduces further layers into the model. Figure 1 depicts the layer structure of ITS-G5. Two so-called basic services, CA and Decentralized Environmental Notification (DEN), are also part of the standard. In this chapter we briefly introduce the Access and Networking and Transport Layers as well as the CA basic service. DEN is not covered as it is not considered in this evaluation.

3.1 Access Layer

The Access Layer encompasses the physical (PHY), MAC and Logical Link Control (LLC) layers corresponding to OSI layers 1 and 2 [9]. The ITS-G5 PHY

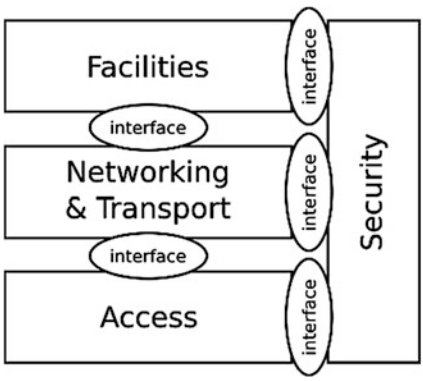


Fig. 1 Layer model of ITS-G5, only the parts relevant to this evaluation are shown (own figure after [9])

operates in the 5 GHz band with three 10 MHz channels currently available for ITS applications, referred to as ITS-G5A. One of them, called the Control Channel (CCH), is dedicated to announcements and warnings. The other ones, called Service Channels (SCHs), can freely be used for security as well as convenience applications. Orthogonal Frequency Division Multiplexing (OFDM) is employed with different modulation schemes, yielding data rates from three to 27 Mb/s [8].[1]

The MAC is the lowest logical layer of the network stack and responsible for direct message exchange between Network Interface Controllers (NICs). The DCC mechanism of ITS-G5, the purpose of which is the management of channel capacity, is located at this layer. DCC constantly monitors the channel load, i.e. the relative amount of time that messages are being sent, at each NIC. Depending on the measured load PHY parameters are altered to keep it—and thereby the probability of collisions–in defined bounds. Four core parameters are controlled by DCC:

- transmit power (Transmit Power Control, TPC)
- message rate (Transmit Rate Control, TRC)
- transmit data rate (Transmit Data rate Control, TDC)
- receiver sensitivity (DCC sensitivity control, DSC)

DCC is organized as a state machine with two states for light load (RELAXED) and overload (RESTRICTIVE). Besides these two states, one or more so-called Active states are defined to gradually adapt to different intermediate load conditions. Different state machines are defined for CCH and SCHs. To avoid continuous switching between states, evaluation periods are defined for changing to a more or less restrictive state [8, 10]. The DCC mechanism was introduced into ITS-G5 as the susceptibility of IEEE 802.11 wireless networks to congestion is one of the aspects most criticized concerning the automotive domain. The last component, LLC, provides information about the upper layer Protocol Data Unit (PDU) encapsulated into the MAC PDU.

3.2 Networking and Transport Layer

The Network and Transport Layer abstracts the Access Layer's low-level hardware view on data exchange and represents OSI layers 3 and 4 [9]. ITS-G5 defines two protocols for wrapping and abstracting higher level messages, like CAMs, and other protocols like TCP/IP. The GeoNetworking protocol [5] defines PDUs for different applications, like Beacons or environmental information like CAMs. Furthermore, it defines a location service which contains the positions of the surrounding ITS-Ss,

[1]For the sake of brevity, references at the end of a paragraph indicate all the information in that paragraph—if not referenced otherwise—is cited from that source.

an addressing mechanism and rules for address configuration [5]. The Basic Transport Protocol (BTP) adds a mechanism for distinguishing different applications sharing the communication link. This achieved using ports, some of which are reserved for distinct applications like CAMs [6].

3.3 The Common Data Dictionary

The common data dictionary (CDD) [11] is an extensive set of basic definitions which are used to define messages for exchanging information between ITS-G5-compliant ITS-Ss. It only defines the basic data structures using ASN.1 syntax [14] but is not intended for independent usage; therefore, further relevant aspects like the encoding mechanism are not covered by the specification. The containers described in the following are built from elements of the CDD.

3.4 Cooperative Awareness Basic Service

The CA basic service is part of the Facilities Layer which represents OSI layers 5 to 7 [9]. Its task is to regularly broadcast information about an ITS-S's parameters to surrounding ITS-Ss. The relevant standard document [7] defines the data structures, called containers, forming a CAM. Compliant ITS-Ss have to generate CAMs regularly according to rules determining the generation rate which are also defined in the standard. The data structures typically consist of mandatory and optional parameters. Depending on their dynamics, parameters are either considered high frequency or low frequency. CAMs therefore consist of high and low frequency containers, the latter of which only has to be sent every 500 ms while the prior is to be present in every CAM. Furthermore, the standard defines common basic and as well as extended data structures for vehicles and Roadside Units (RSUs). Additional containers for special purpose vehicles are specified containing further information like loaded dangerous goods.

An ITS-S is required to generate between one and ten CAMs per second depending on its dynamics. The intent of this mechanism is to quickly communicate abrupt changes of a vehicle's position and motion—like braking or avoiding an obstacle—to the surrounding ITS-Ss. Three conditions are defined mandating the instantaneous generation of a CAM, all regarding the respective values included into the last one generated: a difference in heading exceeding 4°, in position exceeding 4 m or in acceleration exceeding 0.5 m/s^2. CAMs are encoded using the ASN.1 Unaligned Packed Encoding Rule (UPER) [15] which aims to create minimum size bit streams from data sets. They shall be transmitted in the CCH [7].

3.5 Security Services

As information communicated between ITS-Ss is intended to be used for control applications, its reliability is of utmost importance. Mechanisms to prevent unauthorized participants from sending data to the network or altering transmissions are hence required. IEEE 1609.2 defines security services available to all layers of the communication stack [13]. ITS-G5 extends these to incorporate rules for the defined facility layer messages. A PDU, the security header, which encapsulates other PDUs like CAMs, and elliptical curve cryptographic services (encryption and digital signing) are defined in [12]. The specification requires CAMs to be digitally signed by the sending ITS-S so receivers can verify their integrity and detect unauthorized messages. The defined mechanisms are based on a certificate system where the authenticity of an ITS-S is warranted by trusted entities (Certificate Authorities, CAs), not all of which may be known by each ITS-S. Therefore mechanisms for resolving unknown certificates and verifying their authenticity are also defined. The security header is encoded according to the TLS RFC [3, 12].

4 Evaluation Framework and Methodology

For conducting this evaluation a custom simulation framework was developed. It extends Veins [20, 21], a simulation framework targeting IEEE 802.11p VANETs. Veins couples the popular traffic simulator SUMO [17] with the network simulator OMNeT++ [22]. The available framework simulates an IEEE 802.11p PHY, an IEEE 1609.4 MAC and the wireless communication between ITS-Ss. Veins was chosen over other simulation frameworks like iTETRIS [18] as it is actively being developed and the network definition mechanism of OMNeT++ is very intuitive for programmers. The latter point was important as extensions to the framework had to be made to adapt it to ITS-G5. We also developed custom C++ implementations of the data types and structures defined for the CDD, CAM and security header, the GeoNetworking and BTP PDUs and ASN.1 UPER and TLS encoders.

Implementing the data types and structures defined in ITS-G5 was necessary as no C/C++ implementation was available. The encoders were implemented as the ones available as open source did not match our demands. To add ITS-G5 CAM support to Veins the existing MAC was extended to add the DCC mechanism and the CA basic service added. The Security basic service is not implemented completely as of now; the CAMs are encapsulated into a standard Security header containing signer information. Unknown certificate resolution will be added in the future.

Using the developed framework we carried out simulation runs for different amounts of vehicles moving across a straight road with six lanes at a constant speed of 130 km/h (36.11 m/s). Vehicles were placed with a time gap of one second, i.e. 36.11 m apart to ensure each ITS-S is in the communication range of all others.

Table 1 Cumulative size of CAM high and low frequency packets

	High frequency	Low frequency
CAM	330 bit	347 bit
Security header	93 byte	
BTP header	4 byte	
GeoNetworking (SHB)	36 byte	
LLC	8 byte	
MAC	28 byte	
Total	211 byte	213 byte

The generation of CAMs is started when all vehicles have entered the road and reached their target speed to guarantee a constant channel load. Each ITS-S adds a random offset from 0 to 500 ms from this time until it generates the first CAM to minimize the risk of collisions. In a real environment the CA basic services of different vehicles would also be asynchronous. We chose this setup as it best resembles the environment in which CACC applications will typically operate: a highway condition with high vehicle speeds and few obstructions to wireless signals. In this scenario, each ITS-S will generate a CAM approximately every 111 ms. Only the mandatory information is included into the generated messages, yielding the smallest possible CAMs. Table 1 gives an overview of the PDU size as a CAM is passed down through the network stack. Packets are sent via the AC_VI queue.

The DCC mechanism was configured according to the values specified in [10] and shown in Table 2. For the Active state, however, we constrained the packet interval to 0.1 s rather than the 0.5 s default packet interval. This is consistent with the specification as the Active state for the CCH only enforces TPC [10]. This implies that applications are free to override the default values defined for other parameters—like the packet interval—as long as they do not exceed the minimum and maximum values. We chose 0.1 s for the packet interval longer intervals quickly render timing critical applications like CACC impossible. The simulation was run with 10, 20, 30, 40, 50, 60 and 90 vehicles participating in the VANET. Different statistics were recorded, the channel load, DCC state, packet collisions as well as the minimum, mean and maximum age of information from neighbor vehicles. Each run lasts 300 s, the results are presented in the following section.

Table 2 DCC parameter settings for AC_VI used in the simulation; from *left* to *right*: states RELAXED, Active, RESTRICTIVE

	Channel load		
Parameter	<15%	≥ 15%, <40%	≥ 40%
TX power	33 dBm	25 dBm	–10 dBm
Packet interval	0.04 s	0.1 s	1 s
Datarate	3 Mbps	6 Mbps	12 Mbps
RX sensitivity	–85 dBm	–85 dBm	–65 dBm

5 Results

In this section we present statistics of the channel load and CA update rate recorded in the simulation runs. The channel load gives an impression of how much the communication system is stressed by the application. Furthermore, it allows for assessing the effectiveness of the DCC mechanism. Table 3 shows the mean of the channel load values recorded by the individual vehicles as well as the maximum value. Channel load is defined as the amount of time the wireless channel was sensed busy in reference to a measurement interval. The measurement interval is one second with a granularity of 10 μs as proposed in [10].

The runs with 10 and 20 vehicles only generate a light load which does not activate the DCC mechanism. The maximum information age, also presented in Table 3, equals the CAM generation interval. The values reported are the worst ones encountered for all vehicles in the run. The run with 30 vehicles shows a different picture; while the mean information age is only slightly above the CAM generation interval, the maximum age is significantly longer. It is also evident that the maximum measured channel load is above 15% and the recorded DCC state shows that one of the vehicles enters Active state. Examining other recorded statistics reveals that several vehicles in this scenario experience packet collisions, thereby leading to CAMs of those vehicles being lost. The chance of collisions obviously rises as the number of vehicles increases but is also dependent on the random CAM generation offset and MAC backup mechanism. This outlier clearly highlights the shortcomings of the IEEE 802.11p medium access mechanism.

The remaining runs show the DCC mechanism in action. The doubled data rate decreases the channel load below the 15% threshold, causing DCC to constantly fluctuate between RELAXED and Active state for the runs with 40, 50 and 60 vehicles. Figure 2 shows a sequence of the channel load as measured by three different vehicles in the 40 vehicles scenario. The run with 40 vehicles shows that the DCC mechanism effectively reduces packet collision likelihood due to the increased data rate in Active state; maximum and mean information age is close to

Table 3 Measured channel load and update rate statistics

Vehicle	Channel load (%)		Age (s)		
Count	Max	Mean	Max	Mean	StdDev
10	4.7	4.6	0.1109	0.1108	1.6×10^{-5}
20	9.3	9.2	0.1108	0.1108	7.3×10^{-6}
30	18.1	13.7	0.6652	0.1136	0.02
40	16.8	11.8	0.1119	0.1108	4.8×10^{-5}
50	19.1	13.4	1.44	0.1147	0.04
60	21.4	15.1	2.437	0.1141	0.045
90	30.2	21.6	4.4308	0.1134	0.04

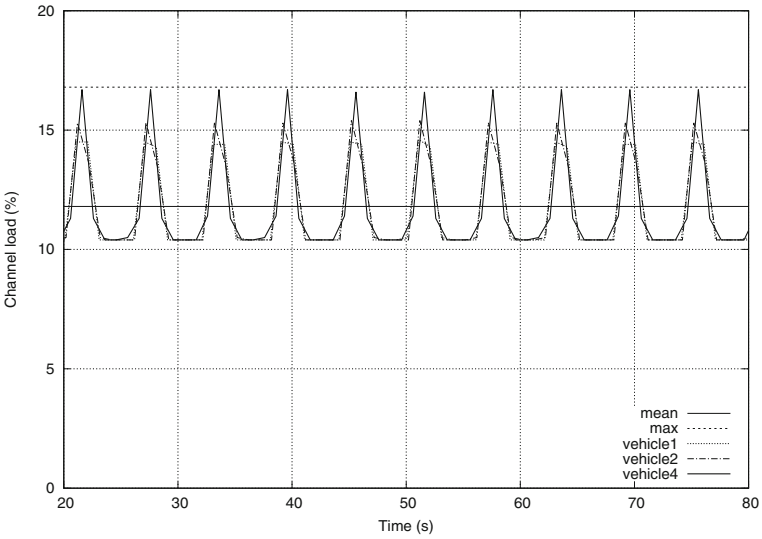


Fig. 2 Channel load measurements (60 s detailed view) of three vehicles in the simulation run with 40 communicating vehicles. The DCC state machine alternates between RELAXED and Active state in six seconds intervals (one second RELAXED, the peaks, five seconds Active). DCC returns to RELAXED state as the channel load drops below the 15% threshold in Active state

the values observed in the 10 and 20 vehicle scenarios. The 90 vehicles scenario shows a mean channel load of 21.6%, DCC remains in Active state the whole time.

Looking at the update rate statistics in Table 3 reveals that, although the maximum information age starts to rise at the 50 vehicles run, the mean and standard deviation show that the majority of messages is successfully delivered. This is even the case for the 90 vehicles scenario, implying that demanding tasks like CACC could still be feasible. It is important to note that the message interval demanded by the CA basic service is still maintained. The presented results, however, do not reveal much insight on their own. To judge whether this should encourage vendors to adapt ITS-G5 and use it to implement next generation driver assistance systems it has to be put in relation. In our scenario, assuming a 1.5 s time gap between vehicles which are 5 m long, 400 vehicles would be in a 4 km section of the highway. For our scenario we showed that 22.5% of the vehicles could likely be equipped with ITS-G5 hardware without impairing timing sensitive tasks like CACC. However, a practical number will be lower as DENMs and the unknown certificate resolution protocol were not taken into account. Evaluating their impact requires an extension of the simulator and a definition of more sophisticated scenarios as their generation mechanisms are not as simple as that of CAMs.

6 Conclusion and Future Work

After performing this evaluation we conclude that ITS-G5 is not quite as bad as many other publications suggest. Nevertheless better technologies have to be developed, especially to better avoid packet collisions which have been observed to become highly probable even at light channel loads approaching the DCC limit. The IEEE 802.11p based standards are hence not viable for an all-connected future but an important enabler for it. To get a better insight into vehicular applications, we plan to extend the simulator to fully conform to the ITS-G5 standard. Our vision is to provide a simulation framework which can be used for evaluating new use-cases of VANETs and compare different networking technologies. We also plan to perform further runs with more vehicles which have been omitted for this publication due to the prohibitively long run times. Additionally, researching higher-level mechanisms to prevent packet collisions seems worthwhile. In our scenario vehicles generating CAMs at the same time face the collision of every message. In one run two vehicles were observed to have lost more than half of their generated messages due to collisions. Furthermore it is planned to alter the realism of the simulation scenario by using real highway sections which would allow us to study situations of varying network load and their impact on different applications.

Acknowledgments This work was conducted in the scope of the KOALA project funded by the German Research Foundation (Deutsche Forschungsgesellschaft, DFG) in SPP1835.

References

1. Autolitano A, Campolo C, Molinaro A, Scopigno R, Vesco A (2013) An insight into decentralized congestion control techniques for VANETs from ETSI TS 102 687 V1.1.1. IFIP Wireless Days, pp 1–6
2. Böhm A, Jonsson M, Uhlemann E (2013) Performance evaluation of a platooning application using the IEEE 802.11p MAC on a control channel Vs. a centralized real-time MAC on a service channel. Available via www.diva-portal.org
3. Dierks T, Allen C (1999) RFC 2246: The TLS protocol Version 1.0. Internet Engineering Task Force
4. Eichler S (2007) Performance evaluation of the IEEE 802.11p WAVE communication standard. In: IEEE 66th vehicular technology conference, pp 2199–2203
5. ETSI EN 302 636-4-1 Vehicular communications; GeoNetworking; Part 4: geographical addressing and forwarding for point-to-point and point-to-multipoint communications; Sub-part 1: Media-Independent Functionality v1.2.1, European Telecommunication Standards Institute, 2014
6. ETSI EN 302 636-5-1Vehicular communications; GeoNetworking; Part 5: transport protocols; sub-part 1: basic transport protocol v1.2.1. European Telecommunication Standards Institute, 2014
7. ETSI EN 302 637-2 Vehicular communications; basic set of applications; Part 2: specification of cooperative awareness basic service v1.3.2. European Telecommunication Standards Institute, 2014

8. ETSI EN 302 663 Access layer specification for intelligent transport systems operating in the 5 GHz frequency band v1.2.0. European Telecommunication Standards Institute, 2012
9. ETSI EN 302 665 Communication architecture, v1.1.1. European Telecommunication Standards Institute, 2010
10. ETSI TS 102 687 Decentralized congestion control mechanisms for intelligent transport systems operating in the 5 GHz range; Access layer part v1.1.1. European Telecommunication Standards Institute, 2011
11. ETSI TS 102 894-2 Users and applications requirements; Part 2: applications and facilities layer common data dictionary v1.2.1. European Telecommunication Standards Institute, 2014
12. ETSI TS 103 097 Security; security header and certificate formats v1.2.1. European Telecommunication Standards Institute, 2015
13. IEEE 1609.2-2012 (2012) Wireless access in vehicular environments—security services for applications and management messages. IEEE
14. ITU-T X.680 Abstract Syntax Notation One (ASN.1): specification of basic notation. International Telecommunication Union, 2015
15. ITU-T X.691 ASN.1 encoding rules: specification of packed encoding rules (PER). International Telecommunication Union, 2015
16. Jiang D, Delgrossi L (2008) IEEE 802.11p: Towards an international standard for wireless access in vehicular environments. In: Vehicular technology conference, pp 2036–2040
17. Krajzewicz D, Hertkorn G, Rössel C, Wagner P (2002) SUMO (Simulation of Urban MObility) - an open-source traffic simulation. In: 4th Middle east symposium on simulation and modelling, pp 183–187
18. Rondinone M, Maneros J, Krajzewicz D, Bauza R, Cataldi P, Hrizi F, Gozalvez J, Kumar V, Röckl M, Lin L, Lazaro O, Leguay J, Härri J, Vaz S, Lopez Y, Sepulcre M, Wetterwald M, Blokpoel R, Cartolano F (2013) iTETRIS: a modular simulation platform for the large scale evaluation of cooperative ITS applications. Simul Model Pract Theory 34:99–125
19. Schmidt-Eisenlohr F (2010) Interference in vehicle-to-vehicle communication networks. KIT Scientific Publishing, Karlsruhe
20. Sommer C, Dressler F (2008) Progressing toward realistic mobility models in VANET simulations. IEEE Commun Mag 46(11):132–137
21. Sommer C, German R, Dressler F (2011) Bidirectionally coupled network and road traffic simulation for improved IVC analysis. IEEE Trans Mobile Comput 10(1):3–15
22. Varga A (2010) OMNeT++. In: Modelling and tools for network simulation. Springer, Berlin
23. Vesco A, Scopigno R, Casetti C, Chiasserini C (2013) Investigating the effectiveness of decentralized congestion control in vehicular networks. In: IEEE globecom workshops

Prototyping Framework for Cooperative Interaction of Automated Vehicles and Vulnerable Road Users

Timo Pech, Matthias Gabriel, Benjamin Jähn, David Kühnert, Pierre Reisdorf and Gerd Wanielik

Abstract The continuous development and implementation of highly automated driving functions for vehicles raise new issues in traffic research, as e.g. the effect of automated vehicles on the driver and the surrounding traffic participants. For the efficient implementation of scientific investigations about advanced cooperative interaction between automated vehicles and other road users, generic hardware and software modules must be available. This paper presents the concept vehicle *Carai3* for automated driving as well as relevant algorithmic components for investigating cooperative interactions of road users. Both are part of the prototyping framework of the Professorship for Communication Engineering from Technische Universität Chemnitz. Finally, applications addressing cooperative interactions between road users which are based on said prototyping framework are introduced.

Keywords Automated driving · Road side unit · Cooperative interaction of road users · Data fusion · Intent estimation · VRU protection · ADAS development · Intelligent transportation systems

T. Pech (✉) · M. Gabriel · B. Jähn · D. Kühnert · P. Reisdorf · G. Wanielik
Chemnitz University of Technology, Chemnitz, Germany
e-mail: timo.pech@etit.tu-chemnitz.de

M. Gabriel
e-mail: matthias.gabriel@etit.tu-chemnitz.de

B. Jähn
e-mail: benjamin.jeahn@etit.tu-chemnitz.de

D. Kühnert
e-mail: david.kuehnert@etit.tu-chemnitz.de

P. Reisdorf
e-mail: pierre.reisdorf@etit.tu-chemnitz.de

G. Wanielik
e-mail: gerd.wanielik@etit.tu-chemnitz.de

T. Schulze et al. (eds.), *Advanced Microsystems for Automotive Applications 2016*, Lecture Notes in Mobility, DOI 10.1007/978-3-319-44766-7_4

1 Introduction

To increase the safety, comfort and efficiency of road transport, the development of individual mobility continues to progress. In this context, the ongoing automation of driving functions raises new questions and issues in traffic research. One of these issues is the impact of automated driving vehicles on the surrounding traffic, especially interactions with vulnerable road users (VRUs).

Most research interests are focused always on the optimal support of the driver as the central user of these new systems, including investigations of effects to her or his comfort and safety. For a safe, comfortable and efficient individual mobility, also the interactions between road users are important factors. The development of automated vehicles and their integration into the road should therefore not only consider the human machine interface between driver and vehicle but also the interactions between automated vehicles and other road users in particular vulnerable road users, such as pedestrians. To address these research issues, the prototyping framework of the Professorship for Communication Engineering from Technische Universität Chemnitz [1, 2] was continuously developed and supplemented by additional hardware and software components. For rapid implementation of new approaches for scientific investigations of advanced cooperative interaction methodologies between automated vehicles and other road users, co-designed software modules and technical components are available. Due to the various sensorial systems as well as communication and localisation devices different applications can be addressed. The generic implementation of software modules support the development of novel algorithms and the close relation to the hardware components enable to evaluate these algorithms in controlled or natural traffic scenarios.

Section 2 first gives an overview of the available sensorial systems and a more detailed description of the test vehicle for automated driving *Carai3* as well as the roadside unit *Protect2*. The used software framework for data acquisition, processing and information fusion is described in Sect. 3. In Sect. 4 an excerpt of relevant algorithms in the research topic of cooperative interaction between road users is presented. Finally, in Sect. 5, applications in research projects are introduced that build up on the described prototyping framework.

2 Prototyping Hardware Equipment and Sensorial Systems

2.1 Overview of Sensorial Systems

To address different applications and studies in traffic research as well as the development of prototype functions in the field of driver assistance systems and automated driving, the professorship of communication engineering has three

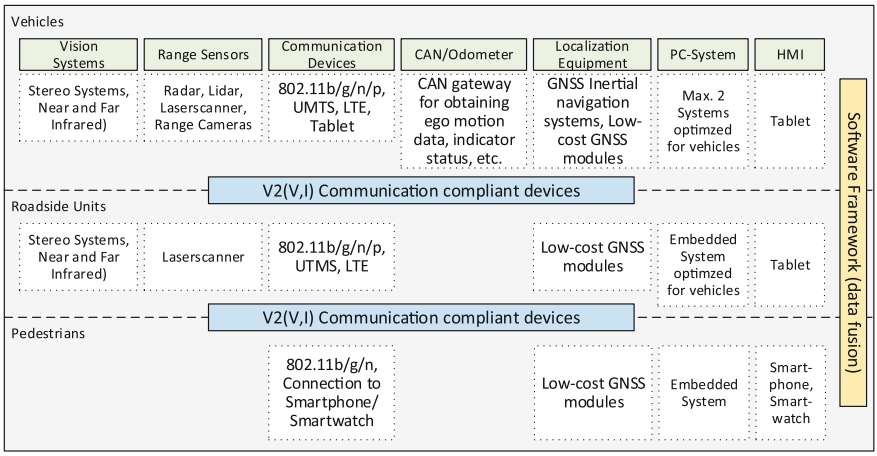


Fig. 1 Overview about test equipment for rapid prototyping

sensor-equipped vehicles *Carai1–3*, two roadside units *Protect1&2* and several mobile devices for pedestrians. The roadside units (RSUs) are divided in a static [3] and a mobile one. Thus, different static and mobile use cases can be handled with the RSUs. All these test beds for traffic research are equipped with various sensors. Figure 1 shows an overview about the different sensor categories. For environment recognition, the test vehicles and the RSUs are equipped with range sensors like Radar or LiDAR and different vision sensors. Precise localization is realized with Global Navigation Satellite System (GNSS). Low cost GNSS modules are also used in combination with different filter methodologies [4]. To realize decentralized sensing and investigate strategies for cooperative mobility communication devices for wireless standards such as Car-2-X (IEEE 802.11p) and LTE (Long-Term Evolution) are available. Interior cameras monitor the driver's glance behaviour and her or his foot position. The current status of the test equipment is a continuous development of the used infrastructure with respect to the state reported in [1, 2]. Due to the available number of test equipment, several use cases can be addressed.

A key factor to use and interact with different hardware is a useful Human Machine Interface (HMI). For this purpose, it is possible to connect the test vehicle with a tablet to inform or warn the driver. Pedestrians can be connected with a smartphone or a smartwatch. Thus, novel HMI concepts can be prototyped rapidly.

2.2 *Research Vehicle for Automated Driving*

In the prototyping framework, it is the test vehicle *Carai3*, which has additional actuators to realize automated driving functions. It is a BMW i3, an electric vehicle for urban mobility. The i3 came with automated features, but does not allow

manipulating these features and using it in a development framework. Therefore, an open and fully customizable system for automated driving was installed. This system can be incorporated in nearly every vehicle with automatic transmission. Based on this demands a system for longitudinal and lateral control was developed out of the emulation of the electric accelerator pedal, a mechatronic brake system and a mechatronic steering system. The automation system is built of different modules, so it is possible to investigate each control function in dependency on the degree of automatization separately. In parallel to the original accelerator pedal a module was installed, which simulates the voltage signals of the two hall sensors of an accelerator pedal. This makes it possible to emulate a synthetic signal for various pedal positions and ramps. The solution also includes the development of a system to realize the operation of the brake pedal. As aforementioned, it was decided to install a mechatronic braking solution, based on a servo drive with a gearbox to ensure the momentum of emergency braking. Fully automated driving necessitates besides longitudinal also a lateral control of the vehicle. In *Carai3* it is also realized with a mechatronic component based on the same type of servo motor, like the braking system, but with a gearbox, that has a suitable combination of momentum and speed. It is able to turn the steering wheel up to 1920 °/s and steer the vehicle while standing. Therefore, the system is able to realize parking situations and emergency turns. The system is mounted directly to the non-moving part of the steering column and is connected with a gear to the cardan joint. There are no restrictions for the driver, like the steering column adjustment. With this solution, an independent and open vehicle control system was realized for the implementation and evaluation of automated driving functions (Fig. 2).

The automated driving Carai3 is additionally equipped with sensors, which detect the driver's actions such as hands on the steering wheel as well as foot on accelerator or braking pedal. If one of the resistive touch sensors is triggered, automated driving will be stopped and the driver gets back the full control of the car. It is additionally possible to switch between automated driving mode and driver mode with control elements at the front, which can be controlled by the driver himself, or at the back of the vehicle by a supervisor.

Fig. 2 Electronic research vehicle Carai3

2.3 Prototyping Testbed—Mobile Road Side Unit

For a more flexible usability of the RSU *Protect1* described in [2], a mobile RSU *Protect2* was developed. With a mobile RSU, the amount of addressable traffic situations can be extended. It is equipped with a battery pack; therefore, for short-term usage for example in supervised research studies it is not necessary to connect the RSU to the public electricity network. The selection of the installed components was made with attention to low power consumption. Like *Protect1* [2], *Protect2* is equipped with vision sensors, localization equipment, communication devices and a range sensor. With the support of different communication standards, such as IEEE 802.11b/g/n/p and LTE, different traffic participants can exchange information via the RSU. If the position of the RSU is measured with a high accuracy GNSS system (sub-meter level) this information can be used to support the localization of the connected road users as a kind of differential Global Positioning System (DGPS) [3].

2.4 Mobile Devices for VRUs

The motivation of using a self-designed mobile sensing system for VRU is the full access to raw sensor data. The VRU device consists mainly of a Raspberry Pi as embedded PC, a low-cost localization device (u-blox) and a small battery pack. Due to the possibility of accessing the raw positioning data from the GPS receiver, the benefits of relative localization methodologies [3–5] can be fully exploited. If the mobile device of the VRU is connected to a near RSU, the high accuracy position of the RSU can be used to improve its own localization. Connected to a Smartphone or Smartwatch, the VRU can be informed about the surrounding traffic situation.

3 Software Framework for Prototyping

3.1 Software Modules Overview

Accompanying the hardware, a software framework is build and used to prototype concrete and high-level applications. For the core functionality, we use BASELABS [2] software that models abstract components and provides interconnection between them in .Net/C#. This interconnection is data-flow oriented and generic. It allows to abstract functionality and reuse acquisition- and processing components as well as user interfaces in different contexts. Further it offers functionality to record and playback data streams, which is useful for designing and testing algorithmic components. Built around this core framework, an extensive library of components is maintained, which consists of data structures and

algorithms that can be reused and adapted for concrete use cases. These include sensor interfaces for various sensor SDKs (Software Development Kits), such as cameras, GNSS, RADAR and LiDAR, as well as communication interfaces such as TCP, UDP, Car-2-X and CAN. Sensors typically offer ready-to-use data structures that can be processed directly by the consuming algorithms.

Generic and common algorithms include:

- Probabilistic filters for object tracking including multiple movement and state models
- GNSS Localization with multipath and non-line-of-sight detection through digital maps
- Data fusion algorithms for multitarget-multisensor tracking
- Generic classifier
- Manoeuvre planning based on splines

3.2 Algorithmic Components

In this section, selected algorithmic components are described which are essential in applications focusing on cooperative interactions between road users in the context of automated driving.

3.2.1 Vehicle Trajectory Representation

In autonomous driving scenarios it is absolutely necessary that the trajectory of the vehicle, which includes the actual path it should take and its velocity along this, is planned in advance. In our scenario we have chosen a spline based approach as the basic mathematical description of the path and the velocity. The path was defined as a 2D parametric spline curve with two polynomials per spline segment, one for each spatial dimension. The velocity profile was separately defined as a spline over the arc length of the path spline.

Following the considerations of [1] it was decided to use polynomials of degree 5 because they provide us the possibility to keep geometrical continuousness at the knot points, until the second order derivative, together with the ability to change individual segments without the need to re-compute the whole spline.

Accordingly, each path segment was defined by its spatial start and end point together with the first and second derivative over the internal curve parameter at these points. In order to avoid numerical problems with high degree polynomials, we internally store rescaled versions of them. There, the geometric extent of each segment was reduced to (–1; 1) and the internal spline parameter was limited to the interval (0; 1). Thus, we stay in an area where the mathematical calculations on the polynomials, using floating point numbers, are numerically uncritical for the required accuracy.

Fig. 3 Path spline, Map data © OpenStreetMap contributors, CC BY-SA

Fig. 4 Velocity profile along the arc length of the path spline (Fig. 3)

Figures 3 and 4 illustrate an example of the splines, the path and the velocity profile. Using the GPS position, the vehicle is able to localize itself on the path spline by calculating the parameter of the closest position. Afterwards the planned velocity can be determined from the velocity profile by calculating the distance on the spline from the beginning to the current position.

This approach permits that each velocity profile can be described in a mathematical way. In the research project DriveMe, the spline approach was extended to model the individual driving style of a driver to find the individually optimal automated driving style [6].

3.2.2 Intent Estimation

By sharing the intent of a road user in cooperative applications with others possible conflicts can be avoided bevor they occur. To get a detailed situational awareness it is important to know what manoeuvre road users plan to perform. The intention of a road user cannot be directly measured and therefore it needs to be estimated from the fusion of different measurement data. To estimate for example if a driver plans to execute a lane change manoeuvre, information about the surrounding environment in combination with the vehicle kinematics and measurements about the driver himself are relevant. Due to the significant relevance of the visual perception, the driver's visual orientation is an important aspect for estimating her or his behaviour [7]. In order to draw conclusions about the behaviour of the driver from his visual orientation, it is mostly sufficient to determine the glance area based on the head orientation. Due to the dynamic and amplitude of head rotations such as shoulder glances head tracking systems are often more stable and solutions with

only one camera are possible. By taking advantage of the correlations between eye and head movement, the viewing direction and the visual field could be estimated by head pose measurement [8]. To detect for example if the driver looks into the left, right or rear view mirror a Bayes classifier can be used [8]. It determines the probability that the driver looks at one of this predefined area. It uses the head yaw movement that is the most relevant feature. Trained with annotated naturalistic driving data of 60 test persons the classifier reaches a mean accuracy of 75 %. With the knowledge about the driver's mirror glances and additional information about the current surrounding traffic situation the intent to execute a lane change manoeuvre can be estimated more reliable.

4 Application Scenarios

In this section of the paper already implemented applications and future research will be introduced that is based on the existing components of the described prototyping framework.

4.1 Manoeuvre Planning for Automated Green Driving and VRU Safety

In the research project NeMoS ("New Mobility at the Sachsenring") [9], among other interactions of road users and prototypical infrastructure components with regard to energy efficiency and safety were investigated. In this context a study with test persons took place on a test track at the Sachsenring in order to make statements about requirements for urban electric mobility of the future and prototypical functions to support an economical driving style. Based on the study an adaptive trajectory planning of the test vehicle *Carai3* with automated longitudinal guidance was demonstrated exemplarily. This was done based on the information exchange of the automated vehicle and a pedestrian via a RSU. For this, the hardware and software components of the prototyping frameworks described in this paper were used in order to realize this application in practice. Figure 5 shows the situation represented in the demonstration.

The test vehicle Carai3 drives on a predefined route. The pedestrian crossing that is located on the route of the automated vehicle can only be observed with on-board sensors at the time if the vehicle is turning to the left. In the case a pedestrian would like to cross, the automated vehicle has not much time to initialize a stopping manoeuvre. This short-term reaction can cause lack of efficiency and comfort. Especially for energy efficiency, a forward-thinking driving style is essential. For an optimal use of the recuperation function of the electric test vehicle, information about dynamical objects on the planned route are needed to be known as early and

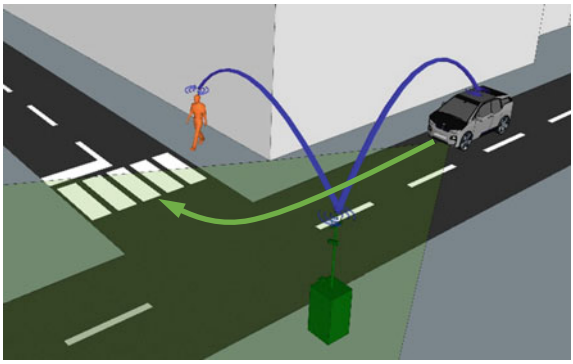


Fig. 5 Traffic scenario of the adaptive manoeuvre planning demonstration for green driving automated vehicles

reliably as possible. In the demonstrated situation, the RSU supports the perception of the vehicle through the transmission of information about pedestrians intending to cross the crosswalk by Car-2-X. This information based on GPS measurements of the pedestrian's mobile device, which are transferred to the RSU via Wi-Fi if the pedestrian is near the crosswalk and the RSU. With this information exchange, the automated vehicle can adapt its planned trajectory by an early initiation of a stopping manoeuvre considering energy-efficiency and the driving comfort.

4.2 Cooperative Interactions Between VRU and Automated Vehicles

What impact do automated driving vehicles have on other road users, especially on pedestrians? This research issue is investigated in the project KIVI [10]. It aims on the development and validation of novel methods for cooperative interaction with vulnerable road users in automated driving. Potential conflicts between automated vehicles and pedestrians in shared traffic spaces, like parking places or intersecting regions, need to be identified and solved in a cooperative way.

The knowledge about the intention of other road users in traffic situations plays an important role in this context. In normal situations, information can, for example, be signalised by using the indicator or by other explicit as well as implicit indication, like observing the trajectory or behaviour of the other road user. The driver and the pedestrian can estimate the intention of the respective other road user based on such indicators. Consequently, a highly automated vehicle should be able to assess if, for example, a pedestrian intends to cross the street. With knowledge about the pedestrian's intent the automated vehicle needs to react in a cooperative way to avoid a conflict situation in advance. The investigation of this research issue such as the development and evaluation of algorithms that realize cooperative interaction between automated vehicles and pedestrian will build up on the introduced prototyping framework.

5 Conclusion

In this paper, a prototyping framework for applications focusing on cooperative interactions between road users in the context of automated driving has been presented. The benefit of generic hardware components is a rapid prototyping of new methodologies and functions for ADAS up to highly automated driving. Novel algorithms can be developed and evaluated quickly using the introduced prototyping framework to investigate actual and future research issues in the field of cooperative interactions between road users.

Acknowledgment The research project NeMoS was funded by the Free State of Saxony (grant number: 100172504) in the scope of the initiative "Elektromobilität verbindet" ["electromobility connects"] as part of the German programme "Schaufenster Elektromobilität" ["showcase electromobility"].

The research project KIVI—"Kooperative Interaktion mit schwächeren Verkehrsteilnehmern im automatisierten Fahren" ["Cooperative interaction with vulnerable road users in automated driving"] is funded as part of the DFG (German Research Foundation)-Schwerpunktprogramm 1835: Kooperativ interagierende Automobile.

The authors would like to thank Michael von der Trenck for his commitment in equipping the *Carai3* and RSUs with all required components.

References

1. Schubert R, Richter E, Mattern N, Lindner P, Wanielik G (2010) A concept vehicle for rapid prototyping of advanced driver assistance systems. In: Meyer G, Valldorf J (eds) Advanced microsystems for automotive applications: smart systems for green cars and safe mobility. Springer
2. Streiter R, Adam C, Bauer S, Obst M, Pech T, Reisdorf P, Schubert R, Thomanek J, Welzel A, Wanielik G (2013) A prototyping ITS station for advanced driver assistance systems and pedestrian safety. In: Fischer-Wolfarth J, Meyer G (ed) Advanced microsystems for automotive applications 2013. Lecture notes in mobility. Springer International Publishing, pp 89–99
3. Obst M, Mattern N, Schubert R, Wanielik G (2012) Car-to-Car communication for accurate vehicle localization—the CoVeL approach. In: Proceedings of the 9th international multi-conference on systems, signals and devices, Chemnitz, Germany
4. Streiter R, Bauer S, Reisdorf P, Wanielik G (2013) GNSS Multi receiver fusion and low cost DGPS for cooperative ITS and V2X applications. In: Proceedings of the 9th ITS European congress, Dublin, Ireland
5. Obst M, Richter E, Wanielik G (2011) Accurate relative localization for land vehicles with SBAS corrected GPS/INS integration and V2V communication. ION GNSS, 2011 Portland, Oregon
6. Scherer S, Dettmann A, Hartwich F, Pech T, Bullinger AC, Wanielik G (2015) How the driver wants to be driven—modelling individual driving styles in highly automated driving. 7. Tagung Fahrerassistenz, München, 25–26 Nov 2015. https://mediatum.ub.tum.de/doc/1294967/1294967.pdf
7. Beggiato M, Krems JF (2013) Sequence analysis of glance patterns to predict lane changes on urban arterial roads. 6. Tagung Fahrerassistenz—Der Weg zum automatischen Fahren, Munich, 28–29 Nov 2013

8. Pech T, Lindner P, Wanielik G (2014) Head tracking based glance area estimation for driver behaviour modelling during lane change execution. In: 2014 IEEE 17th International conference on intelligent transportation systems (ITSC). Qingdao, China, S. 655–660
9. NeMoS Internet Presence (as part of the German project initiative "Schaufenster Elektromobilität Bayern-Sachsen"). http://www.elektromobilitaet-verbindet.de April 2016
10. KIVI Internet Presence (as part of the DFG-Schwerpunktprogramm 1835: Kooperativ interagierende Automobile). http://gepris.dfg.de/gepris/projekt/273258218 April 2016

Communication Beyond Vehicles—Road to Automated Driving

Steffen Müller, Timo van Roermund and Mark Steigemann

Abstract Innovation by semiconductor manufacturers is a key requirement for future cars. In communication architectures of today's high-end cars we see more than one hundred control units, a trend that continues to increase. When we introduce Automated Driving, communications will also have to pass the boundary of the car. A closer look unveils the need for steadily increasing bandwidth, higher fault tolerance, and real-time performance. Data security is an urgent issue, given the prominent car hacks in the US. System costs also come into the equation, as not all solutions are affordable. The following paper shows the state-of-the art car communications and roadmaps its expected evolution showing Vehicle-to-Everything (V2X) as the pre-requisite technology for Automated Driving. It also outlines an approach to categorize the levels of data security in the car and ends with an overview of requirements for secured car communication systems.

Keywords Automated driving · Secure connected car · Secure car communication · Car communication architectures · In-vehicle networking · Automotive ethernet · Switched ethernet · V2X · Car hacks · Security car architecture · Car security requirements · Smart vehicle

S. Müller (✉) · T. van Roermund · M. Steigemann
Business Unit Automotive, NXP Semiconductors Germany GmbH,
Stresemannallee 101, Hamburg, Germany
e-mail: st.mueller@nxp.com

T. van Roermund
e-mail: Timo.van.roermund@nxp.com

M. Steigemann
e-mail: mark.steigemann@nxp.com

T. Schulze et al. (eds.), *Advanced Microsystems for Automotive Applications 2016*, Lecture Notes in Mobility, DOI 10.1007/978-3-319-44766-7_5

1 Trends—Automated Driving and Smart System

Today, many different players are engaging in the field of *Automated Driving*. Carmakers are presenting and testing a variety of prototypes of highly automated vehicles, some introduced over-the-air software update. IT companies have entered automotive value chains with self-driving concept vehicles. Politicians are debating about data security, robot ethics, connectivity and the need for infrastructure.

Figure 1 shows the well-known *Automated Driving* level scheme J3016 by SAE and illustrates how the responsibility for vehicle operation moves step-by-step from driver to vehicle. Starting at level 4, we begin to talk about *Self-Driving*. At this stage, safety critical decisions and operations need consideration and alignment with the (direct) environment. At this point, V2X communication is essential.

Figure 2 gives an overview regarding what level of *Automated Driving* typically requires what set of sensors. With level 4, cocooning for the vehicle starts. Beyond the ability to communicate to the environment (V2X), sensors are to scan the direct neighbourhood. Recent developments in automotive radar lower the system costs, as level 4–5 needs multiple radar sensors for cocooning.

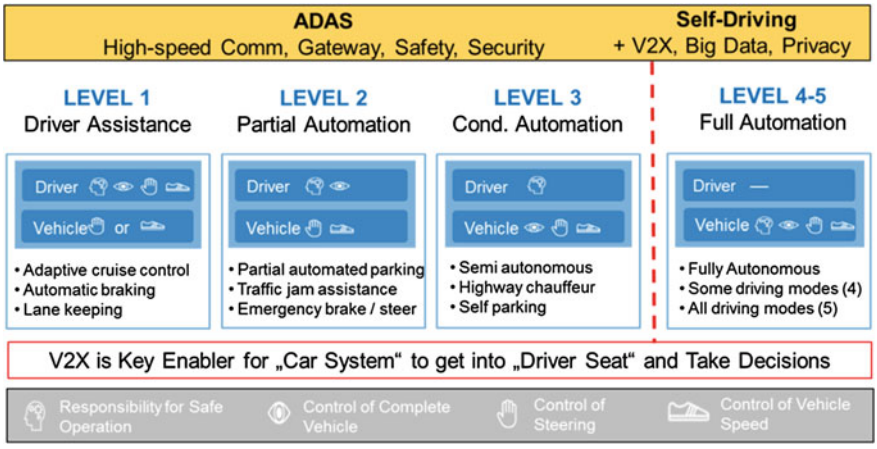


Fig. 1 Driving automation levels as defined in SAE J3016 [1] and application examples

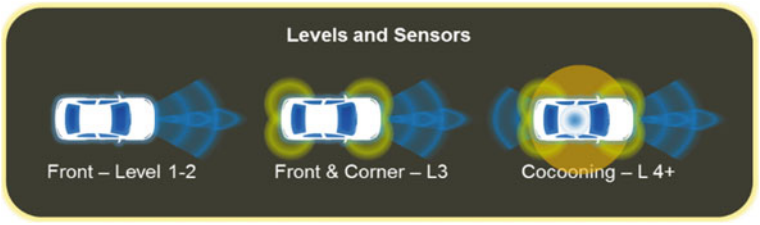
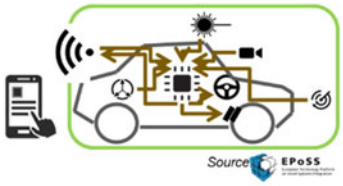


Fig. 2 Levels of automated driving and engaged sensors



Smart System Element	Effect	Means
Sensors, Actuators, and Cognitive Systems	Zero accidents by human error	ADAS
Device Reliability	Zero component failures	Robust Design
System Integration & Functional Safety	Zero accidents by system failures	ISO26262
Communication System & Functional Security	Zero accidents by system hacks	„Secure Comm."

Fig. 3 Elements of the smart vehicle/smart system

2 Robustness—the Need for Smart Vehicles

Figure 3 lists the elements of a *Smart Vehicle* and the related initiatives to ensure its system robustness. While initiatives and activities for the elements *Cognitive Systems*, *Functional Safety, and Device Reliability* (rules and tools for robust design) span enough framework to progress sufficiently, the element of *Functional Security* still needs clarification.

Functional Security focuses on the correct operation of the entire system that will in future include functions outside of the vehicle. The operation failure would be in a worst case scenario the loss of data caused by bandwidth limitations, instability of the communication system resulting from a weak link-up behavior, or intrusion of faulty data initiated by hackers. The *Secure Comunication* ensures that all risks are well considered. This forms the content of the following chapters of this contribution.

3 Evolution—Communication Architectures

Today, the *Connected Car* can be thought of as a networked computer on wheels that can have within its architecture more than one hundred control units and many sensors, interconnected by wires. The vehicle is also rapidly taking more software content on board. Over time, the *Connected Car* increasingly communicates to other vehicles and infrastructure, and to various user devices and cloud services.

Figure 4 shows complementary sensors of a smart vehicle enabling *Automated Driving*. The attached tables refer to levels as shown in Fig. 1, the expected communication protocols and the number of necessary resources bases on an

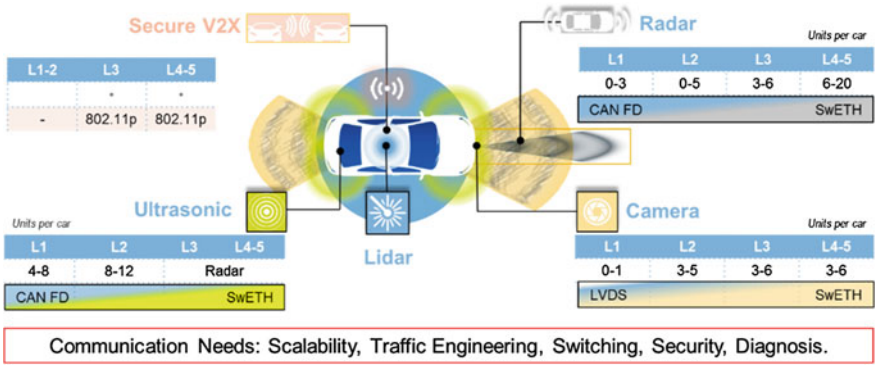


Fig. 4 Complementary sensors in automation

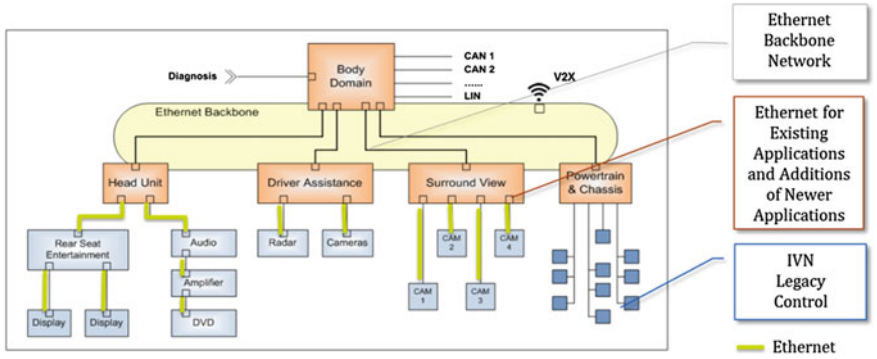


Fig. 5 In-Vehicle communication architecture "2020"

internal study and gives an indication. The cost structure of radar sensors is important as up to twenty sensors may find its entry to the higher levels, assuming that it takes up the role of Ultrasonic as well.

Switched Ethernet (SwEth) prepares the communication systems for *Scalability* (allows adding sensors over the vehicle's lifetime), allows *Traffic Engineering* (parameters latency, bandwidth, throughput), provides *Switching* (time synchronization, prioritization, traffic shaping, admission control) and *Data Security* (authentication, encryption). In view of communication protocols, V2X communication is based on Wifi 802.11p. CAN FD, FlexRay, LIN stay important in powertrain, chassis, and body domain as outlined in Fig. 5.

Ethernet is the first choice for high-speed communication in the vehicle as it is a matured technology based on the IP protocol. The eco-system has existed for a long time and the industry is familiar with its usage. The automotive domain is merely an extension of its application scope. The left side of Fig. 6 shows this eco-system.

The *Ethernet Backhaul* connects the vehicle either to user devices or via access networks to the home, enterprise, business networks, or to data centers.

In this first scenario of future driving, the vehicle is a node in the "bigger" network. Each vehicle is a terminal and there seems to be no limit to the type of transmitted data. Sensor data, control data, infotainment, plausibility and health checks, infotainment, etc. With this setup, it might be possible that a fleet of vehicles is in future somehow centrally controlled and guided through an urban area. Evolution of this technology could lead even to *Automated Driving* only cities.

The right side of Fig. 6 indicates a second scenario. The vehicle itself stays in the control. The local data acquisition by sensors is done within the vehicle and from the nearest environment. Each vehicle is a sensor by itself, each user device can add with information to the actual perception of the environment. At the end, the vehicle takes responsibility for data fusion, decision preparation, clarification with traffic partners and execution of the drive operation. It is all about provision of enough bandwidth and data availability at the right time.

While the question stays open as to which scenario will dominate in our future (the controlling or the controlled vehicle), the *Switched Ethernet* will lead *Automated Driving* into its future as the means for high-speed communication. In parallel, various other communication protocols will stay for good reasons, as there are optimized bandwidth needs versus costs and proven robustness, which is a key factor in the increasing electronic content in the vehicles. Figure 7 sketches the future *Secure Connected Car* with all its aspects.

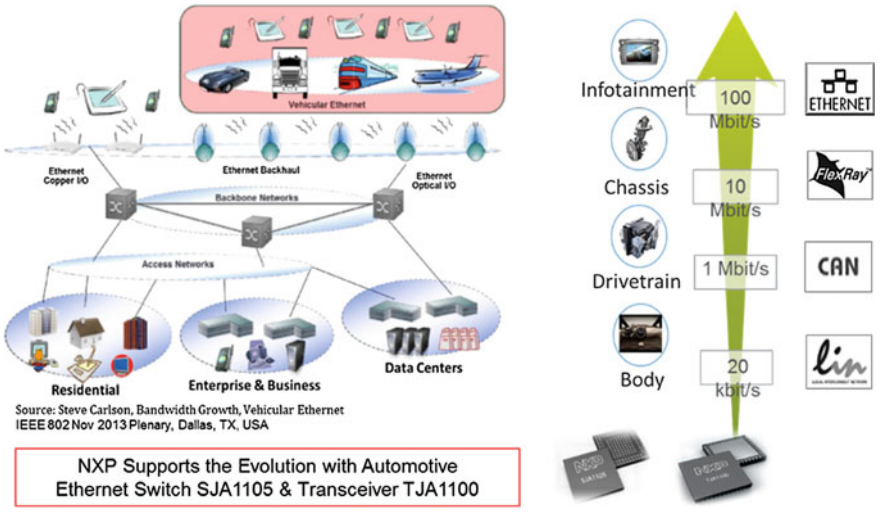


Fig. 6 Evolving trends for secure communication—automotive ethernet [2]

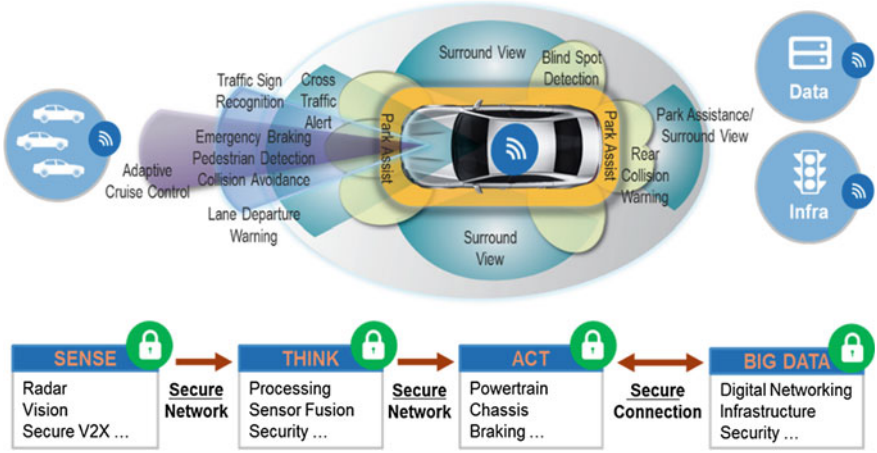


Fig. 7 Tomorrow—enabling the secure connected car

4 Essentiality—V2X Communication

V2X Communication is essential for *Automated Driving*, especially with level 3 and above. With this function, the vehicle transitions from a node that can only sense, to a node that can sense and communicate. The V2X system can be seen as a virtual sensor that increases the safety profile of the vehicle as it enables the beyond-line-of-sight sensing. This enables many new use cases, illustrated in Fig. 8: Do-not-pass warning, Platooning, Road works warning, Emergency vehicle warning.

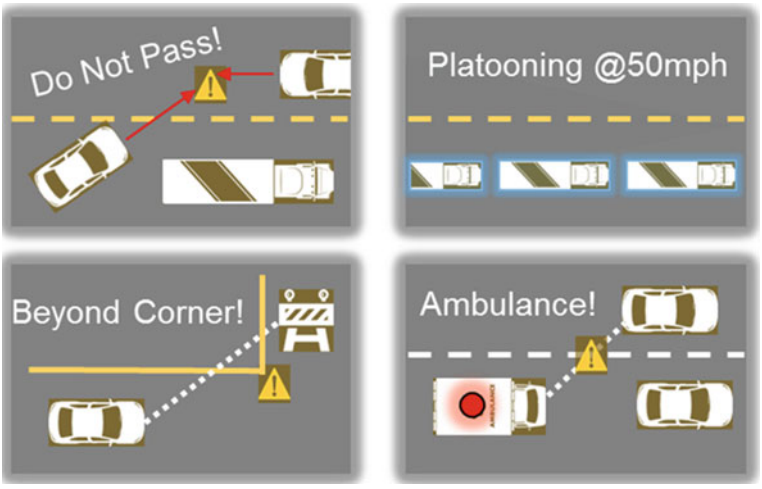


Fig. 8 Scenarios for the V2X beyond-line-of-sight sensing

Figure 9 draws the example of a secure and safe traffic interaction enabled by V2X communication based on WiFi 802.11p. This 11p communication, in the context of V2X also nicknamed *Dedicated Short Range Communications* (DSRC) [3], is a short- to- medium-range wireless ad hoc communications capability. As part of the IEEE WAVE standard, it provides low latency, ad hoc communication links between vehicles, and infrastructure, with high link reliability, priority for safety applications, as well as security and privacy—all of which are critical in communications-based active safety applications.

The *Vulnerable Road User* (VRU) is an important part of the scene. Other traffic partners are informed about the trajectory of the VRU either by a responder carried by the VRU, or by a central and local hotspot, that uses local sensors (e.g. cameras) to trace the VRU.

On top, the safety applications introduced in Fig. 8 are applied between the vehicles. Traffic management and control is applied via *Cloud*, sharing relevant information via the hotspots. Individual vehicles may use *Cloud Service* for travel information and planning. In this scenario, the "hotspot" becomes a stationary, multi-protocol, and multi-sensor service station that may relay to the *Cloud* but also generate valuable data by itself.

Vehicles are becoming increasingly dependent on information provided by their environment. As such, it is essential that the receiver can trust that a received message originates from a trusted entity, and has not been modified while in transit. In other words, *Authenticity* is a key security property for trustworthy communication in the V2X intersection scenario. The V2X standards provide a means to

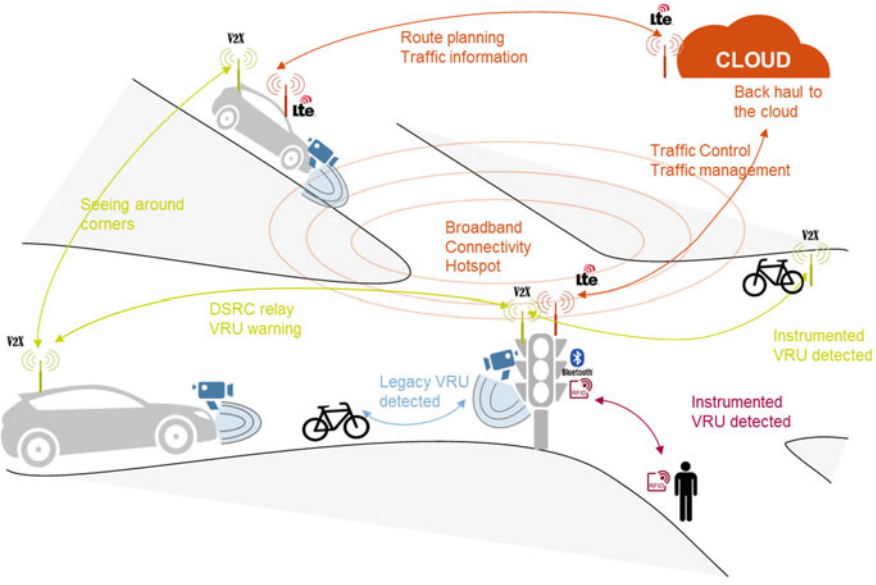


Fig. 9 Secure and safe traffic intersections

verify *Message Authenticity*, based on *Public-Key Cryptography*. The keys for message verification are exchanged using certificates, which are managed in a *Public-Key Infrastructure*.

Because more and more information is shared with the environment, also *Privacy* is a key requirement. A user does not want that others can easily track him by observing V2X communication. To protect the *Privacy* of the users, vehicles use pseudo-identities, which are regularly changed and are not directly linkable to the vehicle's (or the user's) real identity.

5 Urgency—Secured Vehicle Architectures

Several vehicle hacks was published in the media in 2015 [4]. The *Connected Car* is an attractive target for hackers. Hackers need an incentive to hack. The group of researchers follow clearly academic goals and try to warn the industry. Probably, the bigger incentive is financial as modern vehicles contain valuable data.

Modern vehicles are inherently vulnerable due to their complexity. A good indicator for the vulnerability of a system is the complexity of its software. The system of a vehicle may contain up to 100 million lines of code [5]. Assuming that 1000 lines of code contain 1 bug, and out of 1000 bugs 1 leads to a security vulnerability, then modern vehicles might have several hundreds of vulnerabilities. By comparison, a modern computer or smartphone operating system is less complex but security vulnerabilities are often found.

The vehicle's (internet) connections form an easy entry point for hackers and a possible path for remote access to the vehicle system and data. Furthermore, it may enable hackers to exploit systematic vulnerabilities at large-scale, e.g. in complete fleets. Hence, the effort and time for hackers decreases whilst the impact increases. The main reason that vehicles in the past were not hacked was simply the lack of connectivity. This is rapidly changing, as depicted in Fig. 10.

In order to secure the architecture of the vehicle, the four *Security Layers* have to be followed so that with this composition the right level of protection is given:

1. Interfaces that securely connect the vehicle to the external world, preventing unauthorized access to the network of the vehicle.
2. Secure gateways, providing isolation between domains, e.g. safety-critical electronic control units (ECU) and infotainment system.
3. Secure networks, providing secured communication between ECUs.
4. Secure processing units, embedded inside the ECUs, to protect executed software and with this to protect the functions of the vehicle.

This framework provides a *Defence in Depth* approach to secure the vehicle communication. Each *Security Layer* stands for a unique form of protection while adding to the entire defence of the system as a whole.

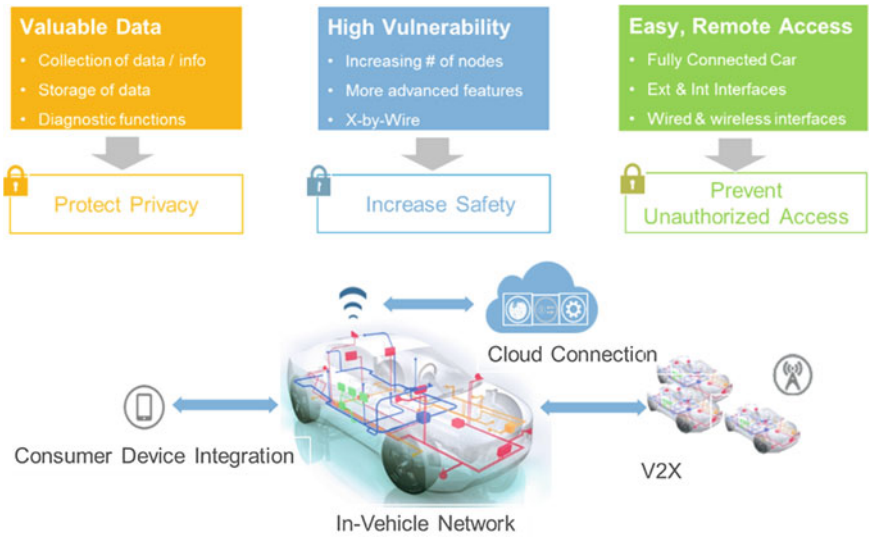


Fig. 10 The connected car is an attractive target for hackers

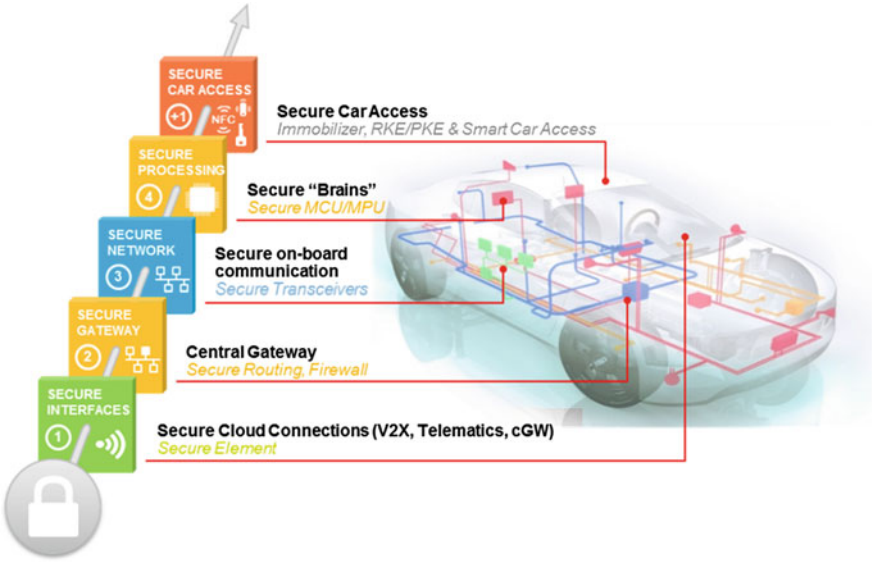


Fig. 11 NXP automotive security 4 + 1 solution

Figure 11 shows these four *Security Layers*, complemented by an extra layer of car access systems. As the leader in automotive security, NXP offers *Security Components* for each of these four *Security Layers*. *Secure Element* for protecting the Interfaces (e.g. V2X), specialized *Gateway Microcontroller* with built-in

hardware security modules (HSM), *Transceivers* with built-in security features to secure the communication on in-vehicle networks, a variety of *Secure Microcontrollers and Microprocessors* for the various ECUs, and a leading portfolio of *Car Access Solutions*. These *Security Components* are pre-requisite to ensure security of the entire system of the vehicle.

6 Outlook—Requirements Secured Car Communication

The requirements of secured car communication are summarized as follows:

- Multiple-protocol data platform—connecting via e.g. V2X, Radar, DAB, 5G, NFC, Bluetooth, 802.11p, Ethernet, CAN, LIN, FlexRay
- Integration of data—from vehicle, user, environment, service providers
- Securing data—personal, routing, infrastructure, traffic, vehicle control
- Built-in privacy—from component to overall system level
- Ad hoc scalability—add sensor, function, actor over vehicle lifetime
- Location independency—multiple e.g. cities, countryside, countries
- Develop further and maintain industry standard interfaces (wireless, wired)—as a large hacker community develops themselves as well
- Secure identities—to secure consumer trust that drives IoT market
- Handshake of politics, industry and public—consistent exertion of security and privacy by transparent design principles

References

1. SAE Levels of automated driving. Source: http://cyberlaw.stanford.edu/loda
2. Carlson S Bandwidth growth vehicular ethernet. IEEE 802 Plenary, Dallas. http://www.ieee802.org/3/400GSG/public/13_11/carlson_400_01_1113.pdf
3. http://www.etsi.org/technologies-clusters/technologies/intelligent-transport/dsrc
4. http://www.wired.com/2015/07/hackers-remotely-kill-jeep-highway/
5. https://www.linkedin.com/pulse/20140626152045-3625632-car-software-100m-lines-of-code-and-counting

What About the Infrastructure?

Jan van Hattem

Abstract What once started as Advanced Driver Assistance Systems (ADAS) has evolved into vehicle control systems that partly or completely take over the driver's task. In doing this, many assumptions are made on the design of the infrastructure that the car will have to deal with. Infrastructure, users and cars should not be looked at separately but in combination. Road operators are faced with these new developments, a larger variety of cars, different user behaviour and are restricted in budget. The areas where infrastructure providers and the car industry should work together more closely are explored in this paper.

Keywords Infrastructure · User understanding · Road operator · Legislation · Controlled road · Electronic data recorder

1 Variation in Vehicles

First of all, the majority of all vehicles driving around today will still be on the road in 2025 [1]. Advanced vehicles have to deal with the existing cars but also with an increasing variation in vehicle design. First of all there are differences in physical shape and performance. The variation in size and weight is increasing as new vehicle concepts are being developed. Small electrically powered 2, 3, and 4 wheel vehicles are being developed. Their shape will be smaller than the present car or motorcycle. It could also result in vehicles driving very close together. These vehicles will show different behaviour, accelerate extremely fast or drive very slow to save energy on a low battery. These new designs will introduce new challenges for the existing automated functions.

Another big challenge is vehicle reliability. Old or badly maintained cars can suddenly break down and present an unforeseen danger [2]. But modern cars with

J. van Hattem (✉)
Rijkswaterstaat, Wegen, Verkeer En Leefomgeving,
Lange Kleiweg 34, 2288 GK Rijswijk, Wegen, The Netherlands
e-mail: jan.van.hattem@rws.nl

T. Schulze et al. (eds.), *Advanced Microsystems for Automotive Applications 2016*, Lecture Notes in Mobility, DOI 10.1007/978-3-319-44766-7_6

more than 30 electronic modules of various suppliers and million lines of code can also break down. The records of the National Highway Traffic Safety Administration (NHTSA) Office of Defects Investigation show that all brands have experienced some issues in this respect. Due to the scale of the automotive industry an issue with one major supplier has its effects on many different brands, remember the Takata airbag inflator call back [3]. Automated systems are heavily dependent on the correct functioning of the whole car. For instance, a fluctuating voltage due to an old car battery could have an unexpected effect on other systems. Influence of the government is limited to vehicle type approval, Periodical Technical Inspection (PTI) and driver's licenses. Standardization and regulation by the Government could help the industry to advance.

2 Evolution in Car Systems

2.1 Introduction

Traditionally Road Operators see it as the responsibility of the manufacturer that the car they produce can cope with whatever comes up its way. The infrastructure is not the same across a country let alone across Europe. The maintenance state also varies considerably concerning faded or missing lines and traffic signs. We see that the function of automated systems is enhanced from cruise control with fixed speed to Adaptive cruise control (ACC). Further development has resulted in Forward Vehicle Collision Mitigation Systems (FVCMS) and Full Speed Range Adaptive cruise control (FSRA) systems. Furthermore, Lane warning systems have evolved in active steering systems. All this is merging into partly or fully automated cars. How these systems 'see' the world and what they need to drive around safely is largely unknown to road operators. Car owners are informed by their dealer and the owner's manual on working and basic limitations. However, users that get used to high levels of automation will soon start to trust on their systems to work wherever they go and could easily forget these limitations.

To prevent a mismatch between infrastructure and car it should be known to Road Operators what expectations are programmed into the systems and when these cannot be met.

2.2 Lateral Assistance Systems

Camera-based lane change warning systems rely on visible lane markings; faded, missing, or incorrect lane markings can present a problem. Temporary line markings during road work could also be an issue. Harsh vibrations could potentially prevent the correct working of the system. These vibrations could occur due to a

defect to the tires, the vehicle stability system or an uneven road surface. The issue with visibility of the lane markings and temporary lining has risks, for both the user as well as road workers.

Advanced Lane Change Decision Aid Systems (LCDAS) do not always require lane markings. They also provide blind-spot warning and closing-in vehicle warning. Research by the AAA in 2014 [4] confirms that:

- Blind-spot monitoring systems have had difficulties detecting fast-moving vehicles—such as when merging onto a busy highway. Alerts were often provided too late for evasive action.
- Road conditions have often been a problem for lane-departure warning systems. Worn pavement markers, construction zones and intersections can cause the lane-departure warning system to lose track of lane location.
- The litany of alerts and warnings could be confusing. Auditory, visual or haptic responses—or a combination between these—could be similar to other advanced driver assistance features that deliver the same warnings.

The ISO 17387 LCDAS standard requires that the following statement is included in the owner's manual: "This system may not provide adequate warning for very fast moving vehicles approaching from the rear."

The dimensions that are used in the test seem to be barely sufficient to detect a small electrical vehicle such as the Renault Twizy with its height of 1.4 m, width of 1.2 m and length of 2.3 m.

The systems are tested under the optimal environmental conditions: The test takes place on a flat, dry asphalt or concrete surface. The ambient temperature during testing shall be within the range of 10 °C and 30 °C. The horizontal visibility range shall be greater than 1 km. Although the user receives instructions, other road users may not be aware of the fact that passing an equipped vehicle with high speed could result in a risk. Furthermore, users of vehicles smaller than the defined test target may not be aware of any extra risk.

2.3 Longitudinal Assistance Systems

Since the introduction of the classic Cruise Control system (CC) that set a fixed speed, major progress has been made. The first step was Adaptive Cruise Control (ACC) which allows the user to set a minimum gap with the target vehicle by controlling the power train and optionally the brake. Further development has resulted in Forward Vehicle Collision Mitigation Systems (FVCMS) and Full Speed Range Adaptive cruise control (FSRA) systems. A FVCMS can warn the driver for a potential collision with another vehicle in the forward path. It could also use a Speed Reduction Braking (SRB) or Mitigating Braking (MB). In this case automatic braking is applied if a collision seems inevitable. The system should

operate between 30 and 100 km/h. The system should work with a target vehicle that travels at least 30 km/h.

The validation methods define the target as a representative passenger's vehicle or representative motorcycle. Since the performance is dependent on the sensor technology, the validation method makes a difference between lidar, radio wave radar and passive optical sensor. It is unclear how these differences affect the performance. Lidar needs a much smaller reflective area than radar. Once more, the driver shall be informed about the systems limitations in the owner's manual.

Full Speed Range Adaptive cruise control (FSRA) systems are intended to provide longitudinal control while travelling on highways (roads where non-motorized vehicles and pedestrians are prohibited) under free-flowing and congested traffic conditions. FSRA provides support within the speed domain of standstill up to the designed maximum speed of the system. The system is not required to react to stationary or slow moving objects.

The system will attempt to stop behind an already tracked and stopping vehicle within its limited deceleration capabilities. The specifications of the test target depend on the type of sensor technology used. The minimum width of the automatic "stop" capability test vehicle is 1.4 m, which is the size of the Renault Twizy. Its lateral displacement should be no more than 0.5 m to the subject vehicle. The familiarity of the daily user with the limitations of the system is unknown.

It is optional to design FSRA systems to respond to the presence of stationary or slow moving targets. In practice, unexpected features of the road system can cause problems. A previous version of the Lexus Pre-collision System (PCS) reacted to metal on bridge parts and started an emergency brake action [5]. It is unclear if this could happen with other systems and if Road Operators have any role in this.

If a given implementation is not intended to respond to stationary or slow moving targets, the driver shall be informed at least by a statement in the vehicle owner's manual. For instance, the Mercedes owner's manual warns the driver that the system might not react to very narrow vehicles or vehicles that are not in the middle of the lane. As the implementation of these systems varies between brands and even between models of the same brand users could get confused when changing cars.

2.4 Automated Cars

Further development and combination of advanced systems brings us highly or fully automated vehicles. For definitions see SEAJ3016 [6]. Depending on the design, these systems display vulnerability towards irregularities like road works, incidents, debris, weather conditions, and disturbances of communication. In the present road systems these disturbances have a great likelihood of occurring. The user has to be able to realize this and regain control in time, should this occur. A minimum standard for robust design of automated cars would make the issue more manageable. Such a standard should be technology independent.

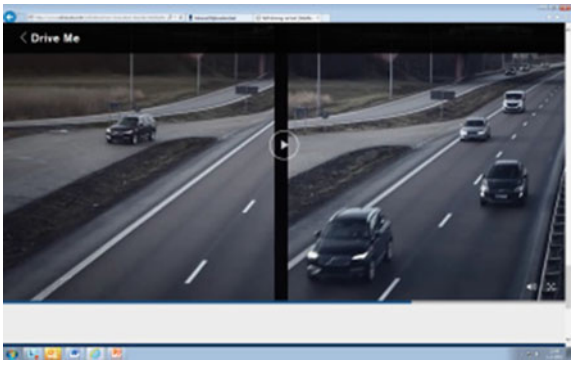


Fig. 1 Safe stop area in Gothenburg, Sweden

Furthermore, the standard for the road design would have to go up and roads should be able to communicate with the car. This requires cooperation between car industry and road operators.

A good example of such cooperation can be observed in Gothenburg, Sweden. The design of the motorway has been altered to accommodate automated driving. Where automated drive is no longer possible the driver is alerted to take back control. If the driver is not responding, the car automatically slows down and leaves the road for a safe stop area next to the highway [7] (Fig. 1).

2.5 Fleets

Systems like ACC that perform satisfactorily when used by an individual car, give problems when they are used by a row of cars. In a row of cars using ACC the last one hardly has any reaction time left to avoid an incident. Using V2V communications all cars can directly react on the first car. ACC and communication together enable platooning. While the platoon itself can be safe there is an issue with the interference with the other road users. The length of the platoon makes it harder for them to enter or leave the road. This is why, on some stretches of road, truck platooning is not permitted by the Road Operator. Communication between cars and trucks and Infrastructure would solve this problem.

2.6 Location, Communication, Maps

The autonomous car needs information from the outside world. Communication between vehicles (V2V) and between the vehicle and the infrastructure (V2I) is expected to resolve many issues such as ACC, detection of road works and incidents and could automatically resolve possible special conflicts on highways and

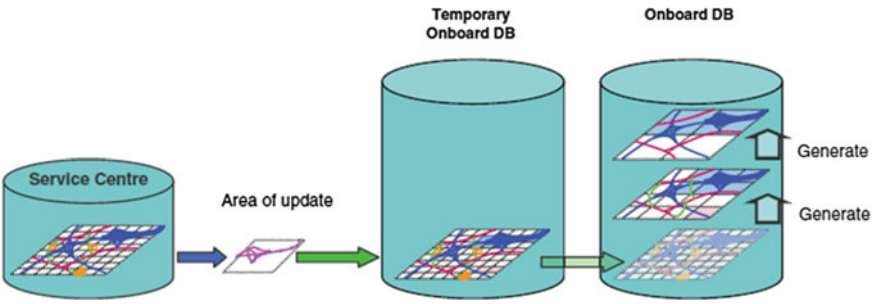


Fig. 2 Map services

intersections. Localization provided by GNSS systems plays an essential part in all this. All communication is sensitive to disturbances. Sources could be jammers, used by drivers that try to fool tracking systems, electromagnetic storms from solar flames or even other users in the same band.

If detected and tracked down, the use of jammers will be fined. This, however, will not solve the problem. The applications have to be resilient and warn users when they are experiencing problems that make save operations impossible.

A special use of GNSS services is "the map as a sensor" development. The needed detail of the map often exceeds the quality of the data that is present with the road operator. The number of lanes, for instance, as well as lane width, position and type of lane marking, traffic signs, special use lanes and traffic rules. Map updates and weather as well as road closures are sent as updates (Fig. 2).

To get the needed information to and from a car in a reliable way close cooperation or maybe even integration on technical and organizational levels between road operators and private parties such as the car industry and map providers is needed.

3 Involved Parties

3.1 The User

The user is being seduced by new functions that can make his journey safer, less tiring and that have a popular hype around it. At the same time he is supposed to know when the functions are not able to take over for him and in such cases he has to be able to quickly regain control over the vehicle. Clear information and education of the user are a necessity. To avoid misinterpretation, a standardization of the user information is desired.

Drivers of conventional cars are confronted with cars that react differently to manoeuvres as opposed to cars with a human in control. Although these differences

can be seemingly small, they could result in dangerous situations. Eye contact during a manoeuvre is not possible, making it harder to predict the other car's reaction. This issue needs more attention than it presently gets.

3.2 Road Operators

Even with automated cars driving around, road operators will still have to deal with broken down cars, incidents, road works and bad weather. Road Operators are confronted with a greater variety of vehicles, partly and highly automated cars. User behaviour also changes due to smartphone addiction.

The road operator's first job is to maintain a safe road system within a tight and, in most cases, decreasing budget. The existing users expect that they travel safely and that any new and allowed system does not put them at risk. On the other hand the road operator is asked to facilitate the development and admit new functions, new behaviour and even automated cars to its roads. In some cases it is clear to the road operator that the use of a system can cause problems. For navigation and cruise control various creative solutions regarding road signs have arisen already (Fig. 3).

This is not manageable for all the new functions since their number is too big, they differ too much and their sensitivity to unexpected circumstances varies too much. Road operators can not anticipate this. Radar systems might not be able to recognise traffic cones used by emergency services, putting them at extra risk.

These issues are hard to tackle. Can the Road Operator expect or demand that all partly automated systems can detect traffic cones as used by emergency services? Should they only allow automated vehicles that are equipped with V2V and V2I communication capabilities to warn for road works by traffic information and V2I communication? By doing so the Road Operator would define the minimal abilities an autonomous vehicle should have.

One other way to go is to define a minimal level of service for roads suited for the use of automated driving systems. This could lead to the definition of a 'controlled road' where radar sensors continuously inspect the road surface for broken down cars, debris, animals and incidents and if these obstacles are detected the oncoming traffic and navigation providers will immediately be informed using road side V-I communications. As most road operators still have the opinion that it is up to the automated car to scan its surroundings, a common minimal standard is not yet foreseen. As a first step car makers should be clear about the limitations of the automated functions and then get together with road operators to determine what it is needed to make real progress in the traffic system. Technically there are also

Fig. 3 How road operator deals with present on-board systems

some hurdles to make the road smarter and more communicative, as the equipment needed is not even defined yet.

3.3 Law Makers

Governments are facilitating trials with (partly) automated cars. To make new applications work in all places the same rules will have to apply everywhere. These rules will lead to new laws and new agreements between countries.

Areas of interest are:

- Definition of quality and performance of new vehicle systems within type approval
- Information and education involving the drivers, possible additions to the driver examination
- Information to other drivers on the reaction of automated cars
- Standardization of the information exchange between road and vehicle
- Definition of suited highways with a minimal level of service
- The use of Electronic Data Recorder (EDR) that contains all relevant data for accident investigation.

In case of an accident with a (partly) automated vehicle there will be uncertainty about the situation before the crash. To help fix operational mistakes and design flaws all relevant information such as the speed, acceleration and positioning data should be directly available for the accident investigation. In this way the interest of the public is served, and the technology can evolve further as well. Incidents should be investigated and serve as input for improvement. The aircraft industry can serve as an example in this respect.

4 Conclusion

The properties of automated cars are largely unknown to Road Operators. System designers should be more aware of the limitations of the users and the position of the road operator. The technological development of cars proceeds much faster than the development of the road. As the diversity of applications increase, the consequences are not clear.

Some road operators hope that the smart cars will take away all their problems. They foresee an empty road. Since the existing cars will still be around for many years, it is time to switch to a more realistic view where smart cars are serviced by an intelligent road. In order to get there, the automated car should be developed with both the user and the Road operators in mind and involved in the dialogue.

References

1. http://www.acea.be/statistics/tag/category/average-vehicle-age, Cars in the European Union are on average 9.73 years old
2. https://www.adac.de/infotestrat/unfall-schaeden-und-panne/pannenstatistik/default.aspx?ComponentId=168800&SourcePageId=47819
3. http://www.nhtsa.gov/Vehicle+Safety/Recalls+&+Defects NHTSA Campaign Number: 16E044000; 16E043000; 16E042000; 14E073000
4. AAA: Blind spot monitoring systems not as effective spotting motorcycles, 9 Dec 2014
5. NHTSA Campaign Number: 15V728000
6. SAE J3016: Taxonomy and definitions for terms related to on-road motor vehicle autmated driving systems. http://www.sae.org/misc/pdfs/automated_driving.pdf
7. https://www.youtube.com/watch?v=asKvI8ybJ5U

Part II
Advanced Sensing, Perception and Cognition Concepts

Towards Dynamic and Flexible Sensor Fusion for Automotive Applications

Susana Alcalde Bagüés, Wendelin Feiten, Tim Tiedemann, Christian Backe, Dhiraj Gulati, Steffen Lorenz and Peter Conradi

Abstract In this paper we describe the concept of the data fusion and system architecture to be implemented in the collaborative research project Smart Adaptive Data Aggregation (SADA). The objective of SADA is to develop technologies that enable linking data from distributed mobile on-board sensors (on vehicles) with data from previously unknown stationary (e.g., infrastructure) or mobile sensors (e.g., other vehicles, smart devices). Data not only can be processed locally in the car, but also can be collected in a central backend, to allow machine learning based inference of additional information (enabling so-called crowd sensing). Ideally, crowd sensing might provide virtual sensors that could be used in the SADA fusion process.

S.A. Bagüés (✉) · W. Feiten
Siemens AG Corporate Technology, Munich, Germany
e-mail: susana.alcalde@siemens.com

W. Feiten
e-mail: wendelin.feiten@siemens.com

T. Tiedemann · C. Backe
DFKI GmbH, Robotics Innovation Center, Bremen, Germany
e-mail: Tim.Tiedemann@dfki.de

C. Backe
e-mail: christian.backe@dfki.de

D. Gulati
fortiss GmbH, Munich, Germany
e-mail: gulati@fortiss.org

S. Lorenz
NXP Semiconductors Germany GmbH, Hamburg, Germany
e-mail: steffen.lorenz@nxp.com

P. Conradi
ALL4IP TECHNOLOGIES GmbH & Co. KG, Darmstadt, Germany
e-mail: info@all4ip.de

T. Schulze et al. (eds.), *Advanced Microsystems for Automotive Applications 2016*, Lecture Notes in Mobility, DOI 10.1007/978-3-319-44766-7_7

Keywords Data fusion · Automotive applications · Sensor crowd · Car-To-X

1 Introduction

Sensors used to be specific for one application, e.g. a driver assistance function. The application was designed, and then operated. The sensor data were processed with exactly this one application in mind, and the processed data could only be used in this application. Consequently, only the absolutely necessary, minimal effort for sensors was spent.

However, with better technology, sensors are used for several applications. Some manufacturers start to install much more powerful sensors than are needed for the basic functionality. If and when the customers want an enhanced functionality, only the software needs to be changed and can even be downloaded from a server (e.g. the autopilot function in the upcoming Tesla Model 3). The car even does not need to go to a workshop to be upgraded. Even functionalities that were not yet been designed at the time of the vehicle's release could be implemented just by upgrading the software. New ways of engineering applications are needed. Future automotive applications will run on different environments and use different sets of sensors to be a lot more flexible than in the past.

Also, an ever increasing number of sensors are being installed in the infrastructure. Here the same need for flexibility and adaptability applies, maybe even more than in the vehicle upgrade scenario. As different vehicles enter a given area, a completely new set of participants in an overall system emerges that has never before been composed in exactly this manner. In order for the individual cars and the overall system to perform at their best, the sensor data in various states of interpretation and aggregation need to be exchanged and put to meaningful use.

SADA plans to provide interfaces and methods to enable the unrestricted dynamic integration and analysis of data from multiple heterogeneous sensors. The focus of the project is the dynamic and fully automated switching between different sensor configurations, including the adaptation of data fusion processes. This will be realized by formal semantic descriptions of sensor properties, operational environment, and necessary algorithms, which are modeled explicitly.

The paper gives an introduction to the SADA project and describes in more detail the system and communication network architecture.

The remaining part of this paper is organized as follows. In Sect. 2 an overview of related work is presented, concerning system aspects as well as enabling technologies. Section 3 introduces the overall system architecture. The system puts high demands on the in-vehicle communication as well as on the Car-2-X communication; the corresponding aspects are discussed in Sect. 4. Section 5 presents preliminary experimental results regarding the communication between distributed modules. In Sect. 6 the current state of the project is wrapped up and further steps are explained.

2 Related Work

From 2007 to 2010, in the research project "Aktiv", funded by the German government (Federal Ministry for Economic Affairs and Energy, BMWi, and Federal Ministry of Education and Research, BMBF), techniques were developed to improve the active safety in vehicles by using Car-2-Car communication [1, 2]. In about the same period, the European research project "coopers" focused on cooperative traffic management, supported in particular by car-to-infrastructure communication [3]. A substantial part of the project was field tests of the wireless communication channels. A different approach was taken by the project "Ko-FAS", funded by the German government BMWi, from 2009 to 2013 [4]. The question was how the use of cooperative components (e.g., active, intelligent radio-frequency identification tags on all traffic participants) could improve traffic safety. The focus was on the technical feasibility of such sensors and their achievable performance features, as well as the integration of this information into an overall scenario. For communication, the standard IEEE 802.11p [5] was used. However, the content that was transmitted to other communication participants was not data on a low level or even raw sensor data, but again aggregated information. The approaches for improved traffic routing were further developed and tested in extensive field trials in further projects:

- In "Drive C2X", from 2011 to 2014, a Europe-wide test was conducted in traffic circulation, on the basis of the emerging standards for the transmission of information about traffic events and traffic situations [6, 7].
- A proof of the successful technical operation of information infrastructure and the effectiveness for safer and more efficient traffic was also provided by the German project "SimTD", from 2008 to 2013. Its emphasis was also on Car-to-X communication, and on the exchange of traffic situations and traffic events [8].
- From 2012 to 2015 the BMWi- and BMBF-funded project CONVERGE dealt with flexible and secure communication for traffic services [9].

Overall, these research activities gave good reasons that a safe, robust communication between vehicles and infrastructure is feasible and leads to safe and efficient traffic. Furthermore, standards for this communication have emerged. However, progress has not only been made with regard to the communication between vehicle and environment. Also inside the vehicle a generic foundation for networking has been laid. The recently developed standard for automotive Ethernet *Open Alliance BroadR-Reach*™ (OABR) brings progress in the reliability of data transmission, as well as a cost-efficient transmission medium (two-wire line). The standard for the physical layer was developed by the industry interest group www.opensig.org and finally became an IEEE standard (IEEE 802.3bw, 100Base-T1). Up to now, 100Base-T1 Ethernet was mainly used for the transmission of image data inside the vehicle. The description of a backbone system in highly distributed networks is new and subject of the architectural work of the SADA project.

Finally, Duchrow et al. [10] present a framework for electric mobility data mining [10]. There, a heterogeneous fleet of vehicles logs data locally and sends it via GPRS to a central backend system. They propose a system for multi-layer filtering and preprocessing and they present analyses from the first year of data collection. Some more SADA-related references with the focus of machine learning in traffic applications and crowd sensing can be found in Tiedemann et al. [11, 12].

3 SADA System Architecture

Sensor data fusion in automotive applications is used to improve the accuracy and robustness of the environment perception. Most of the new driver assistance functions use the knowledge about the location of stationary objects and moving targets around the vehicle. So far, this is done by aggregating the information received from known sensors, previously installed in the vehicle. Today, the integration and fusion of new sensor data on an ad hoc basis is still not possible. What is missing is the possibility to dynamically adapt the system i.e. by adding or removing components (e.g., in case of sensor drop outs), as well as by using data of local or remote sensors (e.g., to increase precision and reliability) for the applications.

3.1 Overview

Despite all the solutions and advances in the area of the communication between vehicle and environment as well as in traffic management, there are still a number of issues that make the development of new automotive applications challenging, expensive and static. SADA addresses some of those issues providing a new solution for dynamic and adaptive data aggregation for any application. We focus on the following requirements:

1. **Dynamic Configuration**: This is the possibility to generate dynamically and online new configurations. The fist requirement is the dynamic and fully automated switching between different sensor configurations for registered applications and services.
2. **Generic and adaptive fusion process**: A data fusion system is usually designed with a particular purpose in mind. In the past, each driver assistance function was implemented completely independently of other driver assistance functions. The classic development of sensor fusion for automotive applications is therefore expensive. The second main requirement of SADA is to adapt the fusion process on the fly, based on generic components and well-known algorithms (e.g., Kalman filter), to meet the quality expectations of any given application.

3. **Formal semantic modeling**: Formal semantic descriptions of sensor properties, operational environment, application quality requirements, necessary components and algorithms are to be explicitly modeled. A third SADA requirement is: The information regarding all the entities involved must be syntactically and semantically described in a formal way. Formal semantic models are the key to enable a fully automated decision process, decisions that today are taken by domain experts.
4. **Sensor Crowd**: Not only local, in-vehicle sensors shall be usable but also those of neighboring cars, infrastructure, and up to a whole crowd of sensors distributed in space and time. Thereby, a different kind of data fusion can be realized where a driver assistance system can use, e.g., sensor data of vehicles that passed a region before. Combined with the requirement of having a central system capable of analyzing all this data even new information could be generated, e.g., leading to predictions of future sensor data.

From these requirements we conclude that instead of the current rather monolithic approach a modular, flexible design is preferable. In the remainder of this section we will describe the architecture choices that were made in the SADA project in order to meet those requirements.

3.2 Distributed System

In this section, we give an overview of the SADA system architecture. The communication aspects are discussed in Sect. 4. To achieve the goal of dynamic and flexible sensor fusion the complete fusion process in SADA is divided and redesigned into logical modules. Modules are designed to be data driven. In the domain of automotive applications, decisions are taken based on the data acquired. With increasingly feasible Car-2-X technologies, the data generation and consumption is not to be restricted to inside the vehicle. This modularization helps to achieve the configuration of the system on the fly, as the modules can be inserted and removed on demand. This also enables the processing to be distributed across various platforms.

Thus, SADA design is based on a distributed architecture (see Fig. 1). SADA is thought to be part of the vehicular electric and electronic (E/E) architecture, as well as to be deployed on the infrastructure i.e. in a traffic camera as fusion node. SADA runs on the backend to aggregate information of a bigger number of sources. Even a small version of SADA could be deployed on any smartphone to allow the user to bring his own sensor information source. The idea behind it is to enable the aggregation of data being today already generated by a heterogonous number of entities. We want to achieve a better perception of the environment without adding extra sensors.

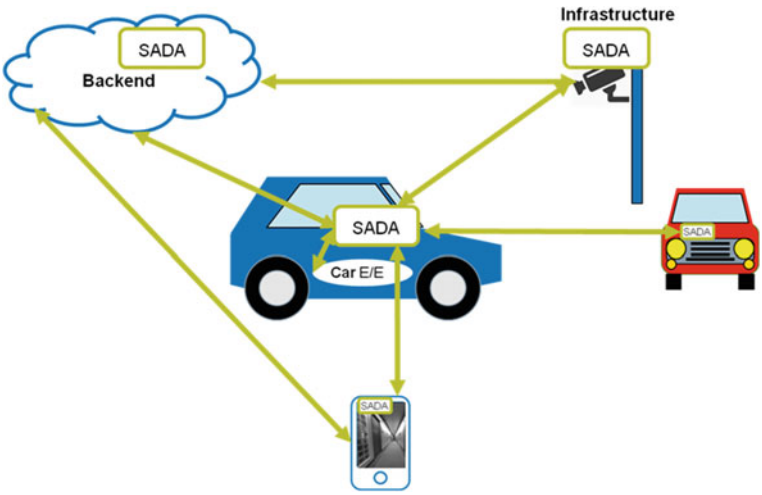


Fig. 1 SADA distributed system

From the point of view of a software layer representation (see Fig. 2) SADA is not to be seen as an independent system. SADA needs to be integrated into its environment. The flow of information between the different layers of Fig. 2 is

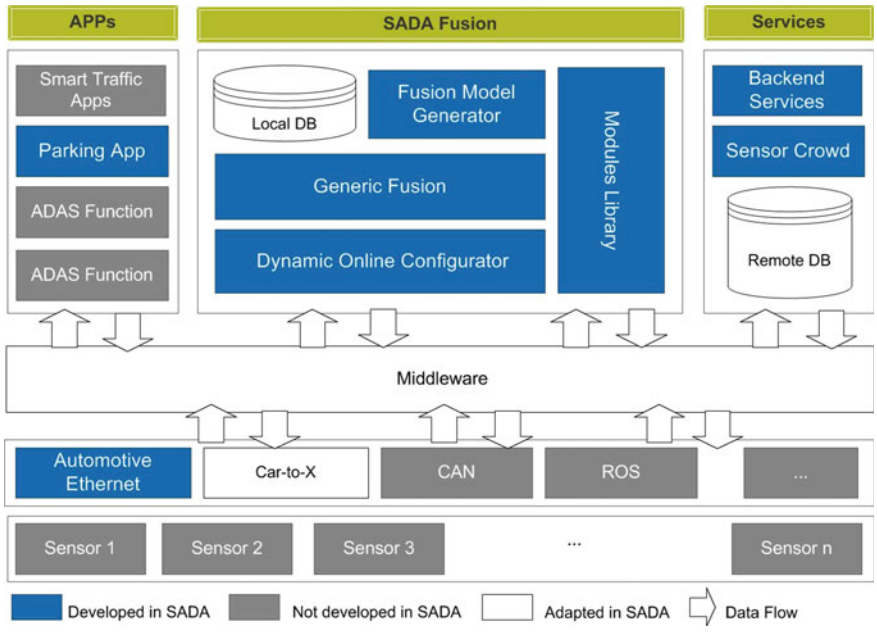


Fig. 2 Software layer architecture

managed by a middleware. It is not in the scope of SADA to develop a middleware (we use ROS [13]). SADA relies on such middleware to orchestrate the data between modules.

In SADA, as mentioned before, the flow of information is not static. SADA adapts the system topology at runtime. A reconfiguration occurs after one of the following events:

(a) Addition/removal of a sensor
(b) Registration/deregistration of an application/service
(c) Change of context scenario (e.g. from an indoor to an outdoor location).

The new configuration is given to the middleware layer. The middleware is used as the "glue" to connect modules. One of the main functionalities of the middleware is to provide a mechanism to agree on the data exchange format and support seamless data exchange. It facilitates runtime adaptation.

In the bottom part of Fig. 2 are the hardware layer, which consists of a diversity of sensors ranging from in-vehicle to infrastructure sensors (e.g., cameras, radars, lasers and ultrasound sensors), and its hardware abstraction layer with the corresponding drivers and communication protocols. As part of the research done in SADA, we gather new requirements for Car-2-X communication (see chapter "Prototyping Framework for Cooperative Interaction of Automated Vehicles and Vulnerable Road Users") to support the aggregation of data from the different entities shown in Fig. 1. Furthermore, SADA uses Car-2-X messages to address the sensors in its proximity.

The upper part of the Fig. 2 shows three vertical layers. The Apps layer represents new automotive applications. Apps can be installed in the car, both in a classic onboard unit, or directly downloaded into the driver's tablet. Apps get information both from the SADA system and/or the service layer. As part of our work the "Parking APP" for a smartphone is being developed. The App helps drivers to find a parking spot and to park the car by aggregating data from sensors outside the vehicle.

The service layer provides data to the applications and formal semantic models to the SADA Fusion. It is also part of the seamless connected components using the middleware. It facilitates offline processing of data, including learning and crowd sensing and delivers the information to the requesting components.

SADA Fusion layer provides the components to achieve flexible and adaptive fusion. Classical fusion systems are developed with only an application in mind and with a predefined set up of sensors and algorithms. Our objective is to come up with a novel approach to data fusion, to adapt the system at runtime to the sensors and the applications available. To do that, SADA Fusion incorporates the following components:

- **Dynamic Configurator**: It reconfigures the system based on the sensors available to meet the app's quality requirements. Decisions are made event driven and regarding the information provided by formal semantic models defined in SADA (requirement 3, not in the scope of this document).

- **Generic Fusion**: In SADA modules can interact with each other irrespective of their locations and can give rise to a more effective fusion strategy. We do not talk about early or late fusion, fusion may occur at every level. Modules can be inserted, removed or replaced, on demand, thereby giving rise to a "*Fluid Fusion*" strategy. The Generic Fusion component manages the final fusion result, adapting the results to the application. It supports a wide range of "internal" representations for the environment (e.g. occupancy grid, object list).
- **Fusion Model Generator**: A fusion model brings together the data fusion algorithm and the properties of the sensors and environment as well as requirements of the application. Currently, the corresponding parameters that define a fusion model can be only compiled by a group of expert developers. The Fusion Model Generator includes learning capabilities to relieve this restriction by automatically calculating the parameters given an explicit fusion topology.
- **Modules Library**: The complete fusion process in SADA is divided into modules. Suitable types of modules are sensors, semantic models, services, preprocessing algorithms, fusion algorithms, etc. The Library provides access to those modules while the Configurator sets the current topology ("chain of modules") to be instantiated by the Middleware.

4 Communication Architecture

The sensor data are communicated between the different modules of the overall system. They can be preprocessed, filtered and aggregated. They can have different requirements on bandwidth. Basis for this is a powerful communication architecture that provides the required bandwidth as well as information on the actual load and channel quality. Even though the SADA system is not linked to a dedicated communication architecture, part of the SADA project is to show the feasibility and requirements for the different communication paths.

Figure 3 shows the communication channels between the modules. They are the communication with the infrastructure, also with other cars, communication with the backend as well as communication with a smartphone.

Inside the vehicle the 2-wire automotive Ethernet (IEEE 801.3bw) will be used, for reasons of weight, space, cost and automotive compatibility. Ethernet provides a high bandwidth as well as a switched network with different priority classes. This will be used to allow distributed computing architectures with a co-existence of highly secure control data, low latency sensor data as well as infotainment data. SADA traffic will only be part of this, but it should be proven that requirements are fulfilled in future E/E architectures.

For the connection of a vehicle to other vehicles or to the infrastructure, the IEEE 802.11p standard (Car-2-X) will be used, as it is implemented e.g. in the on board unit Cohda MK5 OBU. As it is also Ethernet based, it allows seamless

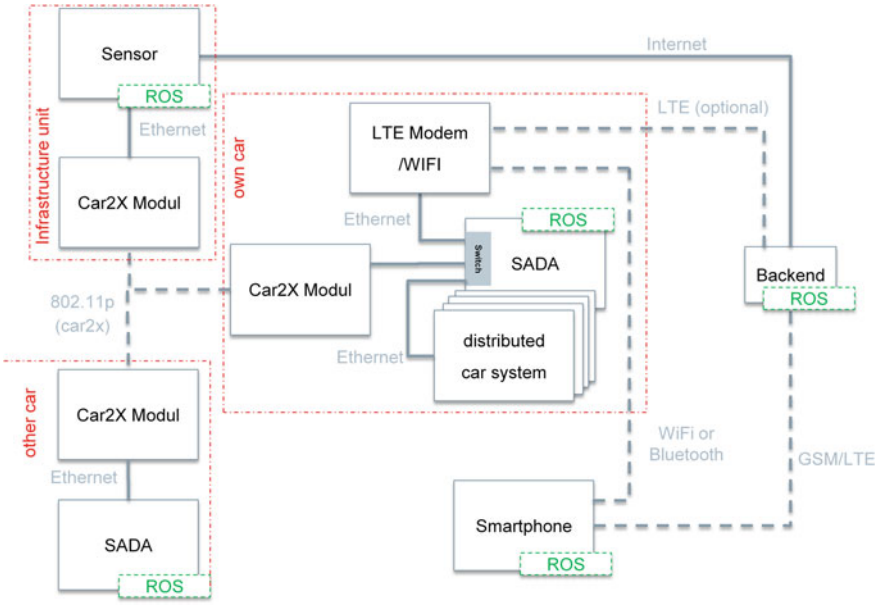


Fig. 3 Communication architecture

integration into the network. Although security aspects need to be considered for the final usage of SADA, this is not in the current focus of the SADA project.

Two features of Car-2-X will be used. On the control channel periodically Service Announcement Messages (SAM) are sent to announce available sensor types, services and service channels. The receiver can switch to the service channel were the contents of the services are transmitted. This feature is not yet supported in the European Car-2-X standard, but generally possible as described in the IEEE 1609.

The same method will be used for sensor fusion with other cars. A car that has a sensor (e.g. camera) could announce this the same way and provide the data to other cars. This would increase the number of available sensors and optimize the sensor fusion for different applications (e.g. autonomous parking).

With an internet connection at the infrastructure unit, also a connection to the backend could be established via the Car-2-X link. This allows exchange between the car and the backend, but also a direct use of the infrastructure sensors for sensor fusion in the backend is possible.

An alternative respectively co-existing communication path between the car and the backend is the use of an LTE connection, e.g. for regular interaction where no infrastructure connection is available. It is part of the project to discuss advantages and limitations of both systems. The preliminary experimental results show a first investigation on this aspect.

As described before, it is also planned to include a smartphone into the SADA system. Sensor data of the smartphone can be provided by WiFi or Bluetooth into the SADA system and be used for the sensor fusion. In the other direction this connection can be used to provide fused data to be visualized by an app (e.g. parking support).

For data from the backend (e.g. parking map) to be used in apps, normal GSM/LTE internet access of the smartphone will be used.

5 Preliminary Experimental Results

Parts of the proposed communication architecture were tested in first experiments. These early experiments were carried out to assess if the proposal is feasible and to obtain first estimates of communication performance metrics. Two setups were tested:

1. IEEE 802.11p communication between two nodes (e.g., vehicle and infrastructure),
2. in-vehicle sensor data transfer and storage in a data base (DB).

These setups are two of the main building blocks for all communication scenarios in SADA. In a full communication use-case, (data flow from a vehicle sensor, via a temporary storage in a local DB, further via Car-2-Backend RF communication, and finally to the backend DB), they represent the whole chain with setup (1) in the middle and setup (2) estimating the two local end points of the communication chain. Furthermore, the latter two communication sections are expected to be the bottle necks for data throughput (RF communication) and latency (massive DB access).

We tested transmission and reception of WAVE (wireless access in vehicular environments, IEEE 802.11p) packets between two Cohda Wireless MK5 units using an on-board unit (OBU) as the transmitter and a road side unit (RSU) as the receiver. The MK5 has two WAVE radios, allowing packets to be transmitted on either radio (single ISM channel), both radios simultaneously (two different ISM channels), or a single channel-switch radio (two different ISM channels, 50 ms per channel). Data throughput limits were explored in the single radio setup for the European control channel 180 and the service channel 172, respectively. OBU and RSU were placed in a distance of approximately 1 m, in an unshielded indoor office environment.

When transmitting packets of 2 kB, packet loss reaches undesirable levels even at moderate packet frequencies (7 percent loss at 100 Hz). Therefore, the experiments were conducted with a packet size of 1 kB. The results are given in Table 1. The packet size was 1 KB in all runs. The number of packets transmitted per second is given in the column titled “Frequency”. The transmission success rate is the number of packets received successfully divided by the number of packets sent.

Table 1 Transmission success rates of first IEEE 802.11p (Car-2-X) experiments

Frequency (Hz)	Throughput	Transmission success rate	
		Control Ch. 180 (%)	Service Ch. 172 (%)
100	100 kB/s	99.93	99.97
200	200 kB/s	99.93	99.90
400	400 kB/s	99.97	99.91
800	800 kB/s	95.84	92.03
1600	1.6 MB/s	52.94	49.45

As can be seen, a data throughput of 400 KB/s is possible without an increase of error. For most critical applications (e.g., using a neighbor vehicle's parking sensor to park your car) this is much more than needed.

While for the data throughput the RF communication is the limiting factor, for latency estimation massive DB access combined with inter-process communication is expected to be the bottleneck. Therefore, we simulated a sensor with one process generating data periodically. A second process collects this data in a FIFO buffer and stored it in a local DB. All experiments were run with a single Intel i7-2620 M CPU (2.70 GHz, 4096 kB cache) on Ubuntu 14.04 (Linux version 3.13.0-85-generic) against a PostgreSQL 9.3 database server.

In our experiments, a sensor with large data packets can be seen either as multiple sensors with small packets or as a single sensor with high frequency, in both cases assuming packets are split up before they are logged and later sent on the network. Since each sensor runs in its own process, and the simulation is carried out on a single processor, the number of sensors is fixed at 100 to normalize operating system overhead. The packet size is still 1 kB (1000 bytes). Thereby, the load of each sensor and, thus, the total throughput of the system only depend on the sensor frequency. We test three orders of magnitude from 1 to 100 Hz, yielding total throughputs between 100 kB/s and 10 MB/s. For higher orders, latencies reach multiple minutes, probably due to swapping. The results are shown in Table 2. Each trial is run for approximately 10 s (1000, 10,000, and 100,000 packets, respectively) and repeated to get 5 runs per configuration. Min./max./avg. latency is computed over all packets of all 5 runs. Latency is defined as interval between data generation and database update completion. The latency measurements of further configurations can be found in [14].

Table 2 Latency measurements of the second set of experiments

Frequency (Hz)	Throughput	Packets	Latency (s)		
			Min.	Max.	Avg.
1	100 KB/s	5000	0.023	0.290	0.062
10	1 MB/s	50,000	0.017	1.986	0.103
100	10 MB/s	500,000	0.028	88.865	29.945

Up to 1000 packets (of 1 kB size) per second the latency is small enough even for critical use-cases like parking a vehicle using distance sensor readings. The latency observed in the high-throughput measurements is much larger than expected. However, it is still small enough for the planned SADA use-cases with large data sizes (i.e. transfer of logged historical sensor data from the vehicle to the backend). Furthermore, it should be noted that these experiments were first tests with a preliminary implementation to obtain quick (worst-case) estimations. Thus, we expect several options to optimize the performance—when the whole SADA system is being integrated.

6 Conclusion

In this paper we give an overview of the system and communications architecture to be implemented in the collaborative research project Smart Adaptive Data Aggregation (SADA). The objective of SADA is to develop technologies to enable adaptive and flexible data aggregation for automotive applications. We describe the first results regarding communication between distributed modules and introduce the four main requirements that guided the design of the architecture. Our Work is still ongoing, central parts of the overall system are currently being implemented and tested. A first app version is under implementation to display available parking slots in a parking lot. Furthermore, the definition of the semantic descriptors and formal modeling of information is being carried out to enable a fully automated configuration of the sensor fusion topology.

Acknowledgments This work was partly funded by the Federal Republic of Germany, Ministry for Economic Affairs and Energy within the program 'IKT EM III', grant no. 01ME14002A.

References

1. Fecher N, Hoffmann J, Winner H, Fuchs K, Abendroth B, Bruder R (2009) "Aktive Gefahrenbremsungen", ATZ – Automobiltechnische Zeitschrift, pp 140–146
2. http://www.aktiv-online.org. Accessed 02 May 2016
3. Toulminet G, Boussuge J, Laurgeau C (2008) Comparative synthesis of the 3 main European projects dealing with cooperative systems (cvis, safespot and coopers) and description of coopers demonstration site 4. In: 11th International IEEE conference on intelligent transportation systems, pp 809–814
4. Ko-FAS-website, see http://www.kofas.de. Accessed 02 May 2016
5. IEEE 802.11p, see https://standards.ieee.org. Accessed 28 May 2016
6. DRIVE-C2X-website. http://http://www.drive-c2x.eu. Accessed 02 May 2016
7. Dokic J, Müller B, Meyer G (2015) European roadmap smart systems for automated driving. Eur Technol Platform Smart Syst Integr
8. Weiß C (2008) V2X communication in Europe from research projects towards standardization and field testing of vehicle communication technology. Comput Netw 55:3103–3119

9. CONVERGE-website. http://converge-online.de. Accessed 02 May 2016
10. Duchrow T, Schröer M, Griesbach B, Kasperski S, Maas genannt Bermpohl F, Kramer S, Kirchner F (2012) Towards electric mobility data mining. In: 2012 IEEE International electric vehicle conference (IEVC), pp 1–6
11. Tiedemann T, Backe C, Vögele T, Conradi P (2016) An automotive distributed mobile sensor data collection with machine learning based data fusion and analysis on a central backend system. In: Proceedings of the 3rd International conference on system-integrated intelligence. Accepted, to be published in SysInt 2016
12. Tiedemann T, Vögele T, Krell MM, Metzen JH, Kirchner F (2015) Concept of a data thread based parking space occupancy prediction in a Berlin pilot region. In: Papers from the 2015 AAAI Workshop (WAIT-2015), AAAI Press
13. ROS, http://www.ros.org/. Accessed 28 May 2016
14. Nerurkar ED, Roumeliotis SI, Martinelli A (2009) Distributed maximum a posteriori estimation for multi-robot cooperative localization. In: IEEE International conference on robotics and automation, ICRA '09, pp 1402–1409

Robust Facial Landmark Localization for Automotive Applications

Manuel Schäfer, Emin Tarayan and Ulrich Kreßel

Abstract This paper introduces a novel system for facial landmark detection using a modified Active Appearance Model (AAM). Traditional AAMs operate directly on the pixel values of the image, leading to problems with inhomogeneously illuminated scenes. Instead of using the gray-level image to detect the facial landmark directly, the Modified Census Transformation (MCT) is performed on the region of interest (ROI) being analyzed, making the system invariant to illumination variations and nonlinear camera characteristics. To achieve efficient and robust fitting with regard to occluded or invisible parts of the face, parameter constraints, coarse to fine fitting and occlusion handling are introduced. The result shows that the new system yields good results even if some areas of the face are occluded or unrecognizable in the image.

Keywords Facial landmark localization · Facial features · Robust active appearance model · MCT AAM · Head-pose estimation · Driver observation · Low-resolution images · Noisy images · Modified census transformation · Inhomogeneous illuminated images

M. Schäfer (✉)
Daimler AG, Hanns-Klemm-Str. 45, 71034 Böblingen, Germany
e-mail: manuel.ms.schaefer@daimler.com

E. Tarayan
Daimler AG, Benzstraße, 71059 Sindelfingen, Germany
e-mail: emin.tarayan@daimler.com

U. Kreßel
Daimler AG, Wilhelm-Runge-Str. 11, 89081 Ulm, Germany
e-mail: ulrich.kressel@daimler.com

T. Schulze et al. (eds.), *Advanced Microsystems for Automotive Applications 2016*, Lecture Notes in Mobility, DOI 10.1007/978-3-319-44766-7_8

1 Introduction

One field of research in the automotive sector is that of video-based systems used for driver observation to increase safety and to introduce new techniques for human-machine interaction. The knowledge of the driver's facial landmark positions allows application of many safety and comfort relevant features such as head-pose or eye gaze. This information can, for example, be used to infer visual attention or the state of drowsiness in order to alert drivers that there is a critical situation out of their field of view or that they are falling asleep. Furthermore, some intentions of drivers can be estimated, e.g., if they want to change the traffic lane, turn at an intersection, or use the infotainment system. To infer all this information from the camera image, the first step is to detect facial landmarks (e.g. corners of eyes, eyelids, nose and mouth) in the image. The challenges in the automotive context are to create a system which runs on low-cost hardware with low-cost cameras and therefore yields low-resolution and noisy images. Further, the system must be robust with regard to major and quickly-varying illumination changes. Especially direct sunlight, deep shadows, reflections, and overexposure lead to strongly inhomogeneous illuminated scenes. Many state-of-the-art techniques for facial landmark localization have difficulties detecting the desired landmarks under these conditions. This paper introduces a novel system for facial landmark detection using a modified Active Appearance Model (AAM).

2 Related Work

The most prominent techniques for facial landmark detection are template-based models such as Active Shape Models (ASM) and Active Appearance Models. A survey over most of the existing methods for head-pose estimation and facial landmark localization can be found in [1]. Template based models like in [2] are built upon similarity matching of certain facial landmarks with a predefined set of templates containing the structure to be detected. These templates are correlated with a region of interest (ROI) to detect the desired facial characteristics. Template-based models have the disadvantage that covering a huge amount of head-poses, different appearances and scaling factors requires a large set of templates. ASMs—first introduced by Cootes and Taylor [3]—have the advantage of describing deformable contours with a few different orthogonal modes. Matching of ASM is an iterative process. In [4] different matching procedures are compared with regard to computation time and their expected accuracy. The advantages of ASMs are their simple structure and computation time. However, they are not that robust and use only a subset of the available information. AAMs—introduced by Cootes, Edwards and Taylor—are an enhancement of the ASMs. A detailed introduction about AAMs can be found in [5]. In addition to the ASM, the whole texture is modelled and used to match an AAM on well-defined objects. This enhancement

makes the model more robust compared to the ASM, but leads to difficulties if parts of the object are occluded or inhomogeneously illuminated. In [6, 7] many AAMs are used to solve the problem of self-occlusion in the facial landmark detection field. In [8, 9] robust error-functions are used to handle occluded areas. For performance optimization in [10] a coarse-to-fine matching with down-sampling and successive refinement is used. All these extensions increase the robustness but still have enormous problems with inhomogeneous illumination in particular.

3 Overview—AAM Framework Using MCT Features

This work introduces a complete framework for facial landmark localization using a modified AAM that is robust with regard to high and inhomogeneous illumination variations, works on noisy and low-resolution images and does not need any preprocessing steps for image optimization. An overview of the complete framework is shown in Fig. 1. The framework consists of three main blocks—the initial guess generation, the new introduced AAM using MCT [11] features, and a head-pose estimation. The functionality of these three blocks will be explained in the following chapters. For each detected face in an image, the framework localizes the positions of all detected facial landmarks in pixel coordinates and an estimation of the head orientation given relative to the image plane. Also a quality criterion for the matching accuracy is provided.

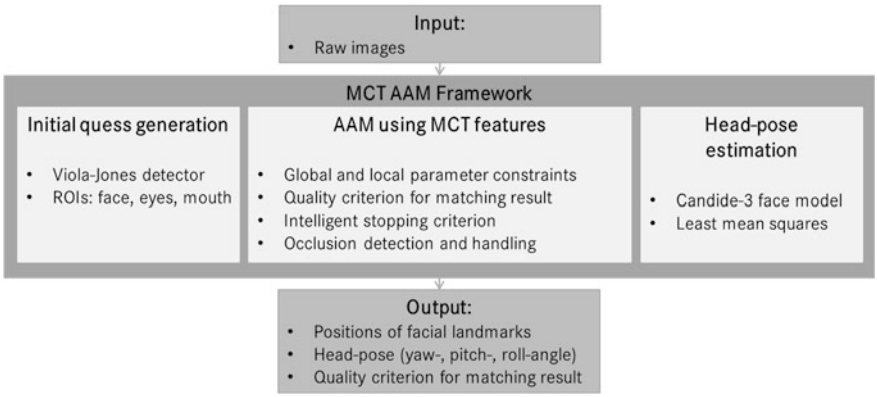


Fig. 1 Overview of the MCT AAM Framework. The first block covers an initial guess generation on raw images using Viola-Jones detectors for detecting faces, eyes and mouths. The second block covers the MCT AAM matching on detected faces. By using the matching results, the third block covers a head-pose estimation. The framework returns the detected facial landmarks, the estimated head-pose and a quality criterion for the matching result

4 AAM Using MCT Features

AAMs combine a linear shape model and a linear appearance model. The linear shape model $\boldsymbol{s} = (x_1, x_2 \ldots x_v, y_1, y_2 \ldots y_v)^T$ spans a mesh containing v nodes. It is defined by $\boldsymbol{s} = \boldsymbol{s}_0 + \sum_{i=1}^{n} \boldsymbol{s_i} p_i$ where $\boldsymbol{s}_0$ (Fig. 2) represents a base mesh configuration, $\boldsymbol{s_i}$ its shape vectors with corresponding weights p_i to enable mesh deformations.

The linear appearance model represents the texture of the AAM and is defined on the base mesh by $\boldsymbol{A}(\boldsymbol{x}) = \boldsymbol{A}_0(\boldsymbol{x}) + \sum_{i=1}^{m} \boldsymbol{A}_i(\boldsymbol{x})\lambda_i$ where $\boldsymbol{x}$ are the pixel coordinates within the base mesh, $\boldsymbol{A}_0(\boldsymbol{x})$ (Fig. 2) the mean appearance image and m appearance images $\boldsymbol{A}_i(\boldsymbol{x})$ with corresponding weights λ_i that enable texture variations of the AAM.

The shape vectors and appearance images represent the principle components with the greatest eigenvalues of a Principle Component Analysis (PCA) computed on a training set. These two models are combined through a warping-function $\boldsymbol{W}(\boldsymbol{x},\boldsymbol{p})$ [5] that transforms the texture of the appearance model by an affine transformation with bilinear interpolation into the shape model with the corresponding weights $\boldsymbol{p} = (p_1, p_2 \ldots p_v)$. The process to match the AAM onto a face in an image can be formulated as a minimization problem to minimize the mean squared difference between the model and an input image. The minimization can be written as

$$\min_{\boldsymbol{\lambda},\boldsymbol{p}} \sum_{\boldsymbol{x}\in s_0} \left[\boldsymbol{A}_0(\boldsymbol{x}) + \sum_{i=1}^{m} \boldsymbol{A_i}(\boldsymbol{x}) - \boldsymbol{I}(\boldsymbol{W}(\boldsymbol{x};\boldsymbol{p})) \right]^2 \tag{1}$$

where $\boldsymbol{I}(\boldsymbol{W}(\boldsymbol{x};\boldsymbol{p}))$ describes the input images warped onto the base mesh. A simplification of this formulation can be found in [12] and is known by "project out". The minimization is done in an orthogonal space respective to the appearance images. Thereby the minimization simplifies to

$$\min_{\boldsymbol{p}} \sum_{\boldsymbol{x}\in s_0} [\boldsymbol{A}_0(\boldsymbol{x}) - \boldsymbol{I}(\boldsymbol{W}(\boldsymbol{x};\boldsymbol{p}))]^2. \tag{2}$$

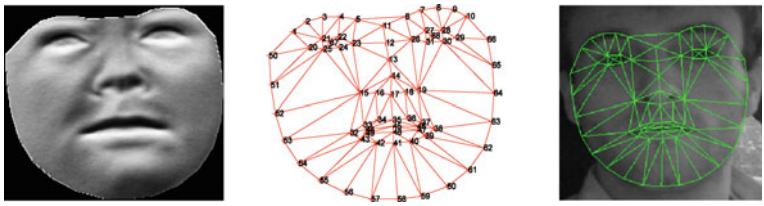


Fig. 2 Mean MCT Appearance Image $\hat{\boldsymbol{A}}_0(\boldsymbol{x})$ (*left*), base mesh $\boldsymbol{s}_0$ containing $v = 68$ nodes (*center*) and the result of a matched MCT AAM model (*right*) on an example face

To minimize this formulation it is assumed that there is a linear relation between a parameter optimization $\Delta \boldsymbol{p}$ and the so called difference image $\boldsymbol{r}(\boldsymbol{p}) = \boldsymbol{A}_0(\boldsymbol{x}) - \boldsymbol{I}(\boldsymbol{W}(\boldsymbol{x};\boldsymbol{p}))$:

$$\Delta \boldsymbol{p} = -\boldsymbol{R}\boldsymbol{r}(\boldsymbol{p}). \tag{3}$$

By doing so, an iterative minimization can be realized with $\boldsymbol{R} = \left(\frac{\partial \boldsymbol{r}^T}{\partial \boldsymbol{p}} \frac{\partial \boldsymbol{r}}{\partial \boldsymbol{p}}\right) \frac{\partial \boldsymbol{r}}{\partial \boldsymbol{p}}$ with $\frac{\partial \boldsymbol{r}}{\partial \boldsymbol{p}}$ being the Jacobian matrix. The problem with this formulation is that in presence of large illumination variations or deep shadows, through the linear relation these influences result in large $\Delta \boldsymbol{p}$ values, even if the model is located perfectly on the subject's face. To make the model invariant concerning these influences, the main contribution of this work is to extend traditional AAM by making use of the MCT. Instead of calculating the difference image on intensity values, the pictured face will be warped onto the base mesh as usual and then transformed by the MCT. Further, the mean appearance image $\boldsymbol{A}_0(\boldsymbol{x})$ will be replaced by a mean MCT appearance image $\hat{\boldsymbol{A}}_0(\boldsymbol{x})$ (Fig. 2). The term $\boldsymbol{I}(\boldsymbol{W}(\boldsymbol{x};\boldsymbol{p}))$ will be replaced by $\hat{\boldsymbol{I}}(\boldsymbol{W}(\boldsymbol{x};\boldsymbol{p}))$, which represents the pictured face warped onto the base mesh, and afterwards be MCT transformed. With these modifications, the formulation for the difference image $\boldsymbol{r}(\boldsymbol{p})$ changes to:

$$\boldsymbol{r}(\boldsymbol{p}) = \hat{\boldsymbol{A}}_0(\boldsymbol{x}) - \hat{\boldsymbol{I}}(\boldsymbol{W}(\boldsymbol{x};\boldsymbol{p})). \tag{4}$$

To compute the matrix $\boldsymbol{R}$, the Jacobian matrix $\frac{\partial \boldsymbol{r}}{\partial \boldsymbol{p}}$ has to be estimated. This is done through numerical estimation of all required derivations over a training set containing different faces of different subjects with varying head-poses. For each training sample, one Jacobian matrix is estimated. To estimate all required derivations, the shape parameters are varied in the range between $[-0.5\sigma_i; 0.5\sigma_i]$, where σ_i represents the standard deviation of the ith shape parameter over the training set. The mean of all estimated Jacobian matrices is used to calculate the matrix $\boldsymbol{R}$. To match the MCT AAM onto a pictured face, an initial guess for the shape model is required. With an initial guess and the iterative formulation for the parameter optimization the MCT AAM model can be matched on a face as shown in Fig. 3.

Fig. 3 From *left* to *right*: Input image with initial guess for the shape model (*1*) warped onto the base mesh (*2*) and transformed by the MCT (*3*). Image (*4*) shows the difference image r(p) which is used to calculate the optimized parameters. The result of the iteration step for matching the MCT AAM onto the face is shown in the last image (*5*)

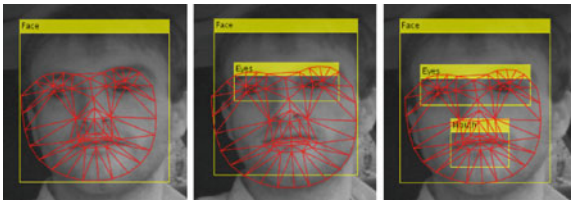


Fig. 4 Face image with initial guess. The more ROIs detected, the more accurate the initial guess computation

4.1 Initial Guess Generation

The iterative matching needs an initial guess for the shape model. That means the model has to be approximately placed on a face in an image. Thus, a Viola-Jones [13] face detector is used to detect the ROI of the face. Within the detected face ROI, other Viola-Jones detectors are utilized to detect the eyes and mouth of the subject. The position and the size of all detected ROIs are used to generate the initial guess by scaling, rotating and translating the base mesh of the shape model. The more ROIs detected, the more information can be used to place the base mesh. In Fig. 4 the same image of a face is shown with three different initial guesses based on the ROIs.

4.2 Model Generation and Parameter Optimization for Matching

The shape and appearance model is generated based on hand-annotated images. The images show different drivers under real driving scenarios captured with low-cost close-to-production cameras. The annotated images cover a wide range of different head-poses and facial expressions and contain images of persons with different ethnic backgrounds. These annotated images are used to calculate the mean MCT appearance model and to generate the shape model. For the shape model, the number of shape vectors (principle components of a PCA) is selected to cover 92 % of the variance over the training set. This value provided an acceptable compromise between accuracy and robustness during the later matching. In this work, the weights for these shape vectors are named as local parameters because they allow different poses and facial expressions to be modelled. To place the model at a certain position within an image, four so called global parameters are required, namely for translation in x-/y-direction, scaling and rotation. The total number of parameters which has to be optimized during matching results to 17 parameters—13 local and 4 global parameters.

4.3 Parameter Constraints and Weighting for Matching

The global and local parameters allow the mesh of the shape model to be adapted to almost every face. But by varying these parameters, net configurations can also be created that do not represent the structure of a face. Experiments on the acquired data set showed that if the mesh has a structure that doesn't have the form of a face, it will not converge to a reasonable form. To suppress obtaining such mesh configurations, the global parameter constraints for the local parameters are introduced as hard limits. The limits for the local parameters are defined so that 98 % of all parameters of the training set are inside these limits. The assumption hereby is that possible outliers in the training set do not unnecessarily enlarge the parameter space. If a parameter exceeds the limit during matching, it will be set back to the predefined limit. Further experiments on the dataset also showed that if the initial guess is not good enough, the model tends to diverge. A closer look at the matching process showed that the local parameters are responsible for that if the model is initialized only approximately. To solve this problem, a coarse-to-fine parameter optimization is introduced. In this context it means that the global parameters are optimized first and the local parameters subsequently. Through this extension, the base mesh is placed as precisely as possible onto the face in the image before optimizing the local parameters. Experiments on the available data set showed that it is sufficient to do just 5 global parameter optimization iterations before starting the local parameter optimization during matching. A further weighting of the local parameters respectively to their standard deviation over the training set can further improve the robustness, as done in the implementation of Luca Vezzaro [14]. These two improvements make the matching process more robust with regard to inaccurate initial guesses. With the assumption that the position of the face and its facial expression do not vary too much in consecutive frames, the global parameter limits can be further limited to local parameter limits, based on the parameters of the face in the preceding image. This improves the robustness of the matching process when parts of the face are not recognizable in some frames of a sequence.

4.4 Occlusion Handling, Quality and Intelligent Stopping Criterion

If some parts of the face are unrecognizable because of occlusion, deep shadows or overexposure (e.g. direct sunlight), these parts prevent a good matching result. To reduce the impact of these parts on the matching, an occlusion handling is introduced. Over the training set, for each triangle of the base mesh a maximum reconstruction error is introduced as the sum of squared distances for the content of each triangle between the model and the image. If the reconstruction error of a triangle exceeds the maximum reconstruction during matching, the content of the difference image for the certain triangle will be set to zero. So that specific triangle

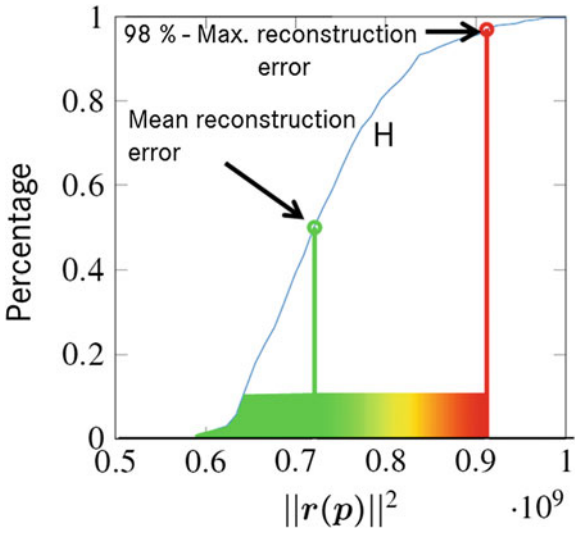


Fig. 5 Cumulative distribution of the reconstruction error over a training set. This distribution is used to rate (*green* good, *red* bad) the matching results

has no impact on the parameter optimization. The reconstruction error over all triangles can further be used to rate a matching result. Therefore the distribution of the reconstruction error over a training set is computed. Through this distribution (Fig. 5), the matching result is rated. If the reconstruction error is greater than 98 % of all reconstruction errors of the training set, it is assumed that the model does not represent a facial structure.

The trend of the reconstruction error during the matching is used to introduce an intelligent stopping criterion. During matching, the reconstruction error should decrease constantly and converge to a minimum. However, there is not necessarily a linear relation between a parameter optimization $\Delta \boldsymbol{p}$ and the difference image $\boldsymbol{r}(\boldsymbol{p})$, as shown in Fig. 6, so it may happen that the reconstruction error increases locally. To avoid an early termination of the parameter optimization here, an intelligent stopping criterion is introduced. The optimization is stopped if the reconstruction

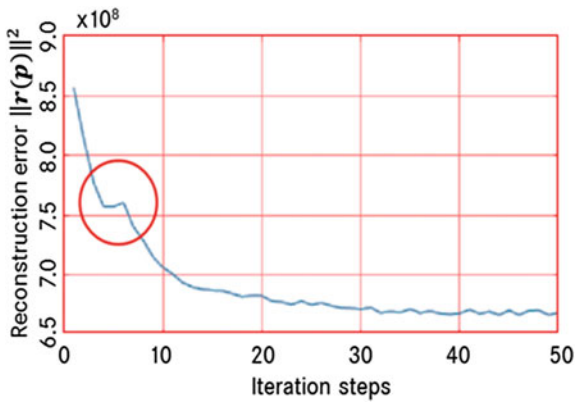


Fig. 6 Reconstruction error during a matching process. Due to the not necessarily linear relation, the reconstruction error may increase locally. To avoid an early termination of the optimization process (*red circle*), an intelligent stopping criterion is used

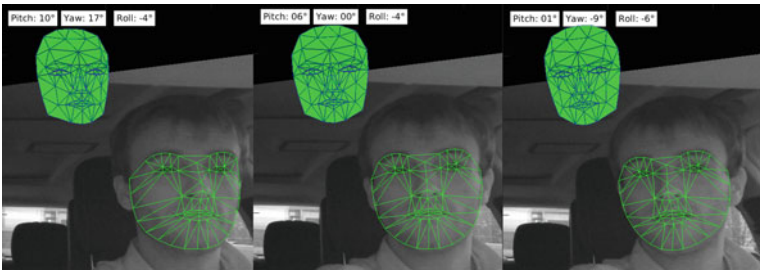


Fig. 7 Examples of the head-pose estimation based on the detected landmarks. The *green* Candide-3 face model in the *upper left* corner shows the estimated head-pose and additionally the estimated angles (yaw, pitch, roll) are visualized

error increases over more than 6 iteration steps or a predefined maximum number of iteration steps is reached.

4.5 Head-Pose Estimation

If the matching of the MCT AAM onto a face in an image was successful, the head-pose is estimated based on the detected landmarks. Therefore a 3D face model is necessary. In this paper the Candide-3 face model [15] is used. The head-pose estimation is formulated as a least squares problem (Eq. 5) between the 2D landmarks $\boldsymbol{p}_{i,2D}$ of the MCT AAM and the corresponding 3D landmarks $\boldsymbol{p}_{i,3D}$ of the Candide-3 model projected by a projection matrix $\boldsymbol{P}$ onto the 2D image plane. Therefore the Candide-3 model is parametrized with 7 parameters, a scaling parameter s_{3D}, a rotation matrix $\boldsymbol{R}$ which is defined by three rotation angles $\phi = (\phi_{yaw}, \phi_{pitch}, \phi_{roll})$ and a translation vector $\boldsymbol{t}_{3D}$. This minimization problem can be solved by any nonlinear least square solver. The results for the parameters can be used directly as head-pose estimation. In Fig. 7 there are examples for fitting results of the least square estimation for the head-pose.

$$\arg \min_{s_{3D},\phi,\boldsymbol{t}_{3D}} \sum_{i=1}^{v} \left|\left|\boldsymbol{p}_{i,2D} - s_{3D} \cdot \boldsymbol{P} \cdot \left(\boldsymbol{R}(\phi) \cdot \boldsymbol{p}_{i,3D} + \boldsymbol{t}_{3D}\right)\right|\right|^2 \tag{5}$$

5 Evaluation

The evaluation was performed on images captured under real driving conditions. The algorithm described in this paper is mostly implemented in MATLAB and runs under Linux on standard PC hardware. Performance-sensitive parts such as image

warping are written in C, without considering parallelization. The achieved frame rate is 12 frames per second. The implementation can easily be modified to enable real-time application (e.g. on embedded platforms). For capturing a low-cost close-to-production IR-camera was mounted in a vehicle and aligned to capture the head motion box of the driver. The dataset contains driving sequences of 42 different subjects with different ethnical background and large inter-individual variation. Out of those sequences 445 images were selected with the focus on having a wide range of different head-poses with respect to a uniform distribution over the subjects. Afterwards the facial landmarks in the selected images were annotated by human experts. Leave-one-subject-out cross validation was performed to evaluate the performance of the proposed framework. To measure prominent landmarks (e.g. nose tip, eye corners, mouth corners) the Euclidean distance in pixels between the detected and annotated landmarks was used. For contour points (eye lid, cheek, nose bridge, lips) a point to counter distance measurement was used. For estimating the annotation variance in the label process, one selected image was annotated by 10 different human experts. The evaluation of the hand annotation process showed that due to the noisy and low-resolution images the positions of the hand annotated landmarks vary within an averaged median of 3.16 pixels over all landmarks. Averaged median means that first the median is calculated for each of the 68 landmarks. Second, that average is calculated over all calculated median values. As a guidance value, the distance between the eye corners of a frontal face in the images is about 45 pixels. For comparison of the novel algorithm, a pre-trained version of IntraFace [16] and the ICAAM-Implementation [14] was trained and tested on the dataset. While IntraFace delivered good results on the dataset, the ICAAM-Implementation [14] diverged for most of the images, probably due to the noisy and inhomogeneous illuminated images. Therefore no meaningful results could be achieved with the ICAAM algorithm. The results for IntraFace and for the novel MCT AAM are shown in Table 1. It can be seen that the novel MCT AAM achieves very good results, close to the accuracy of the hand label process and it even outperforms the IntraFace algorithm on the dataset. In Figs. 8 and 9 some matching results are shown with unseen faces for the MCT AAM. In Fig. 8 it can be seen, that the MCT AAM returns good results, even in the presence of hard shadows, inhomogeneous illumination and overexposure. Figure 9 shows that the MCT AAM generalizes quite well for different persons, different ages, ethnic backgrounds and persons wearing glasses or with heavy beard growth. The accuracy of the head-pose estimation is not evaluated in this work since the focus was

Table 1 Evaluation results of the novel introduced MCT AAM. The results show that the results are close to the accuracy of the hand annotation process. It even outperforms a pre-trained version of IntraFace on the dataset

Method	Averaged median (pixel)
Hand annotation process	3.16
MCT AAM Point-to-Point	**3.51**
MCT AAM Point-to-Contour	**3.18**
IntraFace Point-to-Point	5.29

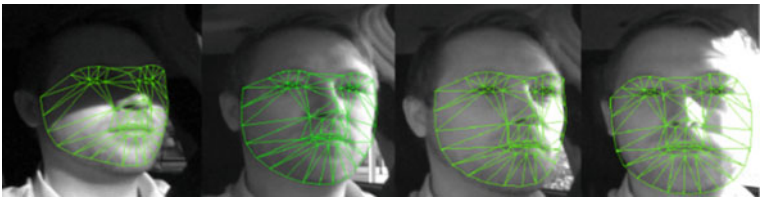


Fig. 8 Examples for matching results of the MCT AAM on major inhomogeneous illuminated images—from hard shadows (*left image*) up to over-exposed areas (*right image*)

Fig. 9 Examples for matching results of the MCT AAM on different previously unseen faces with different head-poses. The MCT AAM returns good results even though some parts are occluded by beard growth, hair or glasses

on robust landmark localization and the fact that beside the hand annotated landmarks, no ground truth for the dataset is available. But as shown, the head-pose estimation returns reasonable values.

6 Conclusion

In this work it is shown that by using the MCT, traditional AAM can be made very robust with regard to major inhomogeneous illumination. In combination with parameter restrictions and occlusion detection and handling, it returns robust results, even if some parts are almost unrecognizable due to noise, over-/under-exposure or due to occlusion by other objects.

References

1. Murphy-Chutorian E, Trivedi M (2009) Head pose estimation in computer vision: a survey. IEEE Trans Pattern Anal Mach Intell 31(4):607–626
2. Cristinacce D, Cootes TF (2006) Facial feature detection and tracking with automatic template selection. In: FGR 2006: Proceedings of the 7th International conference on automatic face and gesture recognition, vol 2006, pp 429–434
3. Cootes TF, Taylor CJ, Cooper DH, Graham J (1995) active shape models-their training and application. Comput Vis Image Understanding 61(1):38–59
4. Zhao M, Chua TS (2006) Face alignment with unified subspace optimization of active statistical models. In: FGR 2006: Proceedings of the 7th International conference on automatic face and gesture recognition, vol 2006, pp 67–72
5. Matthews I, Baker S (2004) Active appearance models revisited. Int J Comput Vis 60 (2):135–164
6. Zhu X, Ramanan D (2012) Face detection, pose estimation, and landmark localization in the wild. In: 2012 IEEE conference on computer vision and pattern recognition, pp 2879–2886
7. Romdhani S, Gong S, Psarrou A (2000) A generic face appearance model of shape and texture under very large pose variations from profile to profile views. In: Proceedings 15th International conference on pattern recognition, ICPR-2000, vol 1, pp 1060–1063
8. Gross R, Matthews I, Baker S (2006) Active appearance models with occlusion. Image Vis Comput 24(6):593–604
9. Theobald B-J, Matthews I, Baker S (2006) Evaluating error functions for robust active appearance models. In: 7th International conference on automatic face and gesture recognition (FGR06), vol 2006, no 2, pp 149–154
10. Cootes TF, Edwards GJ, Taylor CJ (2001) Active appearance models. IEEE Trans Pattern Anal Mach Intell 23(6):681–685
11. Froba B, Ernst A (2004) Face detection with the modified census transform. In: Proceedings of the sixth IEEE International conference on automatic face and gesture recognition, pp 91–96
12. Baker S, Matthews I (2004) Lucas-Kanade 20 years on: a unifying framework. Int J Comput Vis 56(3):221–255
13. Viola P, Jones MJ (2004) Robust real-time face detection. Int J Comput Vis 57(2):137–154
14. Vezzaro L (2011) ICAAM—Inverse compositional active appearance models, MathWorks File Exchange
15. Ahlberg J (2001) Candide-3. An updated parameterised face, LiTH-ISY-R-2326, vol 1, pp 1–16
16. Xiong X, De la Torre F (2013) Supervised descent method and its applications to face alignment. In 2013 IEEE conference on computer vision and pattern recognition, pp 532–539

Using eHorizon to Enhance Camera-Based Environmental Perception for Advanced Driver Assistance Systems and Automated Driving

Hongjun Pu

Abstract Driven by the vision of automated driving (AD), future advanced driver assistant systems (ADAS) will require considerably stronger capability of environment perception compared to the current ones. Among the automotive sensors, camera(s) can deliver the richest environmental information and will be indispensable in AD-scenarios. However, it is practically not easy to match in real time the objects of a camera picture to the original ones on the road. This paper suggests a solution to this problem consisting of simple calculations based on optical and mounting parameters of the camera and road topology data provided by the electronic horizon (eHorizon). Given the fact that the eHorizon is increasingly deployed due to its great potential for optimized fuel and energy consumption, the proposed solution does not require additional budget of material (BOM) and provides large benefit in enabling and enhancing a lot of ADAS and AD applications.

Keywords Automotive camera system · Environment perception · ADAS · AD · ehorizon

1 Introduction

Recently, the research and development of advanced driver assistant systems (ADAS) is experiencing a great push from the emerging desire for automated driving (AD). In this context, future vehicle controls are expected to be more precise, agile and situation-adaptive. This leads consequently to increased demands on the ability of environmental perception of vehicle systems.

For a driver, sight is the sense which brings to him the most environmental information when he is driving a vehicle. Consequently, a system for automated driving must possess functions substituting the driver's eyes. Cameras mounted in and on the vehicle can deliver the richest environmental information compared to

H. Pu (✉)
Continental Automotive GmbH, Philipsstrasse 1, 35576 Wetzlar, Germany
e-mail: hongjun.pu@continental-corporation.com

T. Schulze et al. (eds.), *Advanced Microsystems for Automotive Applications 2016*, Lecture Notes in Mobility, DOI 10.1007/978-3-319-44766-7_9

other automotive sensors such as radar, laser, ultrasonic, infrared, etc. and camera pictures or videos can be used stand-alone or as redundant data sources to other sensors for diverse driving assistance applications.

Given the great benefit of automotive camera systems, it will be an essential task to match the images taken by a camera to the original objects in the real world, i.e. to map the 2D pixel coordinates of a picture to the 3D world coordinates, continuously during the driving.

For example, the current products of TSR (traffic sign recognition) give the driver a signal, when a traffic sign is recognized. Usually, it is also sufficient, since the driver can judge the situation by himself. In case of automated driving, a computer has to decide, if a recognized traffic sign is relevant for the ego-vehicle. To do so, more information is necessary, e.g. where the traffic sign is and how it stands with respect to the ego-vehicle and to the current road. In this case, a mapping from the camera picture to the real world is needed.

In opposite, an algorithm of object recognition can be executed more efficiently, if it is roughly known, in which part of the picture the image of the relevant object, e.g. a pedestrian on the street, is expected. For this purpose, one needs a mapping from the real world to the camera picture.

In this paper, after a short introduction of the electronic horizon (eHorizon) [1] in Chap. Probabilistic Integration of GNSS for Safety-Critical Driving Functions and Automated Driving—The NAVENTIK Project, an approximating solution to the problem of mapping camera picture to real world and vice versa will be presented. The algorithm is based on simple geometric optics and the road topology provided by the eHorizon.

Due to its great potential for the optimization of fuel and energy consumption [2], the eHorizon is expected to be increasingly equipped in commercial vehicles and passenger cars. Under this assumption, the use of eHorizon for the enhancement of camera-based environmental perception would not require additional budget of material (BOM), but just exploit the synergy. On the other hand, the benefit is very large, since a practicable method for the mapping between camera picture and real world would enable and enhance a lot of ADAS and AD applications.

2 The eHorizon

The eHorizon is an emerging technology providing road information to ADAS applications for the purpose of fuel/energy optimization and safety enhancement.

The development of eHorizon was initiated by a group of car manufacturers, vehicle system developers and map data companies who are organized in the ADASIS Forum. The system and interface specifications of the ADASIS Forum, current version ADASIS v2 [1], are widely accepted by the automotive industry as de facto standard. Ress et al. published in 2008 a well structured description of ADASIS v2 [3].

Fig. 1 Elevation of road surface (relative to current position) provided by eHorizon

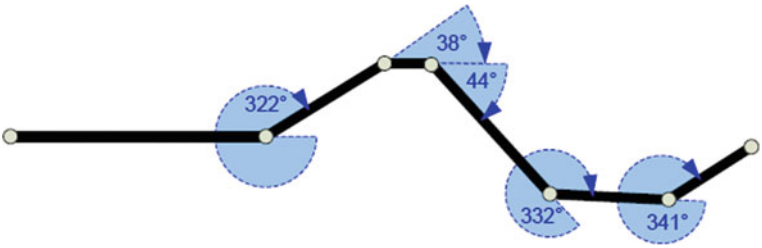


Fig. 2 Heading changes of road segments (relative to previous segment) provided by eHorizon

The core information provided by the eHorizon to the ADAS applications are the so called ADAS attributes. These are mainly the road classes, slopes, curvatures, speed limits, etc. Figures 1 and 2 show examples of two ADAS attributes, i.e. slope and curvature, along with the road.

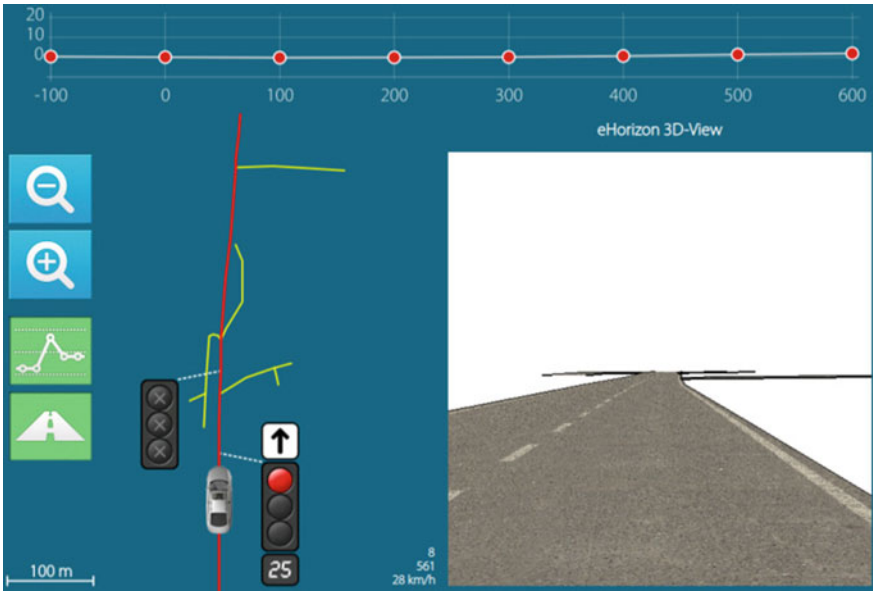


Fig. 3 Traffic light and 3D road topology based on eHorizon data

With data provided by the eHorizon, various ADAS applications can be realized and enhanced, e.g. fuel consumption optimization according to road topology [2], traffic light assistance based on extended eHorizon [4, 5], etc.

Another interesting example of eHorizon application is, as illustrated in Fig. 3, the reconstruction of road topology in 3D space using ADAS attributes provided by the eHorizon [5].

3 The Problem and Solution of Mapping a Camera Picture to the Real World and Vise Versa

3.1 Coordinate System and the Perceptible Range of the Camera

As shown in Figs. 4 and 5, a 3D coordinate system fixed on the vehicle is used to present the camera, the ego-vehicle, the road and objects on the road.

The origin of the coordinate system is fictively fixed exactly on the focus of the camera lens. The camera lens is assumed to have a vertical FOV (field of view) of 2β and a horizontal FOV of 2ϕ. The mounting angle of the camera is α and the mounting altitude is h.

3.2 Analysis Based on Linear Geometric Optics

The following assumptions are made for simplifying the analysis

(a) The camera can be treated with linear geometrical optics, i.e. ignoring any optical distortion

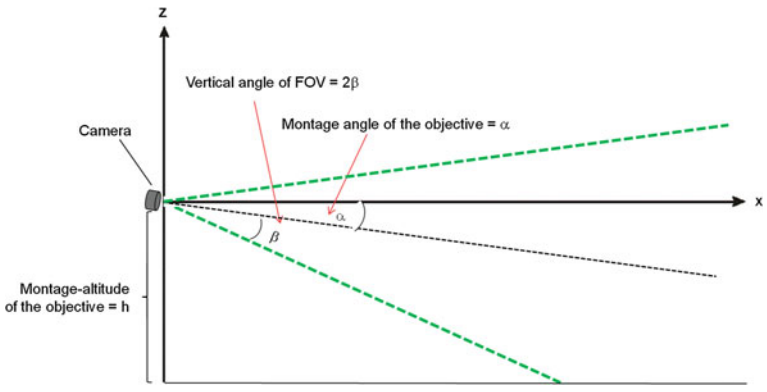


Fig. 4 Perceptible range of the camera projected on the vertical

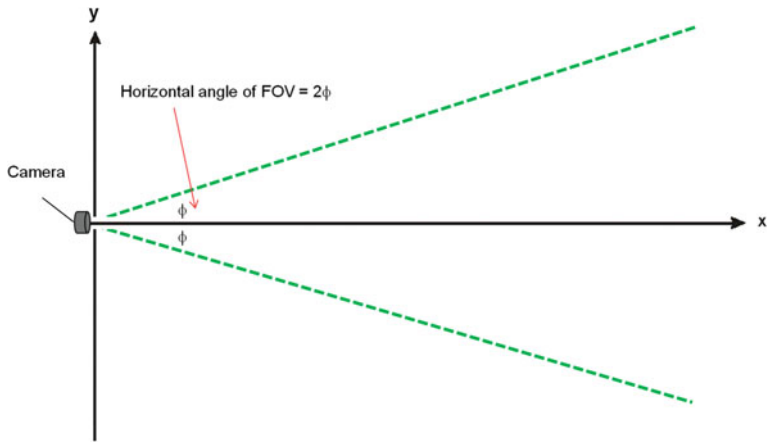


Fig. 5 Perceptible range of the camera projected on the horizontal

(b) The vehicle-body and the mechanical structure of the camera system are rigid, i.e. impact of elastic vibration during the driving is not considered

The dotted green lines in Figs. 4 and 5 present the projections of a pyramid, in which the light-rays go from objects of real world to the camera lens and form an image there. This pyramid contains the perceptible range of the camera and proceeds unlimitedly, until it reaches the road surface or another objects.

Suppose that the camera picture has a resolution of $2n$ pixels by $2m$ pixels. Then a point (x, y, z) within the perceptible range can be mapped to a point (v, w) on the camera picture with $-n \leq v \leq n$ and $-m \leq w \leq m$.

On the other hand, it is more desired in practice to know for a given point (v, w) of the picture the corresponding real world point (x, y, z).

Following linear geometric optics, the pixel coordinates (v, w), i.e. the position on the picture, of the image of a real world point depend only on the vertical entry angel (between $-\beta$ and β) and the horizontal entry angle (between $-\phi$ and ϕ) in which the light-ray from the real world point reaches the focus of the camera.

With a proper linear transformation, the above vertical and horizontal entry angles can be presented as the pitch ($-\alpha - \beta \leq \eta \leq \beta - \alpha$) and yaw ($-\phi \leq \xi \leq \phi$) of the light-ray in the vehicle coordinate (Figs. 4 and 5) and it holds

$$\frac{v}{n} = \frac{\eta + \alpha}{\beta} \tag{1}$$

$$\frac{w}{m} = \frac{\xi}{\phi} \tag{2}$$

From Eqs. 1 and 2 it turns out

$$\eta = \frac{v}{n}\beta - \alpha \tag{3}$$

$$\xi = \frac{w}{m}\phi \tag{4}$$

The above two equations give for each pixel of the picture (v, w) the entry angles η and ξ, in which the light-ray comes from the original object to the focus. By "inverting" the light-ray, i.e. drawing a fictive line from the focus into the real world with the exit angles η and ξ, one must be able to reach somewhere the original object depicted at (v, w). This idea is the essential basis of this paper.

3.3 The Inverse-Light-Ray and Its Construction

Firstly it is defined that

For any given (v, w) *in a camera picture, the inverse-light-ray of* (v, w) *is a line starting from the focus of the camera with a vertical exit angle* η *and a horizontal exit angle* ξ *as expressed in Eqs.* 3 *and* 4

Then it holds

Whenever the inverse-light-ray of (v, w) *reaches for first time a real object, then this is the original object depicted at* (v, w) *in the picture.*

Figure 6 shows the inverse-light-ray of (v, w) for the simplified case with $\alpha = 0$. However, the following analysis is valid for general cases with $\alpha \neq 0$.

Since the focus point is the origin of the 3D coordinates system, the *inverse-light-ray of* (v, w) can also be expressed as $k\vec{r}_0(v,w)$ for $0 \leq k \leq \infty$, where $\vec{r}_0(v,w)$ is a unit vector with the pitch of η and yaw of ξ.

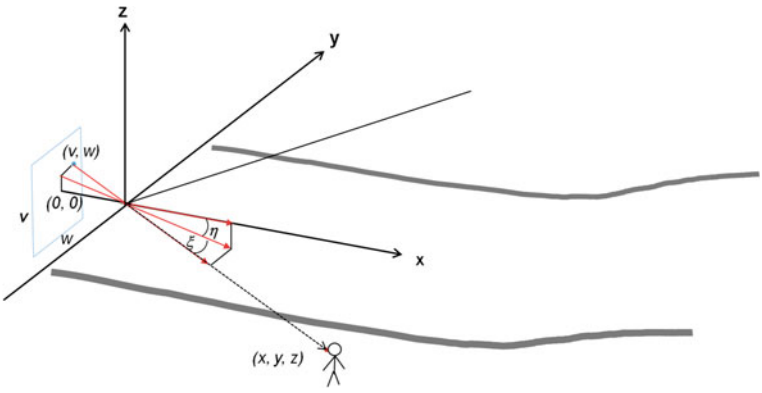


Fig. 6 Illustration of the inverse-light-ray concept

As illustrated in Fig. 6, the $\vec{r}_0(v, w)$ can obtained by rotating the unit vector [1, 0, 0]^T firstly for η degree about the y-axis and secondly for ξ degree about the z-axis, i.e.

$$\vec{r}_0 = \begin{pmatrix} \cos(\xi) & -\sin(\xi) & 0 \\ \sin(\xi) & \cos(\xi) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(\eta) & 0 & -\sin(\eta) \\ 0 & 1 & 0 \\ \sin(\eta) & 0 & \cos(\eta) \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos(\xi)\cos(\eta) \\ \sin(\xi)\cos(\eta) \\ \sin(\eta) \end{pmatrix} \tag{5}$$

So any point (x, y, z) on the inverse-light-ray can be expressed as

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = k \begin{pmatrix} \cos(\xi)\cos(\eta) \\ \sin(\xi)\cos(\eta) \\ \sin(\eta) \end{pmatrix} \quad 0 \le k \le \infty \tag{6}$$

3.4 The Inverse-Light-Ray-Method for Mapping a Picture to the Real Word

Following the analysis of the last section, the inverse light ray proceeds with increasing k from $k = 0$ and reaches somewhere the object matching to (v, w). To determine the crossing point exactly, the road topology is necessary and can be approximately provided by the eHorizon.

For the purpose of this paper the road is reconstructed as simple discrete poly-lines in 3D space based on ADAS data. The reconstruction consists of the following steps

(i) Fix a coordinate system on the ego-vehicle as described in Sect. 3.1
(ii) Build the road or, if possible, each lane of the road as a poly-line consisting of discrete road points in the coordinate system. This is always possible for a known current car position (longitude, latitude and elevation) and the ADAS attributes, especially, slopes and curvatures.
(iii) The above poly-line(s) may be built for a few meters to a few hundreds of meter ahead of the ego-vehicle, according to the function of the camera as well as the performance of the eHorizon.

Note that there are totally p discrete road points are determined from the above steps, and the elevations of these p points vary from $z_{\min}$ to $z_{\max}$.

At first, the interval $[z_{\min}, \quad z_{\max}]$ is equally divided in q sub-intervals. Then, for each $z_i \subset [z_{\min}, \quad z_{\max}]$, $i = 1,\ldots, q$, the corresponding x_i and y_i are calculated according to Eq. 6:

$$x_i = \frac{z_i}{\sin(\eta)}\cos(\xi)\cos(\eta), \quad \eta \neq 0 \tag{7}$$

$$y_i = \frac{z_i}{\sin(\eta)}\sin(\xi)\cos(\eta) \quad \eta \neq 0 \tag{8}$$

Finally, each of the q points obtained above is compared with each of the p road points obtained from eHorizon.

Where the minimal distance among the $p \times q$ pairs occurs, the inverse-light-ray reaches the road surface approximately there.

Many camera-based ADAS applications are designed to track, detect and recognize certain objects such as car, truck, pedestrian, traffic sign, etc. which do not directly lie on the surface of the road. In these cases, the estimated position of the treated objects in relation to the road surface must be considered in the calculation of road points with steps *i* to *iii*.

The computational efforts, but also the precision of the results, increase with p and q, as well as the resolution of the picture. While q can be chosen as large as the system capability and other boundary conditions allow, p is limited by the resolution of road topology the eHorizon can provide.

3.5 The Mapping from Real Word to Picture

Compared to the problem treated in Sect. 3.4, it is quite easy to map an object in the perceptible zone in the real world to the camera picture

Firstly, any point in the real world can be expressed in the coordinate system defined in Sect. 3.1

Secondly, for any given point (x, y, z), one obtains the η and ξ directly from Eq. 6

$$k = \sqrt{x^2 + y^2 + z^2} \tag{9}$$

$$\eta = \sin^{-1}\left(\frac{z}{k}\right) \tag{10}$$

$$\xi = \sin^{-1}\left(\frac{y}{k\cos(\eta)}\right) \tag{11}$$

Finally, one gets v, w from Eqs. 3 and 4

$$v = n\frac{\eta + \alpha}{\beta} \tag{12}$$

$$w = m\frac{\xi}{\phi} \tag{13}$$

In order to use Eq. 11, it is necessary that $\eta \neq \pm\pi/2$ which is almost always true in real use cases.

4 Conclusion and Outlook

This paper presents a solution to map the camera pixel to the real word and vice versa. While the mapping from real world to picture is simply analytical, the mapping from picture to real world has to be calculated approximately, using the road topology data from eHorizon.

The precision of the inverse-light-ray-method introduced by this paper depends strongly on the quality of road information provided by the eHorizon. Fortunately, the resolution of future digital map will be significantly higher than the current ones due to the requirement of automated driving.

Further the solution of this paper is derived under two assumptions made at the beginning of Sect. 3.2.

The assumption (a) can be held, at least approximately, by selecting proper cameras and cutting the images so that optical distortions are reduced to a negligible level. Also software-based pre-processing of the image can help to reduce distortion.

The assumption (b) is made to rule out the impact of elastic vibration of vehicle body and camera system. However, this vibration, usually of high frequency, exists all the time and makes the parameters like h, α, etc. inconstant. This disturbance may be compensated, at least significantly reduced, by some predictive controls and/or parameter estimations making use of the measurement data of the inertial sensor in the vehicle. The compensation of parameter uncertainty is beyond the scope of this paper, but an interesting topic for the future research in the effort to achieve efficient and precise camera-based environmental perception.

References

1. ADASIS Forum, ADASIS v2 Protocol, December 2013. Document can be requested under www.adasis.ertico.com
2. Scania Group, Scania Active Prediction, December 2011. Document downloadable from http://www.scania.com/products-services/trucks/safety-driver-support/driver-support-systems/active-prediction/
3. Ress C, Balzer D, Bracht A, Durekovic S, Loewenau J (2008) ADASIS protocol for advanced in-vehicle applications. In: 15th ITS World Congress. New York, USA, 16–20 November 2008

4. Pu H (2015) Dynamic eHorizon with traffic light information for efficient urban traffic. Springer lecture notes in mobility: advanced microsystems for automotive applications 2015. Springer International Publishing, Berlin, pp 15–24
5. Kuhfuß H, Schürmann J, Bieger S, Pu H (2016) Elektronischer Horizont mit Lichtsignalinformation aus der Cloud am Beispiel der URBAN-Implementierung. Automotive meets Electronics 2016. VDE-VERLAG, Offenbach

Performance Enhancements for the Detection of Rectangular Traffic Signs

Lukas Pink and Stefan Eickeler

Abstract Most countries around the world present regulation and rules, applying on public roads, by putting up traffic signs. Therefore it is useful for driver assistance systems and important for autonomous vehicles to understand the meaning and consequences of those signs. One class of traffic signs that present important information is speed limit signs, which underlie strict norms. In this paper, we will introduce performance enhancing methods for the detection of rectangular traffic signs on the example of speed limit signs in the United States of America (USA). We will show that with a small and acceptable loss of accuracy the number of calculations needed and their complexity can be greatly reduced. Due to that, the energy consumption of the embedded hardware and the processing time per frame are reduced.

Keywords Symmetry detector · Traffic sign detection · Performance enhancement

1 Introduction

As driver assistance systems are advancing ever more, their functionality is expanding too. The latest version of cars can help the driver to drive at constant speed, turn on the light when necessary and even break the speed if the situation requires it. Currently the automatic reaction to traffic sings along the road is researched.

L. Pink (✉)
University of Applied Sciences Bonn-Rhein-Sieg, Grantham-Allee 20,
53757 Sankt Augustin, Germany
e-mail: lukas.pink@smail.inf.h-brs.de

S. Eickeler
Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS,
Schloss Birlinghoven, 53757 Sankt Augustin, Germany
e-mail: stefan.eickeler@iais.fraunhofer.de

T. Schulze et al. (eds.), *Advanced Microsystems for Automotive Applications 2016*, Lecture Notes in Mobility, DOI 10.1007/978-3-319-44766-7_10

The information presented on traffic signs allows anyone and anything on a public street to behave according to the law and decreasing the risk of accidents. For autonomous vehicles on the streets, it is even more important to recognize and understand the meaning of a street sign than for vehicles driven by a human being. This is due to that fact that humans can already react on signs by themselves. But since a human can be distracted, traffic sign recognition in their cars would also be helpful. As a result of that, most vehicles are to be installed with a system, equipped with a camera, facing the street ahead and reading traffic signs.

Independent whether a human being or a computer system bases decision on such traffic sign reading system, there are certain needs.

In order to keep the deciding instance informed correctly, the system must not miss a sign and not give false information. The only exception is, if the system had no chance of recognizing it. That might be the case if the sign is blocked or covered in any way.

Delivering correct information however is not the only important matter. As this information is only helpful when available in time, the system has to interpret the given information fast.

As the vehicle has only a limited amount of energy, while doing the above, the system is to use as little energy as possible.

For a street sign recognition system, recognizing, identifying and understanding a sign gets simpler the more those signs are standardized. That includes shape, color and text. A category of signs that is highly standardized and therefore easily recognizable, are speed signs. In most countries around the world, there are no more than 20 different speed limits. Therefore and due to the norm of size and color, the speed limit signs within a country, differ only by the digits presenting the speed limit. In most cases the number, representing the highest allowed speed, consists of two or three digits. Due to the simplicity of those signs and the importance to the deciding instance, their recognition was tackled early.

2 Related Work

In the last three decades different approaches to recognize traffic signs, where presented. They can basically be divided into two categories. Either the signs where detected using a machine learning approach [1] or with a parametric approach [2–4]. Apart from some exceptions [1] however, the common approach generally consists of two steps. In the first detection step, candidates are determined and in a second recognition step these candidates are either verified or dismissed [2–5]. The step of recognition, takes a detailed look at the candidate and therefore it is costing a considerable amount of calculation time. Even, if this recognition step applied on its own would mostly come to the same conclusion as without preprocessing, doing so generally causes the system to decrease its calculation time per frame. In order to ensure that, the first step needs to be as fast as possible. Thereby, it is allowed to make mistakes, as false detections can be tracked down in the verification step.

Based on the idea of "radial symmetry detection" [2] proposed by Loy and Barnes, which was successfully used in 2008 [5], Eickeler and Valdenegro presented techniques for the reduction of processing time up to an factor of 10 for circular traffic signs [6]. Loy and Barnes also presented a promising radial symmetry approach for rectangular shaped signs. As Eickeler and Valdenegro also proposed that their improvements would also be adaptable for triangle and rectangle detection, we will use their techniques and combine it with other improvements, to enhance the runtime performance and reduce the energy consumption.

3 Radial Symmetry Detection

The assumption behind the algorithm [2] states that based on the edge of a shape, the position of its center is distinguishable. Therefore each edge determined within the image votes for a certain position as the shape's center to which the respective edge may belong to. This center is supposed to be in the direction of the edge gradient. As the exact size of the shape is not known however, this voting is not concentrated on a single point, but on many points along a line, which is orthogonal to the gradient of the edge. As it is also unknown, on which side of the edge the center is, an edge votes for two identical lines, one for each direction of gradient.

The authors made an additional assumption. As recognizable street signs are within a certain range of sizes, the "voting lines" are of limited length. At the end of the positive voting lines, there is a negative line attached, voting this position as unlikely to be the center of a shape. Figure 1 presents a graphical representation of voting lines, for some shapes.

Voting for a position as center, following Loy and Barnes, means voting with a vector. These vectors of length one are based on the edges' gradient vector. In order to ensure that all edges of a sign vote with the same vector, the gradient vectors angle is multiplied by the number of sides of the sign. Figure 2 presents the

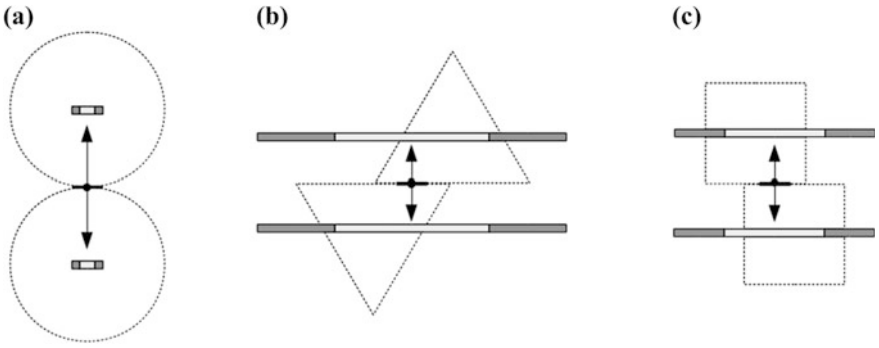


Fig. 1 Exemplary voting processes for a circle (**a**), a triangle (**b**) and a square(**c**), based on an edge and its gradient

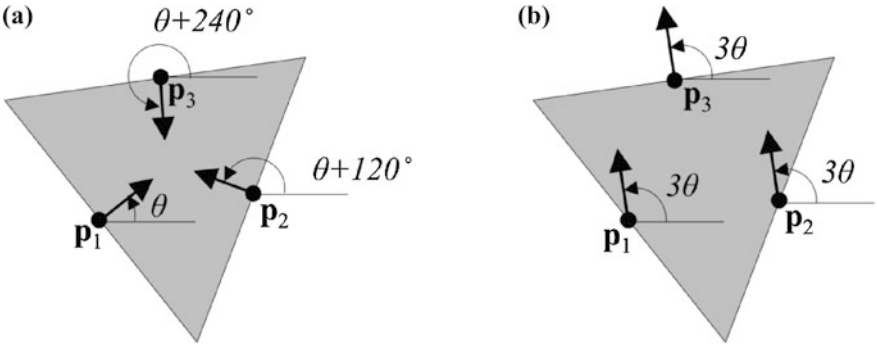


Fig. 2 "Example of three edge points pi on a triangle. Showing (**a**) the angles of the unit gradient vectors, and (**b**) the resulting vectors obtained by multiplying the gradient angles by n (n = 3)." (1, p. 4). When using vector representation with x- and y-coordinate, the multiplication by n can be replaced by a potentiation, if the vectors are of length one and treated as complex numbers

derivation of the voting vector for a triangle. The length of the voting lines and its distance to the edge, are very essential criteria, when applying the algorithm. They depend on each other, as the center of the shape will be further away from the center for larger shapes. The shapes, which will be marked as candidates, are defined by the length and the distance of the voting lines. To overcome the limitation of just a single size, the authors created multiple arrays, each being filled with voting lines of different lengths. That enables the algorithm to detect the same sign in a period of frames, while the car passes it. Thereby it also reduces the cases during which the sign is blocked from the camera view.

4 Improvement Potential

After analyzing the algorithm presented in [2] and briefly summed up in Sect. 3, we saw the potential to improve the processing time, while changing the outcome only insignificantly.

4.1 Scaled Voting Arrays

As described in Sect. 3, there are different radii r which combine to the set of all radii R. Each r is considered while voting. For each edge that is detected in the image, voting lines for each r are created and stored. Votes of the same radius are accumulated. That results in the creation of $n = |R|$ different arrays, containing votes. Each of them is the size of the input image. Detections are determined, by searching the arrays for high values.

As shown in [6], the size of these images can be reduced, assuming that the algorithm is used to detect signs of a certain size. Using scaled voting arrays reduces the allocation time, as less memory is needed. But even more important, the time it takes to evaluate the scaled voting images is smaller too.

4.2 Altered Voting Process

The voting process, implying the processing of all pixels in the input image, consists of the following steps.

1. Calculating the edge of the pixel in question.
2. Calculate the vector with which will be voted.
3. For all $r \epsilon R$
   (a) Calculate the position and length of the voting line with radius r
   (b) Add or subtract this pixels voting vector to the scaled voting images (see Sect. 4.1 and [6]) for each pixel along the two voting lines.

Analyzing this process, the most frequent calculation is the position of the pixels, where voting vectors will be added or subtracted. Therefore to optimize this has been primary goal.

1. **Vote only for relevant edges**
   Observing exemplary input images, we restricted the edges, that are allowed to vote. As the algorithm searches for rectangular traffic signs and signs are placed upright at the side of the street, their edges should appear parallel to the image's border. With respect to smaller graphical distortions edges where the gradient is between 10° and 80° are therefore irrelevant. Using this restriction, the number of votes is reduced to a fraction of its former amount.
2. **Vote with predefined values**
   As a result of the restriction of edges allowed to vote, the voting vectors are of a certain form. As we stored the vector as x- and y-components, instead of degree and length, one component is very close to one, and the other to zero. For an edge on top or on bottom of the sign, the x-component is nearly zero, for the left and right edges, the y-component is. Based on this observation, the exact calculation of the vector is dropped and voting is done, with zero and one.
3. **Vote with a four points only**
   Instead of voting at potentially dozens of pixel, we replaced a voting line by four voting points. To ensure good detection performance, a post processing step has to be inserted after the voting. A box filter is applied to deduce lines out of points. In the filtering process the new value for a pixel is calculated according to Eq. 1, where the box has the dimension $a * b$.

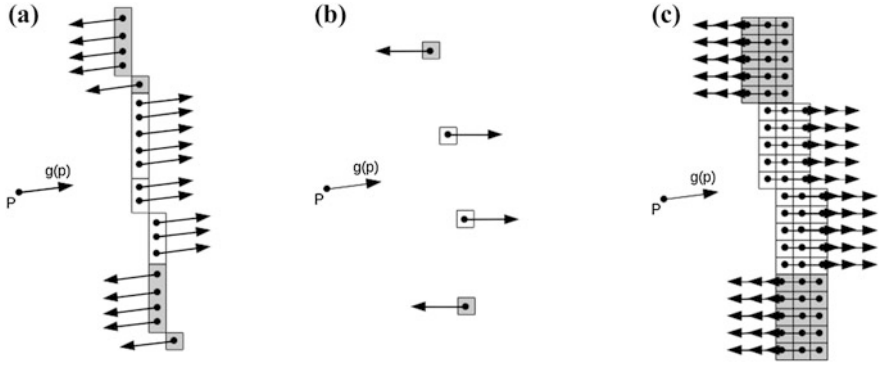


Fig. 3 The original voting process (**a**), the voting with points, instead of lines (**b**) and the voting points after applying the post processing (**c**)

$$VoteValue(x, y) = \sum_{i=-a}^{a} \sum_{j=-b}^{b} voteValue(x - i, y - j) \tag{1}$$

This is calculation is performed using a summed area table as presented in [7], whereby the same result is reached, but instead summing up a times b values, only four values are needed.

Figure 3 presents the altered voting process.

5 Implementation

We used a single core C++ implementation of the rectangle detection algorithm, as described in Sect. 3. Afterwards we applied the changes described in Sect. 4. To improve the performance further, we replaced all the calculations based on *floats* by *integer* calculations. In order not to get to inaccurate, fixed point arithmetic with seven decimal spaces was used. Also we replaced the square root, as frequently used when calculating the position of the four voting points, by a "look up"-table the size of 16 times 16. Thereby, after scaling the gradient vector into a specific range, the table offered fast deduction of the hypotenuses length. As a last step, we examined the scaled images for their range and, as the values where between zero and 5000, used *short* to store the votes.

Thereby the optimized version was created. When ever possible, those changes where also applied to the reference implementation. For both versions, the following part of the configuration was hard coded as follows.

To detect high points in the voting image, a maximum on this image is determined, and then a threshold is calculated as $thr = 0.8 * maximalValue$. And for any value above that, it is checked to be a local maximum. Pixels with a local maximum

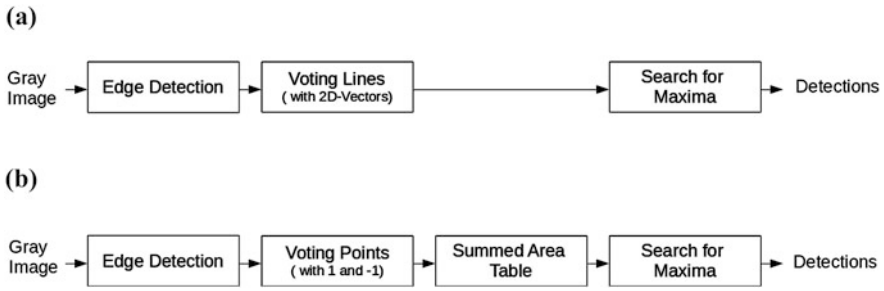


Fig. 4 The original algorithm by Loy and Barnes (**a**), and the algorithm after optimization (**b**)

are assumed to be centers of rectangles. Also the aspect ratio of the signs is set. The height of the signs is defined as 1.48 times the signs width.

An overview of the two algorithms steps is presented in Fig. 4.

6 Results

6.1 Benchmark

We tested the reference implementation and the optimized version with the same data. The benchmark consists of data collected by "The Laboratory for Intelligent and Safe Automobiles" (LISA) [8] and data collected by Dennis Hansen.

The benchmarks part A, which are data from LISA, contained various traffic signs. As we created a rectangle detector for speed signs, we choose only the images containing speed signs. After removing the signs, that either where too large or to small for our use case, 300 images remained, presenting a total of 45 different traffic signs. The resolution of the images is 1024 * 522 pixels.

The data of part B was supplied by Mr. Hansen. It was collected under real live conditions in the United States of America. The resulting videos contain footage of highway and city under different conditions, including rain and bright sunshine. We broke the video down into single images. A total of 24 images contain ten signs, to be recognized. Each image has the videos resolution of 1920 * 1080 pixels. From both sources we added images not containing traffic signs. From LISA these were 300 images and from Hansen 24.

6.2 Qualitative Performance

As the algorithms assignment is to detect rectangles, that are street signs, the quality of each version is established as follows. The position of each traffic sign is

Table 1 Confusion matrix with best achievements for benchmark part A, archived by the reference implementation

	$x \in$ *TrafficSigns*	$x \notin$ *TrafficSigns*
$f(x) \in$ *TrafficSigns*	11	62
$f(x) \notin$ *TrafficSigns*	13	15

Table 2 Confusion matrix with best achievements for benchmark part A, archived by the optimized version

	$x \in$ *TrafficSigns*	$x \notin$ *TrafficSigns*
$f(x) \in$ *TrafficSigns*	18	53
$f(x) \notin$ *TrafficSigns*	6	14

determined and stored in a file. For each image rectangles are detected. If detection is overlapping the traffic sign by at least 50 %, and the sizes are similar, the detection is considered a true positive. Otherwise such detection is declared a false positive. If there is a traffic sign, that was not detected, it counts as false negative. And if there is neither detection nor a sign, it's a true negative. Those values combine to a confusion matrix. As we are looking for traffic signs, but the algorithm detect all kind of rectangles, we are bound to produce false positives.

As the resolution of the images in parts A and B of the benchmark is considerably different, we used two different configurations here. Tables 1, 2, 3 and 4 contain the confusion matrices with the best achieved results.

For part A the minimal size of a rectangle is set to 10 pixels in width and the number of scaled images was ten. For part B the minimal size was 12 pixels in width and ten scaled images where used.

The combination of the number of scaled images and the minimal size determines the range of sizes that detected signs will have. Therefore, when using the

Table 3 Confusion matrix with best achievements for benchmark part B, archived by the reference implementation

	$x \in$ *TrafficSigns*	$x \notin$ *TrafficSigns*
$f(x) \in$ *TrafficSigns*	114	361
$f(x) \notin$ *TrafficSigns*	186	218

Table 4 Confusion matrix with best achievements for benchmark part B, archived by the optimized implementation

	$x \in$ *TrafficSigns*	$x \notin$ *TrafficSigns*
$f(x) \in$ *TrafficSigns*	231	476
$f(x) \notin$ *TrafficSigns*	70	107

detector, one has to decide on the minimal and maximal size of rectangles to be found. That allows the deduction of the number of scaled images.

6.3 *Quantitative Performance*

As the algorithms runtime is below 0.09 s per frame for images from benchmark part A and below 0.03 s for benchmark part B, the quantitative performance is established not by frame, but by all images per part.

As the configuration is crucial for the runtime, we tested several configurations here. But one should bear in mind that the quality will suffer from this. Figure 5 presents the variation of runtime, with the different configurations. The configuration, as stated in Sect. 6.2 creating the best results was used. During the evaluation just one parameter was changed at a time.

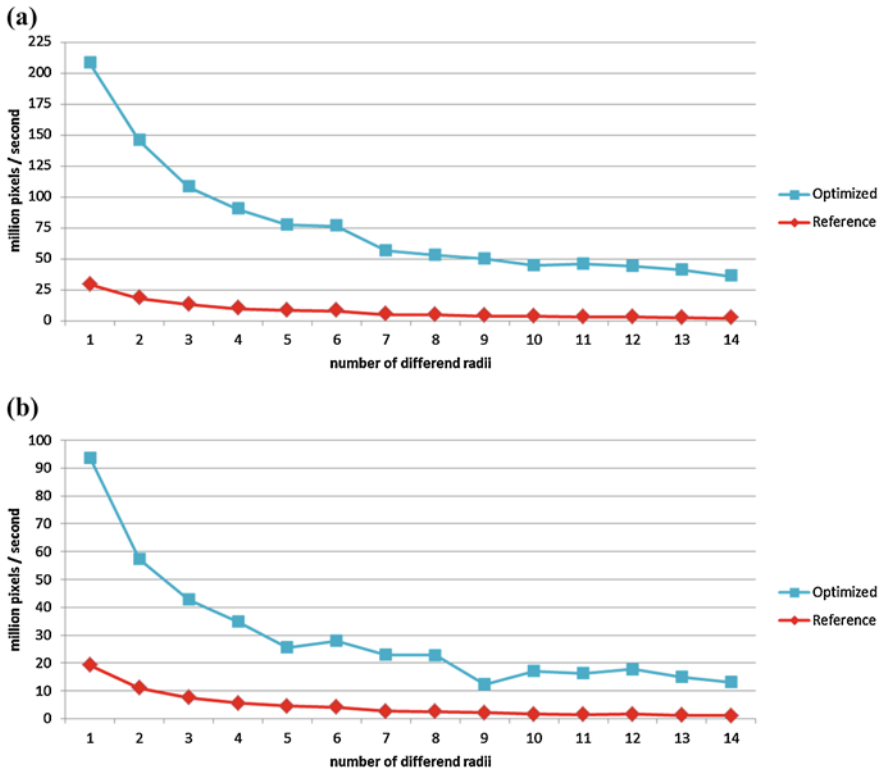


Fig. 5 Performance results, changing the number of radii used, while testing on parts A (**a**) and B (**b**) of the benchmark

Testing was performed on one core of an Intel® i7-6700 CPU. When switching to an embedded system, using six scaled images turned out to be a good configuration. Applied on images of one megapixel, the quality of the result and the calculation time are in balance.

6.4 Conclusion

By altering the process on the algorithmic level and on programmatic level, the algorithm on average performed eight times faster than the reference implementation. This performance improvement is additional to the improvements presented by Eickeler and Valdenegro [6]. At the same time, as we demonstrated, the quality of the detection results is not changed significantly. Thus the energy consumption for the algorithm is reduced, which makes it appropriate for field testing. Also we showed that the same configuration works for all kind of environments, as long as the resolution of the input images is constant.

7 Future Work

Using the techniques we proposed in this paper which enhance the detection of rectangles, we will tackle the recognition of rectangular traffic signs next. Thereby we hope to boost the development of traffic sign interpretation systems so that the deciding instance for automated vehicle operation in a public street can make the right choices faster.

References

1. Sermanet P, LeCun Y (2011) Traffic sign recognition with multi-scale convolutional networks. In: The 2011 International joint conference on neural networks (IJCNN), IEEE, pp 2809–2813
2. De la Escalera A, Armingol JM, Mata M (2003) Traffic sign recognition and analysis for intelligent vehicles. Image Vis Comput 21(3):247–258
3. Barnes N, Zelinsky A, Fletcher LS (2008) Real-time speed sign detection using the radial symmetry detector. IEEE Trans Intell Transp Syst 9(2):322–332
4. De La Escalera A et al (1997) Road traffic sign detection and classification. IEEE Trans Ind Electron 44(6):848–859
5. Loy G, Barnes N (2004) Fast shape-based road sign detection for a driver assistance system. In: IROS: 2004 IEEE/RSJ International conference on intelligent robots and systems, pp 70–75
6. Eickeler S, Valdenegro M (2016) Future computer vision algorithms for traffic sign recognition systems. In: Schulze T, Müller B, Meyer G (eds) Advanced microsystems for automotive applications 2015—smart systems for green and automated driving. Springer International Publishing, Berlin, pp 69–77

7. Crow FC (1984) Summed-area tables for texture mapping. ACM SIGGRAPH Comput Graph 18(3):207–212
8. Møgelmose A, Trivedi MM, Moeslund TB (2012) Vision based traffic sign detection and analysis for intelligent driver assistance systems: perspectives and survey. IEEE Trans Intell Transp Syst

CNN Based Subject-Independent Driver Emotion Recognition System Involving Physiological Signals for ADAS

Mouhannad Ali, Fadi Al Machot, Ahmad Haj Mosa and Kyandoghere Kyamakya

Abstract Supporting drivers by Advanced Driver Assistance Systems (ADAS) significantly increases road safety. Driver's emotions recognition is a building block of advanced systems for monitoring the driver's comfort and driving ergonomics additionally to driver's fatigue and drowsiness forecasting. This paper presents an approach for driver emotions recognition involving a set of three physiological signals (Electrodermal Activity, Skin Temperature and the Electrocardiogram). Additionally, we propose a CNN (cellular neural network) based classifier to classify each signal into four emotional states. Moreover, the subject-independent classification results of all signals are fused using Dempster-Shafer evidence theory in order to obtain a more robust detection of the true emotional state. The new system is tested using the benchmarked MAHNOB HCI dataset and the results show a relatively high performance compared to existing competing algorithms from the recent relevant literature.

Keywords ADAS (advanced driver assistance systems) · Subject-independent emotion recognition · EDA (electrodermal activity) · ST (skin temperature) · ECG (electrocardiogram) · CNN (cellular neural networks)

M. Ali (✉) · F.A. Machot · A.H. Mosa · K. Kyamakya
Institute of Smart System Technologies, Alpen-Adria University Klagenfurt, Klagenfurt, Austria
e-mail: Mouhannad.ali@aau.at

F.A. Machot
e-mail: Fadi.almachot@aau.at

A.H. Mosa
e-mail: Ahmad.hajmosa@aau.at

K. Kyamakya
e-mail: Kyandoghere.kymakya@aau.at

T. Schulze et al. (eds.), *Advanced Microsystems for Automotive Applications 2016*, Lecture Notes in Mobility, DOI 10.1007/978-3-319-44766-7_11

1 Introduction

Advanced Driver Assistance Systems (ADAS) are systems that support drivers during the driving process. The goal of such systems is to increase both driving comfort and road safety. This can be realized by permanently monitoring both the internal state and the external environment of vehicles. The internal state might refer to the engine and other technical parameters, while the external environment refers to identifying and detecting dangers around the vehicle to avoid accidents and collisions. Additionally, ADAS systems consider also the driver's comfort by helping them to find their shortest path, avoid traffic jams and guide them to the best parking place and help to pay toll easily. Moreover, the involvement of emotion recognition into ADAS systems does enable (i.e. it is a building brick of related advanced systems) an online monitoring of both driver's comfort and driving ergonomics besides a forecasting capability of fatigue and drowsiness. This functionality has an evident positive impact on drivers and does contribute to proactively prevent accidents. For illustration, the significance of this idea comes from the observed fact that driving in aggressive or sad mood in urban areas does lead to traffic accidents [1]. Recently, different physiological signals have been widely used in advanced data processing intended for recognizing human's emotion. Thereby, the advanced signal acquisition systems are mainly involving non-intrusive sensors, which can record human physiological signals easily, are sufficiently comfortable for the human subject and might be accepted with respect to social and cultural differences.

Although physiological signals do contain complex patterns to be recognized, they are still more preferable to be used for emotion recognition purposes due to their higher robustness when compared to any other alternative concept, such as visual and audio signals. Such signals tend to be less effective in the ADAS context because of their higher vulnerability (e.g. facial expressions can be faked, and additionally, voice audio data are not always available as the driver is not always talking while driving) [2].

Several relevant studies have been described in the state-of-the-art literature on emotion recognition using speech and physiological signals. Thereby researchers mostly started first with subject-dependent approaches, where the emotion recognition system is applied to only one single user. Subject-dependent refers to the fact that both training and test data of a given classifier have been measured on the same human person. Such so trained classifier performs very weakly if it is tested using data collected on a different human user; thereby it needs either to be retrained or better to be appropriately recalibrated in order to reach an acceptable performance on a different user. Thus, subject-dependent classifiers are not good enough in view of practical requirements.

Therefore, now the focus is shifted more towards subject-independent approaches, where the emotion recognition system is tested with unknown data (speech and physiological signals), i.e. data collected from users, who are fully different from the ones used for the initial training of the classifier.

In the following we briefly comment a representative sample of the related works. For example, in [3], the authors have extracted and used features involving statistical, energy, sub-band spectrum and entropy measures from EMG (electromyography), ECG (electrocardiogram), EDA (electrodermal activity) and RSP (respiration change) signals, and then used the LDA (linear discriminant analysis) classifier to recognize 4 emotions (i.e.: joy, anger, sadness, and pleasure). They have reached 70 % accuracy in subject-independent validation.

A further work used as features the running mean, running standard deviation slope as features extracted from the physiological signals EMG, EDA, ECG, RSP and BVP (blood volume pulse), and involved in a classifier to detect the arousal and the valance states [4]. The maximum performance they reached was 96.58 % for subject-dependent validation.

Furthermore, different statistical features were extracted from BVP, EMG, EDA, RSP and ST (skin temperature) and used as inputs for SVM (support vector machine) and Fisher LDA classifiers to detect six different emotional states (i.e.: amusement, contentment, disgust, fear, sadness, and neutral) [5]. This method has reached 92 % performance accuracy for subject-dependent validation.

Moreover, other statistical features were used in a further work, namely the zero-crossing MFCCs (Mel-frequency cepstral coefficients) from speech and the following physiological signals: EMG, EDA, ECG, BVP, ST. Then, the KNN (k nearest neighbors) classifier has been used to detect arousal and valance states [6]. In this work the authors reached 92 % accuracy for subject-dependent and 55 % for subject-independent validation.

Additionally the authors in [1] have extracted statistical moments from MFCCs, the pitch and the energy from human speech in the car, and used them as features for a classifier to detect different emotional states, namely: fear, sadness, neutral, happiness and anger. They were successful in reaching 80.32 % accuracy for subject-independent validation.

In this paper, our aim is to consider different types of physiological signals to recognize four emotional states. The signals to be used are: the electrodermal activity signal for skin conductance, the skin temperature and the electrocardiogram, which measures the electrical activity of the heart muscle. The proposed system, which is subject-independent, shows promising results when compared to other state-of-the-art approaches in subject-independent classification contexts. In addition to that, an additional fusion of the individual classifiers involving respectively one single physiological signal does further significantly improve the overall classification performance. This fusion is performed using adaptive Dempster-Shafer approaches [7] to solve the issue related to uncertainty handling.

The rest of this paper is organized as follows: Sect. 2 provides a short explanation of all biological signals involved. Section 3 presents the overall system architecture with respect to the extracted features and the classification model used. Then Sect. 4 focuses on presenting and discussing the classification results obtained. Finally, Sect. 5 summarizes a series of concluding remarks along with a brief outlook at future related/subsequent works.

2 Physiological Signals Used

Different physiological signals can be collected or measured from human beings, which can give, after appropriate processing, information about the human (health, emotion, etc.). In this paper, we are considering the following physiological signals:

- **Electrodermal Activity (EDA)**: It refers to skin conductivity (SC) that basically measures the conductivity of the skin, whereby it is known that the conductivity increases if the skin is sweaty. EDA consists of a slowly changing part called Skin Conductance Level (SCL), which is overlaid by short, fast conductance changes (phasic components). The phasic components can be separated in two different types. The first one is the Skin Conductance Response (SCR), where the peak occurs in reaction to a stimulus. The second one is the Non-Specific Skin Conductance Response (NS.SCR), which is empirically very similar to SCR but with the main difference being that it occurs "spontaneously" without any stimulus [8].
- **Electrocardiogram (ECG)**: It refers to a test that measures the electrical activity of the heart. The ECG signals consist of three main waves. The first wave is known as the P wave, which corresponds to the depolarization of the atrium. The second wave is the QRS wave, which indicates the start of ventricular contractions. Finally, after the ventricles stay contracted for a few milliseconds, the third wave known as T wave appears, which occurs when the ventricular repolarizes [9].
- **Skin Temperature**: It is the measure of the peripheral skin temperature. The skin temperature depends on the blood flow in the underlying blood vessels. Since muscles are tense under strain, the blood vessels may be contracted due to a given emotional state and therefore the temperature will decrease [10].

3 Research Methodology

This section provides a comprehensive presentation of the physiological dataset used to train the robust patient-independent classifier developed in this paper. A new dataset has been recorded for the purpose of this classifier. Then, the feature extraction step which extracts and involves different features as indicated by the state-of-the-art methods is explained. Moreover, our proposed novel CNN based classification model is explained and demonstrated. At the end of this section, the high-level classification fusion of the different individual signal type dependent classifiers is explained.

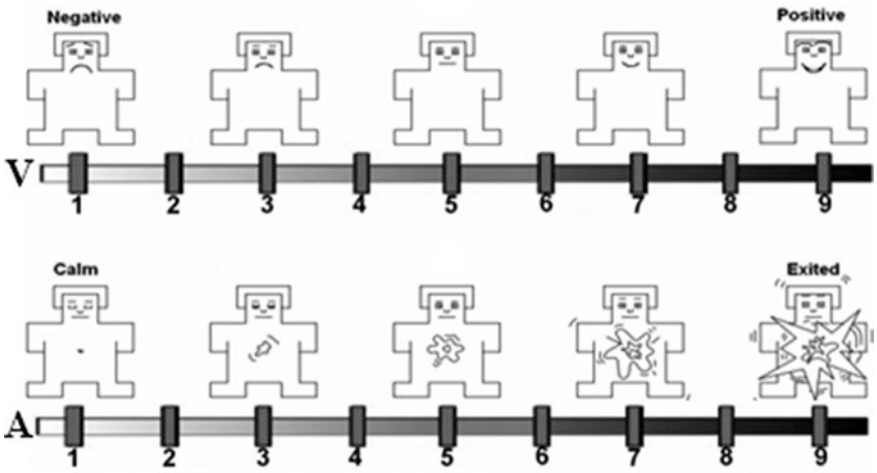


Fig. 1 Self-assessment manikins scales of valence (*higher part* of this figure) and arousal (*lower part* of this figure) [25]

3.1 Physiological Datasets Used

In this paper, a publicly available benchmark dataset for emotion analysis using physiological signals is used in the extensive tests of our new proposed emotion detection approach involving cellular neural networks. That dataset is called MAHNOB and has been collected by Soleymani et al. [11]. It includes different physiological signals from 30 young healthy adult participants, 17 were female and 13 were male; ages varied between 19 to 40 years. All signals were downsampled to 256 Hz and the trends of both ECG and GSR signals were removed. Each subject watched 20 emotional video clips and performed a self-assessment of his respective degree of both valence and arousal by using the Self-Assessment Manikins (SAM) questionnaire [12]. SAM is a distinguished questionnaire that visualizes the degree of valence and arousal dimensions by manikins. The subjects had to select one of the numbers from 1 to 9 which are written below the manikins, see Fig. 1.

In this work, we have mapped the scales (1–9) into 2 levels of each valence and arousal state according to the SAM ratings. The valence scale of 1–5 was mapped to "negative" and 6–9 to "positive", respectively. The arousal scale of 1–5 was mapped to "passive", and 6–9 to "active", respectively.

3.2 Feature Extraction

In this work, the most robust and well-known features (i.e. most commonly used one) have been selected from the state-of-the-art approaches in order to extract

features from EDA, ECG and ST signals. From each of the three physiological signals, we have extracted respectively 12 EDA features, 8 ECG features and 5 skin temperature features, with time-windows of 4 s length, from various analysis domains including time and frequency domains.

EDA Features: The EDA signal consists of the slow changing part (skin conductance level, SCL) and the phasic components (skin conductance responses, SCRs), which are the short-term conductance changes. Statistical measures obtained from the SCL analysis have been found to be well-correlated with emotion [13]. Different statistics are calculated here, especially the following ones: the mean, the standard deviation, the maximum, the minimum, the root mean square, the mean of the first derivation, the mean of the second derivation, and the mean of negative slope.

For SCR analysis, we calculate the rate of SCR occurrences from the very low frequency band (0–0.1 Hz), the response latency of the first significant SCR, the sum of SCR amplitudes, and the sum of SCR areas [13].

ECG Features: The statistical time-domain features are calculated directly from the ECG signals. The time domain measures include the following one: the mean, the standard deviation of the beat-to-beat interval (NN interval), the root mean square of differences of successive NN intervals, the number of successive differences that are greater than 50 ms, and the percentage of total intervals that successively differ by more than 50 ms.

For frequency domain features, we extract the average power of the low frequency range (0.04–0.15 Hz) and the high frequency band (0.15–0.4 Hz), as well as the ratio of power within the low frequency band to that within the high frequency band [14].

Skin Temperature Features: Standard statistical feature moments (e.g. mean, max, STD) are calculated from the skin temperature signal.

3.3 Classification Concept

In this chapter, we propose a CNN based classification; the CNN architecture adapted for this purpose in the one presented in Ref. [15]. A cellular neural network (CNN) is a network of adjacent coupled nonlinear cells. The interaction between the connected cells is modeled by a system of differential equations. CNN was first proposed by Chua and Yang (1988) [16]. CNN combines the strong features of both artificial neural networks (ANN) and cellular automata (CA). CNN does, however, overcome shortcomings of CA by its nonlinear dynamical relation between cells, and surpasses ANN by its local connectivity property. The ordinary state equation of a 3×3 neighborhood (or radius = 1) CNN cell is given by [16], see Eq. (1).

$$\frac{dx_{i,j}(t)}{dt} = -x_{i,j}(t) + \sum_{k=-1}^{1}\sum_{k=-1}^{1} a_{k,l}y_{i+k,j+1}(t) + \sum_{k=-1}^{1}\sum_{k=-1}^{1} b_{k,l}u_{i+k,j+1}(t) + I \tag{1}$$

where $x_{i,j}(t)$ and $u_{i,j}$ are respectively the current system state and the input of the cell (i,j). $A = a_{-1,-1}\ldots a_{+1,+1}$ is the feedback template, $B = b_{-1,-1}\ldots b_{+1,+1}$ the control template, I is the cell bias and $y_{i,j}(t)$ is the output nonlinear function of the state of cell (i,j), see Eq. (2).

$$y_{i,j}(t) = f\left(x_{i,j}\right) = \frac{1}{2}\left(\left|x_{i,j}+1\right| - \left|x_{i,j}-1\right|\right) \tag{2}$$

The nonlinearity behavior of the ordinary CNN can be realized by the relation between the states-outputs connection. However, the ordinary CNN cannot cope with highly nonlinear problems because of the following reasons:

(1) The simple nonlinear function used by Chua (between states and output) cannot fit highly nonlinear problems; and
(2) The relation between inputs and cell states is linear.

To overcome these two above identified drawbacks, we develop an extended CNN architecture to cope with strong nonlinearities. The novel model has the following features:

(1) A nonlinear relation between the inputs and the cell states, which is modeled by Gaussian Radial Basis Functions (GRBFs) [17].
(2) A nonlinear relation between cell states and cell states relation, which is also modeled by GRBFs.
(3) The Chua nonlinear function in Eq. (2) is replaced by a hyperbolic tangent sigmoid transfer function [18].

Accordingly, our proposed extended CNN model is given by Eq. (3) (see also Fig. 2):

$$\dot{X} = A * X + B * Y + C * U + D * GRBF1(U) + E * GRBF2(comb(U,U)) + F * GRBF2(comb(X,U)) + G * GRBF1(X) + H * GRBF2(comb(X,X)) + I \tag{3}$$

$$y_i = \tan\mathrm{sig}(x_i) = \frac{2}{1+e^{-2x_i}} - 1 \tag{4}$$

whereby: $X = (x_1, x_2, \ldots x_n)^T$, $Y = (y_1, y_2, \ldots y_n)^T$, $U = (u_1, u_2, \ldots u_m)^T$ are the state, output and input vectors respectively. Furthermore, the following combinations:

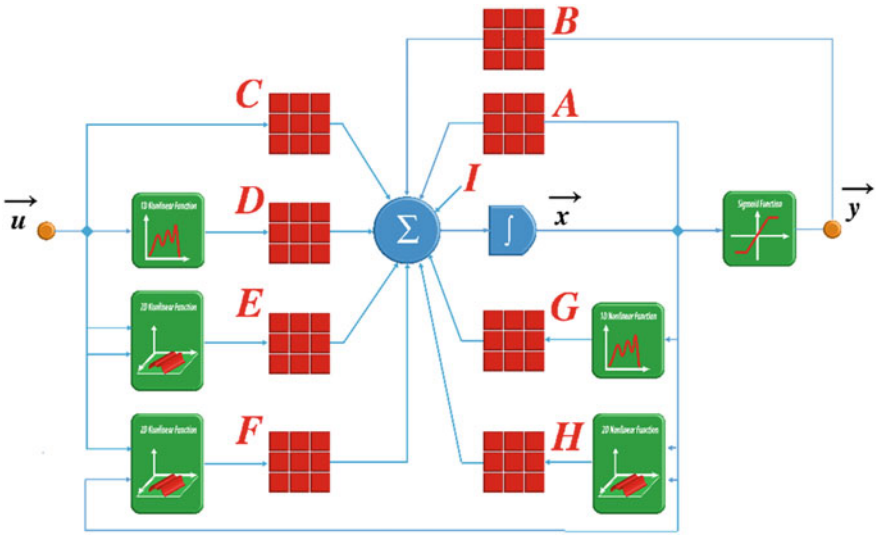


Fig. 2 Our proposed cellular novel neural network architecture, which is used for classification problems

$comb(X,X) = ((x_1,x_1),(x_1,x_2)\ldots(x_n,x_n))$, $comb(X,U) = ((x_1,u_1),(x_1,u_2)\ldots(x_n,u_m))$, $comb(U,U) = ((u_1,u_1),(u_1,u_2)\ldots(u_m,u_m))$ are combinations of 2D vectors.

The terms GRBF1, GRBF2 are respectively one- and two-dimensional Gaussian Radial Basis Functions [17]. And n, m are the number of states/output and inputs respectively. Finally, the terms

$$A = \begin{pmatrix} a_{1,1} & . & . \\ . & . & . \\ . & . & a_{n,n} \end{pmatrix}, \quad B = \begin{pmatrix} b_{1,1} & . & . \\ . & . & . \\ . & . & b_{n,n} \end{pmatrix}, \quad C = \begin{pmatrix} c_{1,1} & . & . \\ . & . & . \\ . & . & c_{n,m} \end{pmatrix},$$

$$D = \begin{pmatrix} d_{1,1} & . & . \\ . & . & . \\ . & . & d_{n,m} \end{pmatrix}, \quad E = \begin{pmatrix} e_{1,1} & . & . \\ . & . & . \\ . & . & e_{n,m^2} \end{pmatrix}, \quad F = \begin{pmatrix} f_{1,1} & . & . \\ . & . & . \\ . & . & f_{n,n*m} \end{pmatrix},$$

$$G = \begin{pmatrix} g_{1,1} & . & . \\ . & . & . \\ . & . & g_{n,n} \end{pmatrix}, \quad H = \begin{pmatrix} h_{1,1} & . & . \\ . & . & . \\ . & . & h_{n,n^2} \end{pmatrix}, \quad I = \begin{pmatrix} i_1 \\ . \\ i_n \end{pmatrix}$$

are the CNN templates to be designed or estimated by considering the training data. Figure 2 illustrates our proposed CNN architecture, in which, the templates A, B, C…, and I are represented in red. The input u feeds the state x through multiple connections: a linear connection, one-dimensional nonlinear RBFs (a function for each of the input entities $u_1, u_2, \ldots, u_m$), two-dimensional radial-basis-function

[RBFs, a function for each of the input entities combinations (u_1, u_1), $(u_1, u_2)\ldots$ (u_m, u_m)], and a further two-dimensional RBFs [a function for each input-state entities combination (x_1, u_1), $(x_1, u_2)\ldots$ (x_n, u_m)]. The state x feeds back itself through a linear connection, one-dimensional nonlinear RBFs [a function for each of the state entities $(x_1, x_2, \ldots, x_n)$], two-dimensional RBFs [a function for each of the state entities combinations (x_1, x_1), $(x_1, x_2)\ldots$ (x_n, x_n)]. And the state x feeds the output y through a sigmoid function; and the output y feeds back the state x through a linear connection. The state equation is a differential one, therefore an integrator is placed before the state x.

The design of our proposed CNN system consists of two main phases:

(1) An unsupervised initial phase, in which the Gaussian Radial Basis (RBF) function centers and widths are estimated. The estimation of the centers is done using K-Means clustering as in [19]. Then the widths are calculated using the heuristic process proposed by [20].
(2) A supervised learning phase where the CNN templates $A, B, C\ldots I$ are estimated. In this phase, we have an optimization problem to deal with. To cope with that, we use the Particle Swarm Optimization (PSO) [21]. In a future work, the reservoir computing and echo-state paradigms will be used to provide a novel, extended and more efficient CNN architecture concept with an easier supervised training involving just least square estimations.

In order to apply the proposed CNN classification concept, the following configuration steps have to be done:

- The number of states and outputs n is equal to the number of target classes (n = 4, four emotional states).
- The number of inputs m is equal to the number features in the features vectors.
- Each Gaussian Radial Basis Function consists of five centers.

3.4 Fusion Concept

Dempster-Shafer theory can effectively represent and process uncertainty and imprecision information. It has been widely used in the field of information fusion [22, 23, 24]. However, in multi-modal sensor networks there are often conflicting sensor reports due to the interference of the natural environment or other reasons. It has been proven [7] that the classical Dempster-Shafer evidence theory cannot deal with the integration of conflicting information effectively. In case of conflicts between the evidences, if one uses Dempster's combination rule to integrate evidences directly, the result does not reflect reality. Therefore, many other improved methods to combine evidences have been proposed. E.g., Ali et al. [2] proposed a combination method by complementing the multiplicative strategy by an adding

strategy; this method shows promising results for evidence combinations compared to other existing approaches.

We use the combination rule proposed by Tazid et al. [7] as it provides more realistic results than the standard Dempster-Shafer rule, when combining conflicting evidence from multiple sources. Equation (5) shows how to calculate the combined probability assignment function:

$$(\boldsymbol{m}_{s1} \oplus \boldsymbol{m}_{s2})(\boldsymbol{e}) = \frac{1 - (1 - (\boldsymbol{m}_{s1}(\boldsymbol{e})) * (1 - (\boldsymbol{m}_{s2}(\boldsymbol{e}))}{1 + (1 - (\boldsymbol{m}_{s1}(\boldsymbol{e})) * (1 - (\boldsymbol{m}_{s2}(\boldsymbol{e}))} \tag{5}$$

where $\boldsymbol{m}$ is the mass function, $\boldsymbol{s}$ is the system and $\boldsymbol{e}$ is the evidence which is calculated based on the performance of classifying each signal type.

4 Experimental Results

Due to preprocessing time constraints, only 10 out of 30 subjects from the MAHNOB dataset have been considered in our experiments. The data records for each individual subject had a duration between 120 and 166 s. The data of each subject were clipped into 4 s time-windows, whereby each window should be classified into one labeled emotion. Splitting the data into "4 s clips" is fixed empirically due to the following reasons: (a) A features' extraction from longer time-windows from the MAHNOB dataset leads to a smaller number of data samples to be used for training (machine learning) the classifiers; and (b) increasing the length of the time-windows does increase the processing time.

In this paper, we classify the data into 4 main emotional states: two valence states and two arousal states (see Fig. 3).

The emotion classes considered in this paper are:

- **High Valence/High Arousal (HVHA)**: This class includes positive emotions such as happy and excited.
- **High Valence/Low Arousal (HVLA)**: This class includes emotions such as **relaxed**, calm and pleased.
- **Low Valence/High Arousal (LVHA)**: This class includes emotions such as angry, fear and distressed.
- **Low Valence/Low Arousal (LVLA)**: This class includes negative emotions such as sad and depressed.

The data of each physiological sensor are divided into 70 % for training and 30 % for testing. Thereby, a subject-independent approach is followed, i.e., testing data originated from subjects different from the ones involved in the training data.

The training and test sets are split from different subjects to insure the independence between both sets as we target a subject-independent classification system. Moreover, the features of each physiological signal are classified using a local classifier for each physiological signal. In order to evaluate the performance of our

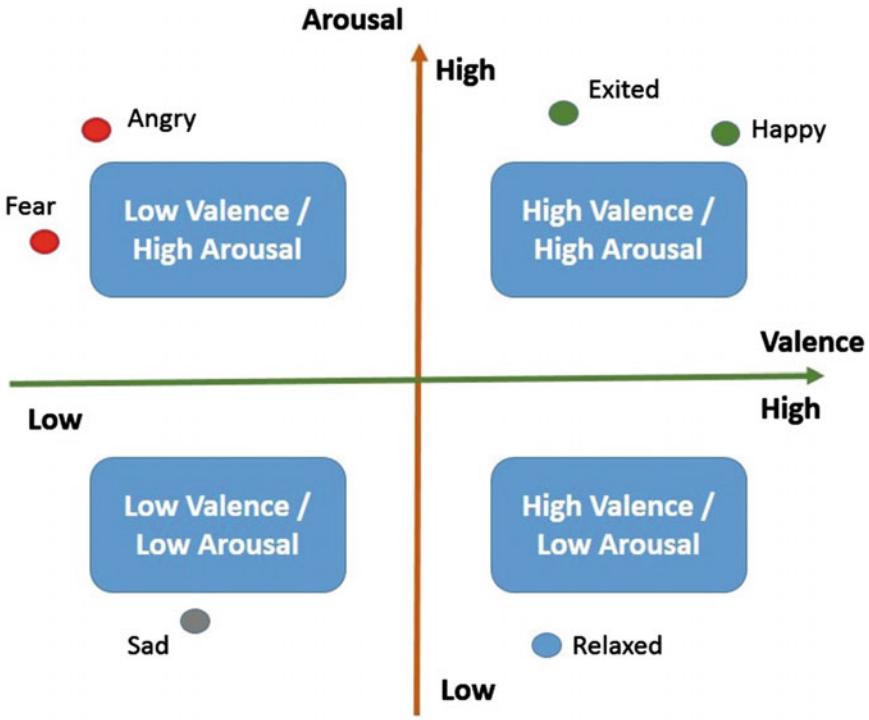


Fig. 3 Arousal-valence based emotional states

novel CNN based classifier, two other classifiers KNN (K-nearest Neighbors) and SVM (Support Vector Machine) are included for benchmarking purposes in order to compare the performance of three different classification concepts.

Table 1 shows the classification accuracy results obtained for each signal type for all 3 classification methods K-nearest Neighbor, Support Vector Machines, and our proposed CNN based classification). The best performance is reached by our proposed CNN classifier, which exhibits a subject-independent accuracy value of 92.42, 53.8, and 72.44 % for EDA, ECG and Skin Temperature, respectively. It is observed that EDA signal provides the best classification results. This is good news because EDA is much more comfortable for practical applications than ECG. For our target demonstration device it is intended to combine EDA and skin

Table 1 Classification accuracy in percentage as reached by the three different classifiers (subject-independent emotion detection)

Classifier	ECG	EDA	Skin temp
KNN	51.2	80.3	67.4
SVMs	32.2	81.5	55.3
CNN	52.98	92.42	72.44

Table 2 Performance measures reached by our proposed CNN classifier (subject-independent emotion detection)

Measure	ECG	EDA	Skin temp
Specificity	75.62	96.78	89.78
Precision	52.43	91.96	74.69
Recall	53.47	91.63	75.05

Table 3 Performance measures reached after fusion of the local CNN based classifier (subject-independent emotion detection; Fusion realized by the method of [7])

Before \after fusion	Accuracy	Specificity	Precision	Recall
Before fusion	72.78	76.34	71.72	87.44
After fusion	82.35	94.37	87.88	93.55

temperature only, both being delivered by a single device carried as a watch on the hand. Also the results underline the clear superiority of the CNN based classifier when compared to KNN and SVM.

In order to illustrate the performance of our proposed CNN classifier more profoundly, Table 2 lists different performance measures (Specificity, Precision and Recall) of each signal type.

Further and to finish, Table 3 shows the performance results obtained after fusion of the considered local CNN based classifiers. In this first phase, we have used the combination rule of [7] to fuse the local classifiers and reach better emotions recognition rate. As we can see the overall performance is improved after fusion.

5 Conclusion

This paper has comprehensively presented a novel CNN based emotion detection concept, which includes a fusion of local classifiers involving three different physiological data types: EDA, ECG, and Skin temperature. This subject-independent concept performs better than comparable approaches from the relevant literature.

Because of the limitations of the standard Dempster-Shafer method, an extended version suggested in [7] has been used for the local classifier fusion process.

As outlook we can briefly describe the very next research steps, which are sub-sequent to the work presented in this paper:

(a) An improved version of the CNN architecture will be developed according to the novel “reservoir computing” and “echo-state” paradigms;

(b) A "radial basis function CNN" based local classifier fusion will be developed and validated—thereby CNN based concepts will be involved for both the "local classifiers" and for the "global classification fusion", for which case a better performance is expected; and
(c) An online real-time working prototype will be implemented and stress-tested, while involving diverse real test persons. The final intention being to realize a validated and stress-tested subject-independent online real-time system with minimum emotion detection accuracy (4 emotions) of 90 %, while involving the two physiological signals EDA and skin temperature.

Driver's emotions recognition is a building brick of advanced intelligent systems for monitoring driver's comfort and driving ergonomics, this additionally to driver's fatigue and drowsiness forecasting. Therefore, information related to driver's emotion and stress is a key enabling component for future driver assistance systems, which will optimize the synergy between human driver and intelligent vehicle. Since in industrial nations most accidents (more than 90 %) are due to human related mistakes/failures according to most recent statistics from the road safety literature, a close online monitoring of the driver's state (emotion being a significant component of it) appears to be of crucial importance in the perspective of significantly improving road safety.

References

1. Al Machot F, Haj Mosa A, Dabbour K, Alireza F, Schwarzlmuller C, Ali M, Kyandoghere K (2011) A novel real-time emotion detection system from audio streams based on bayesian quadratic discriminate classifier for adas. Nonlinear dynamics and synchronization (INDS) and 16th international symposium on theoretical electrical engineering (ISTET) IEEE, 1–5, 2011
2. Ali M Haj Mosa A, Al Machot F, Kyamakya K (2016) A novel real-time EEG-based emotion recognition system for advanced driver assistance systems (ADAS), INDS'15—4th international workshop on nonlinear dynamics and synchronization, pp 9–13
3. Jonghwa K, André E (2008) Emotion recognition based on physiological changes in music listening. Pattern Anal Mach Intell IEEE Trans 30:2067–2083
4. Lætitia LC, Nasoz F (2004) Using noninvasive wearable computers to recognize human emotions from physiological signals. EURASIP J Adv Signal Process 11:1–16
5. Choubeila M, Pruski A (2010) Emotion recognition through physiological signals for human-machine communication. INTECH Open Access Publisher
6. Jonghwa K (2007) Bimodal emotion recognition using speech and physiological changes. INTECH Open Access Publisher
7. Tazid A, Dutta P, Boruah H (2012) A new combination rule for conflict problem of Dempster-Shafer evidence theory. Int J Energy Inf Commun 3(1):35–40
8. Cowie R, Ellen D, Tsapatsoulis N, Votsis G, Kollias S, Fellenz W, Taylor J (2001) Emotion recognition in human-computer interaction. Sig Process Mag IEEE 18(1):32–80
9. Silverthorn DU, Ober WC, Garrison CW, Silverthorn AC (2009) Human physiology: an integrated approach. Pearson/Benjamin Cummings, San Francisco, CA, USA
10. Andreas H, Goronzy S, Schaich P, Williams J (2004) Emotion recognition using bio-sensors: first steps towards an automatic system, ADS, pp 36–48

11. Soleymani M, Lichtenauer J, Pun T, Pantic M (2012) A multimodal database for affect recognition and implicit tagging. Affect Comput IEEE Trans 3(1):42–55
12. Bradley MM, Lang PJ (1994) Measuring emotion: the self-assessment manikin and the semantic differential. J Behav Ther Exp Psychiatry 25(1):49–59
13. Mathias B, Kaernbach C (2010) Decomposition of skin conductance data by means of nonnegative deconvolution. Psychophysiology 47(4):647–658
14. Tarvainen MP, Niskanen J, Ranta-Aho PO, Karjalainen P (2014) Kubios HRV–heart rate variability analysis software. Comput Methods Programs Biomed 113(1):210–220
15. Haj Mosa A, Kyamakya K, Mouhannad A, Al Machot F, Chedjou J (2015) Neurocomputing-based matrix inversion involving cellular neural networks: black-box training concept, autonomous systems 2015, Fortschritt-Berichte VDI Reihe 10, pp 119–132
16. Chua LO, Yang L (1988) Cellular neural networks: applications. IEEE Trans Circ Syst 35 (10):1273–1290
17. Paul Y, Haykin S (1995) A dynamic regularized Gaussian radial basis function network for nonlinear nonstationary time series prediction. Acoust Speech Signal Process IEEE 5:95
18. Vogl T, Mangis J, Rigler A, Zink W, Alkon D (1988) Accelerating the convergence of the back-propagation method. Biol Cybern 59(4–5):257–263
19. Stuart PL (1982) Least squares quantization in PCM. Inf Theory IEEE Trans 28(2):129–137
20. Avijit S, Keeler J (1990) Algorithms for better representation and faster learning in radial basis function networks. In: Advances in neural information processing systems, pp 482–489
21. James K (2011) Particle swarm optimization. In: Encyclopedia of machine learning. Springer, USA, pp 760–766
22. Andino M, Hasan M (2013) The Dempster-Shafer theory algorithm and its application to insect diseases detection. Int J Adv Sci Technol 50(1):1
23. Faouzi S, Chibani A, Amirat Y, Benhammadi F, Mokhtari A (2012) An evidential fusion approach for activity recognition under uncertainty in ambient intelligence environments. In: Proceedings of the 2012 ACM conference on ubiquitous computing, pp 834–840
24. Pashaa E, Mostafaeib H, Khalaj M, Khalaj F (2012) Fault diagnosis of engine using information fusion based on Dempster-Shafer theory. J Basic Appl Sci Res 2:1078–1085
25. Sepideh H, Maghooli K, Nasrabadi AM (2014) The emotion recognition system based on autoregressive model and sequential forward feature selection of electroencephalogram signals. J Med Signals Sensors 4(3):194

Part III
Safety and Methodological Challenges of Automated Driving

Highly Automated Driving—Disruptive Elements and Consequences

Roland Galbas

Abstract The paper aims to give a compact overview of the changes towards the complete system "vehicle" imposed by the introduction of high-level-automation. At first, the changes are focused on three disruptive elements. Each of these elements is analyzed with respect to its requirements which are caused. Finally four main systemic changes are derived from the requirements. The idea of this breakdown is to propose a path aiming to structure the future developments in the area of highly automated driving (HAD). Highlighted are cross-over effects of requirements, the impact of framework conditions and derived consequences mainly from a technological point of view. Within the conclusions, some approaches are proposed in order to cope with these challenges—especially with focus on market deployment.

Keywords Highly automated driving · Verification and validation

1 Disruptive Elements

In order to structure the current field of framework conditions for highly automated driving, three disruptive elements are proposed:

1. "The physical change"—the role of the driver to actuate—"Driver is no more fail-safe backup".
2. "The change of responsibility"—the responsibility for the driving task is transferred from driver to vehicle. The role as percipient, observer and decider turns to the vehicle. The means that also changes, uncertainties and given risks of the open world traffic that in the past were handled by the driver now have to be handled by the vehicle—"The driver is no more responsible".

R. Galbas (✉)
Chassis Systems Control, CC/ENA, Robert Bosch GmbH,
Robert Bosch Allee 1, 74232 Abstatt, Germany
e-mail: roland.galbas@de.bosch.com

T. Schulze et al. (eds.), *Advanced Microsystems for Automotive Applications 2016*, Lecture Notes in Mobility, DOI 10.1007/978-3-319-44766-7_12

Fig. 1 Disruptive elements

3. “The change of the vehicle getting part of a mobile network (Data-driven mobility ecosystem)”—external mobility data will be part of the vehicle functions. The shorter development cycles of the IT-industry will influence today’s longer automotive development cycles as the “vehicles will be part of the digital world” (Fig. 1)

In the following part each disruptive element will be considered in order to analysis its requirements and based on this the derived changes concerning the complete system.

1.1 The Physical Change

Currently all scenarios for possible failures caused by the vehicle are using the driver as an observer and physical backup. In case the driver does not conduct the driving task this assumption is not valid anymore. Depending on the level of automation the driver cannot take over the physical driving task at all or at least not fast enough.

- **For common vehicles**—not highly automated systems—the driver can take over the driving task in case of a system failure. In this case within a safety analysis it is typically sufficient to qualify the vehicle or a subsystem as “fail safe”. This is possible, because in case of a malfunction of the vehicle the driver has the task to bring the vehicle into a minimum risk state. Example: In case of a malfunction of the brake booster, the driver has to enforce the missing braking power and bring the vehicle to a minimum risk state as fast as possible.
- **For future highly automated systems** the driver can’t be considered as a mechanical backup. In this case the safety analysis requests the system to be “fail operational”. Thus, in case of a possible malfunction of the vehicle the vehicle has to bring itself into a minimum risk state.

As a consequence of this “fail operational” requirement the vehicle has to cover a remaining driving task to ensure a minimum risk scenario for certain possible malfunctions as e.g. a single point failure of a system. Example: In case of a failure within the steering system, it is commonly not estimated that the stability system (ESP) could take over the steering task for all relevant driving situations (this case would be considered as functional redundancy). Thus, the steering system itself has to cope for own single point failures e.g. by internal redundancy of the steering system.

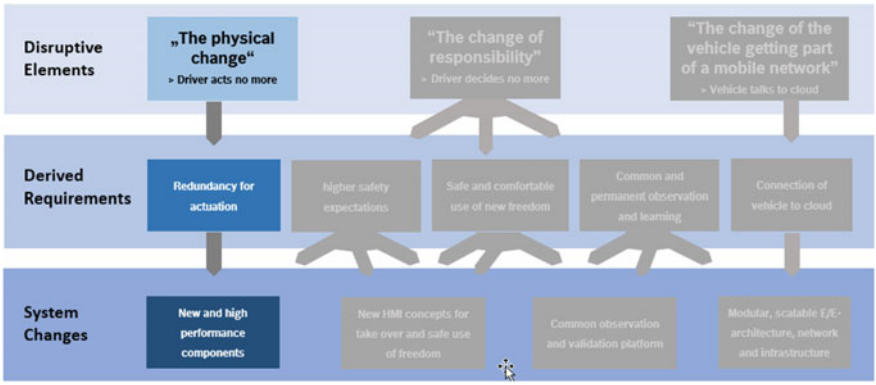


Fig. 2 Brake-down of disruptive elements down to system changes

This redundancy for actuation has to apply to the related subsystems for Braking, Steering and other subsystems as power net, data transfer, etc.

All solutions applying redundancy to the systems and subsystems highly depend on the level of automation. All solutions will be carefully selected between concepts for functional redundancy and internal redundancy. The redundancy distribution of the power net and the data system strongly depends on the concepts for braking and steering systems.

The given necessity for reengineering leads to a considerable complexity of future vehicular systems—especially because every kind of redundancy or even only changes causes additional costs.

– The main consequence of system changes according to “The physical change” is “Redundancy for actuation” and therefore new components (Fig. 2).

1.2 The Change of Responsibility

There are three main requirements driven by the expected change of responsibility:

- 1.2.1 Higher safety expectations.
- 1.2.2 Safe and comfortable use of new freedom.
- 1.2.3 Common and permanent observation and learning.

1.2.1 Higher Safety Expectations

Current public safety expectations relate to a mean value of observed accidents, though there is surely a wide spread over countries and other relevant aspects.

The risk currently taken by drivers is an individual risk which results from a decision between personal comfort and personal accident risk. The current risk is also temporary and depending on many driver-related factors as emotions, health-state, etc. and also on external factors as weather conditions, etc. Also many factors are often not realized by the driver.

The public expectation towards an artificial intelligence algorithm and related technology (e.g. perception) is to avoid such risks. The simple reason behind this, is that people believe that algorithms can permanently minimize risk. The algorithm has always to avoid also very improbable cases (e.g. absolute hidden children spontaneously running fast out of parking vehicles and not even a smallest indication has been given before).

So public expectation is that algorithms are able to do so and if not yet, that technology (also e.g. perception technology) has to be developed further before we let it enter the market.

Another aspect coming up with highly automated driving is that the responsibility for accident occurrence is (mainly) transferred to the manufacturer (OEM). So the OEM has to ensure to follow the public safety expectation because the OEM carries the full liability risk. Every indicated problem relates always to a fleet of vehicles and thus to an aggregated risk for the OEM. So it will be necessary to document the internal system states to provide evidences in case of accidents (data recording).

Furthermore the driver as the responsible person for the safe status of the vehicle and for any kind of change which happens in the open world is taken out of the loop. The world is changing permanently, e.g. by the introduction of new traffic rules, this aspect being part of the "open world problem". All legal and environmental changes as also certain technical checks have then to be accounted for by the vehicle.

In consequence the system needs a replacement which may similarly work as a typical quality assurance system, building up an independent instance within the vehicle itself. Such a system could be considered as a self-observing system which also is in charge to perform data recording.

In order to cope with the higher safety expectations the following subsystems are affected:

1. Newly designed surround sensing systems including localization and networking systems with a significant high performance level (redundancy included).
2. Data processing and fusion systems have to provide highest performance (also with included redundancy) while the argumentation for redundancy follows the same concept as for actuation redundancy. In order to enable the performance of cognition and perception system a change of E/E-architecture towards high performance is also necessary.
3. Self-observing systems as well as data recording systems provision have to be established for increased safety and preservation of evidence in case of accidents.

Public safety expectations increase compared to the current state of the art. Thus the safety of functions has to perform significantly better than current drivers perform especially in order to maintain current driving comfort as e.g. speed. The disappearance of the responsibility of the driver also enforces systemic answers.

– The main consequences of system changes according to "Higher safety expectations" are adapted and high performance components and systems.

1.2.2 Safe and Comfortable Use of New Freedom

One main driving issue for highly automated driving is the use of the won freedom for entertainment, productivity, relaxation, health and other purposes. Driver and passengers expect to get offered a wide range of opportunities similar to what they can perceive in other known conditions.

For individually unavoidable accidents the already established "passive safety systems" will be necessary on long time scales (minimum for the duration of mixed traffic). The technology to mitigate the severity of injuries today consists mainly in crash structures (chassis), belt and airbag systems. Belt and airbag systems only work efficiently if the passengers keep their bodies and extremities within a certain area of their seat. Any displacement of the belt or significant move of the passenger out of a certain area (so-called "out of position" case) decreases the nominal safety margin significantly.

In order to cope with the expectations of the passengers to use the freedom and also keep or even increase the safety margin for individually unavoidable accidents it is necessary to develop new HMI (Human Machine Interface) components. Foreseeable are

- Combined safety systems using passenger monitoring and new restraint systems providing the expected benefit (e.g. new belt/seat combinations).
- Innovative communication systems like windshield-screens, new haptic HMI concepts or combined communication-control technologies like audio and gesture.
- New control concepts for steering and braking providing a safe responsibility transfer and enabling more free space within the wagon.

– The main consequences on system changes according to "Safe and comfortable use of new freedom" are "New HMI concepts" for takeover of control and safe use of freedom.

1.2.3 Common and Permanent Observation and Learning

There is no perfect system, especially not in the beginning. Looking at the extremely complex environment of traffic situations, it is to be expected that after the homologation process of a highly automated vehicle new unpredictable, but risky situations may occur because of a certain environmental change. Uncertainties and risks in field have to improve the performance of all vehicles. This implies that not only one vehicle learns for itself, but it aggregates and shares data. The data of many vehicles will be analyzed (e.g. in case of near misses), and based upon this new knowledge will be created and redistributed to adaptable vehicles. This process could be called "Community learning".

Related to "Community learning" is the establishment of a "Permanent safety procedure" as opposed to today's one-time homologation process. The reason behind this lies in the fact that it has to be ensured that safety-relevant changes are implemented within a safety-proven process (see also Sect. 1.2.1.).

Also related to "Community learning" is the establishment of a "Permanent observation system" within the vehicle which provides e.g. the safety relevant vehicle data for "Community learning" and enables to cope with the safety requirements for an independent instance (see also Sect. 1.2.1) and last but not least to increase the safety performance of the vehicle (e.g. by health monitoring).

Taking also into account that the amount of processible information is growing exponentially it can be derived that

- future mobile systems will be adaptable for changes in order to incorporate new knowledge and changes as e.g. safety updates, and that
- the installation of an in-car observation system is highly probable supporting the community learning and ensuring safe application of updates (permanent safety updates).

Assuming that these driving issues lead to the described system changes then each single OEM has to carry the complete liability risk for content and process of each release/upgrade. Therefore, for the sake of efficiency it is desirable that multiple OEMs and Tiers establish a common standardized safety process (see proposal by SafeTRANS [1].

A probable system in charge for "Common observation and validation" could consist on three main elements, (1) a central platform for data aggregation and analysis as mentioned in Fig. 3, (2) a network layer and (3) an independent instance at each vehicle. This instance could be represented by the already mentioned vehicle-internal observation system which carries similar requirements.

– The main consequences of system changes according to "Common and permanent observation and learning" are standardized interfaces as well as a common observation and validation platform and a procedure for permanent field-observation (Fig. 4).

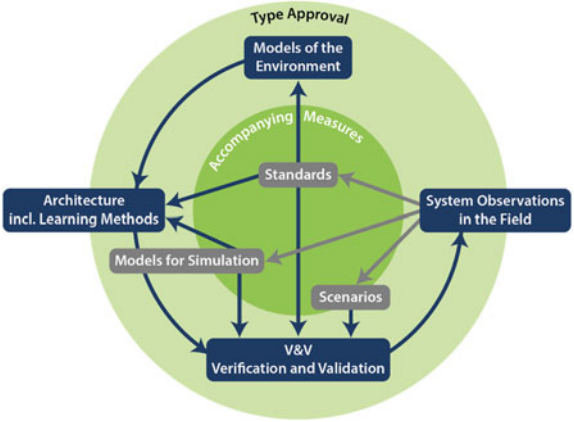


Fig. 3 Common observation and validation platform, key elements of a system of continuous supervision and learning from field observations for highly automated systems [1]

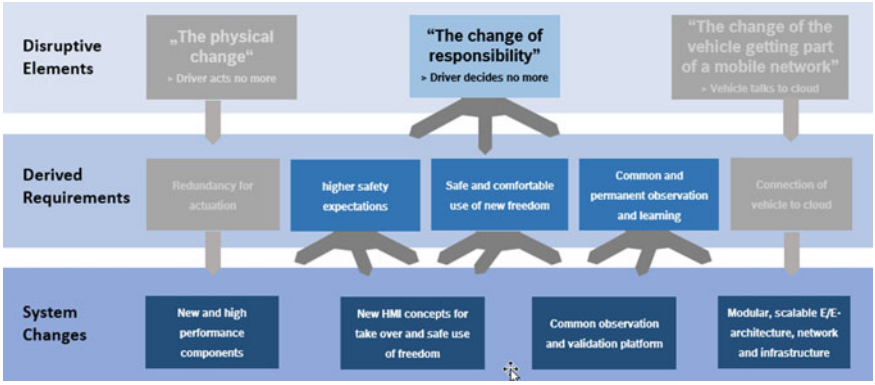


Fig. 4 Breakdown of disruptive elements into system requirements for "The change of responsibility"

1.3 The Change of the Vehicle Getting Part of a Mobile Network (Data-Driven Mobility Ecosystem)

In order to cope with the request for "community learning" and for a "permanent safety procedure" a network unit is surely necessary (see also Sect. 1.2.3). Furthermore the rapid technology change of the digital and telecommunication world requests an adaptable system (e.g. safe flash over the air, scalable HW e.g. network unit—see also Sect. 1.2).

External data will be used for safety relevant vehicle functions. Drivers for external data use are the following aspects:

1. After a certain point of perception performance it is cheaper to use external data than more or improved surround sensing components. The range of this point highly depends on the function and its conditions. So the driving issue for use of external data is the cost—performance ratio (e.g. safe information about a construction zone on a highway could increase the medium speed for this highway function) (Fig. 5).
2. The current accident risk taken by the median of drivers is higher than the risk which society is accepting to be taken by the automated vehicle (see also Sect. 1.2.1 higher safety expectations). This circumstance becomes important for e.g. urban areas where the possible driving speed could be strongly reduced in order to cope with the increased public safety expectation (imagine driving in a crowded area with side-parking cars). One of the possible measures to increase safety and/or keep a comfortable speed is the use of external data (e.g. for localization of pedestrians).
3. For certain vehicle functions external data are the only solution to improve performance. The reason is typically occlusion of objects. Some traffic situations as overtaking with oncoming traffic or e.g. complex urban traffic situations like crossings can be approached in a safe, agile and comfortable way with external data (C2X) (see also Sect. 1.2.1 Higher safety expectations).

The consequences of these driving issues on the vehicle level are:

- The vehicle will require a connectivity unit, probably containing multimodal ability (see also Sect. 1.2.3 “common observation platform”).
- The need for connectivity drives another important issue: The technology providing connectivity and backend services has faster innovation cycles than the automotive industry. Furthermore also the subsystem as e.g. braking, steering and power net will impose an independent need for data exchange e.g. for maintenance reasons or in order to reduce complexity of the complete vehicle system. Also the different subsystems will be developed following their own

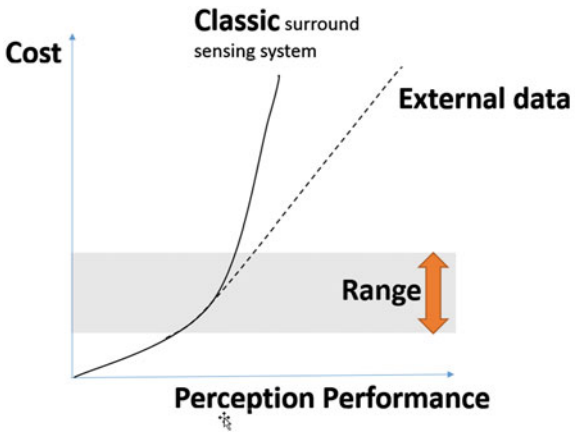


Fig. 5 This graph should present that for certain vehicle functions the use of external data is more effective than the improvement of the classic surround sensing system

independent cycles. Their need for information exchange towards external data and also internal data will develop in a different way. These reasons enforce the E/E-architecture to develop into a service–oriented structure. One approach to cope with different innovation cycles is to develop an appropriate modularity within the E/E-architecture building up vehicle internal and external harmonized service interfaces.

– Thus, to summarize the main consequences of "The change of the vehicle getting part of a mobile network" on system changes are a modular and scalable E/E-architecture with a service structure and a connectivity unit (Fig. 6).

Finally all systemic changes could be rooted to the 3 disruptive elements (Fig. 7)

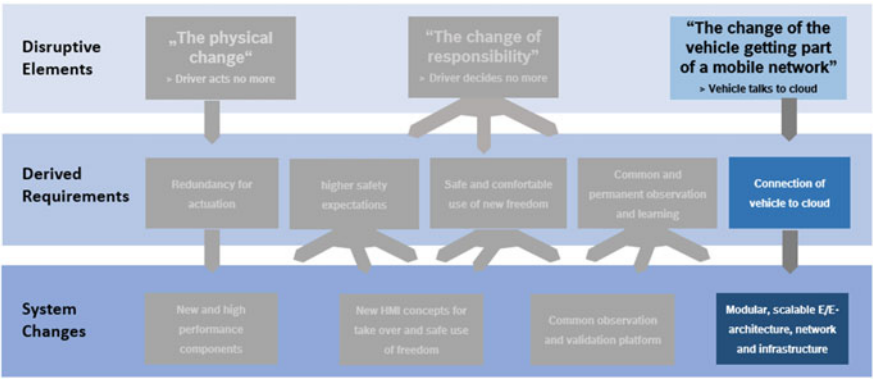


Fig. 6 Breakdown of disruptive elements into system requirements for "The change of responsibility"

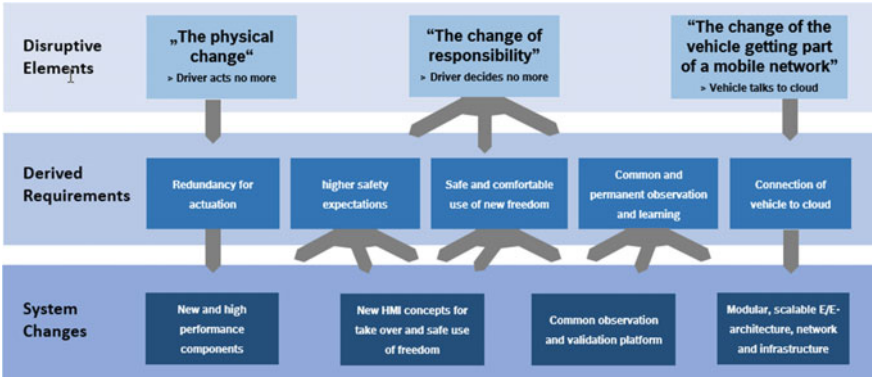


Fig. 7 Complete breakdown of disruptive elements into system requirements

2 Conclusions

It has been shown that there are probable systemic changes with a strong impact on the whole vehicular system. One of the most complex and challenging changes is the expansion of the system towards network and backend layer—shown in Fig. 8. The consequences of this expansion are fare-ranging:

- Regulations and methods have to be harmonized between automotive industry and digital and telecommunication industry. One issue which is already mentioned is the harmonization of the innovation cycles of these industries.
- Underlying systems as e.g. the enhanced integrated safety system now belong also to the network and backend layer with all related constrains and implications that a safety system requires

Costs Another simple but important conclusion is, that the systemic changes itself lead to significant additional costs of the system.

Looking at each derived systemic change only "new HMI concepts" are really visible to the buyer. All the other changes are neither easily explainable nor sellable. Thus the reduction of cost, weight and complexity of these systems is absolutely needed to reach market attractiveness.

Analyzing the 4 systemic changes it is obvious that any of them is highly determined by safety requirements. Thus the assessment of safety is highly cost sensitive, any wrong safety assessment could lead to high investment risks—especially in case of an assessment by only a single stakeholder.

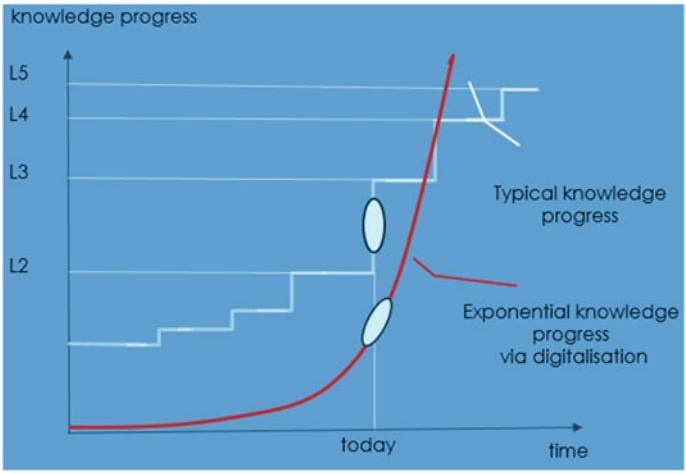


Fig. 8 Schematic picture of the increase of knowledge for typically development processes compared with the knowledge progress using digitalization

Conclusion: It is effective to establish a common safety assessment involving all relevant stakeholders. This also would mitigate the liability risk and accelerate the harmonization process—thus a common activity enables early market deployment.

Time to market Besides the mentioned challenge above, the aspect "time to market" is very important for highly automated driving. Success or non-success of an enterprise or even a whole business region depend on fast market deployment. The reason behind is that with an early market deployment de facto standards are established.

Complexity It can be stated that highly automated driving is a complex field of innovation—especially because of the strong coupling of the "digital-world" and the "safety-world".

Digitalization On one hand "digital" companies force the classic stakeholder to hurry up. On the other hand—digitalization could be considered as a key to cope with these challenges. One example is the significant development of simulation for verification and validation purposes in terms of high depth modelling or the use of big data analysis methods as deep learning for acknowledge gathering in the field of risk analysis. Thus digital instruments could allow an exponential progress of knowledge as compared to typical development cycles. The graph in Fig. 8 intends to present a schematic comparison of the classic path of knowledge progress and a path with significant use of digitalization (where L1–L5 are used as the SAE-levels of automation).

Conclusion If speed or respective "time to market" is important *and* highly automated driving can be considered as a complex field then the progress of knowledge strongly determines the success of enterprises. One of the most promising measures coping with both challenges is the significant use of the digitalization.

3 Summary and Outlook

It has been shown that highly automated driving imposes strong changes to the vehicle and expands the "system vehicle" towards the network and backend layers. This "merging process" of the "Digital world" and "Mobility world" brings up strong challenges as well as tremendous chances for shaping the upcoming future of mobility. In order to cope with the given challenges the following recommendations can be stated:

- The significant use of digitalization within development and product-life cycle will gain advantages with respect to "time to market" and market attractiveness.
- A common European assessment of related safety issues by all relevant stakeholders will enable secure investments and helps to avoid market fragmentation.

Reference

1. Heidl P, Damm W, SafeTRANS Working Group (2016) Highly automated systems: test, safety, and development processes. In: Recommendations on actions and research challenges, safetrans conference

Scenario Identification for Validation of Automated Driving Functions

Hala Elrofai, Daniël Worm and Olaf Op den Camp

Abstract The continuous development and integration of Automated Driving Systems (ADS) leads to complex systems. The safety and reliability of such systems must be validated for all possible traffic situations that ADS may encounter on the road, before these systems can be taken into production. Test-driving with ADS functions requires millions of driving kilometers to acquire a sufficiently representative data set for validation. Modern cars produce huge amounts of sensor data. TNO analyses such data to distinguish typical patterns, called scenarios. The scenarios form the key input for validating ADS without the need of driving millions of kilometers. In this paper we present a newly developed technique for automatic extraction and classification of scenarios from real-life microscopic traffic data. This technique combines 'simple' deterministic models and data analytics to detect events hidden within terabytes of data.

Keywords Automated driving systems (ADS) · Testing and validation · Scenario identification and classification · Real-life data · Microscopic traffic data · Big data

1 Introduction

Automated driving technology is anticipated to be a key aspect for achieving a higher level of road safety, a further reduction of harmful emissions, improving traffic flow, and increasing comfort and ensuring mobility for all, including elderly and impaired

H. Elrofai (✉) · O. Op den Camp
TNO Integrated Vehicle Safety, Automotive Campus 30,
5708 JZ Helmond, The Netherlands
e-mail: hala.elrofai@tno.nl

O. Op den Camp
e-mail: olaf.opdencamp@tno.nl

D. Worm
TNO Cyber Security and Robustness, Anna van Buerenplein 1,
2595 DA The Hague, The Netherlands
e-mail: daniel.worm@tno.nl

T. Schulze et al. (eds.), *Advanced Microsystems for Automotive Applications 2016*, Lecture Notes in Mobility, DOI 10.1007/978-3-319-44766-7_13

car users. In addition to these societal benefits, ERTRAC projected an EU economic impact for automated driving ranging up to €71bn in 2030 [1]. The estimated global market for automated vehicles is 44 million vehicles in the same year.

However the current lack of standardized and systematic validation procedures for ADS is hindering the approval and promotion of these technologies. Due to the increasing complexity and level of automation of these systems the requirements for testing and validation are getting more challenging. The main challenge is to overcome the dilemma of testing the entire complexity of real-world traffic without the need of driving millions of kilometers.

Intensive testing and assessment of automotive technology to ensure implementation of new safety and comfort systems and functionalities has become state of the art in the last decades. Safety driven assessments such as Euro NCAP have promoted the implementation of new functions and strongly enhanced the overall car safety performance, both to prevent accidents and to mitigate the consequences.

Currently the performance of new systems is tested in case-by-case situations i.e. for singular use-cases. For instance, an autonomous emergency braking (AEB) system is tested under critical car to car and pedestrian or cyclist crossing situations that would result in an accident if no AEB response is issued. However evaluating the performance of ADS functionalities on case-by-case situations may lead to incomplete conclusions. It is impossible to judge beforehand the situations that such a system may encounter on the road in daily traffic and needs to respond correctly at each point in time. Furthermore, it is very hard to reveal negative side-effects of a system in case-by-case testing. Maurer and Stiller comment that: 'If testing and assessment methods cannot keep pace with this functional growth, they will become the bottleneck of the introduction of advanced DAS to the market.' [2]. Therefore there is an increasing need for developing innovative and efficient validation and testing methods to pave the way for future ADS.

One way to overcome the limitations of case by case testing is test-driving with autonomous car prototypes on regular roads. This is a very expensive and time consuming alternative. Millions of driving kilometers are needed to get sufficient 'driving events' that trigger the ADS functionality under test. According to Bosch, even for high technology readiness level functionalities such as a highway pilot, several million kilometers' worth of test-driving need to be completed before the function can be released for production [3]. Following more theoretical approach by applying the ISO 26262 ASIL D, it is found that easily over 200 million test driving kilometers are needed for validation of automated diving functionalities.

This is an enormous effort that does not seem to be feasible. Moreover, most of the events occurring during these test drives are not interesting. Straightforward scenarios easily cover millions of kilometers, while other events only happen once over a short distance (in the order of meters). Therefore there is an essential need for constructing and collecting relevant traffic events and situations (scenarios) for testing and validation of ADS functionalities. The collection of these scenarios should in principle represent and cover the entire range of real-life traffic situations that might be encountered by the ADS under test.

In this paper we present a newly developed technique for extraction and classification of scenarios form real-life microscopic traffic data (traffic data collected on the level of individual vehicles). These scenarios are called 'real-world scenarios' Typically they include e.g. road layout, subjects involved, maneuvers, relative distances, speeds, view blocking obstructions, weather and light conditions. The scenarios form the key input for validating ADS without the need of driving millions of kilometers as will be discussed shortly in this paper.

In Sect. 2 validation of ADS functions using real-world scenarios is further discussed. The definition of a scenario is introduced in Sect. 2.1. The construction of the real-world scenarios from microscopic traffic data is discussed in Sect. 2.2 as well as how these scenarios can be used for validation of ADS. In Sect. 3, the proposed/developed technique for extracting real-world scenarios from microscopic traffic data is introduced. In Sect. 3.1, we show an example of automatic detection of 'turn' and 'lane change' scenarios out of test driving in platoons on public road. The setup of these platoon drives is discussed as well as the available sensors and the data logging equipment. The process for the extraction of the scenarios out of these data is shown in Sect. 3.2 and a first classification and parameterization of these scenarios is performed. The paper will be concluded with a vision on how such extraction algorithms contribute to the establishment of a scenario database for testing and validation of automated driving functions. The involved challenges are also discussed.

2 Validation of ADS Functions Using Real-World Scenarios

A scenario can be designed by test engineers for development and validation of ADS functions. In our framework, we aim to extract scenarios from microscopic traffic data for the same purpose. These scenarios are then called real-world scenarios.

Before discussing how to extract real-world scenarios and use them for validating ADS functions, we introduce our definition for a scenario.

2.1 Definition of a Scenario

A scenario is in general a setting or a sequence of events. In our framework a scenario for testing an ADS functionality is globally defined as the sum of all relevant conditions, under which the ADS is examined during the test. Scenarios include three main elements: the host vehicle (including the driver), passive environment (e.g. layout of the roads, obstacles, traffic signs) and active environment (e.g. other traffic participant, state of traffic lights, weather conditions). The scenario is then defined as '**the combination of actions and maneuvers of the host vehicle in the passive environment, and the ongoing activities and maneuvers of the**

immediate surrounding active environment for a certain period of time'. Typically the duration of a scenario is in the order of seconds. The end of a scenario is the first situation that is irrelevant to the scenario [4].

We call the actions and maneuvers performed by the host vehicle and other traffic participant within a scenario: 'events'. Therefore, a scenario can consists of one or more events that are typically performed by the host vehicle (turn, lane change, brake) and/or the other interacting traffic participants (cutting-in, cutting-out, braking). In addition, a scenario can contain information about e.g. the weather condition (sunny, rainy), type of the road, state of the driver and static environment.

2.2 Real-World Scenarios for Testing and Validation of ADS

With the above definition of a scenario, the real-world scenario can be generated from the sensor output of the host vehicle such as accelerometer, camera, radar and Global Positioning System (GPS). This is considered as a minimum set of information that is needed to describe a scenario. It is possible to extend the contents of a scenario by combining further on-board information from the CAN-bus, such as individual wheel-speeds, braking pressure and steering wheel rotation. It is also possible to add external information like sensor output from the infrastructure (e.g. road cameras), roadmaps and centrally collected weather data.

Automated and connected cars produce huge amounts of raw sensor data. This Big Data can be used to extract real-world scenarios and events. This is very challenging as it requires:

- Collection of huge amounts of microscopic traffic data to cover the entire range of relevant scenarios that might be encountered by ADS on the road;
- Development of innovative and efficient techniques for automatic extraction and classification of scenarios.

In the following section, we focus on developing techniques and algorithms for 'automatic event detection' in collected microscopic traffic data. These algorithms/detectors are considered as the first step towards developing scenario identification and classification techniques.

The developed detectors are used to filter the interesting parts out of the data. The next step is the identification and classification of scenarios. Condensation of data can subsequently be performed by discarding all irrelevant data, or data that is sufficiently available in the recorded data. Once relevant scenarios are known, testing and validation of ADS functions is performed by reproducing those scenarios either physically at a lab, a test track or even on the road, or virtually in a simulation environment. In case scenarios are described in a virtual world, enrichment is possible, as any variation can virtually be added to an existing scenario.

3 Detection of Driving Events in Microscopic Traffic Data

During a scenario the host vehicle represents the most central active entity and typically performs certain actions or manoeuvres, e.g. taking a left turn, hard braking, or changing lanes. In addition, the actions of other traffic participants need to be taken into account, such as cutting-in or braking. As mentioned in the previous section, we call the maneuvers and actions of the host vehicles and of the other traffic participants 'events'. In this section we focus on detecting driving events of the host vehicle.

To automatically detect events, a hybrid technique has been developed. This technique combines 'simple' physical/deterministic models and data analytics to detect events hidden within terabytes of data. The newly developed technique uses the domain of expertise of vehicle dynamics modelling as well as data analytics.

Furthermore, it is using only basic in-car sensors data to make the detection algorithms as generic as possible. The detection method not only provides an overview of the type of events, also parameters describing the character of events is stored for classification purposes. For instance, the acceleration level in lateral direction is an indication of the aggressiveness level of a lane change. Such acceleration level is stored with a detected lane change.

In this chapter we will show an example of how to automatically detect relevant 'driving' events from basic in-car sensor data coming from driving on public roads. The setup of these drives as well as the available sensors and the data logging equipment will be discussed. For automatic detection of events occurring during such drives we have developed detectors for various different events, including braking, turns, lane changes and cutting-in. Here we focus on two event detectors to detect host vehicle maneuvers: a turn detector and a lane change detector, and describe their application on the sensor data.

3.1 Data Set

Sensor data was gathered in June 2015 in the Netherlands. During the test drives, three cars were driven in a cooperative platoon on a public road, having cooperative advanced cruise control and vehicle following functionality on board. In the tests, the first car was driven manually, the second car was partially automated (longitudinal control following the lead vehicle) and the third car was either partially (longitudinally) or fully automated (both longitudinal and lateral control to follow the predecessor vehicle on the road). A highway route was followed from Helmond to Eindhoven and back.

From all test drives, both sensor data and video material of all three cars have been collected.

Sensor data includes sensor outputs of in car sensors (e.g. car speed, car acceleration, wheel speed, yaw rate, GPS) and vehicle-to-vehicle (V2V) data

(control inputs and outputs that are exchanged between the cars in the platoon). For the logging of sensor outputs, an Intelligent Vehicle Safety Platform (iVSP) has been used. iVSP is a scalable, multi-sensor fusion and processing architecture developed by TNO [5]. Videos were collected using on-board GoPro cameras.

An assessment of detection performance is essential for developing event detection performance. Consequently a 'ground-truth' needs to be determined. Manual video assessment has been chosen in this particular case. However, for the analysis of large data streams this process needs to be automated as well. The videos for the first car were used to manually annotate the start and end times of both turns and lane changes occurring during the runs.

3.2 Detection Methods

As mentioned in the above subsection for developing *turn* and *lane change* detectors, we limit ourselves in using only basic in-car sensor data to make the detection algorithms as generic as possible. We define a *turning event* as an event where the car goes to the left or right in a crossing/road exit or where the car enters or exits a roundabout. Merely driving on a curvy road is not a turning event. However, as will be shown later in this paper, the turn detector can be used to detect curves as well. We define a *lane change event* as "a segment in which the vehicle starts moving toward another lane" and continues in this new lane [6].

We use a model-based detection approach. The reality is formed by the car and its driver, performing all kinds of actions like changing lanes, turning, braking and accelerating. This process is influenced by various external (stochastical) parameters like other traffic, driver behavior, road type, weather, that often cannot be directly or fully observed. Instead we observe several sensors embedded in the car, measuring physical properties like velocity and yaw rate. These sensors are not perfect; the sensor data may include measurement noise and systematic errors, e.g. offset in the sensors or missing fields due to logging problems. Using these observable sensors as well as underlying physical relationships we construct models that are used to detect turns and lane changes.

We investigated the available sensor data, and aimed towards using only basic sensors, because these will be most readily available in existing and newly developed cars.

For detecting turns and lane changes, it appears to be sufficient to focus on the velocity of the car and its yaw rate (or angular velocity). The available yaw rate sensor turned out to be not accurate enough. For this reason we estimate the yaw rate using the rear wheel velocities as illustrated in Fig. 1.

If we apply this formula to the rear wheel velocities gathered for the first car in run 1, we obtain the plot in Fig. 2.

For the *turn detector* we only use the estimated yaw rate. As this estimation is in general quite noisy, the detector first applies an exponentially decaying weighted averaging filter:

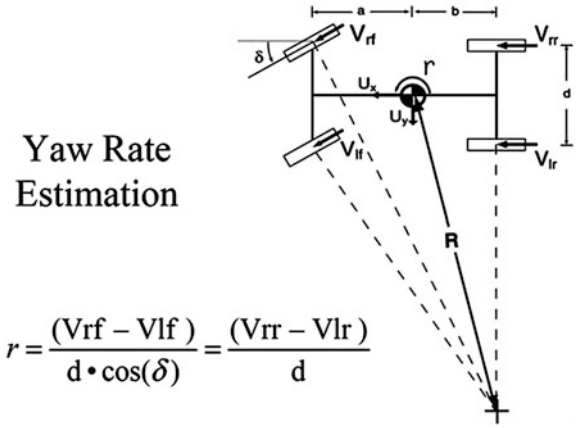


Fig. 1 Estimation of the yaw rate using the wheel velocities

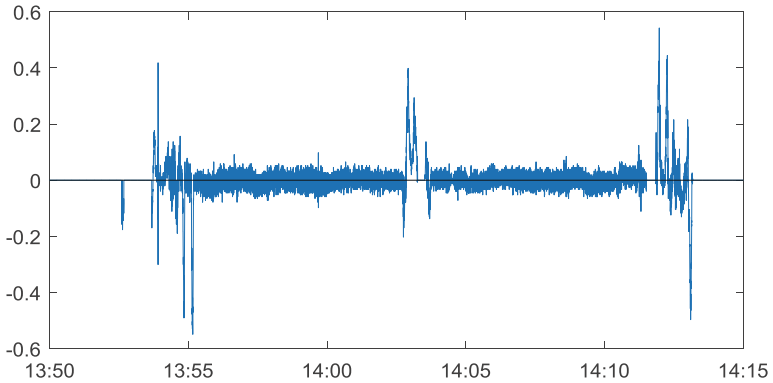


Fig. 2 Estimated yaw rate using the wheel velocities of the first car during run 1

$$W(\tau) = \int_0^{\infty} e^{-bt} S(\tau - t) dt$$

where S(t) is the yaw rate at time t, and b is a positive parameter to be chosen.

The main idea of this filter is to smooth out the random fluctuations, while giving more weight to the yaw rate in the 'present time' than a time in the 'past', see Fig. 3.

As a second step, the turn detector determines whether the filtered signal W crosses thresholds T, −T:

If $W(\tau) > T$ for a certain amount of time, a turn to the left is detected.
If $W(\tau) < -T$ for a certain amount of time, a turn to the right is detected.

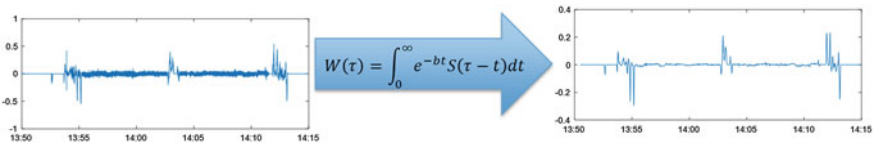


Fig. 3 Smoothing out the estimated yaw rate using a filter

The specific parameter and threshold values of this detector have been tuned using part of the available sensor data including the manually annotated turning events as a training set.

Detecting lane changes is very challenging, since lane changes require only minor changes in direction. Moreover, in this particular data set the lane changes are very smooth, and thus difficult to detect. For instance, a lane change to the left lane consists of a very shallow turn to the left, followed by a very shallow turn to the right. Since the changes in yaw rate are quite minimal during a lane change, combined with the fact that the yaw rate (even if filtered as above) contains noise, simply applying the turn detector with smaller thresholds will not be enough to distinguish between actual lane changes or driving straight.

For our *lane change detector* we use the velocity of the car as well as the yaw rate, again estimated via the rear wheel velocities. From this data we derive two model-based parameters that describe part of the trajectory of the car made in the past x seconds. These parameters are related to the direction change and the position change of the car during the trajectory. It turns out the parameters under investigation are often correlated during a lane change. After some experimentation with part of the sensor data, it appears that x = 8 s gives the best results regarding detection. Note that most lane changes take place within 8 s, and even if it does not, a large part of the lane change maneuver will take place in an 8 s window. Our lane change detector then works as follows: if the correlation crosses a threshold, a lane change is detected.

We combine the lane change detector with the turn detector, since the lane change detector is sometimes triggered during a turn (resulting in a false positive detection). Thus, we only consider those detected lane changes that did not overlap with a detected turn. A disadvantage of this approach is that lane changes occurring during a turning event could be missed. Smart combination of these detectors as well as other detectors, using e.g. machine learning, will be used in future work to further optimize detector performance.

3.3 Results

We tuned the parameters of the turn detector on part of the data and applied it to another part. In Fig. 4 we indicate both the detected turns and the annotated turns over time.

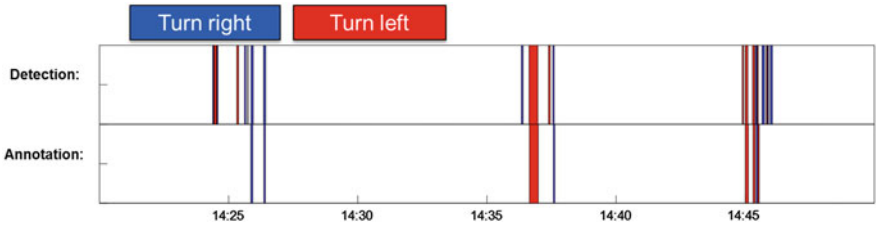


Fig. 4 Turn detector performance

Figure 4 illustrates that all annotated turns have been detected, so the hit rate[1] is 100 %. There are some false positives: the detected turns in the first and last part of the data set are related to turns made from the parking location towards the highway, which were not annotated. The additional detected turns in the middle are actually related to curves in the road. This indicates that the detector actually detects curves very well. Improving the turn detector, for instance by including features such as the curvature, may lead to a better distinction between merely curves in the road and actual turning events.

After tuning the parameters based on part of the data, we applied the lane change detector to another part of the data, after combining it with the turn detector as discussed in the previous subsection. The lane changes have been manually annotated in this part of the data. We obtain a hit rate (see Footnote 1) of 77 % and a precision[2] of 76 %, implying that many of the lane changes are actually detected, with a relatively low false alarm rate. The required hit rate and precision depend a lot on how the events are being used. For instance, in some settings it is important to ensure that (almost) all occurrences of a certain event have been detected, even if this results in more false alarms. This would require the hit rate to be close to 100 %.

Note that these performance results are based on a relatively small number of turns and lane changes. We are investigating and further developing these detectors based on larger data sets.

Each detected event will be completed with an annotation of the typical parameters for classification of the event. In this way all detected events are stored in a database with its characterizing parameters.

4 Discussion and Conclusion

In this paper a newly developed technique for detecting driving events in collected microscopic traffic data has been introduced. The contribution of these events in the identification and classification of the real-world scenario has been discussed. The

[1] *Hit rate* is the percentage of actual events that has been detected.

[2] *Precision* is the percentage of events that has been correctly detected.

concept of "scenario" and "real-world scenario" were defined and discussed in detail. Furthermore, we have discussed how the collected real-world scenarios can be used for testing and validation of ADS function, which represent an efficient approach in addition to test-driving. Ideally, the number of test driving kilometers on the road will be drastically reduced by use of a database with representative scenarios.

Our definition of a scenario gives a visual and behavioral description of all involved entities in the scenario during a certain time (order of seconds). These entities can have an active or passive role in the scenario. The passive entities include layout of the road, traffic signs and obstacles. The active entities include the host vehicle, other traffic participant (e.g. other vehicles, cyclists and pedestrians) and disturbances such as weather and lighting conditions. Although the actions of other traffic participants play an important role in the scenario, the most central active entity is the host vehicle. For this reason we started first with the event detection (and validation) for the host vehicle. Only when the events of the host vehicle are detected and classified, it makes sense to identify the events related to the other active entries from the sensor information of the host vehicle and other sources.

Main challenge here is capturing and collecting the relevant scenarios, both nominal as well as extraordinarily critical scenarios, which cover the entire complexity of real world driving.

The autonomous and connected cars produce huge amounts of real world driving data. This Big Data can be used to extract scenarios and driving events for validation of ADS functions. Beside this, scenarios can be extracted from accident data (e.g. national statistics, CARE [7]), existing naturalistic driving studies (UDRIVE [8]) and Field Operational Tests (FOTs). The extracted real-world scenarios will be stored in a scenario database that will grow over time, as more driving data is collected, to include relevant scenarios required for testing and validation of ADS functions.

Another challenge is to determine the relevant scenarios for testing, and the way in which these tests need to be performed. A reasonable balance between safety assurance and feasibility of testing needs to be achieved. Therefore a limited number of scenarios that are relevant to the SAE level of automation [9], and which assure minimum safety and performance requirements should be identified. With the selection of relevant scenarios, the distribution of the characterizing parameters is known. Using the selected scenarios, the evaluation is then achieved by replication of those scenarios either physically at a lab, a test track or on the road, or virtually in a simulation environment. Virtual testing allows for the simulation of a large number of scenarios for evaluation within a reasonable period of time, where physical tests are needed for specific performance checks and to validate the models used for virtual testing. In the virtual testing environment, not only the scenarios as recorded on the road will be used. The parameterization of the scenarios allows for excursions in virtual simulation beyond the ranges of the recorded parameters. In this way, an extended scenario set will become available to cover both nominal driving conditions, and driving conditions rarely seen on the road, but essential for automated driving systems to deal with.

References

1. ERTRAC, Automated Driving Roadmap, Version 5.0. http://www.ertrac.org/uploads/documentsearch/id38/ERTRAC_Automated-Driving-2015.pdf. 21 July 2015
2. Bengler K, Dietmayer K, Farber B, Maurer M, Stiller C, Winner H (2014) Three decades of driver assistance systems: review and future perspectives. Intell Transp Syst Mag 6:6–22
3. Automated driving, Bosch now conducting tests on roads in Japan. Mobility Solutions, Bosch Media Service. http://www.bosch-presse.de/presseforum/details.htm?locale=en&txtID=7504
4. Elrofai H, Paardekooper J-P, Op den Camp O, Kalisvaart S, van Nunen E, Real-Life Scenario-Based Assessment Methodology for Automated Driving Systems, white paper, TNO, preprint
5. Kwakkernaat M, Bijlsma T, Ophelders F (2014) Layer-based multi-sensor fusion architecture for cooperative and automated driving application development. In: Advanced microsystems for automotive applications. Springer International Publishing, New York, pp 29–38
6. Salvucci D et al (2007) Lane-change detection using a computational driver model. Hum Factors J Hum Factors Ergon Soc 49(3):532–542
7. European Road Safety Observatory, SafetyNet (2007) http://ec.europa.eu/transport/wcm/road_safety/erso/index-2.html
8. UDRIVE, European naturalistic Driving Study. http://www.udrive.eu/index.php/about-udrive
9. Surface vehicle information report: Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems. Issued Jan 2014. http://standards.sae.org/j3016_201401/

Towards Characterization of Driving Situations via Episode-Generating Polynomials

Daniel Stumper, Andreas Knapp, Martin Pohl and Klaus Dietmayer

Abstract For the safety of highly automated driving it is essential to identify critical situations. The possible changes over time of a given situation have to be taken into account when dealing with criticality. In this paper, a method to generate trajectories with polynomials is considered. Thus, the trajectories can be tested and characterized analytically. With this approach it is possible to calculate a huge amount of feasible outcomes of a driving situation efficiently and therefore the criticality of the situation can be evaluated.

Keywords Functional safety · Chebyshev polynomials · Generation of trajectories · Criticality · Episodes · Situations · Critical assessment

1 Introduction

Highly-automated driving depicts a fascinating and innovative technology and has henceforth been an attractive research field for scientists especially in the past few years. Since the 1990s there were projects with experimental vehicles like the PROMETHEUS project [1–3] or the successfully autonomously driven Bertha Benz Memorial Route in 2013 [4]. Most of the challenges of those experiments were to deal with the sensors and the usage and fusion of data. Following, there is

D. Stumper (✉) · K. Dietmayer
Institute of Measurement Control and Microtechnology, Ulm University,
Albert-Einstein-Allee 41, 89081 Ulm, Germany
e-mail: daniel.stumper@uni-ulm.de

K. Dietmayer
e-mail: klaus.dietmayer@uni-ulm.de

A. Knapp · M. Pohl
Daimler AG, Hans-Klemm-Straße 45, 71034 Böblingen, Germany
e-mail: andreas.knapp@daimler.com

M. Pohl
e-mail: martin.pohl@daimler.com

T. Schulze et al. (eds.), *Advanced Microsystems for Automotive Applications 2016*, Lecture Notes in Mobility, DOI 10.1007/978-3-319-44766-7_14

the task of planning the right route and finding a feasible trajectory. Today, there are still unsolved problems or at least unimplemented approaches to deal with before building a highly-automated vehicle for the mass market. One of them is inflicted by the ISO 26262, defined in 2011 [5]. Naturally, all the systems of experimental vehicles were designed and tested to protect the safety of the driver and other road users. However, with the ISO 26262, there exists a tool to analyze and ensure the safety of the electrical and/or electronic systems of passenger cars. Before mass production of a highly-automated vehicle, it has to be guaranteed that all issues of functional safety have been taken into account. One aspect is to make sure, that the vehicle can identify all kinds of critical situations and knows how to deal with them. However, it is first necessary to define what a critical situation is and how to distinguish it from a non-critical situation. The most common way is to cope with thousands of testing kilometers, but for a highly-automated vehicle the amount of driven distance needed on highways is estimated to be $6 * 10^9$ km [6]. This is a likely unrealistic value to reach and even if it was reached there is no guarantee that all possible critical situations have been seen at least once, therefore in this paper another approach is used and explained in the following.

Sections 2 and 3 provide the required definitions and the general procedure for the generation of episodes. Sections 4 and 5 give a survey of how to identify collisions and how to determine the criticality of a situation. Sections 6 and 7 include an evaluation of the presented methods and a discussion of further aspects of the topic. Section 8 summarizes the paper in a conclusion.

2 Definition of Situation and Episode

To evaluate the criticality of a driving situation it is necessary to define what a situation is in this context. Here, a situation $\boldsymbol{S}$ holds information about the amount and the state of all considered vehicles at a fixed moment of time.

$$S = \{s_0, \ldots, s_n\}, s_i = \left(x, y, v_x, v_y, a_x, a_y, \varphi\right).$$

The state $\boldsymbol{s_i}$ of a vehicle $\boldsymbol{i}$ includes the position, the velocity, the acceleration (each divided in x- and y-direction) and the yaw angle (based on the x-direction), illustrated in Fig. 1.

With this definition it is possible to describe any vehicle constellation in a global coordinate system. A situation with three vehicles is shown in Fig. 2.

In this context, an episode $\boldsymbol{E}$ describes one development over time of a given situation $\boldsymbol{S}$.

$$E_S = \{\zeta_0(t), \ldots, \zeta_n(t)\}, \zeta_i(t) = \begin{pmatrix} p_{x,i}(t) \\ p_{y,i}(t) \end{pmatrix}.$$

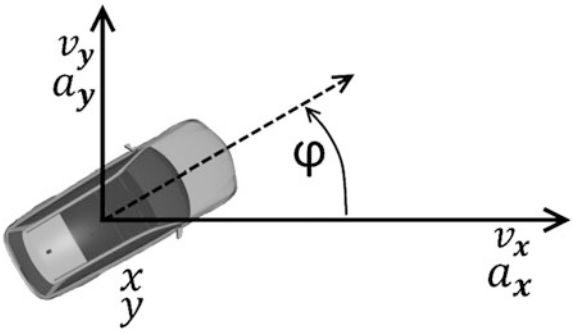


Fig. 1 The variables used for the state of a vehicle are shown here. The point (x, y) is the reference point, φ is the yaw angle and v and a are the velocity and the acceleration in the x- respectively the y-direction

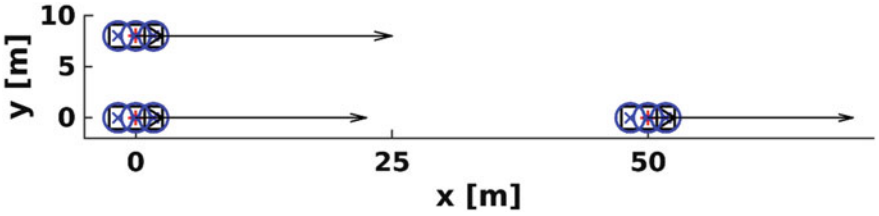


Fig. 2 This situation contains three vehicles. The *arrows* represent the velocity and the yaw angle, respectively

Therefore, it holds information about the trajectory $\boldsymbol{\zeta_i}$ of each vehicle $\boldsymbol{i}$ considered in the situation.

3 Generation and Evaluation of Episodes

If episodes are defined, there remain many possibilities to implement them. To handle a huge amount of trajectories, it is unhandy to represent each single trajectory by saving the consecutive states of all vehicles. If the states are defined by time gaps, one has to set the time interval on a fixed value, which leads to a large amount of data or to a too coarse sampling. So a trajectory is represented by polynomials for easier storage and fast processing. Hermes et al. [7], [8] use Chebyshev polynomials [9] to represent the yaw angle and the velocity of a vehicle. In this paper instead, trajectories are generated using Chebyshev polynomials to define the accelerations of the vehicles. The corresponding coefficients are varied to generate different trajectories.

3.1 Generation

As stated before, Chebyshev polynomials $\boldsymbol{T_k}$ are used to describe the acceleration, divided into the x- and y-direction:

$$\zeta_i(t) = \begin{pmatrix} p_{x,i}(t) \\ p_{y,i}(t) \end{pmatrix} = \begin{pmatrix} \int_0^p v_{x,i}dt \\ \int_0^p v_{y,i}dt \end{pmatrix} = \begin{pmatrix} \int \int_0^p a_{x,i}dt \\ \int \int_0^p a_{y,i}dt \end{pmatrix}.$$

Where the accelerations are given by

$$a_{x,i}(t) = \frac{c_0}{2} + \sum_{k=1}^{n-1} c_k T_k(t), \quad a_{y,i}(t) = \frac{d_0}{2} + \sum_{k=1}^{n-1} d_k T_k(t).$$

Here, $\boldsymbol{p}$ denotes the time period of an episode and the coefficients $\boldsymbol{c_k}$ and $\boldsymbol{d_k}$ are crucial for the generated trajectory. $\boldsymbol{T_{n-1}}$ has degree n − 1, whereby the acceleration polynomials have degree n − 1, too. Therefore, a trajectory is defined by 2 * n coefficients. Afterwards the trajectories are calculated through integration.

As there is no underlying model or any other information used about the situation when determining the coefficients, there is no guarantee that those trajectories are physically plausible. Therefore, the next step is to evaluate them.

3.2 Evaluation

Before comparing the positions of the different vehicles according to their trajectory, it is necessary to ensure that each trajectory on its own is physically plausible. Thus, the limits of the vehicles have to be taken into account. As the trajectories are stored as polynomials, they can be efficiently checked for boundaries. Immediately after generating a pair of accelerations by choosing the 2 * n coefficients, the first step of evaluation is to verify that no acceleration limit is violated. For each direction there is an upper and a lower limit for the acceleration. Let $\boldsymbol{a_{x,\max}}$, $\boldsymbol{a_{x,\min}}$, $\boldsymbol{a_{y,\max}}$ and $\boldsymbol{a_{y,\min}}$ be the boundaries. Now the inequalities

$$a_{x\_\min} \leq a_x(t) \leq a_{x\_\max} \tag{3.2.1}$$

and

$$a_{y\_\min} \leq a_y(t) \leq a_{y\_\max} \tag{3.2.2}$$

must be satisfied. To ensure this, each boundary can be verified separately by checking the difference to zero. This leads back to a root-finding on polynomials, which can be efficiently done, for example with eigenvalue methods [9].

If (3.2.1) and (3.2.2) hold true, it is still possible that the total acceleration violates the circle of forces. Therefore

$$\sqrt{a_x^2(t) + a_y^2(t)} \leq A \Leftrightarrow 0 \leq A^2 - a_x^2(t) - a_y^2(t)$$

has to hold true, where A is the radius of the circle of forces. As squaring and adding of polynomials results in another polynomial (with higher degree), the same root-finding approach as before can be used.

Subsequently, the velocities can be calculated by integration of the accelerations. Depending on the requirements, analog checks can be performed on velocities, too. After those verifications, only the trajectories that are physically plausible are left.

With this approach trajectories for each vehicle can be generated and checked. As there is no information about collisions taken into account up to this point, in the next section their identification is elaborated.

4 Identification of Collisions

To identify a critical episode it is necessary to check it for collisions. In this context an episode is critical when there is a collision between the host vehicle and any other vehicle and it is not critical when there are no collisions involving the host vehicle.

To achieve an efficient workflow, the identification of collisions is done in two stages, similar to the approach in [10].

4.1 Coarse Collision Check

As the trajectory of a vehicle represents the movement of the reference point of that vehicle, it holds no information about its dimension. However, it is possible to identify critical and non-critical episodes just using the trajectories. It is assumed that the vehicles have the same width as well as, that the reference point is located in the middle of the vehicle and the length is greater than the width. The closest distance two reference points can have without a collision of the corresponding vehicles is equivalent to the vehicle width w. This leads to

$$\left(p_{x,i}(t) - p_{x,j}(t)\right)^2 + \left(p_{y,i}(t) - p_{y,j}(t)\right)^2 < w^2.$$

Therefore, all episodes which satisfy this inequation lead to a collision and hence are critical.

On the other hand, if the distance between the two trajectories is greater than a threshold value $\boldsymbol{W}$, there cannot be a collision and hence the episode is not critical. This means, if

$$\left(p_{x,i}(t) - p_{x,j}(t)\right)^2 + \left(p_{y,i}(t) - p_{y,j}(t)\right)^2 > W^2$$

does not hold true for all pairs of vehicles in the episode, it is unclear if there is a collision or not. Therefore another step is needed to check those episodes.

4.2 Fine Collision Check

Up to this point it is possible to perform all checks only with the given trajectories and therefore on polynomials without discretization. But for the fine collision check it is necessary to examine the episode using the contours of the vehicles. Usually, a vehicle is approximated by a rectangle, but in terms of calculation time it is expensive to check for collisions between two rectangles. Thus, in the fine collision check a vehicle will be approximated by circles. As demonstrated in [11] a passenger car with length $\boldsymbol{l}$ and width $\boldsymbol{w}$ can be represented with $\boldsymbol{n}$ circles whose radii $\boldsymbol{r}$ are only dependent on $\boldsymbol{l}$, $\boldsymbol{w}$ and $\boldsymbol{n}$. The radii $\boldsymbol{r}$ and the distances $\boldsymbol{d}$ between the centers of two adjacent circles are defined as:

$$r = \sqrt{\frac{l^2}{n^2} + \frac{w^2}{4}} \quad \textit{and} \quad d = 2\sqrt{r^2 - \frac{w^2}{4}}.$$

With this approach, the size of the vehicle is overrated (see Fig. 2), but this can be counted as a safety margin. The benefit is that the circles are rotationally invariant and only the center points need to be rotated according to the yaw angle.

Hence, for each time step there are n^2 checks between two vehicles, one for each combination of center points.

5 Criticality Assessment of the Situation

In Sects. 3 and 4 it is shown how to generate and evaluate episodes for a given driving situation. To evaluate the whole situation many episodes have to be generated and classified into one of the following three categories:

- Physically implausible
  One of the checks of Sect. 3.2 leads to an exclusion of the episode. This occurs when the vehicle limits are exceeded.
- Critical

A collision between the host vehicle and another vehicle was identified according to Sect. 4. The corresponding episode is rated as critical.

- Without collision
  The tests confirm an episode with no collision involving the host vehicle, therefore the episode is not critical.

Using this classification all episodes assigned to the first category can be rejected immediately. Those episodes are physically implausible in regard to the limits of the vehicles and therefore not relevant for further consideration. All episodes assigned to the other categories are counted towards the criticality of the situation. Let $\boldsymbol{C}$ the amount of critical episodes and $\boldsymbol{K}$ the amount of episodes without collision, then the criticality Υ of the situation $\boldsymbol{S}$ is defined as:

$$\Upsilon(\mathrm{S}) = \frac{C}{C+K}.$$

Hereby the criticality lies in the interval [0, 1] and depends on the ratio of critical episodes to the sum of critical episodes and non-critical episodes.

In Sect. 2–5 definitions of situations and episodes where given. A possible approach to verify a plausible episode and to calculate its criticality was introduced. In the next section an evaluation on example situations using the presented procedure is given.

6 Evaluation of Example Situations

The method described above is illustrated with reference to the following situations. First, the previously used situation is evaluated and 500 episodes without a collision involving the host vehicle are shown in Fig. 3. The calculated criticality for this situation is 15 %. According to the definition from Sect. 5, this means that out of

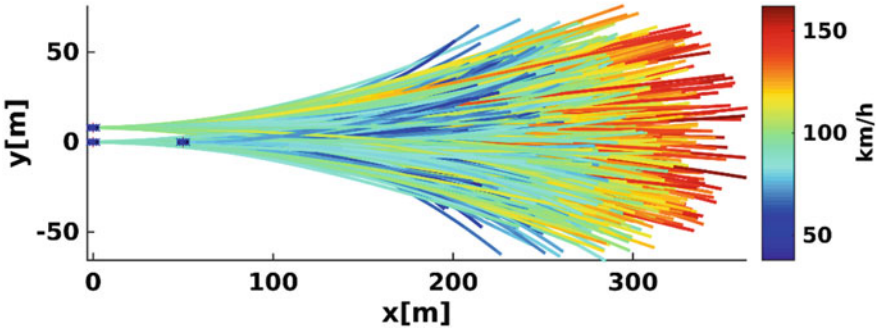


Fig. 3 Three vehicles are considered in this situation. Randomly generated episodes with no collision involving the host vehicle are plotted. The colors of the trajectories represent the velocities of the vehicles

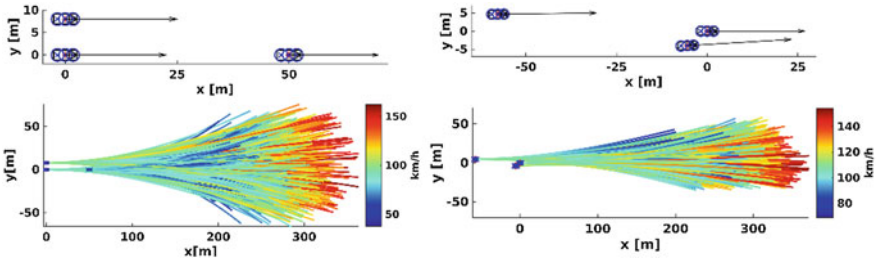


Fig. 4 In the *top row*, the two different situations are plotted. In the *bottom row*, the corresponding episodes are shown. For both situations 500 episodes with no collisions involving the host vehicle are plotted

Table 1 Results of the simulation of the situations shown in Fig. 4. Execution time on a standard notebook with an Intel i7-4710MQ @2.5 GHz

	Situation 1	Situation 2
Tested episodes	1,000,000	1,000,000
Physically implausible	577,464	577,455
Critical	67,566	385,064
Without collision	354,979	37,490
Criticality	15.9 %	91.1 %
Execution time	143 s	143 s

10,000 physically plausible episodes 1500 lead to a collision with the host vehicle. It should be noted that no statement about the likelihood of the episodes is taken.

In comparison, Fig. 4 illustrates two different situations and their evaluation according to the presented procedure. The first situation results in a criticality of 15.9 % and the second situation results in a criticality of 91.1 % (see Table 1).

From Table 1 it is evident that the majority of the episodes is rejected as physically implausible. This is due to the generation of the trajectories and can be influenced by the choice of the intervals from which the coefficients are taken.

7 Discussion

To keep the approach presented as general as possible, no information on the course of the road or the existing lanes is used to represent the situation and to calculate the trajectories. However, it would be possible to take this information into account. To accomplish this the reference line of the road as well as the lane markings can be represented with Chebyshev polynomials, too. The distances between the vehicles and the lane markings can be tracked over the temporal evolution of the episode in a coarse and a fine manner analog to the approach presented in Sect. 4. It would be also possible to use knowledge of the behavior of drivers as well as of their preferences to generate more realistic trajectories. Hereby, less physically implausible episodes would be generated but it could cause critical episodes to be

overlooked. Critical episodes are inherently improbable in terms of the normal behavior of the driver. To ensure functional safety according to ISO 26262 it is necessary to consider all possible critical situations. For virtual testing even the least probable constellation of vehicles should be taken into account. Therefore, no driver model is used in the approach presented.

8 Conclusion

In this paper definitions for situations and episodes are provided. A new approach is shown how to generate episodes with the use of Chebyshev polynomials. By use of an analytic representation, little memory usage per episode is required. In addition, most computations are performed on polynomials, which enable a rapid processing of the episodes. Once the examined episode is classified into a category, no further checks have to be performed, for this reason the more complex calculations are placed at the end of the processing chain. Thus, an efficient and rapid analysis of the generated episodes can be achieved. The possibility of evaluating situations efficiently opens access to identify critical situations which is essential for functional safety of highly-automated vehicles.

References

1. Dickmanns E et al (1990) An integrated spatiotemporal approach to automatic visual guidance of autonomous vehicles. IEEE Trans Syst Man Cybern 20(6):1273–1284
2. Franke U et al (1994) The Daimler-Benz steering assistant: a spin-off from autonomous driving. In: Proceedings of the intelligent vehicle symposium, pp 120–124
3. Dickmanns E et al (1994) The seeing passenger car 'vamors-p'. In: Proceedings of the intelligent vehicle symposium, pp 68–73
4. Ziegler J et al (2014) Making bertha drive—an autonomous journey on a historic route. IEEE Intell Transp Syst Mag 6(2):8–20
5. ISO 26262:2011 (2011) Road vehicles—functional safety. International Organization for Standardization, Geneva
6. Wachenfeld W, Winner H (2015) Die Freigabe des autonomen Fahrens. In: Maurer M et al (eds), Autonomes Fahren: technische, rechtliche und gesellschaftliche Aspekte. Springer-Vieweg, Berlin, pp 439–464
7. Hermes C et al (2009) Long-term vehicle motion prediction. In: IEEE intelligent vehicles symposium, pp 652–657
8. Hermes C et al (2011) Manifold-based motion prediction. Proc 6 Dortmunder Auto-Tag, Dortmund
9. Press W et al (2007) Numerical recipes—the art of scientific computing, 3rd edn. Cambridge University Press, Cambridge
10. Likhachev M, Ferguson D (2008) Planning long dynamically-feasible maneuvers for autonomous vehicles. In: Proceedings of robotics: science and systems IV
11. Ziegler J, Stiller C (2010) Fast collision checking for intelligent vehicle motion planning. In IEEE intelligent vehicle symposium, pp 518–522

Functional Safety: On-Board Computing of Accident Risk

Grégoire Julien, Pierre Da Silva Dias and Gérard Yahiaoui

Abstract Safety estimation for a given driving style is of utmost importance in the perspective of fully automated vehicles. Recent progress on accident estimation measurement made by insurance companies has revealed that correlation approaches based on "severe braking" events are not satisfactory. Here we propose a new solution. Taking into account the dynamics of the vehicle and inputs of different nature (visibility, grip, shape of the road, …) a dimensionless quantity is calculated which detects "near miss" events. This solution, based on deep knowledge of causality relationships, can be used for making data-based services richer, for a more relevant estimation of safety level and a better accident anticipation.

Keywords Safety · Accident avoidance · Driving style · Driver assessment

1 Introduction

Safety estimation is of utmost importance for in the perspective of fully automated vehicles. In particular, recent progress on accident estimation measurement made by insurance companies has revealed that correlation-based considerations do not give satisfaction. Indeed, there is no correlation between accident data bases and data collected by accelerometers such as the so-called "severe braking" that was supposed to estimate anticipation of the driver and thus a "risk of accident". More recently, it has been demonstrated that accidents are rare events being the consequence of the repetition of "near-miss" accidents (e.g. a quasi-accident). It has also been shown that there exists a causality relationship which fully explains "near-miss" accidents.

G. Julien (✉) · P. Da Silva Dias · G. Yahiaoui
R&I Nexyad S.A., 95 Rue Pereire, 78100 Saint-Germain-En-Laye, France
e-mail: gjulien@nexyad.net

T. Schulze et al. (eds.), *Advanced Microsystems for Automotive Applications 2016*, Lecture Notes in Mobility, DOI 10.1007/978-3-319-44766-7_15

2 A New Solution for Measuring the on-Board Risk of Accident

In 2004, the French department responsible for road safety initiated a comprehensive research operation called an "Observatory of trajectories". The goal of the programme was to observe, record, and count the "near-miss" accidents, by assuming that an accident cannot be predicted. An accident is rare: the reflexes of the driver and the other road users, the circumstances, luck, etc. usually allow avoiding the accident to happen. A near-miss accident is the expression of a high risk, statistically speaking, and the probability for an accident to happen increases with the repetition of near-miss accidents. However, the question under which circumstances a near miss accident turns into an accident cannot be answered by causality relationships. It is only statistical: On average, an accident is the consequence of the accumulation of 100–1000 near miss accidents [1]. Figure 1 depicts an illustration of this fact.

The notion of near miss accident is very interesting. Indeed, if one can compute a real time risk function, then one can assess the risk taken by the driver at all time. If the considered car happens to be in risk too often, there is a high probability it will end up having an accident. Hence, having an indication of the risk at real time can be used to control the car.

Based upon results discussed above, NEXYAD has been designing and developing a software module named **SafetyNex** which allows to estimate the driving style and its relevance regarding the road environment, in real time, and assigns a risk score (score usually presented in reverse: Safety) correlated by construction to accident. This solution is the result of over 15 years of R&D, particularly in the context of four French national collaborative research programs [1, 2]. This allowed NEXYAD to extract causal rules of accidents. Those rules have been validated by experts of accidents of the French Road Safety Administration [3]. SafetyNex is a knowledge-based engine, with more than 5000 rules, using "possibility theory" [4] and its implementation into fuzzy sets theory [5]. In the current state, this engine takes the dynamics of the vehicle (acceleration and speed) and an Electronic Horizon (HE) as Inputs. The EH is obtained from map matching algorithms, and it

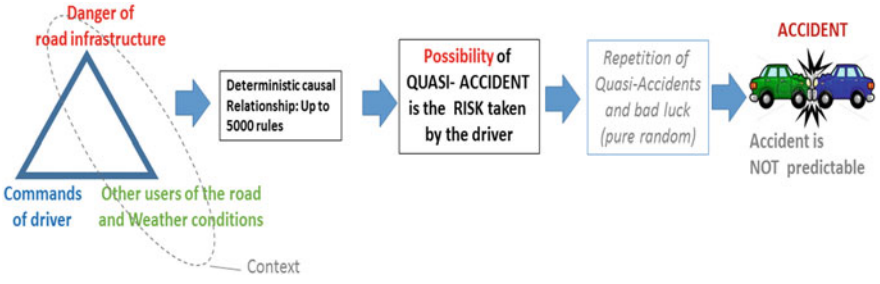


Fig. 1 Scheme explaining how a quasi accident and eventually an accident may happen

corresponds to the most probable driving path in front of the vehicle. A dimensionless quantity $r \in \{0; 1\}$ is the output. It gives an estimation of the risk at real time: if $r = 0$, the risk does not exist, if $r = 1$, the risk is high and a "near miss" accident is detected. Thanks to the Electronic Horizon, this solution is also able to obtain a "risk profile" as a function of the distance along the most probable driving path in front of the vehicle. It is thus possible to fully anticipate the risk in the future considering the current behavior of the vehicle.

3 Results and Discussion of Validation Tests

Safety measurements have only been made by analyzing data collected from accelerometers so far, by assuming that a "severe breaking" (important acceleration variation) reflects a lack of anticipation, and thus, an unsafe behavior. The variation of acceleration is also assimilated to the Eco driving attitude (Fuel consumption is related to acceleration variations). Hence in the severe breaking assumption, the safety is directly correlated to the Ecological driving attitude.

We have imagined four different scenarios to highlight the most common driving behaviors:

1. The "Good" driver
2. The "Quiet—dangerous" driver
3. The "Bad" driver
4. The "Expert Sportive" driver

Driver profiles for those four scenarios are the following:

The "Good" driver does not accelerate much and has a good anticipation when approaching a "Dangerous Point Of Interest" (DPOI), which is defined as a road location where a near miss accident has a high probability of happening. The "Quiet —Dangerous" driver does not accelerate much, but does not stop at all when approaching a DPOI. The "Bad" driver accelerates very often and strongly, and does not slow when approaching a DPOI. The "Expert Sportive" driver accelerates very often and strongly, but slows down when approaching a DPOI.

Each scenario has a duration of about 90s, following the same journey of 1.5 km. The Journey, driven by a professional driver, was mostly composed of traffic lights, pedestrian crossings and intersections, representing a typical journey in an urban environment.

To measure the impact of the individual driving attitude regarding the vehicular energy efficiency ("eco attitude") when executing those scenarios, an extra module called EcoGyser, developed by the Nomadic Solution Company, has been integrated to SafetyNex. The EcoGyser module takes the acceleration and standard NMEA/GPRMC GPS frames as inputs. Then the EcoGyser engine analyses acceleration signals following ten different rules and it outputs a dimensionless quantity varying from 0 to 100 % which quantifies the eco driving style ("eco attitude").

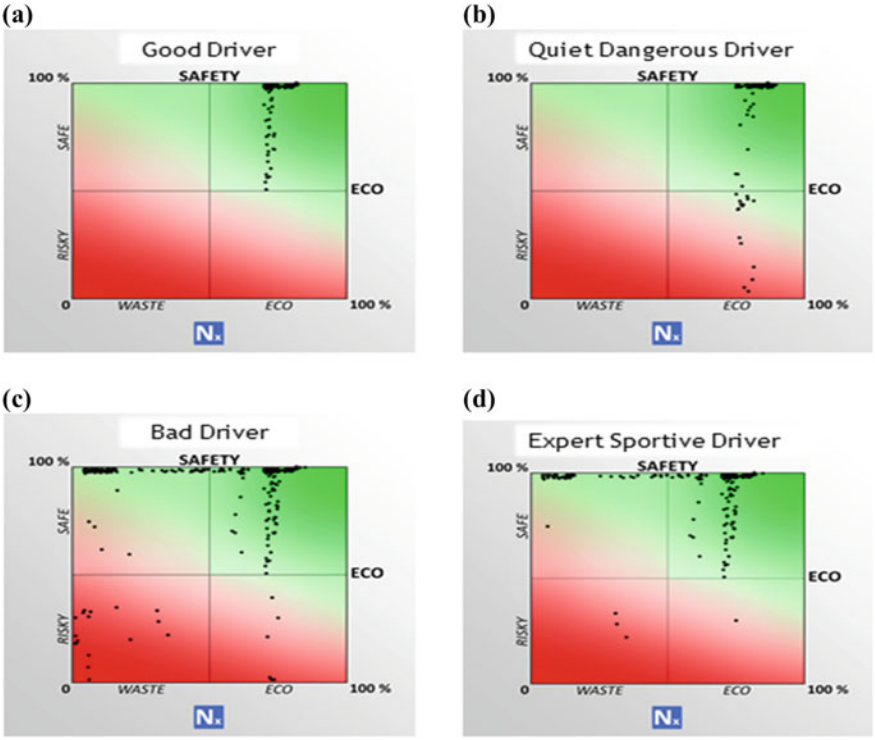


Fig. 2 Point clouds in a safety—eco space for four different scenarios. Each point has been recorded every second with our solution SafetyNex for the safety, and EcoGyser (from our partner nomadic solution company) for the quantification of the energy efficiency of the driving style

The combination of the two modules allows obtaining point clouds in the Safety—Eco Space by collecting data at every second. Point clouds in the Safety—Eco space (scale in percent) for those fours scenarios are given in Fig. 2. On can see that for the "Good" driver case (Fig. 2a), points are only confined in the green area, with Safety values varying between 100 and 50 %, and the Eco values are always around 70 and 75 %, which corresponds to a good Eco attitude: the driver has a global safe and an ecological driving attitude at the same time. In the "Quiet Dangerous" driver case (Fig. 2b), safety values vary between 100 and 0 % and Eco values are around 70 and 75 %. In the "Bad" driver case (Fig. 2c), points spread all over Safety and Eco range, with a rather important number of points confined in the red area: the driver has a global unsafe and wasting attitude Finally, in the Expert Sportive case (Fig. 2d), most of safety values are found between 100 and 50 %, with a few of them below 50 % though. Eco values otherwise vary between 70 and 0 %.

Let us discuss the "Quiet and dangerous" and the "Expert Sportive" driver cases. For the first case, the driver generates dangerous situations because he never slows down when approaching a DPOI, but has a good Eco driving attitude (the driver does not accelerate much, and keeps a relatively constant speed). Hence, as an example, if the driver approaches a pedestrian crossing, he does not slow down to anticipate the presence of a pedestrian who would suddenly appear. So even though it has a good Eco driving signature, the driver is dangerous, which contradicts the idea that a good Eco driving attitude is necessarily safe. Hence, there are in general no correlations between safe and ecological driving styles. This is visible from a formal point of view when observing the cloud point in the Safety—Eco space: one cannot fit the cloud point with a linear curve passing at the origin and having a slope "one half". The Expert Sportive driver case, in the opposite, does not generate dangerous situations most of the time, even if his Eco driving attitude is very bad (the driver accelerates and breaks very often and rather strongly).

4 Conclusion

We show that our approach is more appropriate to estimate the real safety level than accelerometers and the so called "severe breaking" assumption. SafetyNex is a module based on causality relationships, which integrates 5000 rules, all validated by Road Safety Experts, which take into account the road infrastructure. In particular, by studying four different driver profiles, the "Good", the "Bad", the "Quiet Dangerous" and the "Expert Sportive", we show that safe and ecological driving styles (measurement of acceleration variation) are not correlated. Hence, the "severe breaking" assumption falls down. This is particularly visible for the case of the "Quiet Dangerous" driver. Indeed, the driver does not accelerate (Good Eco score), however he does not slow down when approaching a dangerous area (Bad Safety Score) but passes through it at constant speed. Upon the "severe breaking" assumption, the diver would be given a good score, whilst he is clearly dangerous. This is a clear illustration of the fact that the "severe breaking" assumption is not appropriate to measure real safety level.

However, since the EH is obtained from map matching, the map resolution clearly limits the efficiency of SafetyNex. As a further extension of our solution, weather conditions and the grip of the road could be added to inputs. The risk estimation could be also improved by adding a camera in order to merge signals obtained from the road/obstacles detection in front of the vehicle. Finally, in addition to an autonomous car, this solution can be used either for making data-based services richer or for a more relevant estimation of safety level and a better accident anticipation, e.g. by the insurance sector.

References

1. Final document of the SARI program by the French research institut IFSTTAR http://sari.ifsttar.fr
2. Final document of the SARI program by the French research institut CEREMA http://normandie-centre.cerema.fr
3. Brunet J, Da Silva Dias P, Yahiaoui G (2010) Evaluation du risque de sortie de route pour l'aide à la conduite ou le diagnostic d'infrastructure, Session caractérisation du risque routier, conférence PRAC 2010
4. Dubois D, Prade H (2006) Théorie des possibilités. REE 8:42
5. Zadeh A (1978) Fuzzy sets as a basis for theory of possibility. Fuzzy Sets Syst 1:3–28

Part IV
Smart Electrified Vehicles and Power Trains

Optimal Predictive Control for Intelligent Usage of Hybrid Vehicles

Mariano Sansa and Hamza Idrissi Hassani Azami

Abstract An innovative optimal predictive control strategy is proposed in this paper for connected energy management purposes applied to hybrid vehicles, for minimization of energy usage and CO_2 emissions during a given trip, according to the driving conditions that can be predicted by intelligent navigation systems with real-time connectivity to the cloud. The theory proposed for such real-time optimal predictive algorithms is based on the mathematical *Pontryagin*'s maximum principle that provides general solutions for optimization of dynamic systems with integral criteria, under given constraints. Several technical approaches are presented to get feasible real-time computational effort for this dynamic optimization. The calculation of a "trip planning" becomes then possible in embedded controllers synchronized to more powerful servers and computers connected to the vehicle. Significant gains of more than -10 % of CO_2 emissions are demonstrated, maintaining acceptable performances and drivability.

Keywords Optimal control · Pontryagin's maximum principle · Predictive control · Model-based control · Hybrid vehicles · Trajectory optimization · Torque efficiency optimization · Eco-Driving · Electronic horizon (eHorizon) · Human machine interface · Driver assistance · Real-time control · Embedded software

M. Sansa (✉) · H.I.H. Azami
CONTINENTAL Automotive France, 1 Avenue Paul Ourliac—BP 83649,
31036 Toulouse Cedex 1, France
e-mail: mariano.sans@continental-corporation.com

H.I.H. Azami
e-mail: hamza.idrissi.hassani.azami@continental-corporation.com

T. Schulze et al. (eds.), *Advanced Microsystems for Automotive Applications 2016*, Lecture Notes in Mobility, DOI 10.1007/978-3-319-44766-7_16

1 Context of This Development for Connected Vehicles

Automotive applications, with a CO_2 reduction purpose, have to consider today different domains: the first two domains are well known, as the "*from Well to Tank*" path, and the "*from Tank to Wheels*" path, to be complemented now by a third one: the "*from Wheels to Miles*" path, having a significant impact on energy savings when using predicted data from the "Cloud" [3].

Connected devices on vehicles are collected in the generic "electronic Horizon®" concept ("eHorizon"), based on extended and accurate enough geo-localization and navigation system, providing real-time static and dynamic information on the type of road, the indicator signals, the traffic and driving conditions, the predicted events, and any other data that can be detected, estimated and predicted, by any in-vehicle device connected to a backend, and having potential impact on the energy balance during the trip (Fig. 1).

1.1 The "from Well to Tank" Path

The "from Well to Tank" impact mainly depends on the choice of the Energy type in the vehicle. When selecting a hybrid car, that implies a possible mix of the CO_2 impact, from both fuel energy and electric energy consumption. In this case, the results will strongly depend on the Electrical Power plant infrastructures and the related CO_2-emissions, with high differences from different countries, as for instance France with large number of nuclear power plants and Germany with much more coal power plants (Fig. 2).

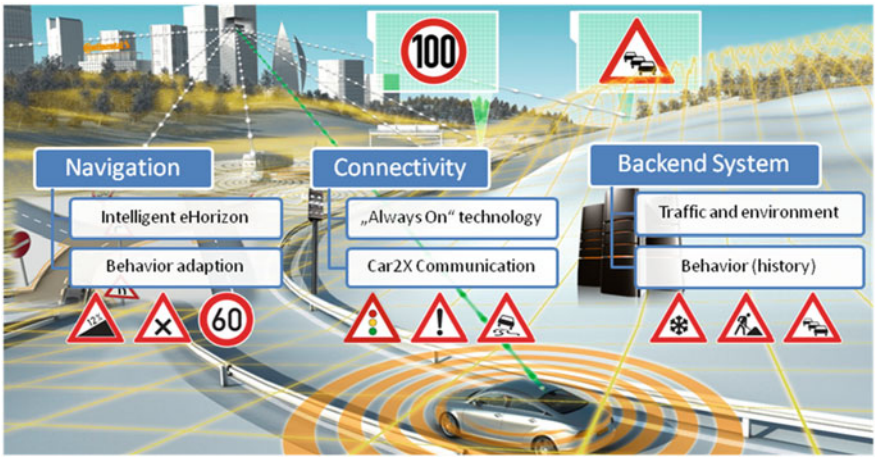


Fig. 1 Connected energy management ("cEM") basic idea at Continental

Fig. 2 Charging station

However, this discussion will not be considered in this paper.

1.2 The "from Tank to Wheels" *Path*

In the "*from Tank to Wheel*" domain, the control obviously has to focus on the optimization of the *efficiency* of power usage in the vehicle powertrain chain, when following the driver's requests, i.e. the requested torque from his or her actions on the gas pedal. This optimization of the so-called "torque-split" makes sense only for hybrid vehicles, where a certain degree of freedom is in the use of the electrical motor (EMA) combined with the standard fuel engine (ICE), and makes possible the optimization of the *power efficiency* that supposes to lead to an *energy* optimization (Fig. 3).

In this purpose, the vehicle can be connected to eHorizon or not. Optimizations can be performed either in a "local optimization", without need for any external data, or in a larger "global optimization" depending on information given on current or predicted driving conditions.

1.3 The "from Wheels to Miles" *Path*

The "*from Wheels to Miles*" domain opens the optimization problem to the global use of the vehicle as a whole system. This question is related to the Eco-Driving

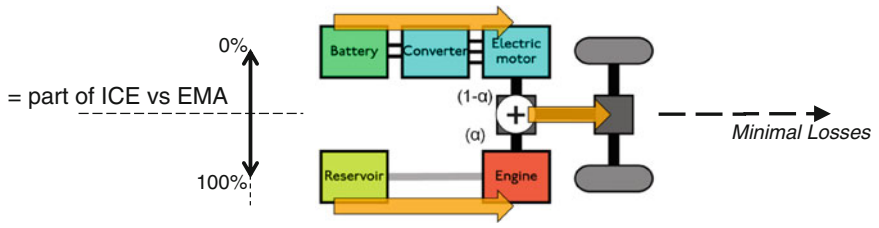


Fig. 3 Torque (or power) split on a typical parallel hybrid configuration

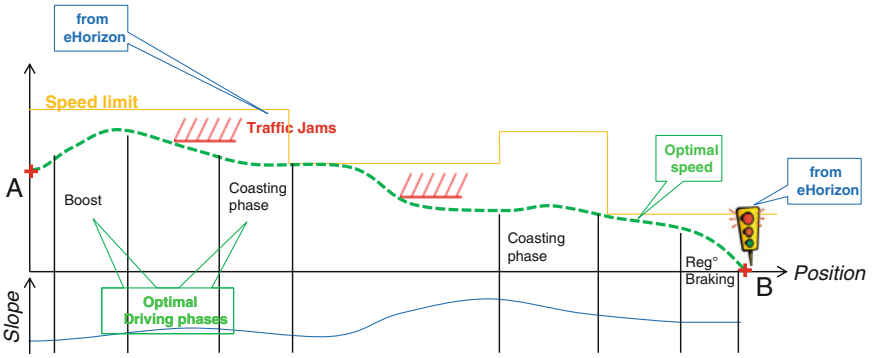


Fig. 4 Optimal speed profile and control phase on a given—or predicted—trip

approach that has typically to provide *an optimal speed profile* or an "optimal trajectory", including all particular phases as: boost, coasting, recuperative braking, etc., and displayed in two ways of actions:

- either to influence the driver, in order to make him follow the optimal trajectory and phases, through appropriate active Human Man Interfaces ("HMI"),
- or to activate the cruise control in order to follow the optimal trajectory, in a more or less partial automated mode control of the vehicle,

so that to achieve energy savings at the end, in comparison with non-optimal statistical driving behaviors (Fig. 4).

In this purpose, connected data from eHorizon are mandatory, obviously to allow calculation of the next speed profile to be followed, with consideration of all constraints in driving conditions and targets to be fulfilled during the trip and at the end of the trip. This predicted time-space approach opens here the field for 'integrative' dynamic optimization methods that can be used and adapted, based on the Pontryagin's maximum principle ("PMP") [1, 2] as described further.

1.4 The Energy Optimization Purpose

The starting point is to consider the state-equations to represent the dynamic (and static) behavior of the system to optimize:

$$\dot{q} = f(q, u, t)$$

where q is the state vector, u the action force (or action torque), and t the time.

In all such control issues, the optimization of energy has to target an integral criteria term, i.e. a cumulated energy consumption, combined with time, as described below in a general manner:

$$criteria\ J = \int_0^T g(u)\,dt + \mu \int_0^T dt$$

where $g(u)$ is the instantaneous power consumption versus the action u, μ a weight coefficient on time, T the final time horizon, and J the criteria to be minimized.

The complete problem must describe the expected 'targets'; that means to get the final values of the state vector to given targets values, as:

$$q(T) = q_f$$

Such a type of criterion fits perfectly with the PMP minimization method that provides a functional solution that minimizes (or maximizes) the integral criteria on the given full time horizon, under given static or dynamic constraints.

1.5 The PMP Method

The PMP method, as well known in Automation communities, is the final expression for optimization of dynamic non-linear systems under constraints, that the Russian mathematician Pontryagin has proposed in the 1950, inspired by the several successive works from Galileo, Liebniz, Newton, Fermat, Bernoulli, Lagrange, Euler, Cauchy, Kantorovich, Dantzig, Karush, Kuhn, Tucker, etc.

The most interesting point is the ability of the PMP method to transform the minimization of a full horizon criterion into a scalar criterion to be minimized at each instant, as explained below.

The PMP method introduces a "Hamiltonian" function that contains both the criteria function and the system model (by the means of its dynamic state equations linked to co-state variable(s) λ):

$$H = g(u) + \mu + \lambda \cdot \dot{q}$$

It is then proven that the minimization of the global criterion J in the full time horizon $[0...T]$ is equivalent to the step-by-step minimization of the Hamiltonian $H(t)$ at each instant, under certain conditions on differential equations called 'optimality conditions'.

The general solution is expressed as:

$$u^*_{opti} = Arg(H_{\min})$$

where 'optimality conditions' are based on differential equations, that link the Hamiltonian and the new co-states, so that logically lead to the possible complete calculation of the necessary solution:

$$\begin{cases} \dot{\lambda} = -\partial H/\partial q & \Rightarrow \lambda(t) = \ldots \\ \partial H/\partial u = 0 & \Rightarrow u_{opti}(t) = \ldots \\ H(T) = 0 & \Rightarrow T = \ldots \end{cases}$$

But the solving of these optimality conditions is the core of the problem, because it cannot always be analytically solved. A particular issue remains on the calculation of the initial values of the co-states λ_o, whose values determine the final targets fulfilling $q(T)$ or not.

Only specific methods can provide acceptable solutions today.

In general, and in the particular case of the hybrid vehicle optimization, these calculations have no analytic solution, as soon as we consider the real non-linear frictions in the system. This complexity led us to start with simplified models that can have admissible solutions from analytic expressions or codable numerical algorithms, as mentioned in this paper.

2 Power and Torque Efficiency Optimization in Hybrid Configurations

2.1 "Local" Optimization

Several solutions are being developed and integrated in actual hybrid demo-cars, based first on a pure "local optimization" of the power efficiency "*from Tank to Wheels*", that is to say selecting the power branch or combination (fuel and/or electric) with the better efficiency at the given instant. Since the battery state-of-charge ("SOC") must be controlled in the meantime, a specific function is introduced, to balance the electrical power impact versus the fuel power impact according to the SOC value: for high SOC values, the electric power will be promoted; for low SOC values, the electric power will be limited.

In this optimization approach, connected data can help, but are not mandatory, as discussed below:

- in a "blind" mode:
  Without any navigation data, the system doesn't know where the vehicle is, what the driving conditions are, and where the vehicle is going to. The torque split can nevertheless be defined directly from a comparison of efficiencies

between the typical 2 cases: in 'motor' mode (electric motor providing torque from battery), or in 'generation' mode (the electric motor recovering energy to battery). Both efficiencies are pre-calculated off-line in two complete look-up tables, and the arbitration can then be made on-line to select the mode with the best efficiency at each instant, according to a given *cost* of the electrical power-source, with a *hysteresis* strategy to avoid oscillations.
The cost of electrical power-source is of course to be related to the battery SOC, so that we get a regulation of SOC versus a given set-point, focusing on limitation of the SOC deviations.
This solution provides good results, but is not able to adapt to variable driving conditions, and needs a recalibration according to the type of trip, to reach the expected SOC targets.

- in an "open eyes" mode:
  Using available static and dynamic data from connected eHorizon allows to adapt the torque-split with a certain degree of prediction, as for instance:
  - self-tuning of the electric power-source cost according to the actual driving conditions (in a city, in a road, in a high-way…)
  - modification of the electric power-source cost to anticipate on next acceleration or deceleration phases;
  - or planification according to next slope profiles;

  All these predictive actions lead to a better SOC regulation and prognosis.

This local optimization, with or without connected data, is a first step, for an already considerable energy saving.

When considering a more global optimization purpose, then deeper optimization methods are required, as typically offered by the PMP method.

2.2 "PMP"-Based Optimization of Torque Split

Simplified equations can be considered as follows (in a first approximation):

$$\frac{\mathrm{d}SOC}{\mathrm{dt}} = -K \cdot M_{EMA}$$

where SOC is the battery state of charge, M_{EMA} is the electric torque, and K a coefficient depending on the characteristics of electric motor and battery.

The criterion remains a function of the ICE torque (fuel consuming):

$$J = \int_0^T g(M_{ICE})dt + \mu \cdot T$$

where $M_{ICE} + M_{EMA}$ = total M given by the driver's request (without any belt ratio or gear ratio considered, for simplification reasons in this demonstration), with a final target $SOC(T) = SOC_F$.

This total M (at the wheels in a given reference) is to be considered as a given *disturbance*, as a function of time, resulting from the driver's action on the gas pedal, and which is not predictable a priori.

This working configuration is called a 'Forced' mode.

The torque split can then be expressed as: $\begin{cases} M_{ICE} = \alpha \cdot M \\ M_{EMA} = (1-\alpha) \cdot M \end{cases}$

with the split factor $\alpha = [0\ldots1\ldots[$.

The different values of this factor α will correspond to the different possible running phases:

$\alpha = 0 \Leftrightarrow$ pure electric traction
$0 < \alpha < 1 \Leftrightarrow$ electric boost
$\alpha = 1 \Leftrightarrow$ pure ICE traction
$\alpha > 1 \Leftrightarrow$ regeneration from EMA

Several preliminary analysis and simulations confirmed to us that usually the energy criterion g can be taken as $g(u) = u^2$

Calculations on the Hamiltonian lead then to the following equation (with a scalar co-state variable):

$$H = \alpha^2 \cdot M^2 + \mu - \lambda \cdot K \cdot (1-\alpha)M$$

The co-state variable λ is a constant, since the state equation on SOC doesn't depend on SOC itself (in this simplified model).

The minimum H will then fulfil:

$$\frac{\partial H}{\partial \alpha} = 2 \cdot \alpha \cdot M^2 + \lambda_o \cdot K \cdot M = 0 \Rightarrow \alpha^*_{opti} = -\lambda_o \cdot K/2 \cdot M$$

as it is a convex function for any $M \neq 0$.

The λ_o coefficient has to be tuned so that the final SOC reaches the expected target SOC_F, what implies to make different "tries" (in off-line simulations) to get the appropriate λ_o value for a given $M(t)$ input profile from the driver on a given time horizon $[0...T]$.

Physical interpretation:

We can observe that the H function has two terms equivalent to primary source powers, from fuel and from battery (taking the constant μ apart):

$$H = P_{FUEL}(\alpha) + \boldsymbol{\lambda_o} \cdot P_{BATT}(1-\alpha)$$

with the (constant) co-state λ_o as a weight coefficient on the battery source Power.

The minimization of the H function becomes then equivalent to the minimization of a *cost* function where the respective source Powers are put in a balance. In other

words, the cost for electric power from the battery is considered to be equal to a fraction of the fuel power cost:

- empirically, the tuned value for λ_o is around ≈ 2.5 for a NEDC simulation, which remains close to the real efficiency ratio between electric power and fuel power.

A solution for SOC regulation results immediately when considering the λ_o coefficient as a table depending on the SOC value, taking:

- high λ_o for low SOC values to promote fuel energy,
- low λ_o for high SOC values to promote electrical energy.

The Hamiltonian then becomes:

$$H = P_{FUEL}(\alpha) + \boldsymbol{\lambda_o}(\boldsymbol{SOC}) \cdot P_{BATT}(1 - \alpha)$$

The local minimization of $H(SOC)$ will then select the torque split α value minimizing the total cost.

2.3 Predictive Complement in Connected Configurations

It becomes obvious that using connected eHorizon providing predicted information can help to adapt the λ_o coefficient, not only according to the *SOC* actual value and target, but also according to the present and future driving conditions, as for instance, taking:

- high λ_o in large roads, to promote fuel energy for high speeds,
- low λ_o in cities, to promote electrical energy for low speeds.

Any other combinations can be considered in λ_o dependencies, extended to:

$$\lambda_o = function\ of\ (SOC,\ type\ of\ road,\ speed\ limit,\ slope\ldots)$$

thus opening up various possible strategies for energy management of connected hybrid vehicles.

2.4 Actual Results

We present here a comparison of results on a NEDC simulation test:

- from a pure local optimization of efficiencies
- from a PMP optimization, including a ‘variable’ co-state dependency

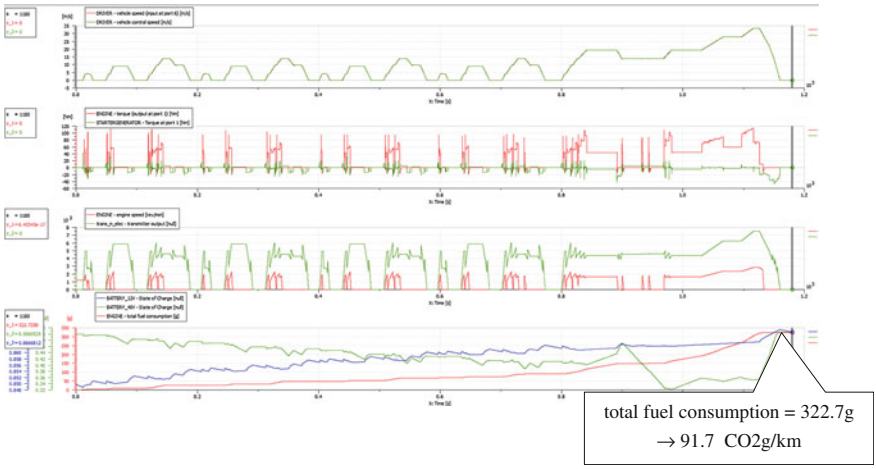


Fig. 5 PMP application to torque-split optimization on NEDC

Simulations have been performed on a co-simulation Matlab®/Simulink®-AMESim® platform, applied to a hybrid FordFocus vehicle model (1L-3Cyl ICE engine and 15 kW electric starter-generator), on a simulated NEDC.

The CO_2 performance (in g/km), resulting from a local optimization of the respective fuel and electric efficiencies, has been found to be 95.4 CO_2 g/km.

The PMP optimization yields 91.7 CO_2 g/km, i.e. an additional 4 % fuel savings. These nearly 4 CO_2 g/km less can represent a huge positive impact for such car, both for the fuel consumption balance as for the CO_2 gain incentives (Fig. 5).

3 Trajectory Optimization on Given Trip

3.1 General Rules for Eco-Driving

Eco-Driving has become a standard today with several on-board systems and displays to influence the driver towards a better driving behavior, both for safety reasons as well as for fuel savings. The main function is acting on the gear shift recommendation, as implemented in several cars today.

A further step is proposed here, to display a recommended speed—not only for speed limitations respect—, but for fuel savings thanks to a complete optimal speed and acceleration profile pre-calculated for a complete given path.

This speed optimization problem can then be readily tackled with the PMP method that provides optimal time-dependent functions giving the minimum or maximum value of a cumulated criterion function.

3.2 "PMP"-Based Optimization

The (simplified) system model to be considered focuses on the kinematics and dynamics of the vehicle:

$$\dot{q} = \begin{pmatrix} \dot{x} \\ \dot{v} \end{pmatrix} = \begin{pmatrix} v \\ u/m - a \cdot v - G \cdot \sin \rho \end{pmatrix}$$

where x is the position, v the speed, m the mass of the vehicle, u the applied force, a the friction coefficient, G the gravity constant and ρ the slope.

We can consider the same integral criterion as in 2.2, so we get the following Hamiltonian equation (with a second-order co-state vector):

$$H = L + \lambda^T \cdot \dot{q} = u^2 + \mu + \lambda_1 \cdot v + \lambda_2 \cdot (u/m - a \cdot v - G \cdot \sin \rho)$$

The optimality conditions lead to the following equations:

$$\dot{\lambda}_1 = -\frac{\partial H}{\partial x} = 0 \Rightarrow \lambda_1 = Cte = C_{10}$$

$$\dot{\lambda}_2 = -\frac{\partial H}{\partial v} = \lambda_{10} - a \cdot \lambda_2 \Rightarrow_2 \lambda_2(t) = \frac{C_{10}}{a} + \left(C_{20} - \frac{C_{10}}{a}\right).e^{a.t}$$

$$\text{with } \lambda_2(0) = C_{20}$$

and with $g(u) = u^2$:

$$\frac{\partial H}{\partial u} = 2 \cdot u + \frac{\lambda_2}{m} = 0 \Rightarrow u^*_{opti}(t) = -\frac{\lambda_2(t)}{2 \cdot m}$$

Some boundary conditions have to be considered in this general optimal control function, depending on the inclusion of phases of engine-off coasting (constant null value for u^*) and of braking (negative constant value for u^*).

Other improvements are necessary, in particular to take into account the real gear ratios, directly given from a separate optimization process; but the PMP method can be extended to the gear optimization, too.

A more complex system will appear when considering a more realistic energy cost $g(u,v)$ depending on both force(or torque) u and speed v. Also, non-linearities due to real slopes depending on position x make the solving much more complex.

Solutions to these new issues are still under development.

3.3 Actual Results

It was conceived possible that simplified models would lead to wrong optimizations, or at least not significantly improved results.

But it fact, we obtain very acceptable results for this eco-driving strategy, as shown below, when comparing statistical fuel consumption when driving in a "usual" way, and when applying the optimal speed profiles as calculated by the PMP method on simplified models as described above.

A "RDE" (Real Driving Emissions) cycle has been taken as 'reference', and the speed profile has been recalculated to run the same distance with the same timing ("Fixed Time method" as indicated in the following picture) and with minimal energy consumption. The optimal speed profile has been simulated with the complete dynamic powertrain and vehicle model on a co-simulation Matlab®/Simulink®-AMESim® platform, applied on a hybrid FordFocus vehicle (1.4L-3Cyl ICE engine and 15 kW electric starter-generator).

The energy fuel savings are significant, more than 10 %, as shown in Fig. 6).

The optimal speed profile can then be used in a *speed assistance system* to induce the driver to follow this speed recommendation as often and precise as possible, by the means of visual displays, sound alerts, and—better—by the means of haptical device, like the Active Force-Feedback Pedal (AFFP) from Continental, already integrated in several series cars in the field, for safety and economic implications.

These significant fuel savings make the PMP method very efficient and useful, to be extended to more complete optimization problems, merging hybrid torque split and optimal speed assistance functions, as described in the next chapter.

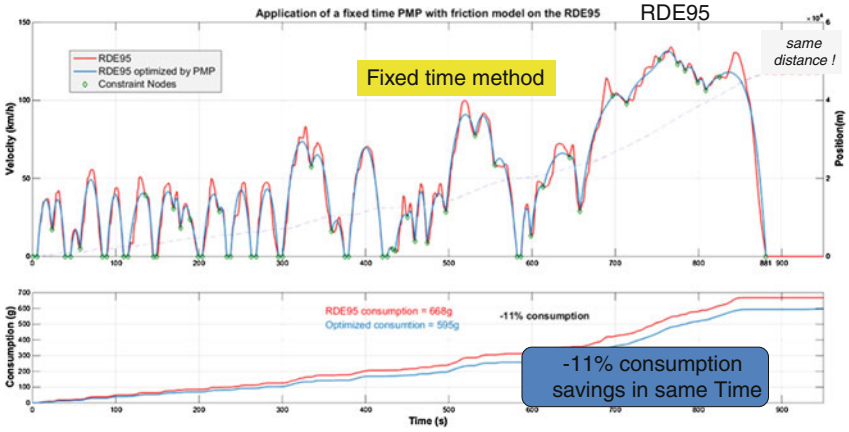


Fig. 6 PMP application to speed profile optimization

4 Merged Optimization

4.1 General Optimization System

The models to be considered can be kept in a simplified representation as follows:

$$\dot{q} = \begin{pmatrix} \dot{x} = v \\ \dot{v} = u/m - a.v - G.\sin\rho \\ \dot{SOC} = -K.(1-\alpha).u \end{pmatrix}.$$

The general action is here extended to the $\begin{pmatrix} u \\ \alpha \end{pmatrix}$ vector, where u represents the total force (or torque) action, and α represents the hybrid split factor, both to be calculated for minimization of the criterion.

The considered criterion remains the same (balance from fuel consumption and time):

$$J = \int_0^T g(\alpha \cdot u)\,\mathrm{d}t + \mu \cdot T = \int_0^T \alpha^2 \cdot u^2\,\mathrm{d}t + \mu \cdot T$$

with a certain expected final target on $SOC(T) = SOC_F$.

The Hamiltonian function becomes then (with a third-order co-state vector):

$$\begin{aligned} H &= L + \lambda^T \cdot \dot{q} \\ &= \alpha^2 \cdot u^2 + \mu + \lambda_1 \cdot v + \lambda_2 \cdot (u/m - a \cdot v - G \cdot \sin\rho) - \lambda_3 \cdot K \cdot (1-\alpha) \cdot u \end{aligned}$$

We get the following conditions for optimality:

$$\dot{\lambda}_1 = -\frac{\partial H}{\partial x} = 0 \Rightarrow \dot{\lambda}_1 = Cte = C_{10}$$

$$\dot{\lambda}_2 = -\frac{\partial H}{\partial v} = \lambda_{10} - a \cdot \lambda_2 \Rightarrow \lambda_2(t) = \frac{C_{10}}{a} + \left(C_{20} - \frac{C_{10}}{a}\right) \cdot e^{a.t}$$

$$\text{with } \lambda_2(0) = C_{20}$$

$$\dot{\lambda}_3 = -\frac{\partial H}{\partial SOC} = 0 \Rightarrow \lambda_3 = Cte = C_{30}$$

and minimization of H leads to:

$$\begin{cases} \frac{\partial H}{\partial u} = 0 \Rightarrow 2 \cdot \alpha^2 \cdot u + \lambda_2/m - C_{30} \cdot K \cdot (1-\alpha) = 0 \\ \frac{\partial H}{\partial \alpha} = 0 \Rightarrow 2 \cdot u^2 \cdot \alpha + C_{30} \cdot K \cdot u = 0 \end{cases}$$

The H function of this general model can be not convex (nor concave), without any local minimum; that implies to consider boundary conditions on the ($\alpha.u$) term (corresponding to the ICE force—or torque-). Optimal solutions have to be found by other numeric methods, as the dichotomy method, or the gradient method.

In addition, the optimal solution (u, α) will not be here fully satisfying, mainly because the electric force (or torque) $(1-\alpha).u$ is found to follow a "bang-bang command" behavior (= min or max). This kind of problems shows the limits of the PMP method when applied to simplified models.

So further developments must be pursued, with integration of more complex models, in particular regarding integration of SOC influence on the SOC variation (that leads to a non-constant λ_3 co-state), or a fuel consumption criterion depending on force and speed $g(u, v)$, etc. The expected results should reach additive CO_2 reductions, thanks to the optimal torque split from one side, and thanks to the optimal speed profile definition from the other side.

The impact of hybrid cars in a connected environment can reach then a considerable level, bringing a significant help in the CO_2 reduction efforts in Transportation.

5 Model-Based Control Impacts on Embedded SW Architectures

5.1 Model-Based Concepts

An interesting consequence of the integration of PMP algorithms for real-time optimization in in-vehicle controllers is the introduction of 'models' within the SW, as a decisive step progress in model-based control design.

The model based control design can be developed in different levels, as discussed below:

5.1.1 "External" Plant Model for Validation

This is the case for the use of plant models for testing the control algorithms under design, by the means of off-line simulations at an early step of development, before integration in real controllers (within the so-called MIL-SIL-HIL step-by-step V-Cycle process).

This MIL phase allows a correct validation of the proposed control solutions, as far as the simulated models are representative enough of the real plant system.

5.1.2 "Internal" Plant Model in SW

Because of some specific requirements on the control architecture itself, a complete or partial plant model is to be implemented into the SW, as described in the following paragraph about the new "MIS" concept.

5.2 Model-in-the-Software ("MIS") Concepts

In the following approaches, the plant model is included in the control modules, becoming thus a part of the code integrated into the software. Several applications can be listed based on this new "MIS" concept.

5.2.1 Delay 'Compensation'

As is well known in the 'Automation' state-of-the-art, simple closed-loop control algorithms (as PIDs for instance) cannot work correctly in presence of direct pure delays or dead-times in the control chain.

Then it is necessary to include a local prediction model (generally called a "Smith Predictor"), in order to apply the direct closed-loop control on the model without pure delays nor dead-times, supported by a secondary adaptive closed-loop applied to the real (delayed) system, with a slower action on it, as shown in Fig. 7 as an example.

The control block-diagram presented here (in orange) is to be included in the on-board control SW, with the plant model split into partial models without—and with—delays and dead-times.

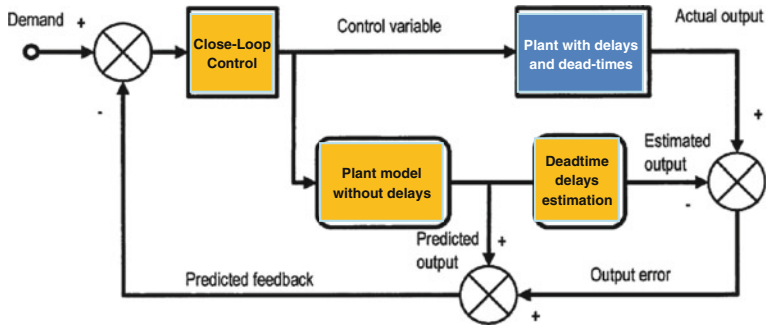


Fig. 7 Conventional model-based "smith predictor" with adaptive closed-loop

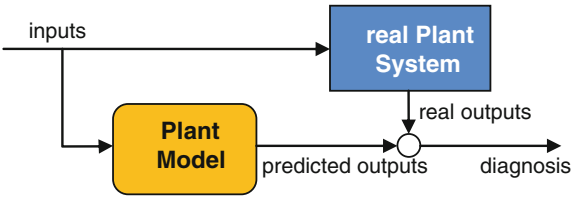


Fig. 8 Conventional model-based on-board diagnosis

5.2.2 Onboard Diagnostic

Obviously, real-time algorithms for diagnosis of a system are usually based on a comparison between real outputs values (from the system under diagnosis), and simulated or predicted ones (from the internal model), as described in the Fig. 8.

Strategies for onboard diagnosis can then consist on auto-adaptation of the inner model, alert signal generation, modification of applied controls, etc.

5.2.3 Model-Based Predictive Control ("MBPC")

A more complex usage of internal plant models within the control SW appears in explicit predictive control algorithms.

A typical example is the Model-Based Predictive Control [4] ("MBPC") or "Predictive Functional Control" ("PFC") that requires to perform internal solving using the modeled equations of the system to be controlled.

Different methods can be applied, as:

- Least Squared method for trajectories solving on predicted horizons (from analytic models)
- Selection on predicted scenarios (from simulated models)

The predictive control is thus able to predict the system behavior and select the optimal case.

5.2.4 Pontryagin Maximum Principle

Another example using an internal plant model in SW is—of course—the PMP method, as explained in this paper, which requires a set of knowledge on the dynamics of the system to be controlled, explicitly integrated in the Hamiltonian function to be minimized, and in some more complex cases, to be implicitly simulated by performing several "shoots" or "tries", to detect the better scenario as optimal solution.

The "Shooting" method associated to a PMP is not always usable in real-time code SW, because requiring long computation times. But this technic remains a solution for off-line verifications, or for control of highly complex systems when being implemented in powerful computers in static or stationary applications.

In any case, the use of an internal plant model knowledge is mandatory for such optimization, which can guarantee the best optimal solution, as demonstrated by Pontryagin in his PMP principle.

5.3 Consequences on Hardware Architecture

A general comment can be given concerning consequences on PMP integration on the electronic hardware architecture of vehicles.

The optimal control based on PMP method is to be considered as a "supervisor" at the vehicle level, getting data typically from eHorizon and providing control commands to subsystems and components of powertrain, chassis and interior parts according to the desired functionalities.

As seen before, the PMP method can be converted into analytic functions to facilitate software coding, but long iterations may also be necessary. In such case, real-time calculations may not be feasible in usual on-board controllers. A partition of calculations must be then organized, shared and synchronized between vehicle controllers and more powerful computers in the backend, with connected interfaces.

Some developments in this area are still ongoing, to be presented in further publications, targeting acceptable industrial integration of real-time coding of the PMP method.

6 Conclusion

Real-time compliant advanced optimization methods as the PMP described in this paper are one of the most interesting options to tackle pending challenges in automotive applications for CO_2 impact reduction in real driving conditions.

The PMP method has been applied to two different cases: on power efficiency optimization "from tank to wheels" in hybrid vehicles, and on speed profile optimization "from wheels to miles" for eco-driving applications.

The significant results that have been already obtained in both cases, with simplified models used at this step for solving constraints, show a quite high potential of such method.

Further developments are ongoing, to insert more real plant models in these algorithms, and to make a more complete merge of all parameters optimization, "from tank to miles", with a clear impact on SW architecture when integrating internal model equations.

We are then "on the way towards the future", where on-board controllers and embedded software architectures start to enable the use of complex and efficient optimizations in real-time, minimizing CO_2 emissions of vehicles.

References

1. Sciarretta A (2014) “Contrôle des Véhicules hybrides”, Institut Français du Pétrole, Filière Expertise R&I, Web-publication
2. Bauer S, Suchaneck A, Puente Leon F (2014) Thermal and energy battery management optimization in electric vehicles using Pontryagin’s maximum principle. J Power Sources 246:808–818
3. Continental Presse Release (2015) Continental’s connected energy management for better fuel efficiency and a smoother ride. Press Portal Corporation
4. Chadli M, Coppier H (2013) Command-control for real-time systems. Automation Control and Industrial Engineering Series, ISTE-Wiley

Light Electric Vehicle Enabled by Smart Systems Integration

Reiner John, Elvir Kahrimanovic, Alexander Otto, Davide Tavernini, Mauricio Camocardi, Paolo Perelli, Davide Dalmasso, Stefe Blaz, Diana Trojaniello, Elettra Oleari, Alberto Sanna, Riccardo Groppo and Claudio Romano

Abstract For the first time in history, the majority of people live now in urban areas. What is more, in the next four decades, the number of people living in the world's urban areas is expected to grow from 3.5 billion to 5.2 billion. At the same time, populations around the world are rapidly ageing. By 2050, the global population of people aged 60 years and over is expected to reach almost 2 billion, with the proportion of older people doubling between 2006 and 2050. This growth and ageing of the population will pose great challenges for urban mobility, which will be addressed within the SilverStream project. In particular, it will develop and

R. John
Infineon Technologies AG, Am Campeon 1-12, 85579 Neubiberg, Germany
e-mail: Reiner.John@infineon.com

E. Kahrimanovic
Infineon Technologies Austria AG, Siemensstrasse 2, Bau 07.4.19,
9500 Villach, Austria
e-mail: elvir.kahrimanovic@infineon.com

A. Otto
Fraunhofer Institute for Electronic Nano Systems ENAS, Technologie-Campus 3,
09126 Chemnitz, Germany
e-mail: Alexander.Otto@enas.fraunhofer.de

D. Tavernini · M. Camocardi
University of Surrey, Guildford, Surrey GU2 7XH, UK
e-mail: d.tavernini@surrey.ac.uk

M. Camocardi
e-mail: M.Camocardi@surrey.ac.uk

P. Perelli
JAC-ITALY DESIGN CENTER S.r.l., Via Torino 21 B, 10044 Pianezza, TO, Italy
e-mail: p.perelli@jac-italy.com

D. Dalmasso
M.T.M. s.r.l., Via La Morra, 1, 12062 Cherasco, CN, Italy
e-mail: D.Dalmasso@brc.it

T. Schulze et al. (eds.), *Advanced Microsystems for Automotive Applications 2016*, Lecture Notes in Mobility, DOI 10.1007/978-3-319-44766-7_17

demonstrate a radically new light and affordable Light Electric Vehicle concept for the ageing population in congested European cities with scarce parking space.

Keywords Smart systems for automotive · Electric vehicles · Ergonomics · Power electronics · Electric drive · Vehicle control system · Torque management system

1 A Comprehensive Approach for LEV Development

The SilverStream project represents a unique approach to urban mobility where a stylish Light Electric Vehicle (L6e category) will integrate a comprehensive set of automotive technologies tailored to the needs of an urban and ageing population. The complete value chain is represented in order to design, manufacture and demonstrate an innovative Light Electric Vehicle (LEV). This approach requires expertise and know-how from multiple disciplines, as covered by the SilverStream consortium, i.e., health and cognitive sciences, power electronics and energy systems, Human Machine Interface (HMI), electric motors, vehicle design, and ride and vehicle dynamics control.

To start with, the vehicle will be designed according to the specific requirements set from experts in the field of medical and cognitive science to maximize its user friendliness. Then, during the course of the project, the most user relevant subsystems (e.g. seats, HMI) will intensively be tested in a clinical environment. Finally the vehicle demonstrator will be verified in realistic scenario (e.g. health

S. Blaz
Elaphe Propulsion Technologies Ltd., Litostrojska Cesta 44c,
1000 Ljubljana, Slovenia
e-mail: blaz@elaphe-ev.com

D. Trojaniello · E. Oleari · A. Sanna
Fondazione Centro San Raffaele, Via Olgettina, 60, 20132 Milan, Italy
e-mail: trojaniello.diana@hsr.it

E. Oleari
e-mail: elettra.oleari@hsr.it

A. Sanna
e-mail: alberto.sanna@hsr.it

R. Groppo (✉) · C. Romano
Ideas & Motion Srl, Via Moglia 19, 12062 Cherasco, CN, Italy
e-mail: riccardo.groppo@ideasandmotion.com

C. Romano
e-mail: claudio.romano@ideasandmotion.com

eco-system at Fondazione Centro San Raffaele (FCSR) and proving ground facilities at MTM). Despite of the vehicle's compact size, it will be characterized by a comprehensive set of innovative features in different domains: electric drive train, ergonomics (interiors and exteriors), electrical energy storage, passive safety and advanced automotive technologies.

2 Multi-disciplinary Investigation and Definition of the Specifications

The definition of the system requirements and the vehicle architecture that will be implemented into the SilverStream light electric car has been headed by FCSR, and will be used to outline the related evaluation framework.

In this preliminary phase of the project, FCSR research was firstly focused on highlighting those specifications which reflect the absolute novelty of the SilverStream vehicle, which are specifically designed for the elderly people. This kind of studies has been conducted with a multidisciplinary and user-centered approach [1], in order to be able to correctly understand the environment the project has to integrate with as well as the corresponding peculiarities.

To this extent, the research team exerted instruments of the "co-creation" methodology, a research methodology implemented in the "City of the Future Living Lab" scenario of San Raffaele Hospital, where user communities (embedded within "real life" situations and environments) actively take part in the SilverStream study and its innovation process from the beginning to the end. Among these instruments, the most relevant used are questionnaires, focus group or interviews with end users (i.e.: the elderly people), personas [2] and brainstorming with field-expert stakeholders (including ergonomists, human factors specialists, gerontologists, physiotherapists, engineers and design specialists).

The results of the application of these methodologies, along with an extensive literature review, have been discussed and used to define an overall project user requirements list targeted for the SilverStream technological solutions. The user requirements list will be updated during the project lifetime, in order to better define and implement user needs once the different vehicle (or its single components) prototypes will be tested with the project population. This iterative process will allow arriving at a final vehicle that responds as well as possible to the needs and expectations of the elderly population living in the European crowded cities.

The vehicle characteristics related to the specific needs of elderly people, intended as improved ergonomics and reduced cognitive stress, will be verified by the FCSR team in San Raffaele hospital lab facilities. The most user relevant subsystems (e.g. seats, HMI, ingress/egress, etc.) will be tested in clinical settings on a representative elderly population, enrolled in clinical validation studies specifically designed by the FCSR team.

In particular, two different strategies will be used to assess the agreement of the tested subsystems with the acceptance criteria outlined in the previous phase of the

project (i.e. requirement analysis): Instrumental evaluation and qualitative evaluation.

The final prototype will integrate different subsystems, described herein below.

2.1 Lightweight Seats

The seats of the LEV represent a distinctive item with particular respect to ergonomics and wellness.

- Optimal posture including lumbar and neck support for comfortable and low fatigue driving—To analyze and assess the new seat in terms of its capability in providing support, capacitive pressure mats (e.g., XSensor PX100) will be used to measure the pressure load distribution of several participants. In particular, an even and wide-spread load distribution is desired to provide a comfortable and low-fatigue seat. In addition, the muscle activity while seating will be registered in order to assess the fatigue and muscle unit enrollment levels. Also, the results will be compared to a standard seat used in the cars of MTM.
- Easy ingress and egress through 90° swivel function—Tests will be performed inside a motion analysis laboratory (MAL) with the involvement of physical therapists and bioengineers. Using specialized cameras (i.e. stereophotogrammetric systems), force-plates, wearable inertial sensors and electromyography systems, motion analysis allows to measure muscle activity, joint motion and forces, and pressure while the subjects perform different motor tasks (i.e. seating, ingress-egress, etc.) (Fig. 1).

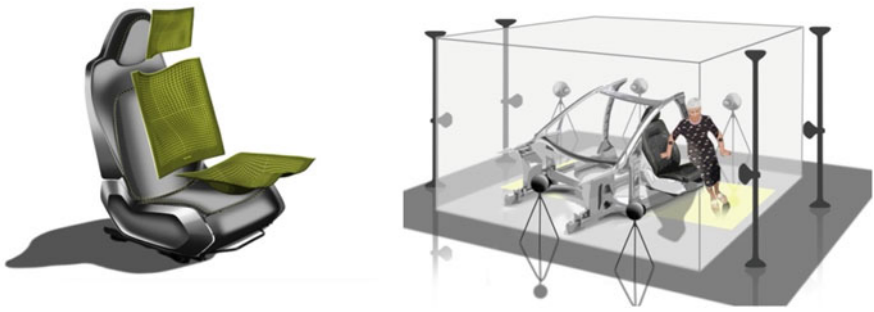


Fig. 1 Lightweight seat prototype and test in the Motion Analysis Laboratory (MAL)

2.2 Assisted Rear e-Lift

An assisted rear e-lift (30 kg payload) for easy loading and unloading of the car has been defined. During the initial implementation phase of the lift, different combinations and initial locations of the goods to be loaded or unloaded will be tested with a wide sample of elderly people (at FCSR premises), who will provide their subjective feedback. The test cases will cover the actual items that can be moved to/from the trunk in real life situations. All those analysis will be performed inside the MAL thus allowing acquiring quantitative data such as muscle load and joint kinematics of the subjects while performing those tasks.

2.3 HMI Based on Gesture Recognition

An innovative HMI based on gesture recognition simplifying the operation of the auxiliary systems has been defined. Acceptance, usability and feasibility of the developed system will be analyzed through questionnaires and interviews and other cognitive tools specifically adapted to the study in order to define the most important functions to be implemented (or removed) into the final prototype of the HMI once integrated into the SilverStream vehicle.

2.4 LEV Test in a Realistic Scenario

The final SilverStream vehicle demonstrator will be tested in realistic scenarios within the health eco-system (i.e. in the "City of the Future Living Lab") at FCSR. The elderly end-users will be involved in the final vehicle validation at FCSR. The final validation will be based on test drives and shadowing, interviews, and focus group in the Living Lab setting already adopted in previous tasks. The focus will be on the improved accessibility and comfort (i.e. in terms of reduction of the backache relating to driving) caused by the rotating seat, its stiffness and adaptability properties, and the design of the doors; the improved maneuverability and steering functions during parking, with attention to the assessment of the HMI aspects; the enhanced performance associated with the rear lift during actual vehicle operation; the gesture control system; the subjective perception of vehicle safety provided by the overall vehicle; the subjective perception of vehicle comfort; the subjective perception of the benefit of the advanced technologies developed.

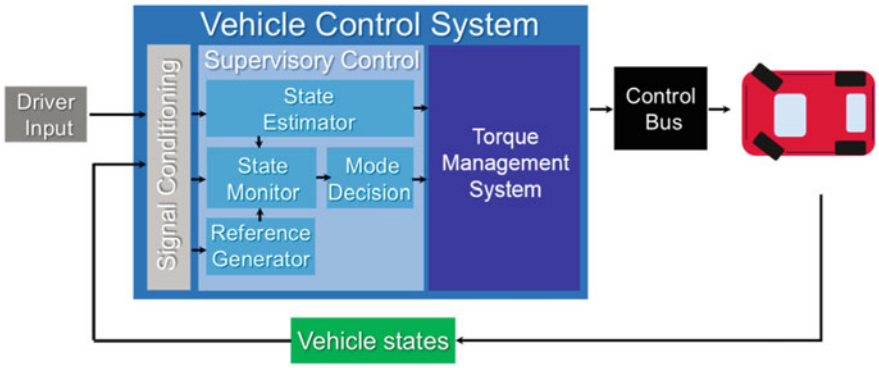


Fig. 2 Scheme of the hierarchical vehicle control system architecture including the TMS

3 Energy Efficient Torque Management System

The L6e quadricycle developed within the SilverStream project represents an over-actuated system, in which the same target in terms of vehicle motion can be achieved by employing infinite combination of inputs (i.e. steering angle and wheel torques). This aspect can be exploited to enhance the overall driving experience of such a vehicle, whilst maximizing energy efficiency. A Torque Management System (TMS) is included in a Vehicle Control System architecture, developed in a hierarchical manner and depicted in Fig. 2. Reading inputs from driver commands and vehicle measured/estimated states, the motion targets corresponding to the driver's intentions are produced in a reference generator, and they are passed to the control unit that computes the output demands for the four electric in-wheel motors. In the context of modern and busy cities, which still have to be safe and accessible for a fast aging population, two main targets are to be achieved by means of the TMS and will be discussed in the following.

3.1 Handling Performance and Energy Efficiency

Vehicle agility is a paramount characteristic to be pursued for the considered vehicle class. This will be achieved by modifying the natural understeer characteristics of the uncontrolled vehicle to make it artificially less understeering, preventing vehicle spin especially in low tire-ground adhesion conditions (i.e. by means of imposed constraints in the control strategy). Moreover the power losses in the drivetrain will be accounted for in the generation of the control strategy. This feature will deliver the best driving experience, overcoming some possible design limits inevitable for this vehicle class, and extend the driving range.

3.2 *Parking Capability*

A second aspect that requires particular attention, since it is among the causes of stress from driving, is parking (especially for elderly drivers in busy cities). The TMS will deliver, at any time of the low speed maneuver, the best electric motor torque distribution between the four corners of the vehicles, by means of an optimized control allocation strategy. In this way we aim to assist the driver, reducing the number of parking steps or the driver effort on the steering wheel for the inevitable ones. A reduced turning radius will be achieved with respect to a similar vehicle of the same class, in conjunction with an increased steering capability delivered by the ad hoc steering/suspension systems also presented in this paper.

4 Advanced Steering and Suspension System Design

The suspension design is based on a pre-existing trailing arm (rear) and a MacPherson (front) used in the L-6 class quadricycles. These two suspension types are characterized by their simplicity, low weight and low cost. The main characteristics of the modified suspension are the high steering angle capabilities (±85°) achieved by means of an innovative steering system and the elimination of the original driveline, replaced by four integrated in-wheel motors. The novel steering system, in combination with the optimized wheel-torque distribution, reduces the turning radius during parking and maneuvering (turning radius < 3.0 m), locating the center of instantaneous rotation of the vehicle closely to the rear inner wheel.

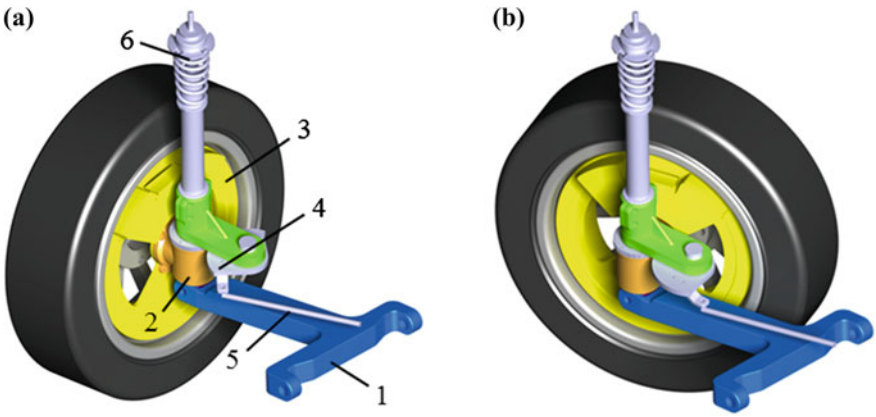


Fig. 3 Front suspension assembly in **a** straight line and **b** 85° steering condition

4.1 Front Suspension/Steering System Design

The modified MacPherson suspension (Fig. 3) comprises a T-shape lower control arm (1), a strut (2) allowing the integration of the in-wheel motor (3), a set of two gears (4) connected to the steering rack through a tie-rod (5) (to actuate the steering motion) and finally a damper/spring system (6).

The T-shape lower control arm was designed to avoid interference at maximum steering angles. Two bearings allow the in-wheel drivetrain to rotate coaxially with the damper axis. The same in-wheel motors are used for the front and rear corners, thus reducing design and manufacturing complexity.

The set of two gears used for the steering operation requires an optimized transmission ratio to achieve the desired rotation and maintain the steering effort within comfortable range.

5 Direct-Drive Air Cooled In-Wheel Motor with an Integrated Inverter

The direct-drive in-wheel motor by Elaphe is designed to satisfy two sets of requirements (for L6e and L7e vehicles), depending on the winding connections. Technical specifications of both versions can be found in Table 1 while torque-speed characteristic and efficiency maps are shown in Figs. 4 and 5. The motor achieves high efficiency even in the most critical operating points, so that the electrical losses at continuous load can be dissipated to the environment at 65–70 K above the ambient temperature. By eliminating components needed for liquid cooling, the motor can be not only produced in a very cost-efficient way, the simplifications also significantly contribute to the competitiveness of the entire vehicle.

Table 1 Technical specifications of the L6e and L7e motor variants

Motor parameters and performance	In-wheel motor for L6e	In-wheel motor for L7e
R—radius at air gap and type	152 mm, outer rotor	
M—active part mass	3.63 kg	
Number of pole pairs	36	
Winding parameters J/VD	4/2	4/4
Torque constant	1.05 Nm/Aeff	0.52 Nm/Aeff
Peak torque theoretical limit	250 Nm	
Boost torque (4s) [requirement]	115 Nm	
Peak torque (20s) [requirement]	68 Nm	
Continuous torque [requirement]	45 Nm	
Max RPMs (no field weakening)	480 RPM	960 RPM

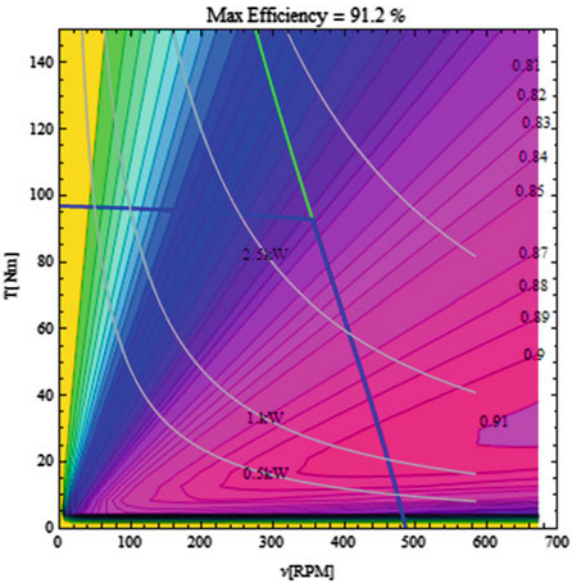


Fig. 4 Motor efficiency and torque-speed map (motor for L6e vehicle). (Due to legislation the continuous power limit is 1 kW per motor. As shown in the above graph (*left*), the speed range is up to 500 RPMs which is suitable for low speed operation of L6e vehicles.)

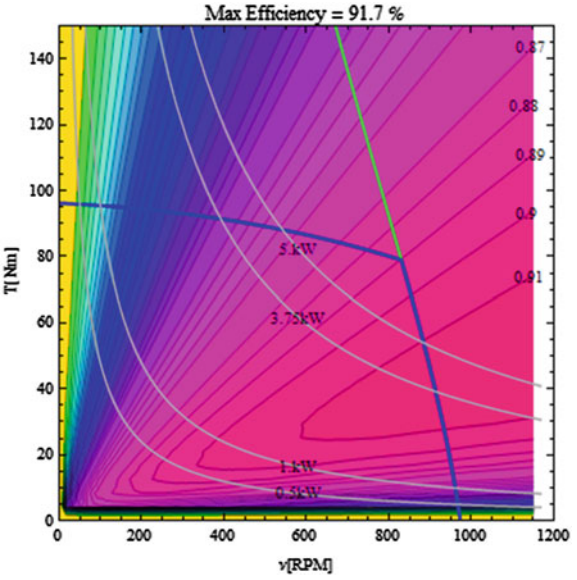


Fig. 5 Motor efficiency and torque-speed map (for L7e vehicle). (Due to legislation the continuous power limit is 3.75 kW per motor. As shown in the above graph (*right*), the RPMs are higher than in the L6e case, up to 1000 RPMs.)

5.1 Flexible Integration

Smart mechanical design of the motor allows the integration on the rear and front axle, on either side of the vehicle without any modification of attachment flanges or

adjustments required for left and right motor. Over 180° steering/turning angle is allowed without any notable impact on cable bending or increase of cable length which would inflate the cost of the cables and increase parasitic inductance.

The inverter can be integrated directly on the motor or it can be mounted on the vehicle, depending on the requirements of the vehicle, with no modification to the motor or its connections, just by changing a single component.

The contactless seals are designed in such a way that they do not affect the motor efficiency in all operating conditions of the motor.

5.2 *Thermal Performance Optimization and Mechanical/Thermo-mechanical Robustness Analysis*

In order to assure a high maturity of the developed system for a fast industrialization after completion of the project, several investigations with respect to the thermal performance of the drive system as well as to the mechanical and thermo-mechanical robustness will be performed. The planned works include (Fig. 4):

- Analyzing of load history from L6e/L7e vehicles in order to perform fatigue and durability simulations and to adapt the respective reliability test parameters.
- Evaluation of air cooling capacity of the motor by performing CFD simulations. By this work the heat transfer coefficients for the motor and inverter housing for different humidity cases shall be obtained in order to serve as input for subsequent thermal simulations and analyses. The goal is to optimize the system with respect to its thermal behavior by deriving design optimizations (e.g. by reshaping the housing or adding additional cooling elements).

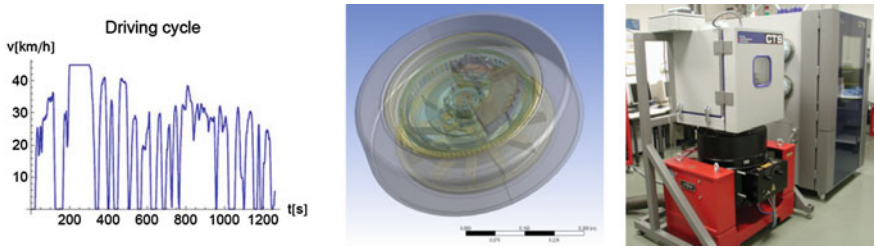


Fig. 6 Exemplary driving cycle of the SilverStream vehicle (*left*), geometry model of electric drive for CDF simulations (*middle*) and test bench (thermal cycling + vibration + humidity) for robustness analysis for the drive system (*right*)

Fig. 7 Preliminary design of the HMI

- Performance of experimental tests to assess the robustness of the electrical drive system with respect to thermal and mechanical loads (Fig. 6).

6 Innovative HMI Based on Gesture Recognition

A breakthrough HMI solution based on "gesture recognition on-air" will be available on the LEV for improved comfort and reduction of cognitive stress.

This smart HMI system is a multi-functional (e.g. infotainment, cluster, HVAC etc.) and multi-control (e.g. gesture, touch, voice, knob) integrated display platform which integrates several vehicle functions in a 12.3 in. screen. The HMI will be controlled by a set of gentle hand wave movements: for each selection action (e.g. navigation, answering the phone, sending a message, entertainment system and other car auxiliary control system) a vocal feedback will be provided. The preliminary design of the HMI is outlined in Fig. 7.

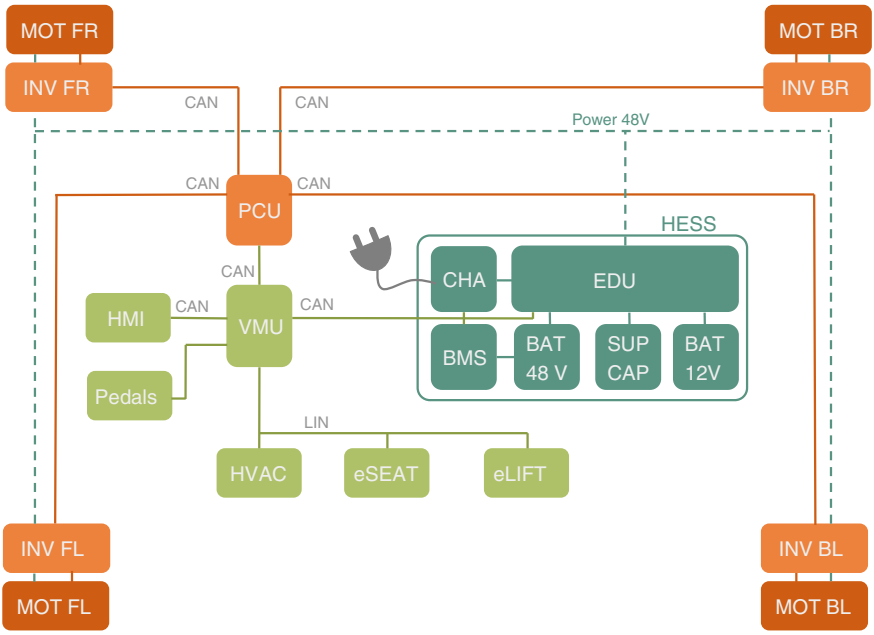


Fig. 8 E/E architecture of the LEV

7 E/E Architecture and Control Systems Development

The E/E system architecture has been developed considering the functional clusters defined in the project. A complex power network for modular and scalable E/E architecture has been designed focused on following three ideas:

- A native dual voltage (12/48 V) power network for safer and simplified E/E architecture;
- A flexible E/E design to allow scalability across the L-category vehicles;
- Multi-core processors for higher computational power and ease of system integration (Fig. 8)

With the objective of reducing the overall system complexity and costs, more functions have been integrated into a single ECU, leveraging the multi-core characteristics of the chosen processor. The mapping of each cluster to one or more specific ECUs is described below (Fig. 9).

This allows on one side to exploit the high performance of the multi-core processor, and on the other hand to achieve functional integration of several features into one ECU through software encapsulation and freedom of interference between software modules. This is utterly important for the VMU, for instance, where several functions are integrated:

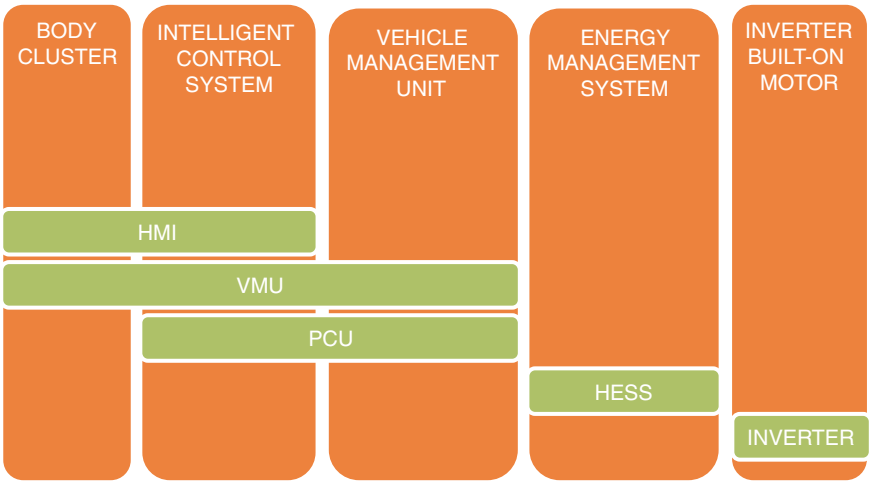


Fig. 9 Mapping of functional clusters to ECUs

- Body cluster
- Intelligent Control System
- Vehicle Management Unit

The functional integration of several ECUs in the car is supported by means of a communication line (CAN bus); in particular, the architecture will provide two separated CAN lines:

- a line connecting the VMU, the HMI, the HESS
- a line for the e-drive (PCU, Inverter)

The VMU, integrating both CAN lines, will provide gateway functionalities between the different domains.

The inverter unit will consist of a control and a power part. The control part will be implemented using an Infineon automotive TriCore AURIX TC277 microcontroller which is equipped with the CAN interface to communicate with the PCU. The PCU will send all inverter related commands to the microcontroller which will execute these commands.

A simplified block diagram of the inverter is provided in the Fig. 10.

The second part of the inverter is the power stage. This stage will be sized and adopted to drive a single in-wheel motor. The specifications are as follow:

- Input Voltage Range: 48 V_{nom}; 36–60 V DC range
- Topology: 3 kW @ 3phase in wheel motor drive (x4—for 4 motors)
- PCB: copper IMS
- Firmware: FOC
- Output Power: 1.5 kW continuous; 5 kW peak
- Continuous current: 43 A (cont.)

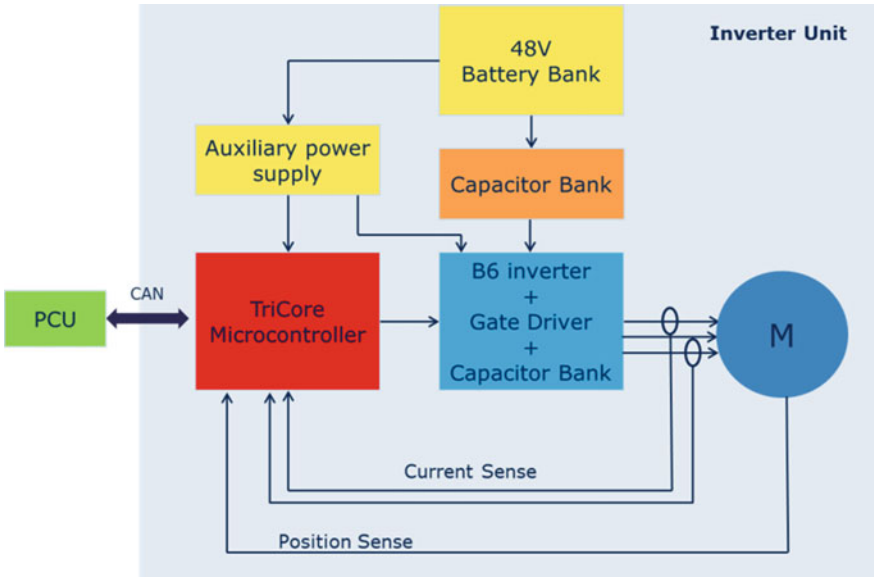


Fig. 10 Preliminary block diagram of the inverter

- Peak current: 65 A (30 s)
- Boost Current: 110 A (5 s)
- Cooling: convection

Main components of the inverter:

- 1 TriCore AURIX TC277—multi-core processor for the control of the in-wheel motor
- 6x IPT015N10N5—OptiMOS power devices for the driver of the in-wheel motor
- 3x 2EDL23N06PJ Eicedriver C—intelligent gate-driver for the Optim-MOS power devices
- TLI4971 phase current sensor for the control loop of the in-wheel motor

8 Conclusions

The SilverStream will developed a LEV for personal sustainable mobility with particular regard to aging population. The key innovation relies in the tight and effective interaction between different skills, ranging from cognitive science to automotive engineering. Preliminary prototypes of the HMI and dedicated seat will allow the experimental activities at FCSR supported by the "co-creation" methodology. The results will be useful for the further development of the LEV. In parallel the electronic and mechanical subsystems are under development,

according to the project time schedule. Potential product exploitation of the results are envisaged through the active participation of industrial partners.

Acknowledgments The SilverStream (Social innovation and light electric vehicle revolution on Streets and ambient) project is supported by funds from the European Union's Horizon 2020 Programme under Grant Agreement No. 653861. The SilverStream will developed its research activities over three years and started in June 2015.

References

1. Reed WA (2014) User-centered design: a guide to ROI with ISO 9241-210. Texology Sci
2. Grudin J, Pruitt J (2002) Personas, participatory design and product development : an infrastructure for engagement. Design:144–161. doi:10.1.1.92.687

Next Generation Drivetrain Concept Featuring Self-learning Capabilities Enabled by Extended Information Technology Functionalities

Alexander Otto and Sven Rzepka

Abstract With the introduction of electrified drive-trains and autonomous driving features, the requirements for electronic systems in vehicles have rapidly become more and more demanding. In order to face these complex requirement scenarios, a paradigm shift from closed-loop-controlled to truly self-deciding and self-learning automata is strongly needed. In this paper, a novel concept for drive-train platforms enabling self-learning capabilities based on sensor integration, micro and power electronics and secure cloud communication directly integrated into the electric motor will be introduced.

Keywords Electric vehicle · Electrical drive-train architecture · Smart systems · Power electronics · Self-learning

1 Introduction

After the invention of the electric motor in the 1820 and 1830s, first electrical drives for traction applications were first realized in the second half of the 19th century. Even tough major improvements have been achieved since then, the fundamental system architecture of today's electrical drive-train systems is still very similar to these early developments. Meanwhile, electrical drives do not only regain importance in the automotive sector, but have greatly been established in industrial automation applications (e.g. in production processes). In fact, almost 50 % of the overall worldwide produced electrical energy is consumed by electrical machines. Although the specific requirements are different in each of the application fields, there is a universal need for reaching maximum energy and power density, energy efficiency and reliability at

A. Otto (✉) · S. Rzepka
Department Micro Materials Center, Fraunhofer Institute for Electronic Nano Systems ENAS, Technologie-Campus 3, 09126 Chemnitz, Germany
e-mail: alexander.otto@enas.fraunhofer.de

S. Rzepka
e-mail: sven.rzepka@enas.fraunhofer.de

T. Schulze et al. (eds.), *Advanced Microsystems for Automotive Applications 2016*, Lecture Notes in Mobility, DOI 10.1007/978-3-319-44766-7_18

minimum costs, which all areas of application have in common. Moreover, the new requirements and technologies created in the field of modern electric mobility can also trigger very useful developments in the industrial sector.

In this paper, a novel concept will be introduced for drive-train platforms addressing the challenges of optimum performance, efficiency, and functional safety by enabling self-learning capabilities based on a joint integration of sensors, micro and power electronics as well as secure communication concepts directly into the housing of the electric motor. In addition, the requirements of the technological building blocks will be discussed including heterogeneous integration of micro-electro-mechanical systems (MEMS) based sensors and modular and ultra-compact wide-bandgap (WBG) based power electronics as well as electromagnetic compatibility (EMC) and reliability issues.

2 State-of-the-Art for Electrical Drive-Train Systems

A typical electric drive-train system of state-of-the-art fully electric vehicles consists of an electric machine, an inverter for converting the DC voltage, e.g. provided by the traction battery, into the three-phase motor currents, an optional DC/DC converter for providing a constant output voltage being independent of the battery state of charge and an energy storage system. In the simplest version, such a system is realized as a set of individual, separate building blocks, which are remotely linked by cables.

In a next step, power electronics and cooling system can be brought closer to the electrical motor or even integrated into its housing. This allows a reduction of construction volume, weight and cabling effort, with the consequences of less interfaces and thus lower fault susceptibility, improved energy efficiency, and lower system costs. Furthermore, by avoiding long AC phase cables within the vehicle, parasitic inductances as well as EMC issues can be reduced dramatically. For example, within the German research project EMiLE [1], such an integrated drive-train has been developed. Primarily, it targets automotive applications but can easily be adapted to industrial automation technology applications as well (Fig. 1 left).

A further system improvement can be made by bringing additional control and health monitoring features into the drive-train. Within the European research project COSIVU for example [2–4], an integrated, modular and compact drive-train system architecture featuring health monitoring of mechanical components (e-motor, transmission) by means of solid-borne sound sensing [5] and of the power electronic modules by means of thermal impedance monitoring [6] was developed. Its feasibility and functionality has been successfully demonstrated on a test bench by Volvo and on a converted passenger car (VW Sharan) by project partner Elaphe Propulsion Technologies. However, although the driver electronics are already integrated within the inverter system, main (sensor) data fusion and control strategy algorithms as well as external communication devices are still kept outside of the system.

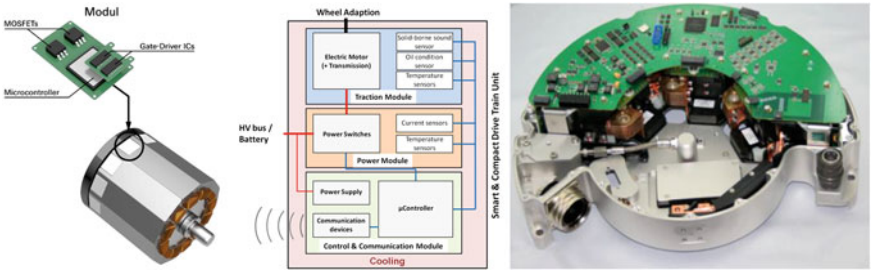


Fig. 1 Schematic description of integrated motor drive aimed for by the EMiLE project (*left*); Block diagram (*middle*) and photographic image (*right*) of developed inverter system by the COSIVU project (Fraunhofer IISB)

3 Novel Concept for Drive-Train Architecture

With the broader goal of introducing (semi-)autonomous driving features, it becomes increasingly important to integrate the drive-train system into the vehicle's holistic IT architecture and even into the Car2X communication network. In other words, a paradigm shift from closed-loop-controlled to truly self-deciding and self-learning automates is strongly needed. The required underlying system concept will be described in the following in more detail.

3.1 System Architecture and Design

Adding to the development history outlined in the previous chapter, a next generation system architecture for electrical drives will be introduced, in which the information and communication technologies involved will widely be extended. These drives will not only be relevant for traction application but can also be adapted to industrial applications easily.

In Fig. 2, the concept of the new drive-train system architecture is depicted (d) and compared to the state of the art system solutions (a–c). The core of the novel approach is a highly integrated and modular drive-train consisting of the electric motor with the power-level sub-system (inverter module) including motor control, driver stage, power modules, DC-link capacitors and cooling system. The modularity enables an easy adjustment of the power range according to the operational requirements. A dedicated sensor network, for example for sensing motor vibrations, solid-borne acoustic and power electronics temperature, together with signal processing and implemented health monitoring algorithms allows an enhanced condition monitoring of critical elements. An example of such critical elements constitutes the power modules of the inverter, in which, by means of thermal impedance spectroscopy, degradations (e.g. cracks and delamination) in the thermal

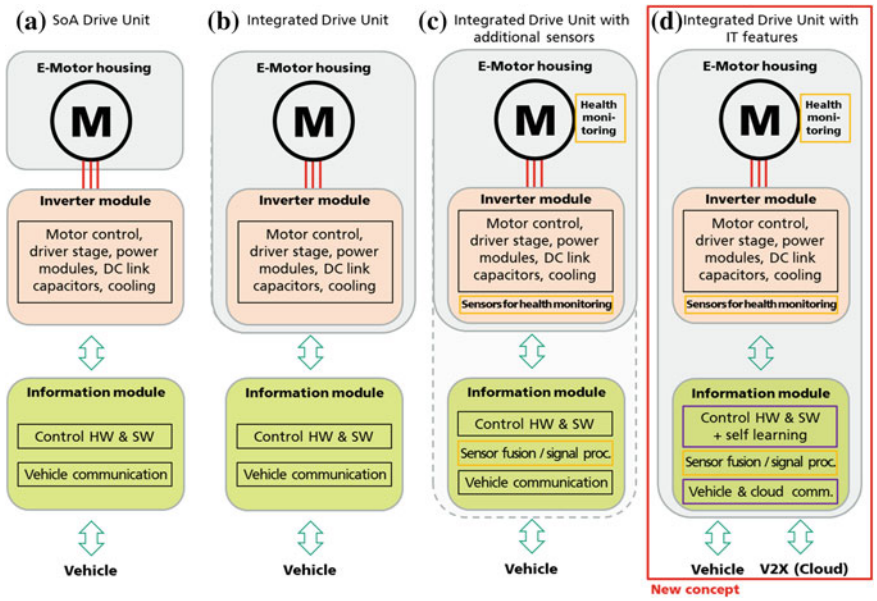


Fig. 2 Comparison of different drive-train system architectures reaching from state-of-the-art (SoA) system with loosely linked building blocks (*left*) to the new proposed system design with self-learning capabilities (*right*)

heat path can be detected. Based on this, the remaining lifetime can be calculated by applying both—data based failure models as well as physics-of-failure based models. This way, an accurate triggering of maintenance on demand can be realized in order to finally ensure maximum reliability and uptime.

A further main part is the information technology sub-system (information module), which comprises a performance central control unit with sufficient memory and computational performance that allows the implementation of intelligent algorithms for the parameter optimization controlling the actual drive operation. These algorithms will be based on cognitive abilities such as self-learning—for example for a systematic accumulation and use of experience gained during operation by the system itself as well as through information received from similar systems in the field. By means of pattern recognition schemes applied to the information from the sensor signals characterizing the current situation, it will be possible to find the optimal operating point much faster or to swiftly react to unforeseen conditions and thus to adjust and improve the overall performance of the entire vehicle during its service life.

A potential application scenario is the detection of the actual road conditions and the triggering of pro-active drive-train actions. In this scenario, the wheel-road interaction would be monitored based on measured torque, slip and rotation speed values for all individual wheels (e.g. paved/unpaved, handy/humpy/muddy,

dry/wet/icy, …) for example. With this set of information, in addition with further information—for example from mounted tires (tire mixture, pressure etc.), where information is either stored in a database of the vehicle's central control unit or derived from the measured data (e.g. from rotation speed)—the maximum applicable power of the motor can efficiently be limited to the level allowing maximum traction in the safe driving mode, i.e. without reaching an excessive range. This will be made possible by comparing the patterns of the current set of input information to those available in the on-board library. Based on the best match or some fast interpolations between the closest neighbors, the optimum motor output parameter can be determined. The new set of information will be sent to an external cloud data server. There, the full algorithm will be applied for computing the best optimum motor parameters. The temporary motor parameters determined by interpolation can afterwards be updated with those obtained by the full-scale computation send back to the drive for enlarging the local library. Next time, the current event happens, the optimum control parameters are readily available. In addition, the database of sample cases for the interpolation scheme is improved. More and more often, the neighboring known situations will be so close that the determination of the motor parameters by interpolation can safely be executed by the actual drive mode and only minor adjustments will later be introduced by the full-scale calculation—yet improving the system further.

This approach of self-learning will make a major improvement of the drive-train behavior possible. It will add pro-active and anticipative components to the state of the art procedures, which only allow reacting to events that occur and try to prevent or to mitigate unwanted effects.

The mentioned example of self-learning, in which experience is accumulated and used, can be implemented by adapting or interpolating the locally generated and stored input and control data as well as by using data sets obtained from the cloud and thus utilizing the full experience gained by similar systems in the field.

In total, the approach of self-learning allows the systems to act and to adapt to the actual needs much faster and with a much higher degree of precision than today. Furthermore, the computational power that is available at the local vehicle level will be sufficient since the complex calculation activities are mainly outsourced on the cloud level. Of course, the communication connection to the cloud is not part of the real-time loop, as temporary interruptions cannot be avoided. That means, the vehicle will always be able to drive autonomously based on the locally available data. However, the cloud communication is certainly stable enough for gathering and providing additional control parameter in advance, i.e., for future situations. This includes the scenario of downloading the information for the next foreseeable part of the current journey when the connection to the cloud is stable. Worth to mention that connection and bidirectional information exchange with the cloud must be secure. Furthermore, the transmitted data needs to be double-checked for plausibility by each side in order to ensure maximum safety.

3.2 Main Technological Challenges

In order to realize the envisaged system architecture, several technological challenges need to be addressed at the same time. The most important ones can be summarized as follows (see also Fig. 3):

- Highly-efficient and ultra-compact WBG based power electronics,
- Robust and heterogeneously integrated MEMS based sensors,
- Secure multi-level communication connections (to the cloud),
- Functional safety/fail-safe operation concepts,
- Cost-efficient mass production,
- and EMC and reliability concerns.

WBG materials, such as silicon carbide (SiC) and gallium nitride (GaN), are gaining more and more importance for power semiconductors due to their superior electrical and thermal properties as compared to conventional silicon materials, such as increased breakdown voltages, melting temperatures, thermal conductivity and saturation velocity of the charge carriers. This allows higher switching frequencies und thus improved energy efficiency. Therefore, the WBG devices will be used in many of the inverters of future drive train applications. However, they also give rise to two distinct challenges, EMC and thermo-mechanical reliability.

The increased EMC concerns are caused by the high switching edges (dv/dt) made possible by the WBG devices. They will even be aggravated by the fact that optimized signal forms will be applied, which include pre-distortions intended to counter-balance the intrinsic imperfections of the motor coils for maximum over-all

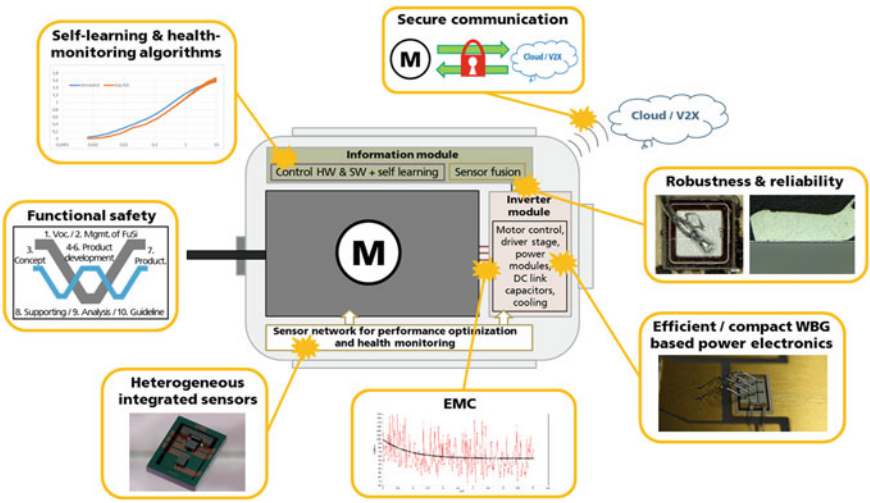


Fig. 3 Overview on selected technological challenges to be addressed in order to realize the envisaged novel system architectures for drive-trains

motor efficiency [7]. These steep pulses of high frequency and power can easily lead to severe distortions in the driver and controller electronics. The risk of interference is even higher for the IT electronics part where sensors generate signals in the range of just a few μV to be processed, evaluated, and memorized. Therefore, the protective measures and strategies for assuring EMC need to be developed further and implemented when designing smart drives of the next generation.

The thermo-mechanical risks are caused by the higher operational temperature that the WBG devices allow. This is very advantageous as it reduces the cooling efforts leading to significant reductions in size and weight of the power modules. However, it will also expose the IT electronics to temperatures close to 200 °C and beyond when they are integrated in close vicinity to the power module as foreseen by the new architectural concept. Consequently, there is a need for developing new system-in-package (SiP) solutions for the information modules, which contain the sensors, the signal processing, the memory, and the communication units for the envisaged motor control, the health monitoring and the performance optimization based on the self-learning capabilities. New integration and assembling technologies, e.g. based on silver or copper sintering or on solid-liquid interdiffusion bonding (SLID) technologies [8] for large joints (like die attach and power connections) but also for the many small joints of the IT devices (IC and passives), new molding and potting materials, new cooling concepts with a reduced number of interfaces between heat source and sink as well as solutions for inter-layer heat removal seem to be among the most pressing needs to be responded to by upcoming research projects in order to master the technological challenges.

In parallel to the hardware platform created this way, algorithms and schemes need to be developed capable of realizing the health monitoring and performance optimization based on self-leaning strategies. These algorithms and schemes need to be transferred into software codes executable on the new hardware platforms. Hence, the concept of next generation drive train opens the door to many different research activities. In total, they will contribute valuably and substantially to the realization of visions like automated driving. They lead to smart drive-trains with an unprecedented high functional safety achieved by smart health monitoring and triggering for maintenance well before the failure occurs. In addition, these drivetrains will be able to characterize and to flexibly act according to the actual service conditions—even involving self-learning strategies for improving this performance continuously. These are essential enablers paving the way to automated driving.

4 Conclusion

In this paper, a novel concept for next generation drive-train architectures has been introduced based on a modular and ultra-compact design with extended information and communication technologies. On the one hand, due to a novel sensor network and implemented health-monitoring algorithms, fault susceptibility and up-time can be significantly improved. On the other hand, pattern recognition capabilities will

enable performance optimization based on self-learning and cloud computing. Furthermore, technological challenges such as WBG devices, efficient cooling as well as reliability and EMC issues have been discussed.

References

1. Günstigere Elektrofahrzeuge durch motorintegrierte Leistungselektronik – Start für Projekt EMiLE, press release, Fraunohfer IISB, 5th Aug 2013
2. Rzepka S, Otto A (2013) COSIVU—compact, smart and reliable drive unit for fully electric vehicles. Micromater Nanomater 15:116–121
3. Gustafsson T, Nord S, Andersson D, Brinkfeldt K, Hilpert F (2014) COSIVU—compact, smart and reliable drive unit for commercial electric vehicles. In: Fischer-Wolfarth J, Meyer G (eds) Advanced microsystems for automotive applications 2014—smart systems for safe, clean and automated vehicles. Springer, pp 191–200
4. Andersson RD et al (2016) COSIVU—compact, smart and reliable drive unit for fully electric vehicles. In: Proceedings of SMTA pan pacific microelectronics symposium, Big Island, HI, USA, 25–28 Jan 2016
5. Kern J, Thun C, Herman J (2014) Development of a solid-borne sound sensor to detect bearing faults based on a MEMS sensor and a PVDF foil sensor. In: Fischer-Wolfarth J, Meyer G (eds) Advanced microsystems for automotive applications 2014: smart systems for safe, clean and automated vehicles. Springer International Publishing, pp 201–211
6. Frankeser S, Hiller S, Wachsmuth G, Lutz J (2015) Using the on-state-Vbe,sat-voltage for temperature estimation of SiC-BJTs during normal operation. In: PCIM Europe 2015, 19–21 May 2015, Nuremberg, Germany. VDE VERLAG GMBH, Berlin Offenbach, pp 132–139
7. Muhsen H, Hiller S (2015) A new simplified space vector PWM scheme for two-level voltage-source inverter. In: PCIM Europe 2015, 19–21 May 2015, Nuremberg, Germany, pp 472–478
8. Dudek R, Sommer P, Fix A, Rzepka S, Michel B (2013) Reliability issues for high temperature interconnections based on transient liquid phase soldering. In: Proceedings EuroSimE 2013, art. no. 6529908, Wroclaw, Poland, 15–17 Apr

Embedding Electrochemical Impedance Spectroscopy in Smart Battery Management Systems Using Multicore Technology

Eric Armengaud, Georg Macher, Riccardo Groppo, Marco Novaro, Alexander Otto, Ralf Döring, Holger Schmidt, Bartek Kras and Slawomir Stankiewicz

Abstract Improving the range of electric vehicles is a key requirement to achieve market acceptance—one important aspect is the efficient management of the electric energy stored into the cells. Within the INCOBAT (INnovative COst efficient management system for next generation high voltage BATteries, http://www.incobat-project.eu/.

E. Armengaud (✉) · G. Macher
AVL List GmbH, Hans List Platz 1, 8020 Graz, Austria
e-mail: eric.armengaud@avl.com

G. Macher
e-mail: georg.macher@avl.com

R. Groppo · M. Novaro
Ideas & Motion SRL, Via Moglia 19, 12062 Cherasco (CN), Italy
e-mail: riccardo.groppo@ideasandmotion.com

M. Novaro
e-mail: marco.novaro@ideasandmotion.com

A. Otto
Department Micro Materials Center, Fraunhofer Institute for Electronic Nano Systems ENAS, Technologie-Campus 3, 09126 Chemnitz, Germany
e-mail: Alexander.Otto@enas.fraunhofer.de

R. Döring
Chemnitzer Werkstoffmechanik GmbH, Technologie-Campus 1, 09126 Chemnitz, Germany
e-mail: doering@cwm-chemnitz.de

H. Schmidt
Infineon Technologies AG, Am Campeon 1-12, 85579 Neubiberg, Germany
e-mail: holger.schmidt2@infineon.com

B. Kras · S. Stankiewicz
Impact Clean Power Technology SA, UL Mokotowska 1, 00640 Warsaw, Poland
e-mail: bartek.kras@imotive.pl

S. Stankiewicz
e-mail: slawomir.stankiewicz@imotive.pl

T. Schulze et al. (eds.), *Advanced Microsystems for Automotive Applications 2016*, Lecture Notes in Mobility, DOI 10.1007/978-3-319-44766-7_19

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 608988.) project, we propose innovative and cost efficient battery management systems for next generation high voltage batteries targeting a more accurate estimation of the battery state over the lifetime. Target of this paper is to present the project's achievements with respect to (a) deployment of safe and secure multicore computing platforms, (b) embedding of electrochemical impedance spectroscopy algorithms, (c) combined thermo-mechanical robustness tests, and finally (d) outlook towards integration of INCOBAT technology into a vehicle demonstrator.

Keywords Multi core · Electric vehicle · Battery management systems · Electrochemical impedance spectroscopy

1 Introduction

The introduction of multicore computing platforms for automotive applications is opening new opportunities for both evolutionary and disruptive innovations. In the context of the INCOBAT project, we target to make electric vehicles more attractive for the end user by optimization of the high voltage (HV) battery and more especially of the battery management system. Hence, the goal of INCOBAT is to provide innovative and cost efficient battery management systems for next generation HV-batteries. The proposed innovations are relying on the introduction of multicore CPU for the automotive domain, which enable advances for model-based estimation of the battery status, as well as embedding of complex analysis methods such as electrochemical impedance spectroscopy (EIS). The improvement—respectively embedding—of these control strategies leads to higher estimation accuracy of the energy available in the battery, and therefore higher real vehicle range as well as cost reduction by correct sizing of the components.

The proposed paper is focusing on the "Which smart systems maximize electric range?" topic of AMAA 2016, including safety, security and reliability considerations for next generation batteries. In particular the following project achievements are presented: Sect. 2 introduces the deployment of safe and secure multicore-based computing platforms for next generation control strategies for battery management systems (BMS)—respectively for a higher degree of system integration, while taking into consideration non-functional requirements such as safety and security. Then, Sect. 3 focuses on embedding of electrochemical impedance spectroscopy algorithms in order to provide more comprehensive information on cell status during operation, finally leading to a more accurate use of the energy available in the cells over lifetime. In Sect. 4, combined thermo-mechanical stress tests are introduced in order to increase the accuracy of robustness evaluation in more realistic scenarios. Section 5 provides an outlook towards the integration of INCOBAT technology into a vehicle demonstrator and the expected benefits, while Sect. 6 concludes this work.

2 Deployment of Safe and Secure Multicore-Based Computing Platforms

2.1 Migration of BMS Control Strategies to Multicore Platforms

The deployment of multicore systems for embedded automotive systems with real-time constraints and safety-critical software functionalities introduces a multitude of fundamental concurrency issues and requires additional efforts to migrate software functions (mostly developed for single core systems) to parallel processing environments [1]. Multicore systems are parallelizing execution of functions, which can jeopardize safety-critical application's most important characteristic: determinism. Therefore a rigorous mitigation strategy for using multicore systems in a safety-critical context is required [2] and skilled personnel and comprehensive understanding of parallel programming techniques are essential.

To maximize the benefit from multicore computing platform, efficient partitioning of software is a key. Therefore, the functionality of the software has to have a parallelizable nature or must be integrated in such way as to not interfere with functionalities of other control units.

Besides this, correct software implementation is the key to operation of any safety-relevant system. Thus, multicore systems definitely increase the significance of already existing shortcomings. Appropriate software development includes settled software design processes, suitable software tooling, and a good understanding of processor architecture and resulting application dependencies coming from the parallelization.

2.2 INCOBAT Multicore Development Framework

The main objective for the definition of the development framework is to provide a common basis for the application software development within the project. This includes the definition of a layered software architecture, including components and the interfaces of the layers, providing access for the applications to the underlying hardware capabilities as e.g. processor cores of the microcontroller, memory, IOs and communication blocks. The design of the architecture is strongly influenced by established automotive software approaches like AUTOSAR and OSEK.

The INCOBAT software system follows the basic concepts of AUTOSAR, in particular by the definition of a layered architecture, enabling as far as possible an abstraction of the BMS application software from the controller hardware (e.g. processor cores, memory, inputs/outputs and communication capabilities). Further, a development constraint is the usage of an OSEK compliant real-time operating system. Figure 1 shows the basic structure of the INCOBAT software architecture, consisting of the following layers:

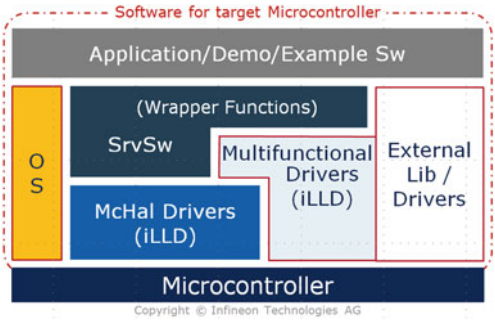


Fig. 1 Overview software architecture

- ***Microcontroller Hardware abstraction layer (McHAL)***: Realized by the Infineon Low Level Drivers (iLLD) which have been specifically developed for the AURIX™ family to provide a—hardware independent—access to the MCU.
- ***Service software (SrvSW)/Wrapper***: The SrvSW offers services for wrapping application specific calls to lower layer software components, e.g. for memory, CPU states or communication.
- ***External Drivers***: Similar to the AUTOSAR complex drivers concept, drivers based on an external library can be used to access the controller hardware if the iLLD McHAL and multifunctional drivers functionality is not sufficient.
- ***Operating system***: The SW architecture will be based on ERIKA-OS, a free of charge, open-source, OSEK/VDX, hard real time operating system which supports the AURIX multicore capabilities.

The process of customizing existing application software for multicore systems and allocating the software onto different cores is the second key topic in multicore development and still a tough and tremendously important task of multicore system migration. Automatic supporting tools, such as parallelizing compilers or automatic schedule generators, are generally in early development stages and still frequently not sufficient in parallelizing codes [3].

A first intuitive approach of parallelizing the software based on identification of the type of parallelism (data and/or task parallelism) and decomposing the software architecture into independent parallel parts would be easily applicable for BMS functions (due to the easily parallelizable nature of estimating battery cell states for multiple cells). Instead, a more general approach for automotive systems (migration of multiple control functions onto one multicore system) has been chosen for the INCOBAT project. This approach also features multicore synchronization efforts, freedom from interference argumentation of cores and hosting of mixed- criticality software functionalities and enables the optimal usage of computation power for computationally intensive functions.

2.3 Hardware Safety and Security Approach

Within the scope of the ISO 26262 safety standard, certain E/E systems such as the BMS will be classified in upper safety categories ASIL C to ASIL D. This corresponds to a fault detection rate of at least 97–99 %. The safety- and security requirements for the BMS represent new challenges for the development and the application. Consequently this matter was considered continuously from requirements definition to verification. According to the objectives of the INCOBAT project the safety item was defined as "Battery Management System (BMS) for purely electric vehicle".

The safety and security concept of the CCU for the defined safety item is developed on basis of the INCOBAT system architecture specification which includes the assumption that the BMS communication (powertrain CAN) is not directly connected to the external environment of the vehicle. These connections are handled via a dedicated gateway which secures communication. For ensuring secure charging and billing with smart-chargers no specific standards or requirements are needed for the vehicle BMS nowadays because the charging station handles this completely. However specifics for unambiguous identification of systems/components and encryption features are considered for future application of the BMS in connected vehicles. The iBMS CCU is prepared to support these requirements by the on-chip security module of the AURIX™ multicore processor offering features such as firewall, AES encryption and protected memory.

According to the bespoken agreements a security aware hazard analysis and risk assessment was performed. For 6 operating conditions such as parking, driving, etc. hazards were identified and evaluated. In total, 52 hazardous situations were identified and classified, and safety goals fully in line with the ISO 26262 standard were assigned. Additionally, 37 security threats have been identified using the SAHARA approach [4], 18 of those security threats have been classified to having possible impacts on safety concepts.

Following to this, the system FMEA has been executed. For that purpose the 35 system level failure modes most likely to occur have been identified and their impact on component and vehicle level was determined. After calculating the Risk Priority Number (RPN) of these failure modes a mapping to counter measures has been done. Components, such as the battery monitoring ASIC (BALI), BALI satellite boards, AURIX microcontroller and safety power supply ASIC were developed as safety element out of context and not detailed further in the system FMEA.

Based on the findings of the FMEA a comprehensive set of countermeasures was defined and implemented. Figure 2 shows the hardware of the iBMS CCU, which was designed to support manifold safety measures, and which comprises the chosen AURIX multicore processor prototype (TC275 in Fig. 2), its safety power supply companion ASIC prototype and the integrated high voltage area with galvanically isolated connections. It is ensured that there is at least one countermeasure for each safety goal and security target.

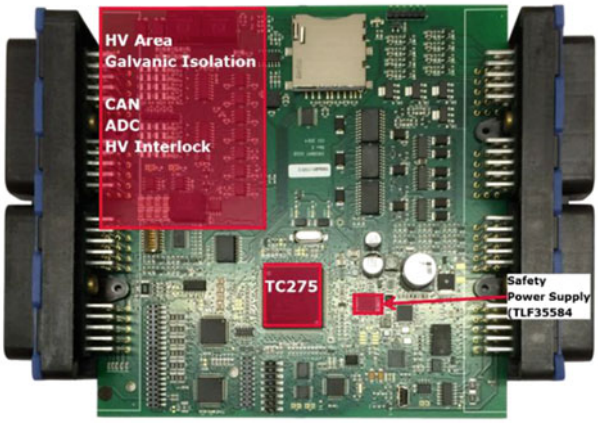


Fig. 2 iBMS CCU hardware

The main cornerstones of the safety concept are the TC275 AURIX multicore microcontroller and its safety power supply companion ASIC TLF35584. AURIX as a microcontroller element covers generic processing related safety requirements through appropriate hardware safety measures. The TC275 was chosen because it offers 3 cores, two of which providing a so-called lockstep core. Thus the different applications can be assigned to the different cores according to their criticality. For the INCOBAT BMS it was decided to run the most critical part, the BMS control on core 0, less critical parts like the monitoring applications on core 1 while the uncritical EIS is assigned to the third core which does not have a lockstep core (see Fig. 3). The three parts do not interfere with each other because of the AURIX memory and port isolation features.

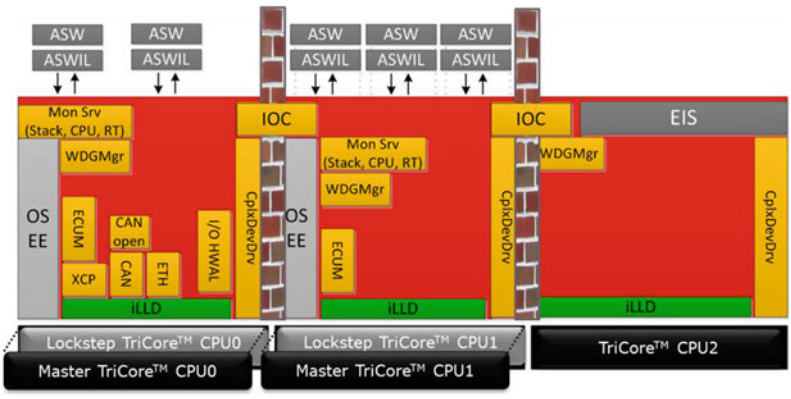


Fig. 3 AURIX cores and software isolation capabilities

The most dangerous fault sources in battery systems are danger from undetected erroneous high voltage at the vehicle chassis and danger from fire or explosion of the high-voltage battery. With regard to safety, the main switch (main relay) plays an important part in avoiding high voltage related accidents and ensuring that the BMS electronics have an adequate fault reaction. The iBMS CCU integrates a control of the HVIL to address this requirement. As for most of the critical situations it is required to monitor information of the high voltage grid of the vehicle and to disconnect the HV battery. To achieve this safe state the iBMS CCU main board is equipped with the necessary hardware such as high voltage switches and isolated connections to the high voltage grid of the vehicle.

The design of the iBMS CCU ensures that the non-critical fail-safe condition is met. This condition is always characterized by the fact that if the BMS microcontroller (MCU) fails, an independent external safety element such as a window watchdog provided by the safety power supply companion ASIC TLF35584 ensures that the main switch relay also reliably opens both high-voltage contacts to the inverter (plus/minus), even if the controller logic fails completely.

3 Embedding EIS in Automotive Control Units

The electrochemical impedance spectroscopy (EIS) methodology measures dielectric properties of a medium as a function of frequency. In particular, as applied to the battery cells, the goal of the EIS is to determine the impedance parameters, and consequently to enable an assessment of the state of health (SoH) of the cells via an analysis of the impedance, see [5, 6]. Challenges for the integration into an embedded platform are (a) the generation of an appropriate stimulus, (b) the appropriate sensing of the response and finally (c) the computing-intensive analysis of the response.

The EIS algorithm injects a known current stimulus into the battery cell, while reading the voltage response at the same time. The resulting waveform from the battery typically has a DC offset, harmonic distortion components, and noise components generated by the cell. Nevertheless, all of the spurious components of the measured signal need to be rejected so that accurate measurements of the fundamental signal at the generator frequency can be made. The effectiveness of the EIS approach on a BMS platform relies on different attributes such as calculation performances, real time capabilities, accuracy of the sensing circuit, signal generation, and the stability of the battery cell during the measurements, in terms of temperature and chemical reactions.

The measured system output is multiplied by both the sine and cosine of the test frequency ω. The results of the multiplications are then fed to two identical integrators, where they are averaged over T seconds. As the averaging time increases, the contribution of all unwanted frequency components vanishes, and the integrator outputs become a constant value which depends only on the gain and phase of the system transfer function at the test frequency. The result of the correlation process is

made up of two components one of which is referred to as the Real (or in phase) component, the other is the Imaginary (or quadrature) component. By performing simple mathematical operations on these raw measurement results, it is possible to obtain the magnitude and phase of the impedance.

The research in the INCOBAT project led to the downscaling of the EIS algorithm, allowing the execution of the EIS analysis on embedded processors, with simple calculations that can be performed in real time. This approach will overcome some of the drawbacks with classical algorithms, based on FFT computations, for instance:

- "Leakage" effects at the beginning/end of the measurement window
- The large amount of data coming from FFT results

Furthermore, the validation of the EIS algorithm in the INCOBAT project has been supported through the development and implementation of dedicated hardware boards, see Fig. 4. Although most of the research on this topic has been done with the target of the lab testing, a parallel research branch was carried out, to investigate possible industrial implementation of the needed hardware interfaces inside the chip managing the active balancing. In fact, the proposed hardware architecture based on heterodyne modulation has the advantage of being easily integrated into the next generation of smart drivers for cell-balancing: this will provide a very cost effective solution and could be considered the first attempt to embed EIS on practical automotive solutions.

First tests of the EIS algorithm demonstrate that the module impedance variation with the System on a Chip (SoC) is extremely low (i.e. about 0.1 mΩ) in the low frequency range (about 1 Hz): therefore, it is nearly impossible to be measured. This variation is also dependent on the cell technology and thus the possibility to extract useful information about the SoC through EIS is rather limited. Instead, the EIS algorithm could be useful to extract the cell core temperature as there is a

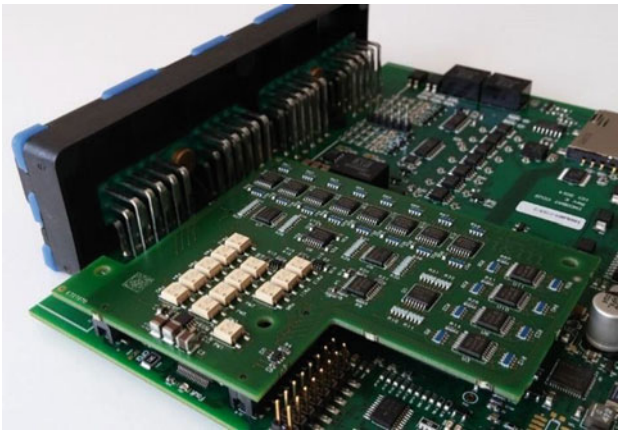


Fig. 4 Dedicated expansion board for electrochemical impedance spectroscopy

significant correlation; the expected accuracy could be in the range of ±5 °C. The effective cost saving of such a solution should be assessed, but this is a promising solution.

On the other hand, with a better estimation of the battery state of health, the EIS algorithm is expected to deliver several benefits, becoming an important enabler for more accurate use of the HV battery over the lifetime:

- reduction of number/size of cells by better battery health management while achieving the same lifetime and mileage
- added value by advanced diagnosis and predictive maintenance
- large scale creation of knowledge database on battery behaviour for further optimization of cell technology
- 2nd life: higher accuracy of battery, module and cell health assessment, therefore higher (more accurate) remaining value

4 Thermo-Mechanical Stress Investigations

Besides function development, an important focus within INCOBAT is set onto the capability for industrialization of the results. This aspect is mainly related to the comprehensive understanding on the maturity of the different components as well as the associated methods to assess the component behaviour and non-functional requirements (e.g., safety, reliability). The introduction of combined thermo-mechanical stress investigation enables faster and more realistic robustness tests of the electronic components.

Electronic components used in automotive applications are known to be exposed to much rougher conditions as for example consumer electronics. Consequently, the extraction of reliability information for these components is an important part of the development process. Therefore, the determination of reliability risks and mechanisms and their allocation to specific construction situations is an essential part of the research.

Derived design rules are an important goal. With the acquired data, reliability models are generated and validated which allow reliable predictions and risk evaluations for the given requirements.

Key points of the reliability assessment during the design phase are:

- Ensuring functionality of the modules during development phase;
- Environmental tests by considering different load cases (temperature, vibration and humidity);
- Lifetime tests to ensure the electrical functionality during the module life time.

The results are not only valid for this specific project but can be applied to all harsh environment packages in principle.

4.1 Ensuring Functionality of the Modules During Development Phase

The generation of derived design rules for the development of modules is a central target. An important method is the use of numerical multi-field simulations based on sophisticated material and damage modelling. In order to achieve trustworthy thermo-mechanical reliability predictions, highest standards are to be observed in the simulation:

- Realistic geometry model with consideration of the existing assembly technology;
- Acquisition of temperature-dependent material properties;
- Considering realistic environmental conditions and factors of influence;
- Intensive adjustment of the simulation results with experimental tests.

For example, the influence of the balancing resistors (i.e., their power losses) on the measurement chip was simulated for the BMS satellite board. Balancing resistors (Fig. 5) are used to dissipate energy from cells with higher charge levels. The temperature behaviour of the overall BMS board depends on the location of the balancing resistors, balancing power losses and number of cells as well as on the intrinsic power losses of the measurement chip. With this information the thermo-mechanical stress within the package of the measurement chip (which has influence on measurement accuracy) was investigated and optimized [7].

The simulation results were verified and adjusted by experimental tests (Fig. 6). This method supports the goal of "first time right" during the design phase.

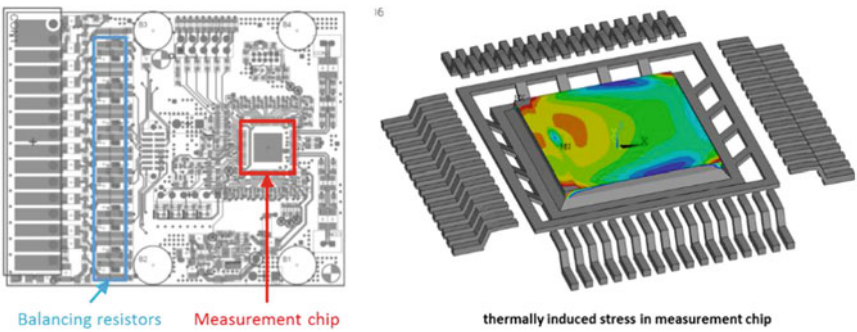


Fig. 5 Thermal-mechanical behavior of the BMS satellite board during operation

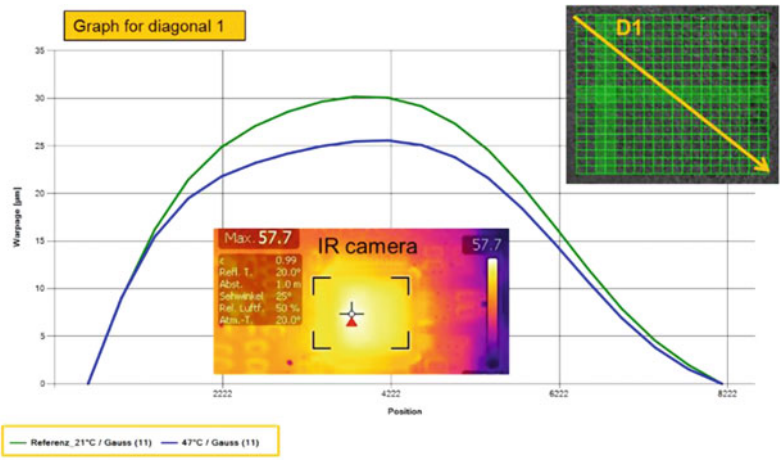


Fig. 6 Experimental verification of numerical simulation

4.2 Environmental and Lifetime Testing

Environmental and accelerated lifetime testing are an essential part of the technological qualification process in order to check and validate the functionality over lifetime by considering all relevant external (and internal) load conditions. Besides the performance of standard testing procedures such as thermal shock (−40/125 °C) or vibration tests the projects aims to also investigate new multiple-load testing strategies. The goal is to induce more realistic failure mechanisms and at the same time to reduce the testing period by applying multiple load factors, such as thermal cycles, vibration and humidity, simultaneously. The main development activities are currently performed on dedicated test boards with various daisy-chain structures for in situ monitoring of e.g. solder joints during testing. Finally, selected tests schemes will also be applied on a limited number of INCOBAT BMS electronic boards in order to verify the test results gained on the test boards.

5 Outlook: Demonstrator Vehicle Integration

Validation of the INCOBAT technology with a vehicle demonstrator is an integral part of the INCOBAT project to demonstrate the project outcomes in a realistic environment. For that purpose, a Renault Twingo 2004 had been modified as eVehicle running with battery pack, see Fig. 7. The battery pack layout has been adapted for the purposes of the INCOBAT project. Especially, the battery has been split into two parts located at front and rear of the vehicle. There were two reasons for that: first, there was not enough space in any of available compartments for the full battery, and there was an internal requirement for not installing batteries inside the cockpit. Second, the

Fig. 7 Battery pack in demonstrator vehicle (*left* front, *right* rear)

center of gravity of the vehicle needed to be shifted comparing to combustion propulsion. The demonstrator vehicle is an illustration that the INCOBAT designing procedure can be easily implemented into existing constructions.

The new INCOBAT BMS enables the setting up of larger battery systems due to the high robustness of data transmission line against electromagnetic compatibility (EMC) disturbances, therefore a division of the battery into several distinct packs becomes possible. Dividing the battery is very important for medium-sized and large energy storage systems, and high demand on battery capacity leads to more distributed constructions which are more vulnerable to EMC issues. Uninterrupted data flow is necessary for estimation algorithms and safety.

Elements inside the housing according to the United Nations Regulation No. 100 (UNR100) are deployed to ensure adequate electric insulation. Connections between the components were made in maximally simplified way. Cable cross sections have been selected in compliance with PN-EN 60664-1 standard. To meet a high degree of IP protection the battery pack housing has been sealed. The estimated degree of protection is IP55. Battery packs have been designed to meet the demands of the regulations (UNR100); detailed tests are ongoing.

The multicore technology as used for the main controller is perfectly suited for implementation in vehicles with respect to new eAutomotive safety standards as well as for other customized installations. Especially the memory and port isolation (inherited from the safety mechanisms and software architecture) is important when many parties are building the code together. Assigning separate resources and memory areas makes collaboration easy, and tasks can be completed remotely. IP protection can be also obtained in a simple way.

6 Conclusion

Accurate state estimation is required for optimal use of the energy stored in the battery. At the same time it is a challenge since it requires appropriate monitoring of the cell status and comprehensive models to be computed in real-time in the

embedded computing platform for cell analysis. The introduction of a multicore computing platform as well as the implementation of electrochemical impedance spectroscopy are two important enablers for increasing the observability of the cell status and enable the computation of more accurate, while more complex, control strategies. The introduction of thermo-mechanical stress investigations is an important method to effectively evaluate the robustness of core components such as the BMS satellite unit, responsible for measuring the cells directly within the battery pack.

References

1. Macher G, Hoeller A, Armengaud E, Kreiner C (2015) Automotive embedded software: migration challenges to multicore computing platforms. In: 13th IEEE international conference on industrial informatics
2. Brecht B, Luys S (2011) Application of multicore CPUs in a safety-critical context Barco Defense and Aerospace
3. Reichenbach F, Wold A (2010) Multicore technology—next evolution step in safety critical systems for industrial applications? In: 13th Euromicro conference on digital system design (DSD) pp 339–346
4. Macher G, Hoeller A, Sporer H, Armengaud E, Kreiner C (2015) SAHARA: a security-aware hazard and risk analysis method. In: DATE'15: proceedings of the conference on design, automation & test in Europe
5. Tröltzsch U, Kanoun O, Tränkler HR (2006) Characterizing aging effects of lithium ion batteries by impedance spectroscopy. Electrochim Acta 51:1664–1672
6. Richardson R, Ireland T, Howey A (2014) Battery internal temperature estimation by combined impedance and surface temperature measurement. Department of Engineering Science. University of Oxford, UK
7. Otto A, Schindler-Saefkow F, Haase S, Hofer G, Armengaud E, Rzepka S (2016) Thermo-mechanical stress investigations on newly developed passive balancing board for battery management systems. AmE conference, Dortmund, Germany

Procedure for Optimization of a Modular Set of Batteries in a High Autonomy Electric Vehicle Regarding Control, Maintenance and Performance

Emilio Larrodé Pellicer, Juan Bautista Arroyo García, Victoria Muerza Marín and B. Albesa

Abstract This work proposes a method for improving performances of the energy storage system of an electric car with high autonomy, and analyzes the influences of a system with unbalanced batteries versus different types of traction control of the vehicle. The experimental procedure consisted in conducting a series of test vehicle driving while modifying the traction control. The aim is to analyze the power system behavior versus the use of different strategies vehicle traction, and to study the behavior of each battery. Furthermore, it has been developed a protocol for selective charging of the unbalanced batteries in order to optimize the charging of the whole energy storage system.

Keywords Electric vehicles · Battery · Prototype · Simulation

1 Introduction

Electric vehicles (EV) are seen by many as the cars of the future as they are highly efficient, produce no local pollution, are silent, and can be used for power regulation by the grid operator [1]. In order to fulfill vehicle power demand and driving range, battery systems in electric vehicles have to be equipped with numerous cells

E. Larrodé Pellicer (✉) · B. Albesa
Department of Mechanical Engineering, University of Zaragoza, C/María de Luna 5, 50018 Saragossa, Spain
e-mail: elarrode@unizar.es

J.B. Arroyo García
Department of Electrical Engineering, University of Zaragoza, C/María de Luna 3, 50018 Saragossa, Spain
e-mail: jbarroyo@unizar.es

V. Muerza Marín
Aragon Institute of Engineering Research, i3A, C/Mariano Esquillor s/n, 50018, Saragossa, Spain
e-mail: vmuerza@unizar.es

T. Schulze et al. (eds.), *Advanced Microsystems for Automotive Applications 2016*, Lecture Notes in Mobility, DOI 10.1007/978-3-319-44766-7_20

connected in series to achieve the required voltage level and in parallel to achieve the required energy capacity [2].

Precisely, the major drawback of EVs lies in its autonomy. Nowadays improving the accuracy of the charging and discharging model of power batteries, especially lithium-ion batteries, is a significant objective, and to estimate the batteries' state of charge and state of health [3–6]. It is therefore necessary to find the best way of proceeding to give maximum performance to the EV power system. One of these systems' main problems is the possible power imbalances of the batteries. As a consequence of these imbalances, there are limitations in the entire power system. Charge is limited by the higher voltage battery causing the others not reach their maximum voltage. Furthermore, discharge is limited by the battery with lower voltage provoking other batteries cannot continue its discharge although they have yet enough voltage.

This work analyzes the behavior of the feeding system of a high autonomy performance EV. In turn, it looks for possible improvements to achieve optimum performance of the power system to provide the best performance. This EV, called "Gorila EV", has been developed as collaboration between the University of Zaragoza and the Spanish company Zytel Automotive S.L. (ZYTEL) to participate in the ZERO RACE, where the electric vehicle will go around the world with zero CO_2 emissions. The sizing of the vehicle was made keeping in mind the following demands of career [7]:

- The estimated daily travel distance is in the range of 400–450 km. The vehicles must have a range of at least 225 km fully charged until the first refueling of outage batteries, traveling at an average speed of 80 km/h. Following a reload of 4 h, it should travel another 200 km.
- The vehicles will be charged during lunch and overnight stops. 220 V outlets and 16 A will be made available, and occasionally even a 380 V in Europe. A vehicle can have up to 3 power outlets.

The vehicle is equipped with 106 batteries connected in series; the performance of each battery is examined in charge and discharge to get a set acting as uniformly as possible, by using the Battery Management System (BMS) device. This device allows reading the state of the battery cells individually. Furthermore, the engine-controller UQM program is used to register the set overall voltage for any test performed, comparing with other parameters such as speed, temperature and energy consumed.

This paper proposes a method for improving performances of the energy storage system of an electric car with high autonomy, and analyzes the influences of a system with unbalanced batteries versus different types of traction control of the vehicle. The experimental procedure consisted in conducting a series of test vehicle driving while modifying the traction control. The aim is to analyze the power system behavior versus the use of different strategies vehicle traction, and to study the behavior of each battery. Furthermore, it has been developed a protocol for selective charging of the unbalanced batteries in order to optimize the charging of the whole energy storage system.

2 Modeling of Gorila EV's Batteries

2.1 Batteries Operation

A battery is a device consisting of one or more electrochemical cells, connected in series or parallel that converts the chemical energy contained in its active materials directly into electric energy by means of an electrochemical oxidation reduction reaction.

Modeling the behavior of batteries is complex, because non-linear effects occur during discharge. As [8] explains, the ideal capacity of a battery would be constant for all discharge currents, and all energy stored in the battery would be used. This differs in a real battery, where the voltage slowly drops during discharge and the effective capacity is lower for high discharge current and the so-called recovery effect occurs.

The battery models usually used in EVs can be divided into three kinds [9]: (i) the simplified electrochemical model, based on the electrochemical theory that could fully describe the characteristics of the power battery by using mathematics to describe the inner action of the battery; (ii) the neural network model, which took the weights of neurons into account instead of the state variables; and (iii) the equivalent circuit model, which is based on the dynamic characteristics and working principles of the battery, was developed by using resistors, capacitors and voltage sources to form a circuit network.

Many factors influence the operational characteristics, capacity, energy output and performance of a battery. These factors include: the voltage level, the current drain of discharge, the mode of discharge (constant current, constant charge, constant power), the temperature of battery during discharge, the service life, the type of discharge (continuous, intermittent, etc.), duty cycles (intermittent and pulse discharges), voltage regulation, charging voltage, the effect of cell and battery design (Electrode Design, Shape and Configuration, Volumetric Efficiency vs. Energy Density, Effect of Size on Capacity), and Battery Age and Storage Condition. The effect of these factors on battery performance is discussed in [10].

A parameter to be considered is the batteries connection. Cells connected in series provide a sum of each cell voltage to get a set with higher voltage level. This is the case of the vehicle traction batteries analyzed in this research. However, cells connected in parallel provide a battery with higher current level. The vehicle possesses two auxiliary batteries for the operation of the automaton, the cooling pump, the power steering, and so on.

2.2 Choice of Batteries

The choice of the batteries was conducted in a previous stage according to criteria of weight, maximum discharge, and engine power and cycle life of the batteries [7].

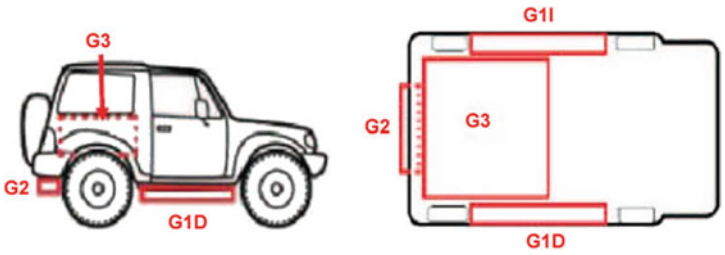


Fig. 1 Distribution of traction batteries

The selected battery was the manufacturer SE180AHA Sky Energy. These were lithium batteries 3.4 V DC, a capacity of 180 Ah and a weight of 5.6 kg. The vehicle has been equipped with 106 batteries, making a total of 360.4 V DC, with a minimum voltage of 212 V DC (considering a 2 V minimum battery voltage). The total weight of the batteries is 593.6 kg.

Due to the high number of batteries, these were distributed in four blocks or coffers (see Fig. 1), two side skirts (G1I and G1D), a rear flap (G2) and the inner drawer, in the area inside the vehicle (G3). Figure 2 shows the connection diagram of the batteries.

Furthermore, batteries are grouped into modules. Table 1 shows the location of each module, the number of batteries and the correspondence with the number of battery.

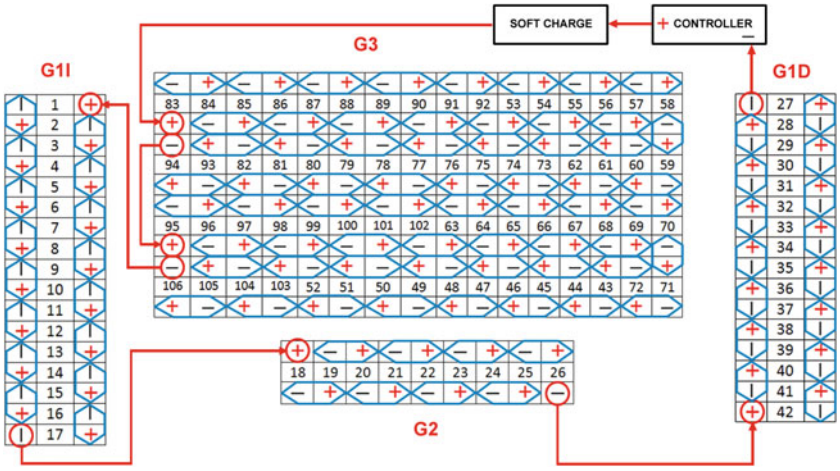


Fig. 2 Batteries' connection diagram

Table 1 Identification of the vehicle batteries

Module number	Location	Batteries amount	Number of batteries
0	G1I	10	01–10
1	G1I	7	11–17
2	G2	9	18–26
3	G1D	8	27–34
4	G1D	8	35–42
5	G3	10	43–52
6	G3	10	53–62
7	G3	10	63–72
8	G3	10	73–82
9	G3	10	83–92
10	G3	4	103–106
11	G3	10	93–102

3 Methodology

The methodology followed to conduct this research involves the following stages:

- **Stage 1**: Documentation about the power system installed in the vehicle.
- **Stage 2**: Study of the current situation of the vehicle's state.
- **Stage 3**: Driving tests. Behavior analysis of the feeding system versus the use of different vehicle parameters.
- **Stage 4**: Design of an action procedure in regards to imbalances in batteries: (i) balance, (ii) equalizing voltage in batteries, (iii) charging procedure.
- **Stage 5**: Full charge of the batteries to study their behavior.
- **Stage 6**: Vehicle test. Analyze whether there is a positive influence of the batteries balancing in the feed system.
- **Stage 7**: Design improvements.

4 Testing and Analysis

4.1 Definition of Parameters

The parameters to be considered in the tests performed in the vehicle are four:

- **Circuit**: the circuit chosen is located in one of the parking of the Faculty of Engineering and Architecture of the University of Zaragoza (see Fig. 3, route marked in red). One-lap circuit distance is 410 m.

Fig. 3 Real image of the circuit (*Source* Google Maps)

- **Temperature**: the temperature is an important factor to consider since the different tests carried out will take place under different weather conditions. Lithium batteries used are sensitive to this effect.
- **Cooling system**: in a first contact with the vehicle without using the refrigeration system, too high temperatures were reached compromising the good operation of the motor-controller system. In order to solve this problem to carry out the different tests, the operation of the cooling system was checked and the corresponding connection to the PLC was performed.
- **Gear selection**: the vehicle has two gears, long and short. The long gear allows reaching a higher speed; meanwhile the short gear provides extra power. The gearshift is performed through a lever.

4.2 Test Characteristics

Tests consist of different driving tests carried out in the aforementioned circuit with the vehicle in two possible states: (i) Vehicle in initial state; and (ii) Vehicle after a balancing of batteries and its corresponding full charge.

4.2.1 Vehicle in Initial State

These have been short-term tests because no battery behavior is known. They all have a similar duration and route, and try to simulate the driving behavior in daily life, varying both controls (torque or speed) and gear, long or short. Table 2 shows the characteristics of the three tests carried out.

Table 2 Tests with the vehicle in initial state

Test	Trial	Number of laps	Vehicle gear	Type of control	Maximum engine speed limit (r.p.m)	Temperature (normal conditions) °C
1	1	21	Long	Torque	2000	26
2	1	10	Long	Torque	2000	35
	2	10	Short	Torque	2000	35
	3	5	Short	Torque	4000	35
	4	6	Long	Torque	4000	35
3	1	6	Short	Torque	4000	33
	2	6	Short	Speed	4000	33
	3	6	Long	Speed	4000	33
	4	6	Long	Torque	4000	33

4.2.2 Vehicle After a Balancing of Batteries and Its Corresponding Full Charge

The performance of this test takes place after the full power system load, the test therefore having long-term character. It has been carried out one trial, driving a distance of 30.22 km (75 laps), during 1 h and 10 min with an average speed of 25.6 km/h. The long gear was used, with torque control, and introducing a maximum engine speed limit of 4000 rpm. The temperature was around 23 °C.

5 Results

This section shows the results obtained in the tests described above. Following the same methodology it is distinguished between the analysis of trials before and after batteries charging. The objective of these trials is to analyze the level of voltage required depending on the different kind of control used and the different gear choice, in order to see which was the less favorable condition (the maximum demand of voltage).

5.1 Vehicle in Initial Stage

Figure 4 one shows the relationship between the engine speed and the average voltage thereof in test 1. As expected, it follows an inverse relationship, when the

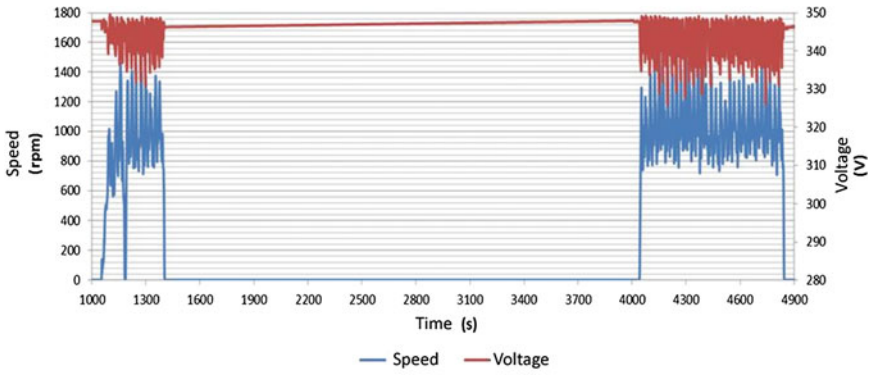


Fig. 4 Voltage curves and engine speed versus time (Test 1)

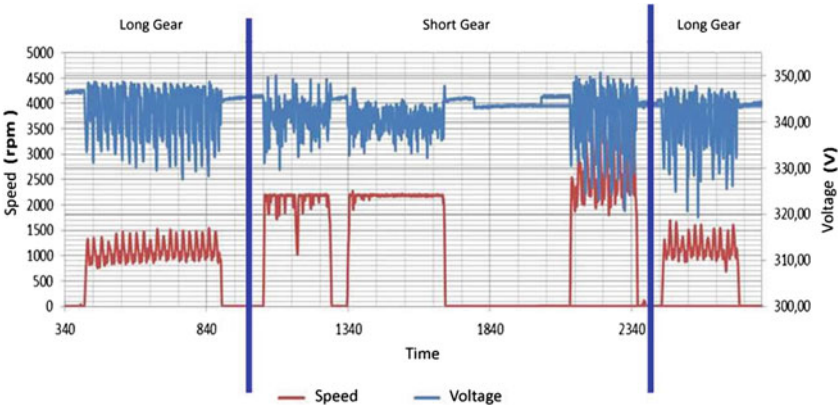


Fig. 5 Voltage curves and engine speed versus time (Test 2)

speed increases, the required current or the battery discharge is higher, and therefore the voltage in the engine decreases.

Test 2 includes a new variable: the gear change. A short gear provides more power, reaching higher rotation speeds, while a long gear allows achieving higher speed with lower engine speed (Fig. 5). Besides, it can be observed periods where the vehicle is stationary, which correspond to changes in gear and speed limit. This effect is also seen in Fig. 6.

Test 3 shows similar results than Test 2 (see Fig. 6). Higher engine speeds are given in short gear (and torque control in comparison with speed control).

Furthermore, similar results are obtained in the two trials carried out with long gear (see Table 2).

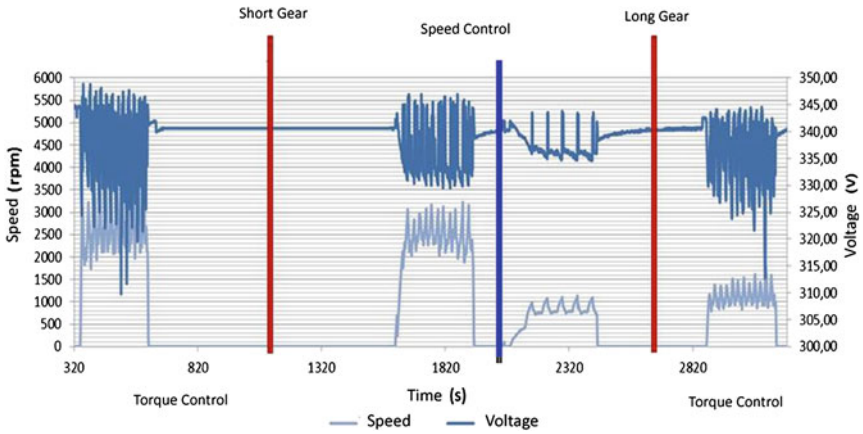


Fig. 6 Voltage curves and engine speed versus time (Test 3)

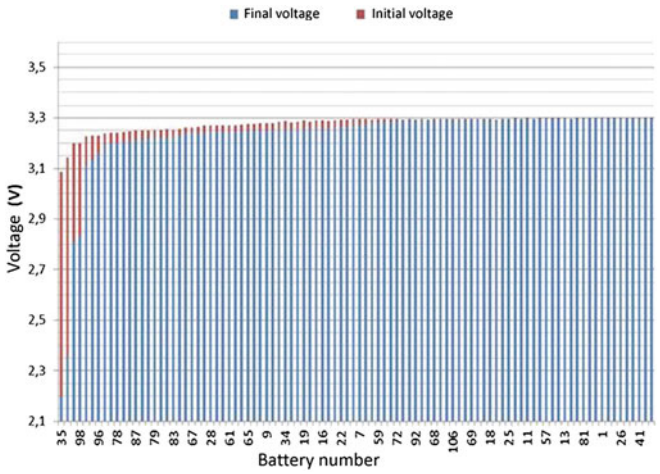


Fig. 7 Display of battery voltages during the test of the vehicle (discharge process). Ordered from lowest to highest final voltage

Likewise, there were analyzed the percentage differences in voltage of the vehicle's batteries set. Results show the existence of two different types of batteries: (i) batteries with normal behavior, and (ii) unbalanced batteries.

Unbalanced batteries increase discharge rate and recovery rate, the latter further. See for example batteries 35, 100, 98 and 84.

The remaining batteries have acceptable performance, being discharge and recovery percentages much smaller and can be used for a longer time without the need to carry out some special maintenance.

The aforementioned differences lead to imbalances (see Fig. 7).

As can be seen in the Fig. 7, some cells presented very low values of voltage. At the beginning, we did not know if the cells were damaged or some cells were deeply discharged for unknown reasons. In the next steps, the objective is to determine what the cause is.

5.2 *Vehicle After a Balancing of Batteries and Its Corresponding Full Charge*

There were analyzed the percentage differences in voltage of the vehicle's batteries set during charge. Figure 8 shows voltages values before and after charging and imbalances in some batteries with a higher initial voltage.

Furthermore, batteries were also analyzed after full charge. Because traction batteries are balanced and charged, the test carried out is longer in distance and time (30 km, 1 h 10′).

The test has been performed with torque control and long gear, providing the highest vehicle speed.

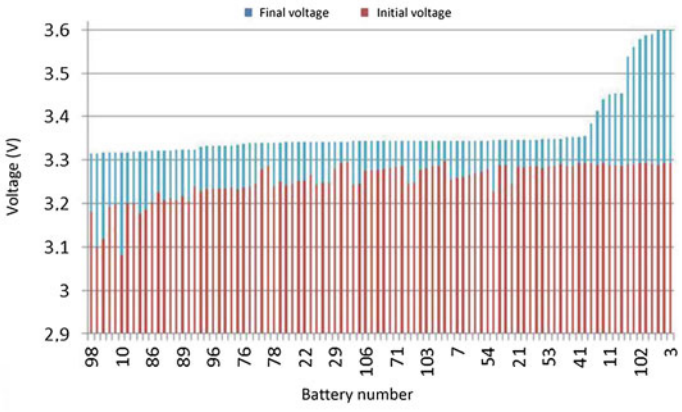


Fig. 8 Display of battery voltages during the charge (initial and final values). Ordered from lowest to highest final voltage

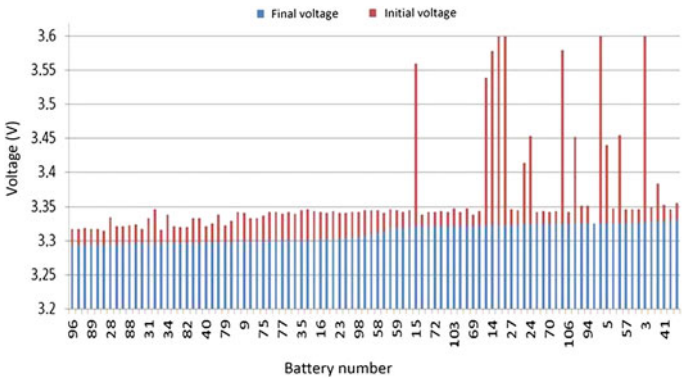


Fig. 9 Initial and final battery voltages. Ordered from lowest to highest final voltage

Figure 9 shows initial and final batteries voltage in the test performed. Results show no battery has undergone major discharge values. Also, there are no large voltage differences, being not necessary to balance batteries.

6 Protocol for Selective Charging of the Unbalanced Batteries

The results show the great importance of a uniform battery performance, so that when a battery limits discharge or load, others have similar voltage values. This section presents the development of a protocol for selective charging of the unbalanced batteries. The protocol follows 4 steps (see Fig. 10):

- **Step 1**: Analysis of the batteries voltage by using the BMS. Also, check the BMS reading by measuring the batteries voltage with the multimeter.
- **Step 2**: It is estimated a voltage value from which a battery must be matched to the remaining batteries. For doing this, it is taken as reference the set maximum value, and if this difference exceeds a percentage (fixed for example in a 14 %), its charge will be performed. However, the decision to replace one battery is taken in a conservative way, i.e., when percentage differences of 40 % are achieved in two consecutive discharges.
- **Step 3**: Locating the position of the batteries that are failing; two options are possible according to step 2 (charging or battery replacement). For the battery charging option it is considered the maximum voltage of the module in which the battery is placed. In case of a battery replacement, it is necessary the full charge of the power system since the new battery has the maximum voltage.
- **Step 4**: In case of voltage differences lower than 14 %; two options are possible: (i) Vehicle full charge (106 batteries), or (ii) Continue the discharge (in case of a high capacity in the remaining batteries).

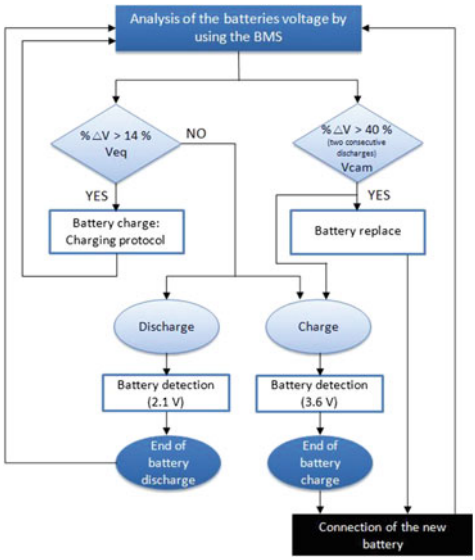


Fig. 10 Diagram of the procedure for balancing the EV batteries

7 Summary and Conclusions

The objective of this paper was to develop a method for improving performances of the energy storage system of an electric car with high autonomy, analyzing the influences of a system with unbalanced batteries versus different types of traction control of the "Gorila EV". The experimental procedure has consisted in conducting a series of tests vehicle driving while modifying the traction control. The aim was to analyze the power system behavior versus the use of different strategies for vehicle traction, and to study the behavior of each battery. Finally, the paper presents a protocol for selective charging of the unbalanced batteries in order to optimize the charging of the whole energy storage system.

From test driving and its analysis it can be drawn the following conclusions: (i) by using the same gear, when the rotation speed increases supply voltage decreases due to the necessity of more discharge current; the vehicle reaches higher engine speeds with short gear. In turn, the higher engine revolutions are given by the torque control, as in speed control the torque is limited by the value of a multimeter; the slope of the energy consumed was higher in torque control, since there is greater discharge intensity. The differences between long and short gears are not significant in torque control, the most common. However, in speed control and short gear is seen an increase of these energies. After performing these tests, we realized that the higher difference in the cells voltage occurred when we had greater discharge intensity, and this situation appeared in short gear when using torque control. (ii) There were imbalances in some batteries of the vehicle, limiting the performance of the entire set. After the studies carried out, we observed that if we

avoid imbalances between cells voltage, we obtain better charging efficiency. Future research will be focused in determining a procedure to automatically equalize the level of the cells voltage.

References

1. Schaltz E (2011) Electrical vehicle design and modeling. In: Soylu S (ed) Electric vehicles—modelling and simulations. InTech
2. Shi W, Hu X, Jin C, Jiang J, Zhang Y, Yip T (2016) Effects of imbalanced currents on large-format LiFePO4/graphite batteries systems connected in parallel. J Power Sour 313:198–204
3. He HW, Xiong R, Chang YH (2010) Dynamic modeling and simulation on hybrid power system for electric vehicle application. Energies 3:1821–1830
4. Dennis D, Evren G, Daniel A (2008) Electrochemical modeling of lithium-ion positive electrodes during hybrid pulse power characterization tests. J Electrochem Soc 155:A603–A613
5. Wang J, Chen QS, Cao BG (2006) Support vector machine based battery model for electric vehicles. Energy Convers Manag 47:858–864
6. Craig L, Paul M (1997) Development of an equivalent-circuit model for the lithium/iodine battery. J Power Sour 65:121–128
7. Larrodé E, Fraile A, Arroyo JB (2015) Design and systems integration in the electrification of an electric vehicle for long distance travel. Hierarchical multi-criteria analysis for designing the vehicle architecture. In: Schulze T, Müller B, Meyer G (eds) Advanced microsystems for automotive applications 2015—Smart systems for green and automated driving. Springer, Berlin, pp 211–222
8. Jongerden MR, Haverkort BR (2008) Battery modeling. University of Twente, Centre for Telematics and Information Technology, Technical report TR-CTIT-08–01
9. He H, Xiong R, Fan J (2011) Evaluation of lithium-ion battery equivalent circuit models for state of charge estimation by an experimental approach. Energies 4:582–598
10. Linden D (2002) Factors affecting battery performance. In: Linden D, Reddy TB (eds) Handbook of batteries. Mc Graw-Hill, New York

Time to Market—Enabling the Specific Efficiency and Cooperation in Product Development by the Institutional Role Model

Wolfgang H. Schulz and Matthias Müller

Abstract This paper discusses major and effective aspects that in future will be important for R&D management in the automotive industry and that will have an enormous impact on the entire value creation chain. Firstly, there are efficiency potentials for reducing development time in design and management of the entire product development process (for example, implementation of an IT system with real-time information, predictive analytics and total networking of all value creation partners). Secondly, the complexity involved in the new technologies and the associated networking being set up between business environment and automotive industry. For example car-to-customer, car-to-car or cooperative intelligent transport systems increasingly demand horizontal and agile cooperation among various industry branches (for example, car industry, telecommunications industry, over-the-top players) and the legislatures of the different markets. Moreover, it becomes clear that the particular public or private highway operators either support or hinder the market success of the new car technologies. Here, the urgent need for close cooperation has been recognized and already exists. However, the process with its milestones is still in a very early phase of product development. Nevertheless, the companies leading the market and the government agencies will cooperate intensely in future and thus also make application of the new technologies possible. For this reason, a form of business architecture was chosen that permits competitively neutral and non-discriminating cooperation with the stakeholders. The theory of the Institutional Role Model (IRM) created a multifunctional approach that permits a holistic type of cooperation and creates very good prerequisites for improving efficiency. The concept of the Institutional Role Model was used successfully for three national research projects (e.g. CONVERGE; Market Design for C-ITS). Within these projects the IRM was used for market phase introduction and penetration. This paper integrates the IRM early into the agile digital product development process.

W.H. Schulz (✉) · M. Müller
Mobility, Trade and Logistics, Zeppelin University Friedrichshafen,
Am Seemooser Horn 20, 88045 Friedrichshafen, Germany
e-mail: wolfgang.schulz@zu.de

M. Müller
e-mail: matthias2.m@web.de

T. Schulze et al. (eds.), *Advanced Microsystems for Automotive Applications 2016*, Lecture Notes in Mobility, DOI 10.1007/978-3-319-44766-7_21

This process redesign will enable managers in the automotive industry to generate unique optimization opportunities in future.

Keywords Agile digital product development process · Market entry · Spheres of activity · Development time · Institutional rule model

Abbreviations

BPR	Business Process Reengineering
CEO	Chief Executive Officer
CIM	Computer Integrated Manufacturing Method
C-ITS	Connected Intelligent Transportation System
DAPDP	Digital Agile Product Development Process
GM	General Motors
ICT	Information Communication Technology
IRM	Institutional Role Model
IT	Information Technology
MIT	Massachusetts Institute of Technology
OEM	Original Equipment Manufacturer
R&D	Research and Development
SCM	Supply Chain Management
US	United States

1 Introduction

With the steadily increasing expectations made of new technologies and innovations the pressure on companies and on industry networks also increases. Increasing competition and the change in customer behavior call for the manufacturer to develop products in increasingly shorter times [1]. What will this change, which will have such a great effect on people's lives, look like? Throughout the world the innovation drivers will start a revolution, the scope of which we cannot estimate today [2]. The availability of data and information will cause a paradigm shift in society and will restructure the working world. Historic antiliberal circumstances will not be compensated with antiliberal prohibitions in the future [3, p. 18]. The digital change offers many chances especially for the internet and telecommunication industry and the automobile branch, these developments are another decisive step towards connected self-driving cars. As Mary Barra, Chairman & CEO GM stated at the 16th Car Symposium in Bochum: "The automotive industry will see more changes in the next five years than in the previous 50 years" [4]. Today's automotive world is characterized by new mobility expectations on the part of customers and by intelligent

business models [5]. These include new technologies such as electric drive systems, fuel cells and the networking of cars and the environment. These changes demand infrastructure activities as well as additional investments in hardware and software as well as in organizational models for the participating value adding partners. For example, Daimler is planning to invest € 500 million in networked systems by the year 2020 [6]. Important success factors on the path to high product returns are the utilization of the various skills and expertises of all the value adding partners involved in the product development process. The ongoing digitalization in the industry and the coming data flood will make it necessary to create a new design for the business and cooperation processes. Available knowledge and information must not be lost again, and expensive time spent looking for documents must belong to the past. That might sound radical, but digitalization has made business reengineering indispensable in the product development process [7, p. 32]. The question arises whether redesign of the product development process will lead to an acceleration of innovations in the automobile industry and thus to an acceleration of market introduction. From the authors' point of view, two phenomena have to be considered. On the one hand, the number of events that lead to innovations must clearly increase in future and, on the other hand, only by shortening the innovation phase and thus permitting earlier market entry can we ensure additional chances to create value. The latter phenomenon will, namely by means of shorter product life cycles/market phases, drive the manufacturers to reduce their development times by adopting a reengineered and digital development process. The motive is relatively simple: both drivers have a strong influence on pricing as well as on the time window for the realization of profits and thus ultimately on the company's profit or even its survival [8]. The study "Industry 4.0: How do IT systems change in purchasing and SCM?" [9, p. 4] showed that companies are planning to increase their implementation of IT systems for planning, coordination, transparency and control and are also planning to intensify process automation. According to researchers, 47.8 % of companies have no industry 4.0 strategy and only 9 % have a strategy. Technological progress in the form of digitalization is not to be stopped and will connect cars with their environment. Audi CEO Ruppert Stadler said in early 2016 that not the emissions scandal will decide the future of Audi, but Audi's ability to change with digitalization. He said: "One day we will indeed earn half of our turnover in these new fields" and also said the coming competition will be brutal [10]. What procedure models will permit the necessary change to be made in the economy and what are the indicators for a necessary redesign of product development? One approach to answering this question could be the comprehensive digitalization and fundamental redesign of the product development process under consideration of cross-functional cooperation. This would lead to optimization of the company processes and thus possibly offer chances to enter the market sooner. Suitable approaches that are discussed in scientific literature in this context will be examined within the following chapter.

2 Institutional Economic Role Model as Methodological Approach

The Theory of Institutional role models is mainly based on the principles of institutional economics [11], system theory [12] and theory of system dynamics [13]. The theoretical concept of the institutional role models is based on Schulz [14]. Within a research project concerning operator models for the introduction of C-ITS for the German Federal Highway Administration, the approach of Institutional Role Models was conceptually enlarged to a theory, that was completely applied within the CONVERGE project [15]. In order, ambiguities and wordplays like "the enterprise is an institution with an organization" is to be avoided and the definition by Schneider [16] is constantly used instead: Institution is the generic term for monitoring systems (orders) and action systems (organizations). Figure 1 illustrates the structure and the process for the development of a role model.

Product development is one of the key roles within the IRM-process because product development is ambiguous: it is both, a technical role and an economic role. Therefore, the number of relevant institutions is high whereby trade-offs between technical and economic objectives exist. Altogether, the complexity of product development processes increases. This paper is a first step to give up a deeper in-sight, which institutions are relevant and why these institutions should coordinate their activities. Therefore, this paper is a first step for the IRM-process to co-ordinate the activities of various stakeholders to accelerate the product development process.

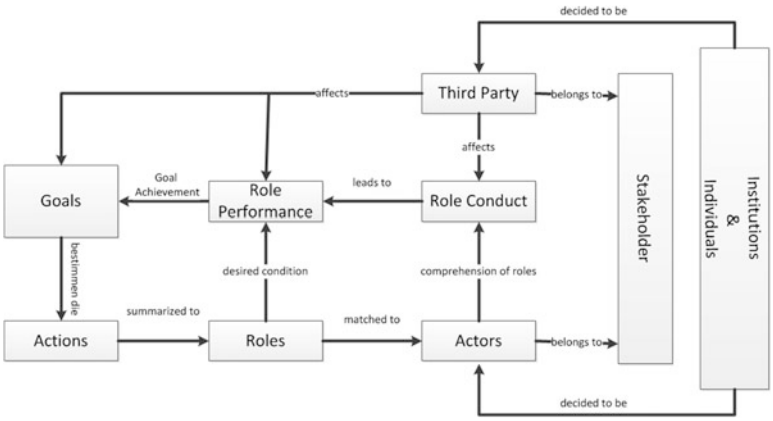


Fig. 1 Structure and process for the development of a role model. *Source* own research author Schulz, W.H.

3 Literature Review—Hypothesis Development

Already in 1984 the Massachusetts Institute of Technology (MIT) was commissioned to study industry use of ICT: "Why does the use of IT result in such small increases in productivity in the industry" This study ("The Corporation of the 1990s—Information and Organizational Transformation") by Michael S. Scott Morgan was published in 1991. Unfortunately, the reengineering study published as "Business Reengineering" only came into public view in 1993 [17, p. 14]. "Simultaneous Engineering" is another method for increasing the efficiency of development projects that has been subject to scientific scrutiny. The aim of this model is to optimize the development time, development costs and quality demands involved in product development [18, p. 126]. The following figure illustrates the advantage to productivity:

Effective implementation of the method shown in Fig. 2 requires up front the elimination of barriers to cooperation for cross-functional teams. The work of Sherman et al. [20] entitled „Differential effects of the primary forms of cross-functional integration on product development cycle time" examined the influence of five important fields (marketing, customer, manufacturing, supplier and strategic partnerships) and their effects with regard to product development time. The results show that R&D integration and knowledge management from previous

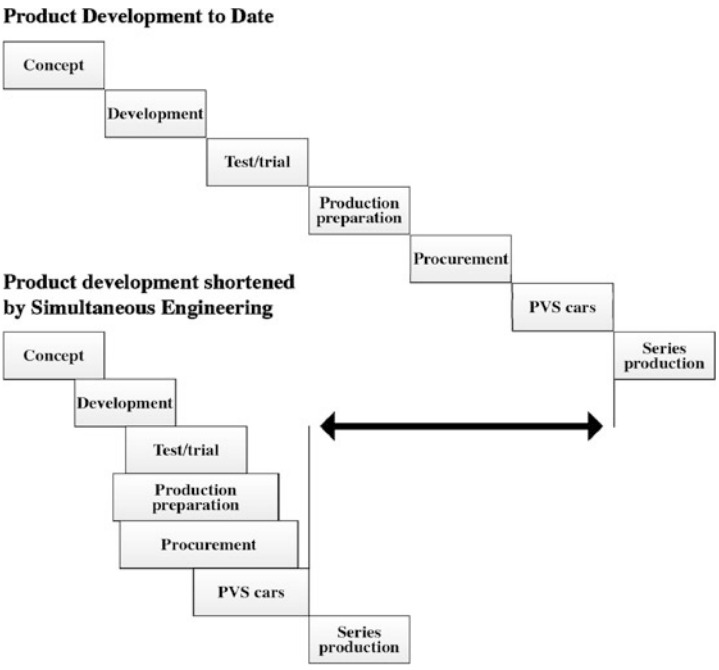


Fig. 2 Advantages of product development achieved by simultaneous engineering [19, p. 18]

projects were the greatest efficiency levers for product development time. Another interesting research approach is afforded by the Computer Integrated Manufacturing Method (CIM), which aimed to sensibly network all company data over all functions of the product development process [21, p. 18]. At that time, implementation was difficult for larger companies due to the large amount of data to be handled and the historically grown data processing systems. This change is taking place in the framework of industrialization 4.0 calls for a radical rethinking of business process reengineering by car manufacturers. In markets with a high rate of technological change the product life cycles are characterized by increasing competition and as a rule become shorter and shorter. Where do we find decisive, economic levers and how can theoretical models (such as CIM or BPR) promote new strategies in product development? Decisive factors in setting the course for shortening the innovation phase and thus the time to market are time, organization, communication and resources [22, p. 80]. Optimal interplay of the input variables can lead to significant increases in the efficiency of product development. Basically, once the decision is made for the "right" innovation, an increase in efficiency in the development projects can be expected. These will no longer be simple cost savings in one's own particular branch, but instead the quality of added value in products, processes and technologies will cover all aspects of life. In this context industrial networks will increase in importance, because they can be expected to produce numerous advantages, for example in time, know-how, expanding market access, resources and competition as well as cost advantages for company projects [23, p. 788]. What mechanisms must be at work in R&D in the automotive industry if ideas are to give rise *faster* to innovations incorporating the digital world? Is there enough research input on how a future development process can look and where will the costs be decided? In this connection it is scientifically proven that already in the early phase of the product development process a considerable portion of the costs have already been set [24, p. 20]. The costs over the life cycle of a product change exponentially. For example, a change of € 1—during task definition means a change of € 10—during design, € 100—during process engineering, € 1000—during production and € 10,000—after delivery [25, p. 14]. According to Wildemann [26, p. 6], an increase in the market efficiency of products can have two main causes, for example setting the right goal or achieving a goal well. In this context, we can also mention effectiveness and efficiency. The following figure once again shows the difference and the possible optimization levers.

The areas marked (R&D Management) in Fig. 3 show where efficiency can be improved in product development. In the following we will take a closer look at these efficiency levers. What are the product development processes currently used in the automobile industry and where can we see possible efficiencies? Today's product development processes do not vary significantly among themselves, but the various car makers differ with regard to task sharing, milestones and flow times. The following table gives an overview of how various car makers have structured their product development processes.

Table 1 demonstrates various aims of product development processes and hereby is showing the differences within the automotive industry. This table clearly

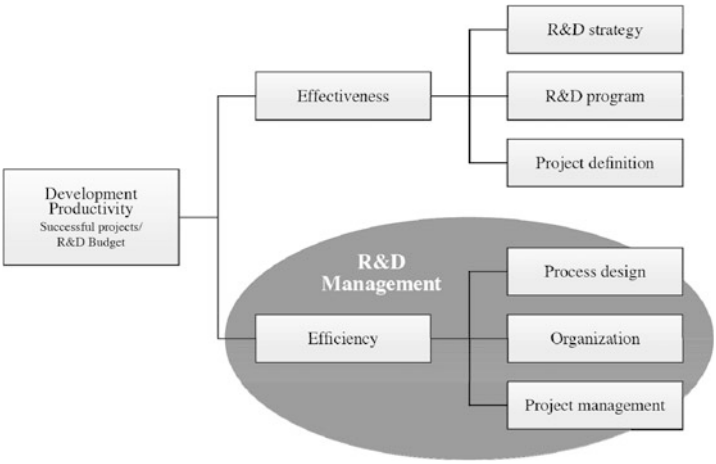


Fig. 3 Effectiveness and efficiency [27, p. 7]

Table 1 Overview of product development processes

	Organization of product development	Organization of development process	Milestones	Time frame
Brand A	• Matrix organization • Project description • Cross-functional cooperation	• Standard process • Phases	Design, develop series	55 months from serial development
Brand B	• Matrix structure • Project definition • Interfunctional product teams	• Standard process • Milestone sequence	Product definition, draw up concept, develop series	48 months from serial development
Brand C	• Matrix organization • Project description • Cross-functional teams	• Standard process • Phase process	Define offer, Analyze concept, develop series	48 months from project start
Brand D	• Matrix structure • Project definition • Multifunctional teams	• Standard process • Phase process	Definition, concept, develop series, SOP	36 months from serial development

Source Author Müller, M., 2016, adapted from Pfaffman [28, p. 66]

shows, that product development times still lean too one-sidedly toward hardware development and that ICT is not yet integrated enough in the product development processes. Empirical studies have shown that time management in R&D is a great, if not the greatest, efficiency lever. Delays in planned market launches are thus often associated with negative results [29, p. 223]. Shortening production time by six

months can bring an approximately 30 % reduction in costs for the company [30, p. 2]. Loss of profit can be caused by higher development costs, actual product costs, smaller volumes or lower prices as demonstrated in the following figure:

Figure 4 clearly shows that shortening development time is the largest lever for increasing added value within product development. All the models mentioned in this chapter offer different approaches to this question and are thus included in the hypothesis discussion of this paper.

Hypothesis

Within the product development process, the necessity to take a closer look beyond the particular product towards the overall process has to be stated clearly [32, p. 40]. In addition, in the "disruptive" development process new actors play a role, for example suppliers as development service providers, internet providers or also network operators in the value added chain. Their know-how lets them support the process holistically, and they thus have a noticeable influence on the reduction of development time. Other important stakeholders are the markets with their political and legal framework conditions as well as various financial services providers, who make funding opportunities accessible. Within the hypothesis developing, the following criteria were considered: industry networks, global orientation, joint trust and harmonization of interests of all actors, definition of common goals and implementation strategies, determination of roles/rights and responsibilities and creation of technical prerequisites in the dynamic system landscapes. The following figure shows the formulated hypothesis:

In the following, the hypotheses formulated in Fig. 5 shall be briefly discussed.

H1 The new technologies, for example 3D printing of prototypes or a worldwide, digital, interconnected development team with robot support and participating system partners, can in future speed up the processes in product development

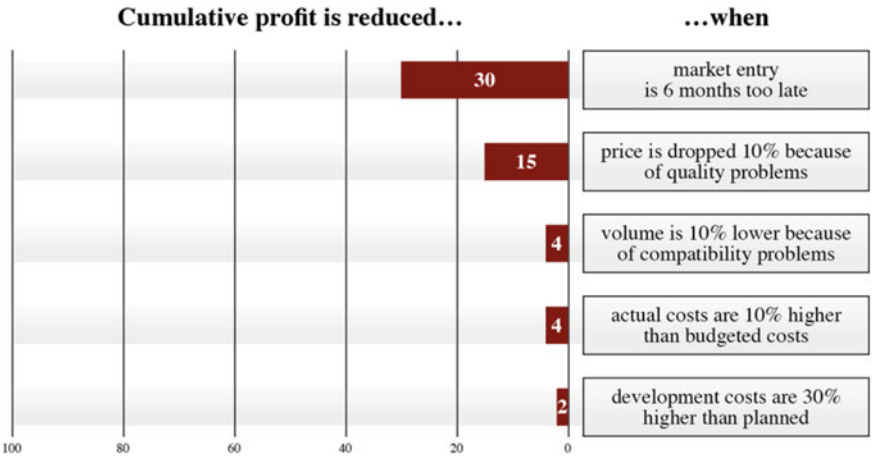


Fig. 4 Influence on earnings caused by deviations in planning within R&D [31, p. 223]

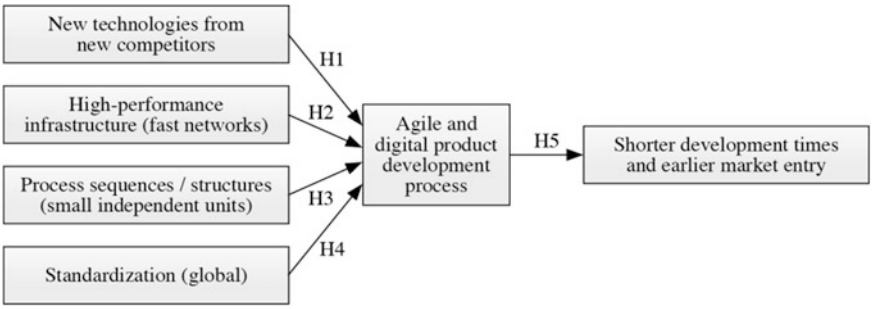


Fig. 5 Hypothetical research model. *Source* own development author Müller, M.

(around the clock, 24 h/day). In addition to saving time, this can also improve cost efficiency.

H2 Realization of the digital processes and structures in the manufacturing companies requires an adequately fast infrastructure, for example a 5G network with approx. 10 giga/bits as enabler for stable networking with the other actors.

H3 The creation of cooperation models such as, for example, small flexible units makes processes and structures leaner and increases their implementation speed.

H4 Global standardization creates scale effects and permits fast networking and continuity of communication.

H5 Finally, a digital and agile product development process ensures reduction of development times and earlier market entry. It also permits knowledge to be saved, data to be analyzed, a learning system to be established and autonomous decisions to be made.

From the mentioned hypotheses (H1—H5) the following research question can be formulated: "Does earlier market entry also mean shorter development times?" Now, that the study has been completed and the hypothesis have been formulated, a deeper insight on the research method within the next chapter.

4 Methodology—Research Design

The survey methods developed in recent years not only call for quantitative analysis models, but also open survey methods using interpretative instruments [33, p. 9]. In order to derive an explanatory model for the hypothesis to „shorten development time by means of a digital and agile product development process", a descriptive approach is chosen, from which possible recommendations can be drawn [34, p. 47]. The descriptive approach aims to use field research methods and a cautious understanding and explaining of the new for the "purpose of content" analysis [35, p. 10].

The empirical study in this paper employs a combination of telephone interviews and an online questionnaire as its survey method. The interviews were conducted by telephone with the aid of a specially designed guideline. The interview results were repeated after the end of the telephone call to confirm the correctness of the recorded information. The online version of the questionnaire was presented to various groups on the Xing platform. The questionnaire included five closed and one open question. A section with general questions to be answered on voluntary base has been added. The results of the studies (interviews and online questionnaire) will be presented in Chap. 5.

5 Statistical Analysis and Results

The results of the empirical study show which efficiency levers could have great optimization potential. The analysis of two empirical studies is given as follows: The first survey, namely by telephone interview, was conducted by phone with six major international car makers from June to August 2015. All seven participants were decision makers in their companies. The following Figs. 6 and 7 provide information on the interviewed parties.

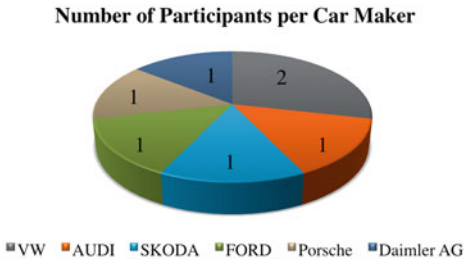


Fig. 6 Number of participants per car maker [36]. *Source* Interviewed by author Müller, M., 2015

Fig. 7 Survey participants in departments [37]. *Source* Interviewed by author Müller, M., 2015

The interview guideline was divided into the following five sections: Strategic product development, Organization of the product development process, Methods and instruments for product development, Communication and Spheres of activity.

From the interview questionnaires the respondents were grouped into various spheres of activity:

1. Interaction/communication/decision behavior
2. Product definition/properties
3. Volume/price/offer complexity
4. Procurement/budget/cost management

The responses given by the experts in the various spheres of activity were weighted according to a school grading system as follows:

1. Little need for action
2. Medium need for action
3. Acute need for action

The analysis of the responses given by the company experts delivered the following results (Fig. 8).

The interview results indicated a need for action in the following spheres of activity (score 3.0):

1. Interaction/communication/decision behavior, and
2. Procurement/budget/cost management.

This first survey revealed, that the sphere of activity „interaction, communication and decision behavior" contains enormous chances for optimization of product development. The choice of the spheres of activity with the score 3.0 is based on the amount of discrepancies and the changes in potentials. The differences in score between the spheres of activity with score 3.0 and 2.5 (Median 2.75) represent up to 20 %. For a prioritized implementation plan of the identified levers of efficiency, the differentiation indicators offer a resource-based approach for a reduction in development time.

A second study on the identification of efficiency levers in the product development process was with an online questionnaire in April 2016. The main aim of this survey was to identify the significance of a reduction in development time

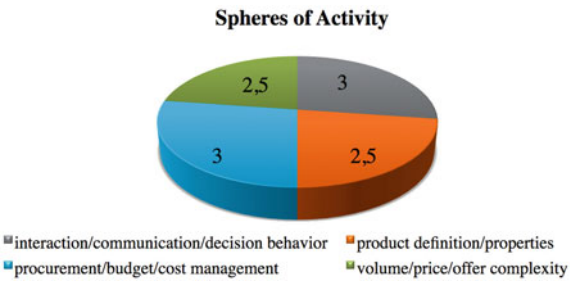


Fig. 8 Spheres of activity. *Source* Interviews by author Müller, M., 2015 [38]

Table 2 Results of online survey 2016

Questions	Results
1. Increasingly shorter product life cycles have a negative effect on the period of time in which a company can make a profit	3, 8
2. The shorter product life cycles mean new products must enter the market sooner	4, 2
3. If a product is to enter the market sooner, the development time in Research and Development has to be reduced	4, 4
4. Reducing the development time in the product development process causes a decrease in development costs	2, 4
5. The possibility to influence costs is greatest in the early phase of the product development process	4, 6

Source own research author Müller, M.

influences for earlier market entry. The survey consisted of five closed questions on product development. The sample included 23 responses from six countries (Germany, USA, Switzerland, Austria, China and Great Britain). The responses were graded on a school grading system as follows:

1. Strongly disagree
2. Disagree
3. Neither agree nor disagree
4. Agree
5. Strongly agree

Analysis of the responses to the online survey gave the following results (Table 2).

The results from questions 2 and 3 very impressively show how important earlier market entry is for new products and how important it is for a company to reduce development times in R&D. This survey confirms the formulated hypothesis, that reduced development time is urgently needed for earlier market entry of new products. Considering these survey results, we conclude that the optimization levers such as “shortening development time” and “cooperation” (interaction, communication, decision behavior) are of high importance. Furthermore, it has to be questioned what efficiency increasing potentials exist in these spheres of activity in the product development process and how a procedure model for a “digital & agile“ product development process can look like. This is followed by remarks in Chap. 6.

6 Approach for a Procedure Model

According to the theoretical model in Fig. 5, we would like to illustrate a first process idea specially designed for the “Digital + Agile Product Development Process (DAPDP).” For this purpose, various development times (*depending on the development effort* involved in the intended innovation) are considered. With this

differentiation (for example low, medium and high effort) efficiency advantages can be made transparent and realized by shortening development time in the particular projects. For a course of action accepted by all actors it is recommended that the particular experts in the various departments be involved in determining the amount of effort and thus contribute their expertise. The following figure shows a first design for DAPDP:

From the splitting method divided into degrees of effort (high, AG3; medium, AG2; low, AG1) as shown in Fig. 9 we see what serious differences in development time are associated with the various degrees of effort. Moreover, the time resources (A + B, Fig. 9) that become available for further, additional innovations, can also be derived. However, how can we ensure that the value adding potentials shown under "A" and "B" are correctly recognized, described and used to best advantage? The draft model in Fig. 10 shows how the various levels interact. It shows how the dimensions "business process," "milestones/time" and "supported IT systems" can work with each other in a networked process. The following schematic drawing shows a draft for the possible procedure model:

This figure shows how the necessary networking and cooperation between the various dimensions could look in the future product development process. From an enormous quantity of information from the various departments "knowledge" must ensue with the help of digital management strategies. Digital communication in the IT system of tomorrow transforms future products, processes or also persons into data sets. The US physicist Larry Smarr even speaks of "mapping" the body [39]. In a later step the information and knowledge will be prepared in a user-friendly manner for autonomous, rational decisions to be made by computer (without having to ask for advice from specialists for interpretation). The involved departments are strongly networked and communicate in real-time via the intelligent, self-learning IT system of the product development process. In a later development stage (totally automated car manufacture including idea generation and product development) networking will have the central role from research & development via new products to the machines [40, p. 25].

The development of new mobile online services in the automotive industry for example is addictive on an integrated, digital platform and agile implementation.

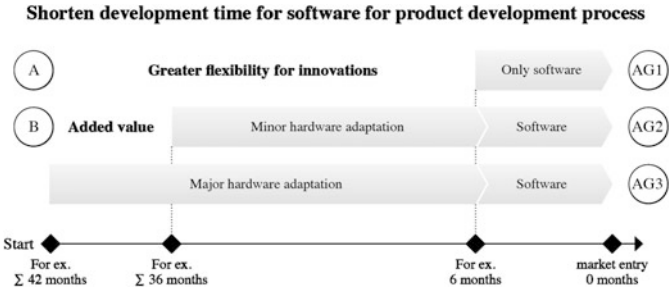


Fig. 9 Digital + Agile product development process. *Source* own development author Müller, M.

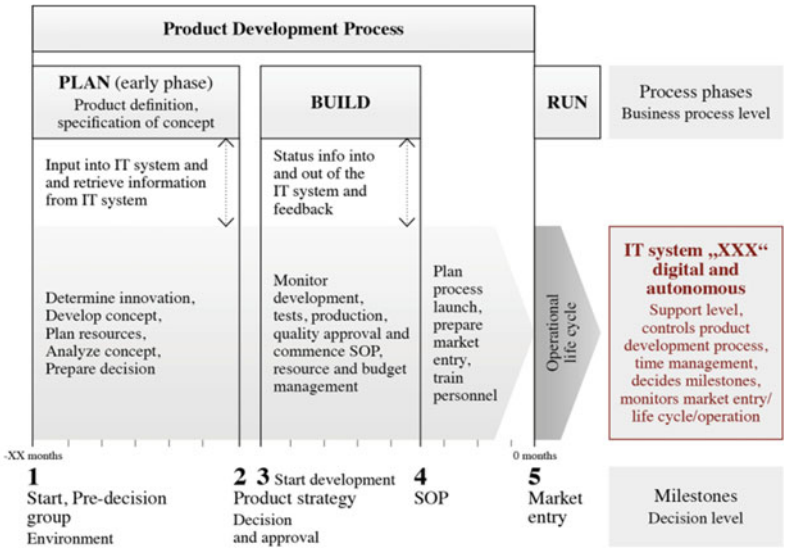


Fig. 10 Concept for procedure model for agile product development process with dimensions. *Source* own development author Müller, M.

The cooperation between the specialist divisions plays a crucial role. There might be significant advantages for increasing efficiency, if products in a common unit consisting of sales, development, IT and on a process-steered platform, depending on the development expenses (agile) will be planned and developed together.

7 Conclusion

The Introduction gives a comprehensive overview of the current challenges and significant opportunities for value creation in the field of product development process in the automotive industry.

Chapter 2 deals with the theory of Institutional Economic Role Model which is mainly based on the principles of institutional economics in terms of acceleration of the product development process. This paper is a first step for the IRM-process to co-ordinate the activities of various stakeholders to accelerate the product development process.

Thereafter a review of the current research literature and formulation of hypothesis is followed.

Chapter 4 looks at the research design. The empirical study employs a combination of telephone interviews and an online questionnaire as its survey method.

Chapter 5 presents two studies over efficiency levers in the product development.

The important influence factors identified are, for example, cooperation and communication, decision-making behavior, transparency or compulsory early integration of procurement into product development. The results of online study very impressively shows how important the earlier market entry for new products is and how important it is for a company to reduce development times in R&D. The findings culled from the studies gave a rise to the digital & agile product development process (DAPDP) approach.

The last part of the paper shows a more detailed description of the agile product development model. As a management method it offers a multifunctional approach for effectively increasing efficiency in product development within the automotive industry.

With its theoretical method of the Digital Agile Product Development Process, this paper can only be a first step towards future optimization of development time in research and development. Further empirical studies and subsequent analysis are thus needed to enhance the model.

References

1. GENCO. Shorter product life cycles impacting supply chains. [Online] Available from: http://www.genco.com/Logistics-Articles/article.php?aid=800900305. Accessed 12 March 2016
2. Dörner K, Mohr N, Schumacher T, Meffer J (2016) Vom Aktenordner zur Cloud. Vom Risiko, den digitalen Umbruch zu verschlafen. Handelsblatt, [Online] Monday 18 Jan 2016. Available from: http://www.handelsblatt.com/adv/digitalatscale/vom-aktenordner-zur-cloud-vom-risiko-den-digitalen-umbruch-zu-verschlafen/12819484.html. Accessed 10 Feb 2016
3. Noser R (2016) *Open Data als Standortfaktor*. Neue Zürcher Zeitung, [Online] Wednesday 3 Feb 2016. Available from http://www.nzz.ch/meinung/kommentare/entwicklung-im-werbe-und-datenmarkt-open-data-als-standortfaktor-ld.4846. Accessed 25 March 2016
4. Schmidt N (2016) GM Chairman & CEO Mary Barra to highlight the future of personal mobility at 16th CAR-Symposium. Automotive industry undergoing biggest change in last 50 years. [Online] Available from: http://media.opel.com/media/intl/en/opel/news.detail.html/content/Pages/news/intl/en/2016/opel/02-08-mary-barra-16-car-symposium-bochum.html. Accessed 15 Feb 2016
5. Malorny A (2016) Das Ende der Autoindustrie, wie wir sie kennen. Die Welt, [Online] Tuesday 29 Sept 2015. Available from: http://www.welt.de/motor/article146698673/Das-Ende-der-Autoindustrie-wie-wir-sie-kennen.html. Accessed 27 March 2016
6. Bay L (2016) Das Ende der Elefantenrennen. Handelsblatt, [Online] Tuesday 22 March 2016, Available from: http://www.handelsblatt.com/unternehmen/industrie/daimler-vernetzt-lkw-das-ende-der-elefantenrennen/13352788.html. Accessed 21 March 2016
7. Bullinger H-J (1994) Neue Impulse für eine erfolgreiche Unternehmensführung. Springer-Verlag, Berlin, p 32
8. Vogl W (2016) Wichtige Kennzahlen bei der Finanzierung von Softwareentwicklungen. Kritische Erfolgsfaktoren der Finanzperspektive. 13 Aug 2012, Weblog, Available from: http://www.innovationsjournal.de/wichtige-kennzahlen-bei-der-finanzierung-von-softwareentwicklungen/. Accessed 2 June 2016
9. Bogaschewsky R, Müller H (2016) Industrie 4.0—Wie verändern sich die IT-Systeme in Einkauf und SCM?. Bundesverband Materialwirtschaft, Einkauf und Logistik e.V. (BME), Bundesverband Materialwirtschaft, Einkauf und Logistik in Österreich (BMÖ), p 4

10. Brower-Rabinowitsch G, Fasse M, Murphy M (2016) Handelsblatt, [Online] Monday 29 Feb 2016, Available from: http://www.handelsblatt.com/my/unternehmen/industrie/autokonzern-setzt-auf-digitalisierung-audis-aufbruch/13028504.html. Accessed 29 Feb 2016
11. Schneider D (1995) Betriebswirtschaftslehre. Grundlagen, 1st edn. De Gruyter, Oldenbourg
12. Luhmann N (2002) Einführung in die Systemtheorie. Carl-Auer Verlag, Heidelberg
13. Schulz WH, Joisten N, Mainka M (2013) Entwicklung eines Konzeptes für institutionelle Rollenmodelle als Beitrag zur Einführung kooperativer Systeme im Straßenverkehr. Bundesanstalt für Straßenwesen, Bergisch Gladbach
14. Schulz WH (2011) Institutional economic role model—a new approach for non-discriminatory cooperation. Mimeo
15. Schulz WH, Wieker H, Kichniawy J (2014) Research joint ventures as a European policy instrument beneath directives and action plans: transitions, interlocking and permeability of political, Technological and Economical Requirements. Chicago, 6 April 2014
16. Schneider, Op. Cit
17. Goldstein B (1999) Modellgestützte Geschäftsprozessgestaltung in der Produktentwicklung. Herbert Utz Verlag, Munich, p 14
18. Kleinaltenkamp M, Plinke W (1999) Markt- und Produktmanagement. Springer-Verlag, Berlin, p 126
19. Bullinger H-J (1992) Marktgerechte Produktentwicklung. Springer-Verlag, Berlin, p 120
20. Sherman DJ, Sounder WE, Jenssen SA (2000) Differential effects of the primary forms of cross functional integration on product development cycle time. J Product Innovation Manage 17(4):257–267
21. Schömann SO (1992) Produktentwicklung in der Automobilindustrie. Springer-Verlag, Berlin, p 18
22. Kleinaltenkamp, Op. Cit. p 80
23. Horwath P (2011) Controlling, 12th edn. Verlag Franz Vahlen, Munich, p 788
24. Röh C (2013) Ressourceneffizienz im verarbeitenden Gewerbe—Fokus Automotive. In: Presented at the IHK Conference „Effiziente Produktion—Materialeffizienz ein Wettbewerbsfaktor“. Munich, 16th May 2013
25. Ehrlenspiel K, Kiewert A, Lindemann U, Mörtl M (2014) Kostengünstig Entwickeln und Konstruieren, 7th edn. Springer Vieweg Verlag, Berlin, p 14
26. Seidemann W (1999) Portfoliomanagement zur Steigerung der Entwicklungs-effektivität. PhD thesis, University of Munich, p 6/7
27. Ibid
28. Pfaffman E (2001) Kompetenzbasiertes Management in der Produktentwicklung, 1st edn. Deutscher Universitäts-Verlag GmbH, Wiesbaden, p 66
29. Backhaus K, Voeth M (2010) Industriegütermarketing. Verlag Franz Vahlen, Munich, p 223
30. Wildemann H (2004) Kundenorientierte Produktentwicklung in der Automobilindustrie, In: Schwarz ES (ed) Innovationsmanagement. Wiesbaden, pp 381–408
31. Backhaus, Op. Cit. p 223
32. Ehrlenspiel K, Kiewert A, Lindemann U, Mörtl M (2014) Kostengünstig Entwickeln und Konstruieren. 7th edn. Springer Vieweg Verlag, p 40
33. Mayring P (2010) Qualitative inhaltsanalyse, 12th edn. Beltz Verlag, Weinheim and Basel, p 9
34. Müller-Stewens G, Lechner C (2001) Strategisches management. Schäfer Pöschel, Stuttgart 2001:47
35. Mayring Op. Cit. p 10
36. Müller M (2015) OEM Survey by telephone interview
37. Müller, M. OEM Survey by telephone interview, 2015
38. Müller M (2015) OEM Survey by telephone interview
39. Kaeser E (2016) Digitale Technologien formatieren unsere Körper: Glücklich bis unter die Haut. Neue Zürcher Zeitung [Online], Tuesday 23rd Feb 2016, Available from: http://www.nzz.ch/meinung/debatte/gluecklich-bis-unter-die-haut-1.18699494 Accessed 10th March 2016
40. Möller K-H (2016) Quellen des Lebens, 1st edn. publiq Verlag GmbH, Berlin, p 25